UNIVERSITAT DE BARCELONA

Adquisición de la entonación de las oraciones interrogativas del español por parte de los sinohablantes: patrones entonativos y propiedades fonético-acústicas

Peizhu Shang

Adquisición de la entonación de las oraciones interrogativas del español por parte de los sinohablantes:

Patrones entonativos y propiedades fonético-acústicas

Peizhu Shang

Tesis doctoral presentada para optar al grado de

Doctora en Lingüística

en el programa de doctorado

Ciencia Cognitiva y Lenguaje

Departamento de Filología Catalana y Lingüística General,

Universitat de Barcelona

Bajo la supervisión de

la Dra. Wendy Elvira-García y del Dr. Paolo Roseano

Tutor: Faustino Diéguez Vide

UNIVERSITAT DE BARCELONA

Mayo de 2022

Peizhu Shang.

致我的父母亲，

感谢你们把我带到了这个美好的世界，

感谢你们的谆谆教导和殷殷嘱托。

Agradecimientos

Esta tesis no hubiera sido posible sin el apoyo de varias personas e instituciones a las que quiero dar mis sinceras gracias.

Primero y antes que nada, deseo expresar mis más sinceros y profundos agradecimientos a mis directores de la tesis. A la Dra. Wendy Elvira-García, por aceptar dirigir esta tesis y abrirme las puertas a un nuevo mundo de posibilidades, pese a que no tuviera una formación en fonética antes de iniciar el doctorado. Por brindarme la oportunidad de participar en sus cursos de Fonética y Fonología de Laboratorio impartidos en esta facultad. Por enseñarme a utilizar las herramientas de análisis de habla y los programas estadísticos desde cero y por prestarme todo el apoyo posible durante el desarrollo del trabajo de investigación. Muchas gracias también por su amabilidad, su entusiasmo y la confianza que siempre ha tenido en mí desde que llegué a la Universitat de Barcelona.

Asimismo, quisiera agradecer al Dr. Paolo Roseano, quien con sus amplios conocimientos y acertadas orientaciones me guio en cada una de las etapas importantes de este proyecto para alcanzar los objetivos planteados. Gracias por su incondicional ayuda y por la dedicación y el rigor con los que ha facilitado la elaboración de esta tesis. No existen palabras suficientes para expresar mi gratitud a mis dos directores de tesis. Sin ellos, sin su paciencia y constancia, no hubiera podido alcanzar los resultados esperados. Ha sido una gran fortuna haber trabajado con ellos durante los cuatro años que se ha dedicado al doctorado.

Además, quisiera expresar mi agradecimiento al tutor de la tesis, el Dr. Faustino Diéguez Vide, por su continua atención a mis consultas sobre los estudios de doctorado. A la Dra. Montserrat Nofre Maiz, por haberme incluido como alumna en la asignatura de Lingüística Aplicada y por brindarme toda la información sobre cursos de estadística que me han resultado imprescindibles para mis investigaciones. Igualmente, quiero expresar mi agradecimiento al Dr. Fernando Polanco Martínez, en cuya asignatura de Pragmática del Español he podido adquirir nociones fundamentales para las etapa iniciales de la tesis.

Deseo también dar las gracias al Laboratorio de Fonética de la Universitat de Barcelona por la cálida acogida y el apoyo académico y técnico que me ha ofrecido para la realización del trabajo de la tesis. De manera especial quiero expresar mi agradecimiento al Dr. Eugenio Martínez Celdrán y a la Dra. Ana María Fernández

Agradecimientos

Planas por su generosa ayuda en el proceso de admisión en el doctorado y en la solicitud de beca. Gracias también a los otros miembros del Laboratorio por hacer que mis experiencias en la Universitat de Barcelona fueran inolvidables.

Gracias a todos los informantes chinos y españoles que participaron en este proyecto de investigación y permitieron el uso de sus datos con fines estrictamente académicos. A mis queridos amigos y compañeros que están en diferentes rincones del mundo, por compartir horas de estudio y por su apoyo siempre que lo he necesitado. En particular, me gustaría hacer mención de Yuanyuan Zhang, Xinyu Guo y Xinyi Li, por sus múltiples palabras de aliento cuando mis ánimos las necesitaban. A Yiqiu Shi, Lingqiao Su y Songlin Wang, por su importante ayuda en el proceso de búsqueda de informantes. A la Dra. Weiqi Li por sus valiosos comentarios y sugerencias en el diseño de los experimentos.

Gracias a mi familia por estar siempre a mi lado en los momentos más difíciles durante mis estudios de doctorado y por su respaldo incondicional en todo lo que he emprendido a la largo de mi vida. En especial, quisiera darle las gracias a mi madre por su interminable paciencia conmigo y a mi padre por darme la fuerza y la energía necesarias para superar los retos y desafíos de la vida. A mis hermanos y hermanas por haber sido siempre mis mejores guías en el estudio y el motor que impulsa mis sueños y esperanzas.

Por último, pero no menos importante, debo un agradecimiento a la institución de China Scholarship Council por su apoyo económico, que ha posibilitado que pudiera dedicarme completamente a la elaboración de este trabajo.

A todos, muchas gracias.

Barcelona, 2022

Peizhu Shang

Resumen

A diferencia de lenguas entonativas, en las que se utilizan patrones melódicos distintos para vehicular la información postléxica, el carácter tonal del chino mandarín hace que la entonación consista en cambios en el rango tonal, mientras que las diferencias en los contornos de la F0 se utilizan para expresar distinciones del significado léxico. Para los sinohablantes, las diferencias entre los dos sistemas prosódicos plantean serios desafíos a la hora de aprender lenguas entonativas. Estas dificultades están relativamente bien estudiadas en el caso del inglés; en cambio, hasta la fecha ha habido una falta general de estudios sobre la entonación en situaciones en que el chino está en contacto con otras lenguas entonativas, como el español. Por lo tanto, la presente tesis tiene como objetivo fundamental investigar la adquisición de la entonación (en particular, de las oraciones interrogativas) del español peninsular por aprendices cuya L1 es el chino mandarín.

Para ello se ha dividido la tesis en tres partes que abordan 1) las diferencias en la expresión de la interrogatividad en chino y en español, 2) la producción y 3) la percepción de los patrones entonativos de las oraciones interrogativas del español por parte de sinohablantes.

Más en concreto, en la primera parte, a partir del carácter interidiomático de las funciones pragmáticas, se ha planteado una nueva taxonomía de clasificación de la modalidad interrogativa que tiene la ventaja de ser aplicable a las dos lenguas estudiadas. En base a este marco teórico, se ha realizado una comparación interlingüística de los recursos lingüísticos usados en chino y en español para expresar las diferentes finalidades pragmáticas de las oraciones interrogativas.

La segunda parte consiste en describir la representación fonológica y la implementación fonética de la entonación del español como LE partiendo del modelo métrico-autosegmental (y, en particular, de la anotación con el sistema Sp_ToBI) y de varios parámetros tonales y temporales. Además, se ha establecido una comparación del uso de los elementos categóricos para expresar los significados pragmáticos de las oraciones interrogativas entre los hablantes nativos y no nativos del español. Asimismo, se ha realizado un análisis de los factores lingüísticos y extralingüísticos que pueden influir en la producción de la entonación del español como lengua extranjera.

La tercera y última parte, dedicada a la percepción, incluye dos tipos de test relacionados con la percepción fonológica y fonético-acústica de las declarativas e interrogativas totales del español. El primero aborda la identificación de diferentes patrones entonativos que se generan segmentando la secuencia melódica de la oración. Mientras en el segundo, se analizan cómo los oyentes chinos y españoles utilizan cambios en las tres dimensiones acústicas (F0, duración e intensidad) para reconocer la categoría entonativa, y cómo interactúan los distintos correlatos acústicos en el proceso perceptivo. En esta parte, también se han tenido en cuenta las características lingüísticas de los estímulos y las propiedades individuales de los oyentes para ofrecer una evaluación completa del desarrollo perceptivo de la entonación de ELE.

En conjunto, esta tesis supone una importante contribución en dos áreas poco exploradas dado que, por un lado, profundiza nuestra comprensión del aprendizaje de la entonación de ELE por parte de los aprendices sinohablantes mediante el análisis prosódico en diferentes dimensiones y, por otro, pone en evidencia las mayores diferencias expresivas respecto a la modalidad interrogativa del chino y del español, proponiendo una nueva solución para clasificar las oraciones interrogativas que podría servir de base teórica para futuros estudios interlingüísticos en este ámbito.

Abstract

Unlike intonational languages, where distinct melodic patterns are used to convey post-lexical information, the tonal property of Mandarin Chinese makes that its intonation consists of changes in pitch range, while differences in F0 contours are used to express distinctions in lexical meanings. The differences between the two prosodic systems pose serious challenges to Chinese speakers in learning intonational languages. These difficulties are relatively well studied in the case of English, but to date there has been a general lack of studies on intonation in situations where Chinese is in contact with other intonational languages, such as Spanish. Therefore, the fundamental objective of the present thesis is to investigate the acquisition of intonation (in particular, of interrogative sentences) of Peninsular Spanish by learners whose L1 is Mandarin Chinese.

To this end, the thesis has been divided into three parts that address 1) the differences in the expression of interrogativity in Chinese and Spanish, 2) the production and 3) the perception of the intonation patterns of Spanish interrogative sentences by Chinese speakers.

More specifically, in the first part, based on the interidiomatic nature of pragmatic functions, we have proposed a new taxonomy of classification of the interrogative modality, which has the advantage of being applicable to the two languages studied. Based on this theoretical framework, we have conducted a cross-linguistic comparison of the linguistic resources used in Chinese and Spanish to express the different pragmatic purposes of interrogative sentences.

The second part consists of describing the phonological representation and the phonetic implementation of intonation in Spanish as a FL based on the autosegmental-metrical model (and, in particular, on the Sp_ToBI annotation) and on various tonal and temporal parameters. In addition, a comparison of the use of categorical elements to express the pragmatic meanings of interrogative sentences has been established between native and non-native speakers of Spanish. We also have analyzed the linguistic and extralinguistic factors that may influence the production of intonation in Spanish as a foreign language.

The third and last part, dedicated to perception, includes two types of test related to the

phonological and phonetic-acoustic perception of Spanish statements and yes-no questions. The first concerns the identification of different intonation patterns generated by segmenting the melodic sequence of the sentence. The second analyzes how Chinese and Spanish listeners use changes in the three acoustic dimensions (F0, duration and intensity) to recognize the intonation category, and how the different acoustic correlates interact in the perceptual process. In this part, we also have taken into account the linguistic characteristics of the stimuli and the individual properties of the listeners in order to provide a comprehensive assessment of the perceptual development of SFL intonation.

Overall, this thesis makes an important contribution in two under-explored areas since, on the one hand, it deepens our understanding of SFL intonation learning in the context of Chinese speakers through prosodic analysis in different dimensions and, on the other hand, it highlights the major expressive differences with respect to the interrogative modality of Chinese and Spanish, proposing a new solution for classifying interrogative sentences that could serve as a theoretical basis for future cross-linguistic studies in this area.

摘要

语调型语言通过不同的旋律模式来传达后词汇信息，而汉语普通话的声调特性使得其语调主要借助音域变化来表达，相反地，其 F0 音高轮廓上的差异则被用来区分词汇含义。这两种语调系统之间的差异给汉语使用者学习语调型语言带来了严重的挑战。这些困难在英语方面的研究相对较多，但截至目前，普遍缺乏对汉语与其他语调语言（如西班牙语）接触情况下的语调研究。因此，本论文的基本目标是研究汉语母语者对伊比亚半岛西班牙语语调，特别是疑问句语调的习得情况。

为此，本论文分为三大部分，分别涉及到：一）汉语和西班牙语中疑问句表达的差异，二）汉语学习者西班牙语问句语调模式的产出以及，三）感知。

具体来讲，在第一部分中，基于语用功能的跨语言特性，我们提出了一个新的疑问语气分类系统，其优点是适用于所研究的两种语言。基于这一理论框架，我们对汉语和西班牙语中用于表达不同语用目的语言资源进行了跨语言比较。

在第二部分中，基于自主音段-节律模型（特别是基于 Sp_ToBI 的标注）以及若干音高和时间参数，我们对西班牙语作为外语语调的音系表征和语音实现进行了描述。此外，我们对西班牙语母语者和非母语者如何使用结构性元素来表达语用意义的情况进行了比较，同时也对可能影响西班牙语语调习得的语言和非语言因素进行了分析。

第三部分，即最后一部分，包括两类感知测试，这两类测试分别与西班牙语陈述句和一般疑问句的音系和声学感知有关。第一类测试是关于对不同语调模式的识别，这些模式是通过对句子旋律曲线的分割得到的。第二个测试主要分析了汉语和西班牙语听众如何利用三个声学维度（F0、时间和强度）上的变化来识别语调类别，以及不同的声学相关线索在感知过程中如何相互作用。在这一部分中，为了全面评估西班牙语作为外语的语调感知发展，我们还考虑到了语音材料的语言特性和听众之间的个体差异在感知过程中的影响。

总的来说，本论文在目前两个尚未充分探索的领域内做出了重要贡献。一方面，

它通过对不同维度的语调分析，加深了我们对汉语母语者学习西班牙语语调的理解；另一方面，它指出了汉语和西班牙语在疑问语气表达方面的主要差异，并提出了一个新的问句分类方案，为未来该方向的跨语言研究提供了理论基础。

Índice

Índice de Figuras

Índice de Tablas

Lista de Abreviaturas

ADV	Adverbio
AM	Autosegmental-Metrical/Métrico-Autosegmental
AQ	Autism Spectrum Quotient
AUC	Area Under the Curve
CLF	Clasificador
CONJ	Conjunción
CYN	Confirmation-seeking *yes-no* question
DCT	Discourse Completion Task/Test
DELE	Diplomas de Español como Lengua Extranjera
DJ	Information-seeking disjunctive question
ELE	Español como lengua extranjera
ERB	Equivalent Rectangular Bandwidth
F0	Frecuencia fundamental
FL	Foreign language
GLM	Generalized Linear Model
INF	Infinitivo
INTERJ	Interjección
L1	Primera materna
L2	Segunda lengua
LE	Lengua extranjera
LILt	L2 Intonation Learning Theory
LTD	Long-term distributional (method)
MCER	Marco Común Europeo de Referencia para las lenguas
NEG	Negación, negativo
OBJ	Objeto
OGLM	Ordered Generalized Linear Model
OR	Odds ratio
PAM	Perceptual Assimilation Model

PDQ	Pitch Dynamism Quotient
POLR	Proportional-Odds Logistic Regression
PREP	Preposición
PRF	Perfecto
Q	Partícula interrogativa
ROC curve	Receiver Operating Characteristic Curve
SD	Standard Deviation
SFL	Spanish as foreign language
SLA	Second language acquisition
TAG	Confirmation-seeking tag question
ToBI	Tones and Break Indices
VOT	Voice Onset Time
VP	Sintagma verbal
WH	Information-seeking wh-questions
YN	Information-seeking yes- no question

Capítulo 1

Introducción

La entonación, entendida como uno de los rasgos suprasegmentales más destacados del habla, se considera uno de los aspectos más difíciles de dominar para los que aprenden una lengua extranjera (LE) en la edad adulta. La dificultad para adquirir una entonación comparable con la de los nativos no solo se relaciona con la mayor susceptibilidad de la entonación a ser objeto de transferencias de los patrones de la primera lengua (L1), sino que pueden atribuirse, en el caso de los estudiantes cuya L1 es el chino, a la falta de atención al fenómeno prosódico en el proceso de enseñanza-aprendizaje de lenguas extranjeras entonativas como el español.

Esta tesis tiene como objetivo fundamental investigar la adquisición de la entonación del español por aprendices cuya L1 es el chino mandarín. En particular, el trabajo se centra en la adquisición de los patrones entonativos de las oraciones interrogativas del español peninsular.

Los antecedentes que fundamentan la elección de dichas oraciones tienen que ver con las diferencias tipológicas de interpretación y expresión lingüística de la modalidad interrogativa entre el chino y el español, que dificultan seriamente la comprensión y la asimilación de las mismas por parte de aprendices chinos. En vista de esto, se hace indispensable contar con análisis contrastivo del sistema interrogativo de las dos lenguas que permita comparar las estrategias lingüísticas usadas en cada una de ellas para el mismo tipo de pregunta. El principio general en el que se sustentan los análisis

comparativos es que las funciones pragmáticas del lenguaje son compartidas por todas las lenguas, y esto es también el enfoque teórico que se adopta en esta tesis.

La diferencia tipológica más importante para esta tesis entre el chino y el español reside en que el chino mandarín explota en menor medido los recursos entonativos y privilegia, en cambio, el marcaje mediante partículas modales. Este hecho, junto con las diferencias entre las estructuras prosódicas y los mecanismos entonativos del chino y español, hace que los aprendices chinos a menudo produzcan –incluso después de muchos años de aprendizaje del español– unos patrones entonativos que se desvían de los que utilizan los hablantes nativos. Estas diferencias pueden observarse o tener consecuencias tanto en el ámbito de la producción como en la percepción, y se reflejan específicamente tanto en el nivel fonológico como en el nivel fonético-acústico de la entonación. Además de la influencia del sistema fonético y fonológico de la L1, existen un conjunto de factores lingüísticos y extralingüísticos (o individuales) que pueden incidir en el proceso de aprendizaje de la entonación.

Esta tesis parte de la relación entre pragmática y prosodia para investigar la producción y la percepción de los patrones entonativos del español por parte de aprendices chinos y evalúa los efectos que pueden tener distintos factores en la adquisición de la entonación de la interrogatividad, con el fin de ofrecer orientaciones que guíen en la enseñanza y el aprendizaje de la entonación del español en el contexto chino.

1.1 Objeto de estudio

El objeto de estudio que se ha elegido para investigar la adquisición de la entonación del español por hablantes del chino mandarín (también llamados sinohablantes) se centra en la modalidad interrogativa, una categoría lingüística frecuentemente utilizada en la comunicación, pero propensa a errores de expresión en las prácticas orales. Las dificultades de adquisición de este fenómeno lingüístico provienen, en su mayoría, de las diferencias tipológicas entre la lengua china y la española, que no se circunscriben al nivel prosódico, sino que se interesan también el plano morfológico, sintáctico, semántico y discursivo. Para describir e interpretar las características del español como lengua extranjera (ELE), se realiza en primer lugar un análisis comparativo de la modalidad interrogativa del chino y del español, ya que se ha comprobado que la

adquisición de lenguas extranjeras está influenciada por el sistema lingüístico de la L1, ya sea de forma positiva o negativa (Gabriel & Kireva, 2014; Gass, 1988; Rasier & Hiligsmann, 2007).

La modalidad interrogativa ha sido definida y clasificada desde muy diversas perspectivas lingüísticas tanto en la gramática china como en la española, pero que ninguno de los marcos teóricos que se han propuesto para cada una de ellas es aplicable de manera perfecta y paralela a ambas lenguas. Por consiguiente, para poder realizar un estudio contrastivo de la modalidad interrogativa, hemos necesitado plantear un nuevo sistema de clasificación y descripción de la modalidad interrogativa a partir de los valores funcional-pragmáticos que, según Belchí (1994), tienen un carácter interidiomático y, por lo tanto, permiten comparar las estrategias lingüísticas de expresión de cada tipo de interrogativa en castellano y en mandarín.

Tal como se ha expuesto anteriormente, esta tesis tiene como objeto principal el estudio de la producción y percepción de la entonación interrogativa del español por parte de sinohablantes. Frente a la imposibilidad de abarcar toda la gama de la modalidad interrogativa, la tesis centra su atención en las interrogativas que desempeñan la función de pedir información y confirmación, es decir, en las así llamadas preguntas parciales informativas, en las preguntas totales informativas, en las preguntas disyuntivas informativas, en las preguntas totales confirmatorias y en la preguntas confirmatorias de apéndice. La selección de dichas variedades de interrogativas responde a las necesidades comunicativas fundamentales en la interacción conversacional por parte de aprendices de ELE.

Los informantes chinos de esta tesis disponen de un nivel lingüístico general que va desde el intermedio (B1, B2) hacia el avanzado (C1) según el Marco Común Europeo de Referencia para las lenguas (MCER) (Consejo de Europa, 2002). Sin embargo, el carácter tonal del chino mandarín hace que sus hablantes experimenten más dificultades sistemáticas a la hora de adquirir los patrones melódicos de una lengua entonativa en comparación con los aprendices cuya L1 comparte con la lengua meta la propiedad de ser entonativa (Yuan et al., 2019). Asimismo, la escasa atención que se ha presentado en China a los fenómenos prosódicos de ELE hace que los sinohablantes a menudo posean una habilidad entonativa inferior a sus conocimientos en otros sectores de la gramática.

Algunos de ellos, aun teniendo un nivel de B2 de español, no saben producir y distinguir correctamente, por ejemplo, la entonación de las preguntas totales, disyuntivas y confirmatorias que, según lo que ha establecido el Plan Curricular del Instituto Cervantes (2007), se tendría que adquirir en la etapa correspondiente a los niveles A1, A2, B1 y B2.

Por estos motivos, la investigación se centra en los patrones entonativos que se corresponderían al nivel básico e intermedio de ELE, concretamente, en aquellos que se emiten para solicitar una información nueva o una confirmación dependiente del contexto comunicativo. Mediante la aplicación de métodos de análisis fonológico y fonético-acústico de la entonación, se identifican los patrones melódicos presentados por los sinohablantes y se analizan las desviaciones de estos en comparación con la realización de los hablantes nativos del español.

1.2 Motivación y objetivos

Durante los últimos años, por primera vez en la historia china el español se ha convertido en la segunda lengua (L2) extranjera más estudiada, precedida solo por el inglés. De acuerdo a las estimaciones realizadas (Yu, 2021), esta tendencia tendrá continuidad a lo largo del tiempo debido a la promoción del español en las escuelas secundarias chinas y a la Iniciativa de la Franja y la Ruta[①]. El desarrollo de los estudios hispánicos en China, así como las relaciones culturales y económicas cada vez más intensas entre ese país y los países hispanohablantes, hacen que los aprendices chinos de ELE necesiten disponer de competencias comunicativas cada vez más altas. Sin embargo, en China –donde las universidades y escuelas dan una mayor importancia a las destrezas escritas y al conocimiento gramatical– se suele aplicar el método tradicional de gramática-traducción para la enseñanza del español, dejando en un segundo plano la formación de las competencias orales. A pesar de que la metodología didáctica en el ámbito de lenguas extranjeras ha experimentado una profunda reforma en la última década poniendo cada vez mayor relevancia en la capacidad comunicativa,

[①] La Iniciativa de la Franja y la Ruta (en inglés One Belt One Road Iniciative) es un megaproyecto con el que la República Popular China planea impulsar el desarrollo e integración de la zona Eurasiática, basado en un sistema comercial y de comunicaciones. Como uno de los países europeos que integrados en este proyecto, España ha conseguido establecer una cooperación cada vez más profunda con China en términos de inversión, comercio e infraestructura.

el actual enfoque de las prácticas orales se limita a los elementos segmentales del habla (p.ej., la articulación de las consonantes y vocales y la discriminación de los pares mínimos), faltando directrices didácticas científicamente fundamentadas que sirvan de orientación para la enseñanza-aprendizaje de los rasgos suprasegmentales. Como resultado, los aprendices sinohablantes de ELE a menudo concluyen su formación, incluso después de muchos años de estudios, con un marcado "acento extranjero" que puede incluso afectar la interpretación del mensaje y fomentar la posibilidad de malentendidos en la comunicación intercultural.

Aparte de la imprecisa realización de los segmentos, la prosodia –sobre todo la entonación– se ha considerado como uno de los factores que más contribuyen a conformar un acento percibido como extranjero. Un estudio de Jilka (2000b), por ejemplo, demuestra que el grado de acento extranjero mostrado por los hablantes estadounidenses de alemán venía determinado, en buena medida, por la entonación que utilizaban en la lengua que estaban adquiriendo. Debido al papel destacado que juega la entonación en la comunicación, en las últimas décadas se han desarrollado numerosos estudios que investigan los patrones melódicos de las LE y los aspectos entonativos en que los aprendices de una determinada lengua difieren de los hablantes nativos de esta lengua. A pesar de que existen trabajos recientes sobre la entonación utilizada por aprendices sinohablantes de lenguas entonativas, la mayoría de ellos se centra en el contacto lingüístico con el inglés, de modo que hasta ahora han sido muy escasos los estudios sobre la entonación del español hablando por el mismo colectivo de aprendices. Contribuir a llenar este hueco en la descripción entonativa de ELE hablada por sinohablantes constituye la motivación fundamental del presente trabajo.

Además de contribuir al estudio del caso concreto del español hablado por aprendices sinohablantes, este trabajo pretende también contribuir al desarrollo del debate teórico en el sector de la adquisición de lenguas extranjeras. De hecho, desde los estudios de aprendizaje de segundas lenguas se suele considerar la entonación como un elemento prosódico altamente susceptible a la interferencia lingüística, poniendo de relieve la influencia de la L1 en la entonación de la lengua meta (Alharbi, 2020; Mennen, 2015; Piske et al., 2001). En este sentido, la investigación sobre la entonación del español hablado por sinohablantes puede suponer una aportación a la teoría de la transferencia

prosódica en una situación del contacto lingüístico entre lenguas tonales y no tonales y, más en general, al conocimiento de la adquisición de lenguas extranjeras.

En general, los objetivos de esta tesis son cinco:

1) El primer objetivo, de carácter más general es el de establecer un marco teórico para la interpretación y la clasificación de la modalidad interrogativa a partir de la función pragmática.

2) El objetivo teórico anterior pretende ser la base para alcanzar un objetivo más concreto, que consiste en establecer una comparación interlingüística de las estrategias usadas en chino y en español para expresar las finalidades pragmáticas de los diferentes tipos de oraciones interrogativas.

3) Al mismo tiempo, siempre desde el punto de vista teórico, se pretende probar las teorías científicas en el campo de la adquisición del habla de lenguas extranjeras. Por ejemplo, la teoría del aprendizaje de la entonación de L2 (*L2 Intonation Learning Theory*) (Mennen, 2015), la teoría de la transferencia lingüística (Choi, 2022; Gabriel & Kireva, 2014) y la teoría de la *phonetic trading relation*② (Repp, 1982), entre otras.

4) El objetivo de tipo descriptivo central de esta tesis consiste en la caracterización fonológica y fonética (tanto desde el punto de vista de la producción como de la percepción) de la entonación de ELE hablado por los sinohablantes.

5) Finalmente, también se pretende explorar los posibles efectos que pueden tener los factores lingüísticos y extralingüísticos (o individuales) en la adquisición de la entonación de ELE y aclarar las causas de las diferencias que se observan en el comportamiento entonativo de los aprendices sinohablantes.

Con estos objetivos en mente, la presente tesis pretende configurarse como un trabajo de

② *Phonetic trading relation* es un fenómeno perceptivo planteado por Repp (1982). Según el autor, el cambio de valores en una dimensión acústica se puede compensar por un cambio contrario en la otra para recuperar la percepción de los estímulos fonéticamente o fonológicamente ambiguos, simple y cuando los correlatos acústicos de estas dimensiones contribuyen a la identificación del contraste categorial. Dado que no existe una traducción establecida en español, en este trabajo, se mantiene el término en inglés para describir el fenómeno de compensación perceptiva entre los diferentes correlatos acústicos.

interfaz entre la pragmática, la entonación y el estudio de factores lingüísticos y extralingüísticos para entender mejor los problemas surgidos durante el proceso de enseñanza-aprendizaje de la entonación de ELE por los sinohablantes.

1.3 Hipótesis

Para este trabajo, se han formulado cinco hipótesis generales que se relacionan con el tema tratado en las publicaciones que componen la tesis. La primera, que se abordará en el capítulo dos (§2), tiene un carácter general y está relacionada con el marco teórico de la modalidad interrogativa que se desarrollará para poder comparar las diferencias en las estrategias de expresión de la interrogatividad en chino y español. La segunda, que es el objeto del capítulo tres (§3), predice la producción de la entonación de ELE a partir de diferentes puntos de vista. La tercera, alrededor de la cual se vertebra el capítulo cuatro (§4), concierne la existencia de un desarrollo general en la producción de las propiedades tonales y temporales del habla de la LE. La cuarta, alrededor de la cual se articula el anexo A, predice la percepción de categorías entonativas a partir del patrón entonativo presentado, la experiencia lingüística y las características individuales de los oyentes. La quinta, que se abordará en el capítulos cinco (§5) y el anexo B, determina las diferencias de percepción de los correlatos acústicos por parte de oyentes nativos y no nativos del español y las relaciona con unas variables individuales y lingüísticas. A continuación, se presenten las cinco hipótesis generales y las fuentes que las respaldan.

1.3.1 Hipótesis 1

La existencia de un sistema de clasificación que permita la comparación es el prerrequisito para poder contrastar las oraciones interrogativas del chino y del español. Sin embargo, los estudios previos que abordaron el tema de la modalidad interrogativa en cada una de esas dos lenguas parten de perspectivas lingüísticas muy diversas tanto en chino como en español. De ellas se pueden destacar las siguientes corrientes: 1) la gramática escolástica, que clasifica las oraciones interrogativas basándose en la caracterización sintáctico-prosódica de cada lengua (cf. Real Academia Española, 2009 para el español peninsular; Lü, 2002; Xing, 2002 para el chino mandarín); 2) la teoría de la relevancia (Sperber & Wilson, 1986; Wilson & Sperber, 2002), a partir de la cual se han dividido las interrogativas del chino y del español en dos categorías generales

dependiendo de la jerarquía interpretativa del contenido proposicional: preguntas auténticas (o estándares) y preguntas eco (cf. Escandell-Vidal, 1999, 2002 para el español peninsular; Chen & Wen, 2001, Shao, 2014 para el chino mandarín). Según esa perspectiva, las preguntas eco difieren de las preguntas estándares en su doble uso interpretativo, es decir, son interpretaciones de representaciones atribuidas (Escandell-Vidal, 2002); 3) el grado de interrogación, un criterio propio de la lingüística tradicional china que clasifica las interrogativas basándose en el grado de incerteza del hablante hacia el contenido proposicional (p.ej., Guo, 2000; Huang, 1986; Shao, 2012, 2014); 4) la gramática funcional, que partiendo de las funciones metalingüísticas del chino se propone una distinción general entre preguntas totales neutras, preguntas totales sesgadas y preguntas parciales (Halliday & McDonald, 2004); 5) la teoría de los actos de habla, a partir de la cual Haverkate (2006) plantea cuatro funciones ilocutivas (exhortativa, asertiva, comisiva y expresiva) de la modalidad interrogativa y clasifica las secuencias interrogativas del español según el acto ilocutivo.

Además de las cinco perspectivas principales, existen otros estudios que clasifican la modalidad interrogativa en función de la partícula modal utilizada en chino (Jin, 2011; Wang, 2016), el tipo de respuesta esperada (Fan, 1982; Shao, 2014) o el patrón entonativo de las preguntas (Font-Rotchés, 2005; Roseano et al., 2015). Sin embargo, ninguna clasificación de las que se han revisado ha demostrado ser aplicable de manera paralela y compatible a la modalidad interrogativa del chino y español. Por tanto, es necesario construir un marco teórico que permita relacionar distintas lenguas a partir de un mismo concepto interlingüístico.

Uno de los principios fundamentales del planteamiento del intercambio conversacional (Eggins, 2004; Fawcett, 2008; Halliday et al., 2014) y la teoría de los actos de habla (Searle, 1969) sugiere que las funciones comunicativas o pragmáticas del lenguaje sean interidiomáticas, es decir, sean las mismas entre distintas lenguas. Este aspecto nos lleva a proponer la primera hipótesis general de que, partiendo del papel discursivo del hablante, el tipo de producto que se intercambia en la conversión y el acto ilocutivo específico realizado por el hablante, se puede postular una taxonomía detallada de la modalidad interrogativa que sea válida para las preguntas tanto del chino como del español.

Como hipótesis derivada, se espera que el español pueda diferir del chino en las estrategias (p.ej., morfológicas, sintácticas, prosódicas o discursivas) empleadas para expresar una misma función pragmática o un determinado tipo de pregunta, debido a que los recursos lingüísticos utilizados en cada sistema lingüístico son distintos. En particular, el hecho de que el chino es una lengua tonal nos hace hipotetizar que la entonación sea una de las estrategias lingüísticas en las que más difieren las dos lenguas a la hora de codificar la modalidad interrogativa. Concretamente, las oraciones interrogativas del chino, lengua en que la modalidad oracional suele estar marcada mediante elementos léxicos y sintácticos, deberían presentar un menor grado de dependencia de la modulación de F0 que las del español.

1.3.2 Hipótesis 2

Uno de los mayores retos para los aprendices adultos de lenguas extranjeras es hablar con una entonación que suene como la que utilizan los nativos de la lengua meta. Las dificultades en la adquisición de la entonación se atribuyen no solamente a las diferencias de los sistemas prosódicos de la lengua materna y de la lengua meta, sino que se relacionan también con la complejidad y multidimensionalidad del aprendizaje de la entonación. En línea con el modelo de métrico-autosegmental (Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988), los estudios interlingüísticos suelen analizar la entonación de la LE desde el punto de vista de la representación fonológica y la implementación fonética (Mennen, 2004, 2007, Li, 2020). Además de esas dos dimensiones frecuentemente examinadas, Mennen (2015) propone que las dificultades de los aprendices en la producción entonativa de una L2 pueden encontrarse en el uso de elementos categóricos para expresar el significado semántico y la frecuencia de uso de estos elementos de la lengua meta. Partiendo de esa idea, se ha desarrollado el modelo de LILt (*L2 Intonation Learning Theory*) que reconoce cuatro dimensiones a lo largo de las cuales se pueden identificar de forma sistemática el comportamiento entonativo de los aprendices de una L2 (Mennen, 2015). Las cuatro dimensiones son:

1) Dimensión sistemática/fonológica: el inventario y la distribución de los elementos fonológicos categóricos.

2) Dimensión de realización/fonética: la implementación fonética de los elementos

categóricos.

3) Dimensión semántica: la funcionalidad de los elementos categóricos o contornos melódicos.

4) Dimensión de frecuencia: la frecuencia de uso del inventario de la lengua y la distribución de los primitivos de la entonación.

A partir de los principios del modelo de LILt (Mennen, 2015), se plantea la segunda hipótesis general de que la entonación del español producida por sinohablantes pueda desviarse de los patrones usados por hablantes nativos en varias dimensiones interrelacionadas. De esta hipótesis general se han derivado cinco hipótesis específicas que se agrupan según la dimensión del aprendizaje de la entonación que se investiga:

1) Desde el punto de vista fonológico, la entonación del español es un fenómeno postléxico y consiste en la combinación de distintos acentos tonales y tonos de frontera. Mientras en chino mandarín, el inventario tonal está definido en el nivel léxico y la modulación de F0 vehicula primariamente distinciones del significado léxico, por lo que la entonación propiamente dicha se realiza como una variación del rango de F0 que se sobrepone a elementos tonales categóricos. Estas diferencias tipológicas del sistema entonativo y elementos estructurales del chino y del español puedan influir en la producción de la entonación de ELE, dado que la adquisición del habla de una LE suele implicar cierto grado de interacción entre la lengua materna y la lengua meta (Graham & Post, 2018). En particular, de acuerdo con los resultados de estudios previos sobre la adquisición del inglés por parte de aprendices chinos (Barto, 2015; Hong, 2012; Ji et al., 2009, 2012; Xu, 2009), se espera que los errores de los sinohablantes sean explicables por las características de su L1. Más en concreto, esperamos que haya dificultades para realizar tonos bajos en posición nuclear, ya que el mandarín carece de un tono bajo estable en su inventario tonal.

2) Desde el punto de vista fonético, aunque diferentes lenguas han demostrado tener distintas propiedades tonales (p.ej., Hanley et al., 1966 para la comparación entre el inglés, el español y el japonés; Keating & Kuo, 2012, Yuan

& Liberman, 2014 para el mandarín y el inglés; Ng et al., 2010 para el cantonés y el inglés; Shi et al., 2014 para el mandarín y el japonés), puede existir un camino de desarrollo compartido entre los aprendices de diferentes comunidades lingüísticas en lo que concierne a la producción del rango tonal (*pitch range*). Los estudios de adquisición sugieren que los aprendices tienden a usar un rango tonal comprimido y/o un contorno de F0 menos variable en la L2 (p.ej., Backman, 1979 para aprendices españoles de inglés; Juffs, 1990, Yuan et al., 2018 para aprendices chinos de inglés; Zimmerer et al., 2014 para aprendices franceses de alemán y aprendices alemanes de francés). En vista de estas observaciones, parece lógico hipotetizar que los sinohablantes tengan una compresión del rango tonal y una reducción de la variabilidad de F0 (*pitch variation*) en comparación con los nativos de español. No obstante, en cuanto al nivel tonal (*pitch level*), es esperable que los sinohablantes utilicen una altura tonal más alta que los nativos debido a los efectos acústicos del aumento del esfuerzo cognitivo al hablar una LE (Peters, 2019) y la influencia de su L1 (la cual tiene una F0 media más alta que el inglés americano, el japonés y el español según Hanley et al., 1966; Keating & Kuo, 2012; y Shi et al., 2014).

3) Desde el punto de vista pragmático, se puede esperar que los sinohablantes tengan dificultades respecto al uso de patrones entonativos adecuados para la discriminación de la función informativa y confirmatoria de las preguntas totales del español, debido a que los dos tipos interrogativos se distinguen en su L1 mediante la partícula final (ma^{0} 吗 vs. ba^{0} 吧) en lugar de la modificación del contorno entonativo (Shang et al., 2021).

4) Desde la frecuencia del uso de elementos categóricos, dado que el ascenso final suele considerarse un marcador universal de la interrogación según el código de frecuencia (*Frequency Code*) (Ohala, 1994), se puede esperar una sobreproducción general de los tonos de frontera altos en las interrogativas del estudio en que los hablantes nativos usarían otro tipo de movimiento final en función del contexto comunicativo.

5) Además de identificar dónde se producen las desviaciones, el método de LILt

permite comparar los aprendices de L2 con diferentes perfiles, en diferentes estilos de habla o cualquier variable que pueda ser relevante en el proceso del aprendizaje (Mennen, 2015). Teniendo en cuenta esto, planteamos la última subhipótesis de que la adquisición de la entonación del español puede verse afectada por las características de los individuos y los materiales del habla. En concreto, esperamos que las desviaciones en las dimensiones citadas disminuyen a medida que aumente el nivel de competencia de los sinohablantes. Por otro lado, estudios previos demuestran que los hombres pueden diferir de las mujeres en el nivel tonal y la variedad tonal que se ha usado, debido no solo al factor fisiológico (Van Dommelen & Moxness, 1995), sino también a las diferencias en el estatus sociocultural (Urbani, 2012; Van Bezooijen, 1995) y la expresión de emociones (Brebner, 2003). Por tanto, también esperamos un efecto del género en la implementación de los rasgos tonales del español. Además, de acuerdo con las observaciones de estudios anteriores (Cortés Moreno, 2004; Yuan et al., 2019), es esperable que los diferentes tipos de pregunta impliquen diferentes niveles de dificultad en la adquisición de la entonación. Finalmente, como el acento llano es el patrón acentual no marcado y más frecuente en español que suele implicar menos dificultades en el procesamiento y aprendizaje (Defior & Serrano, 2017; Roca, 2019), se espera que los sinohablantes presenten mejores resultados del rango tonal en las oraciones finalizadas con una palabra llana que las acabas con una palabra aguda.

1.3.3 Hipótesis 3

Como se muestra en el apartado anterior (§1.3.2), los estudios interlingüísticos que examinan las propiedades fonéticas de la pronunciación de una LE suelen centrarse en la cuantificación del rango tonal a nivel oracional medido en Hz o semitonos. Por tanto, ha sido poco explorada la producción del rango tonal de los aprendices a nivel silábico. Uno de los pocos estudios que se han realizado sobre este tema encuentra que el rango tonal a nivel silábico presentado por aprendices chinos de inglés es más amplio que el de los hablantes nativos debido a una transferencia negativa de los tonos léxicos de la L1 (Ding et al., 2016). Puesto que en chino mandarín cada sílaba tiene un tono inherente, los sinohablantes pueden tener la tendencia a asignar un tono a la mayoría de las sílabas

de su inglés L2, lo que daría lugar a mayores fluctuaciones medias en el nivel silábico (Yuan et al., 2018).

Además de las desviaciones tonales, el habla de la LE se encuentra frecuentemente acompañado de una disminución de la fluidez oral. Por un lado, los estudios de Ding et al. (2016), Lee and Sidtis (2017) y Peters (2019) muestran que los aprendices de una LE tienen una reducción significativa en la velocidad de habla y/o la tasa de articulación en comparación con los hablantes nativos. Por otro, existe un debate sobre las características de la velocidad del cambio de F0 o tasa de cambio tonal (*pitch change rate*) en el habla de la LE, en particular cuando se investiga en situaciones interlingüísticas entre lenguas tonales y entonativas. Por ejemplo, el estudio de Yuan et al. (2018) reporta una velocidad del cambio de F0 más lenta para los sinohablantes de inglés L2, mientas que según Ding et al. (2016), los nativos de inglés y los aprendices chinos de inglés tienen una tasa de cambio tonal similar.

De acuerdo con las observaciones anteriores, planteamos una tercera hipótesis general: que los aprendices chinos se desvían de los hablantes nativos no solo en la producción de las propiedades tonales, sino también en la implementación de los perfiles temporales. En concreto, la primera subhipótesis formulada es que los aprendices chinos tienen un mayor rango en el nivel silábico que los nativos de español, aunque el rango y la variedad tonal de los sinohablantes serán menores en el nivel oracional. Además, es de esperar que los valores de los tres parámetros temporales (la velocidad de habla, la tasa de articulación y la velocidad del cambio de F0) producidos por los sinohablantes sean inferiores a los de hablantes nativos en las oraciones interrogativas del español.

1.3.4 Hipótesis 4

Los estudios de percepción realizados con oyentes nativos de español demuestran que los patrones entonativos en distintas posiciones dentro de una oración pueden contribuir –en diferentes medidas– a la correcta identificación de una categoría entonativa (Face, 2005, 2007, 2011; Sensui, 1995, 2003). Por ejemplo, el movimiento final ha demostrado ser el indicio más fuerte para distinguir las declarativas de las preguntas totales del español, a pesar de que los oyentes nativos son capaces de identificar correctamente una oración en el 95% de los casos escuchando solo el movimiento tonal inicial de la

oración (Face, 2007). En el caso de la percepción de la entonación de ELE, estudios previos demuestran que los oyentes no nativos perciben de una manera distinta los contornos entonativos de la lengua meta (Li, 2020; Zarate-Sandez et al., 2015), dado que, según la mayoría de las teorías de percepción, la identificación de los elementos segmentales y suprasegmentales de una LE puede estar implícitamente o explícitamente limitada por las estructuras fonológicas y las propiedades fonéticas de la L1 (Alexander & Wang, 2016; Best & Strange, 1992; Best & Tyler, 2007; So & Best, 2010).

Además de la influencia de la L1, la mayoría de los estudios interlingüísticos reconocen la existencia de una serie de variables que pueden incidir en las diferencias individuales a la hora de aprender y de realizar una evaluación subjetiva de la entonación de una LE (ver Barreiro Bilbao, 2002; Piske et al., 2001; Trimble, 2013a para un resumen de los factores de la adquisición de L2). Las variables más estudiadas en los estudios perceptivos centrados en las LE, además de la L1, pueden ser el nivel de competencia en la lengua meta (Archila-Suerte et al., 2010, 2012), la edad de adquisición de la misma (Baigorri et al., 2019; Richards et al., 2018), el género (Bishop et al., 2020), las propiedades psicométricas (Bishop et al., 2020; Jun & Bishop, 2015) y la experiencia musical (Alexander et al., 2005; Wiener & Bradley, 2020).

De acuerdo con estas observaciones, se puede plantear la cuarta hipótesis general de que la percepción de los patrones entonativos del español por parte de sinohablantes no solo está determinada por la L1 como se ha afirmado en el primer párrafo de este apartado, sino que está modulada por un conjunto de factores individuales que abarcan el nivel de competencia, la habilidad pragmática y la experiencia musical de los aprendices. A continuación, se exponen las dos hipótesis específicas derivadas:

1) En primer lugar, se hipotetiza que los sinohablantes tienen una exactitud perceptiva inferior que los oyentes nativos del español en la identificación de los patrones entonativos que no se presenten en el inventario tonal de su L1. Sin embargo, las investigaciones anteriores indican que tener una gran experiencia con una lengua tonal puede fomentar la sensibilidad auditiva de los oyentes hacia el contraste de los eventos tonales de una LE (Chang et al., 2017; Hallé et al., 2004; Ortega-Llebaria et al., 2017), por tanto, es de esperar que los sinohablantes se comporten mejor que los aprendices provenientes de una lengua

no tonal documentados en los estudios previos paralelos (p.ej., Trimble, 2013b).

2) En segundo lugar, se espera que los aprendices sinohablantes con un alto nivel de competencia, alta habilidad pragmática y experiencia musical tengan mejor rendimiento en la identificación de las categorías entonativas de la lengua meta.

1.3.5 Hipótesis 5

Los principios básicos de que parte esta hipótesis es que la percepción de cualquier contraste fonético o fonológico del habla abarca informaciones acústicas en múltiples dimensiones (Toscano & McMurray, 2010). En el área de la percepción de la entonación, por ejemplo, se ha observado una integración de diferentes parámetros acústicos (F0, duración e intensidad) en la percepción por parte de hablantes nativos de inglés (Feng et al., 2019; Peng et al., 2012), de chino mandarín (Yuan, 2006), de cantonés (Ma et al., 2008, 2011), de alemán (Niebuhr, 2007) y de coreano (Chang, 2013). Aunque los oyentes pueden aprovechar cada uno de los parámetros acústicos que se han citado para reconocer una categoría entonativa, la contribución o el peso perceptivo de dichos parámetros acústicos es diferente (Ballesteros, 2014; Holt et al., 2018; Kuang & Cui, 2018). Para la mayoría de las lenguas del mundo, la interrogación se percibe prioritariamente gracias a la modulación de la F0, que a manudo se manifiesta en las lenguas entonativas como una subida final y en las lenguas tonales como una expansión del rango tonal (Bolinger, 1978).

Además, varios estudios interlingüísticos indican que los hablantes provenientes de diferentes comunidades lingüísticas pueden percibir de forma distinta los correlatos acústicos de un contraste segmental o suprasegmental. Por ejemplo, se ha encontrado que la percepción de las declarativas e interrogativas del inglés por parte de oyentes nativos resulta afectada de forma significativa por cambios en la duración y la intensidad de la palabra nuclear, mientras que los aprendices chinos de inglés no son sensibles a las variaciones de estas propiedades acústicas (Feng et al., 2019).

Por tanto, en el caso que nos ocupa, parece legítimo hipotetizar en términos generales que los sinohablantes y los hablantes nativos utilicen de una manera distinta los tres correlatos acústicos (F0, duración e intensidad) para percibir la modalidad declarativa y

la interrogativa del español. Concretamente, partiendo de los hallazgos anteriores y la característica prosódica del chino y del español, se formulan cuatro hipótesis concretas en este apartado:

1) La primera subhipótesis planteada es que los oyentes nativos son más sensibles a las modulaciones de la duración y la intensidad en la parte final de la oración que los aprendices chinos del español.

2) En cuanto a la percepción de la F0, las investigaciones recientes muestran que los oyentes de una lengua tonal y no tonal comparten el mismo mecanismo cerebral bilateral para el procesamiento de las informaciones tonales, a pesar de la realización específica entonativa de la L1(Chien et al., 2020). En vista de este resultado, es esperable que los aprendices chinos y los nativos de español tengan una sensibilidad comparable a las modulaciones de F0 procesadas como entonación en este estudio.

3) Partiendo de la teoría de la *phonetic trading relation* (Repp, 1982), también esperamos que los oyentes nativos y no nativos puedan aprovechar –en diferentes medidas– la compensación perceptiva entre los diferentes correlatos acústicos para llegar al reconocimiento de la modalidad oracional.

4) Por último, varios estudios perceptivos muestran que las características individuales, como la edad y el género, pueden afectar a la identificación del contraste fonético y fonológico a través de la afinación de la capacidad auditiva y la habilidad pragmática de los oyentes (Bishop et al., 2020; Patel & Grigos, 2006; Ryalls et al., 1994). Por tanto, se hipotetiza que la percepción de la declarativa y la interrogativa mediante los tres correlatos acústicos citados puede estar influenciada por variables individuales como la edad y el género, además de la propia L1 de los oyentes.

1.4 Estructura de la tesis

Esta tesis se presenta en formato de compendio de publicaciones. En el primero de los seis capítulos principales se realiza una introducción global acerca del objeto de estudio, su motivación, sus objetivos y su estructura. En el capítulo 1 están incluidas también las

hipótesis generales de este estudio y una breve reseña de los estudios previos en que nos hemos basado para formular dichas hipótesis.

Los capítulos de 2 a 5 y los anexos A y B incluyen las publicaciones originales que constituyen la parte esencial del cuerpo de la tesis. Las seis publicaciones aportadas conforman una línea de trabajo de investigación bien definido y orientado, en términos generales, a la caracterización de la expresión de la modalidad interrogativa y la adquisición de la entonación interrogativa del español por parte de sinohablantes.

De las seis publicaciones que constituyen el cuerpo de la tesis, dos han sido publicadas (§2 y §3), dos están aceptadas para su publicación (§4 y §5), y dos están en proceso de revisión, por lo que se presentan en el anexo (en particular, en anexos A y B) y no forman parte del cuerpo principal de la tesis. A continuación, se presenta la lista de publicaciones (Tabla 1.1) que conforman la tesis doctoral y se recapitulan la temática y las conclusiones generales de cada una de las publicaciones.

Tabla 1.1. Lista de publicaciones derivadas de la tesis doctoral.

Capítulo 2	Shang, P., Elvira-García, W., & Roseano, P. (2021). La modalidad interrogativa en español y en chino: un enfoque funcionalista. *Círculo de Lingüística Aplicada a la Comunicación*, *88*, 235-255. doi: https://doi.org/10.5209/clac.78313
Capítulo 3	Shang, P., & Elvira-García, W. (2022). Second language acquisition of Spanish prosody by Chinese speakers: Nuclear contours and pitch characteristics. *Vigo International Journal of Applied Linguistics*, (19), 129-176. doi: https://doi.org/10.35869/vial.v0i19.3762
Capítulo 4	Shang, P. (en prensa). Cross-Linguistic Comparison of the Pitch and Temporal Profiles between L1 and Chinese L2 Speakers of Spanish. *Loquens*.
Capítulo 5	Shang, P., Elvira-García, W. & Li, X. (en prensa). Cue weighting differences in perception of Spanish sentence types between native listeners of Chinese and Spanish. *Proceedings of the 11th International Conference on Speech Prosody*. Lisboa, Portugal.
Anexo A	Shang, P., Roseano, P. & Elvira-García, W. (en revisión). Perception

	of Spanish statements and questions by native Chinese speakers: Effect of language experience and individual features.
Anexo B	Shang, P., Roseano, P. & Elvira-García, W. (en revisión). Dynamic cue weighting in the perception of Spanish intonation contrasts: Differences between tonal and non-tonal language listeners.

Capítulo 2: ***La modalidad interrogativa en español y en chino: Un enfoque funcionalista.*** Este artículo parte de una visión funcional-pragmática para plantear una nueva taxonomía de clasificación de la modalidad interrogativa que sea aplicable al chino y español y, asimismo, llevar a cabo un análisis comparativo de los recursos lingüísticos usados por las dos lenguas para expresar cada tipo interrogativo con una función pragmática específica. Los resultados arrojan las siguientes conclusiones:

1) La modalidad interrogativa de las dos lenguas no desempeña solo la función prototípica de pedir información, sino que sirve para finalidades pragmáticas muy diversas que incluyen la función de confirmar hipótesis, formular ruegos u órdenes, ofrecer bienes o sugerencias, expresar sorpresas y etc.

2) Desde el punto de vista de la realización lingüística, el modo verbal, la entonación y el orden sintáctico son los elementos lingüísticos más usados para marcar la interrogatividad en español. En chino, sin embargo, el carácter analítico y tonal de la lengua hace que las preguntas se expresen en su mayoría mediante partículas modales y construcciones sintácticas, de modo que la modulación de F0 desempeña un papel menor.

Esta diferencia en los recursos lingüísticos empleados para expresar la interrogatividad, junto con las conocidas diferencias tipológicas entre lenguas tonales y entonativas, plantean grandes desafíos para los sinohablantes a la hora de aprender la entonación interrogativa del español. En vista de esto, se ha decidido indagar en lo siguiente las dificultades de los sinohablantes en el proceso de adquisición de la entonación del español como lengua extranjera.

Capítulo 3: ***Second language acquisition of Spanish prosody by Chinese speakers:***

Nuclear contours and pitch characteristics. El capítulo 3, se centra en la producción de la entonación del español por parte de los sinohablantes en cinco tipos de pregunta. El objetivo es describir las producciones entonativas de los sinohablantes en comparación con las de hablantes nativos a lo largo de la dimensión fonológica, fonética, pragmática, y de frecuencia. Asimismo, se pretende analizar los efectos de las variables lingüísticas (la posición acentual y el tipo de pregunta) y extralingüísticas (L1, el nivel de competencia y el género de los hablantes) en la producción de la entonación. Mediante un análisis que se enmarca en el modelo métrico-autosegmental de la entonación, se han descrito los patrones nucleares de 555 preguntas producidas en castellano por informantes chinos y españoles. Además, se han caracterizado las propiedades tonales de estas oraciones a través de seis métricas distintas. De los resultados obtenidos se extraen las siguientes conclusiones:

1) El comportamiento entonativo de los sinohablantes depende, en una buena medida, de los parecidos y las diferencias prosódicas entre la lengua materna y la lengua meta. En concreto, los buenos resultados de los aprendices chinos en la producción de las preguntas disyuntivas y de las preguntas de apéndice se puede atribuir a una transferencia positiva de su L1. En cambio, el error más destacado –que consiste en reemplazar los acentos nucleares bajos/descendentes por tonos altos/ascendentes– se explica por la falta de un tono bajo estable en el inventario tonal del chino. Otra característica frecuente en el habla de los sinohablantes es la sobreproducción de los tonos de frontera altos al final de aquellos tipos de preguntas donde, de acuerdo con el modelo proporcionado por los hablantes nativos, se esperaría un movimiento tonal descendente. Además, se observa que los sinohablantes tienen serias dificultades en el uso de patrones entonativos adecuados para distinguir la función informativa y confirmatoria de las preguntas totales del español.

2) Los aprendices sinohablantes cuentan con un nivel tonal más alto (tanto de F0 media como de F0 mínima), un rango oracional más comprimido y un contorno de F0 menos variable que los hablantes nativos de español.

3) Los sinohablantes son capaces de producir patrones entonativos y perfiles tonales más cercanos al objetivo a medida que aumenta su nivel de competencia

del español. En cuanto al efecto de género, los resultados son consistentes con estudios anteriores (Urbani, 2012; Van Bezooijen, 1995) mostrando que las mujeres tienen un nivel tonal más elevado y una mayor variedad tonal que los hombres en el habla. Por otro lado, los resultados de diferentes tipos de pregunta sugieren, en síntesis, una jerarquía de dificultad del aprendizaje de la entonación de las oraciones del español. Además de la influencia del sistema prosódico de la L1 y las características individuales, se ha encontrado que el tipo acentual de la palabra nuclear también puede afectar a la implementación del rango tonal de las oraciones de la lengua meta.

Capítulo 4: ***Cross-linguistic comparison of the pitch and temporal profiles between L1 and Chinese L2*** [③] ***speakers of Spanish.*** Este capítulo, como extensión de la publicación anterior (§3), investiga sobre todo el rango tonal a nivel silábico y oracional (medido en ERB, en inglés *Equivalent Rectangular Bandwidth*) y la caracterización temporal de las oraciones producidas por los aprendices chinos de ELE. El corpus y los informantes que conforman este capítulo son idénticos que en el capítulo 3. Mediante el cálculo de seis parámetros tonales y temporales, se han revelado importantes diferencias de producción tonal y temporal entre los hablantes nativos y no nativos. En concreto, de los resultados obtenidos se extraen las siguientes conclusiones:

1) El rango tonal oracional cuantificado en ERB coincide con los resultados calculados en semitono, es decir, los sinohablantes tienen una compresión significativa del rango oracional en comparación con los hablantes nativos.

2) Contrariamente a lo esperado, los sinohablantes presentan un rango tonal más reducido que los hablantes nativos en el nivel silábico. Una explicación de este resultado es que los sinohablantes son demasiado cautelosos con la articulación de los sonidos del español, de modo que no les queda suficiente tiempo de planificación para alcanzar los objetivos tonales y producen con un contorno relativamente plano de F0.

[③] A pesar de que los informantes chinos de este estudio pueden empezar a estudiar el inglés a una edad más temprana que otras lenguas extranjeras, el español siempre ha sido la segunda lengua extranjera dominante por ellos. Por tanto, se ha utilizado en algunos capítulos (en particular, §3 y §4) la expresión de aprendices chinos de español L2 para referirse a los sinohablantes cuya segunda lengua dominante es el español en vez de la segunda lengua estudiada según el orden temporal.

3) Los resultados temporales son coherentes con los estudios previos del habla de lenguas extranjeras, en el sentido de que los sinohablantes tienen una reducción significativa en la velocidad de habla, la tasa de articulación y la velocidad del cambio de F0 en comparación con los hablantes nativos de español.

4) Por último, se documenta que, en la mayoría de las preguntas, los perfiles tonales y temporales de los sinohablantes con un nivel avanzado de español se acercan más a los de hablantes nativos, excepto en el caso de preguntas parciales. Además, concordando con los resultados del capítulo 3, encontramos un efecto del tipo acentual y del género en la variedad tonal evaluada mediante el parámetro de PDQ (*Pitch Dynamism Quotient*).

Capítulo 5: ***Cue weighting differences in perception of Spanish sentence types between native listeners of Chinese and Spanish.*** El capítulo 5 está constituido por un artículo de congreso internacional en que se propone un enfoque alternativo para examinar la percepción acústica de la entonación del español por oyentes chinos y españoles. Las cinco categorías de respuesta auditiva (declarativa, más declarativa que interrogativa, declarativa o interrogativa, más interrogativa que declarativa e interrogativa) se codifican como 0, 0,25, 0,5, 0,75 y 1 según el grado de interrogatividad. Los valores recodificados se interpretan como la probabilidad de 0 a 1 que un enunciado se perciba como una interrogativa. Este estudio se implementa el modelo sigmoidal y el algoritmo de Levenberg-Marquardt[④] para extraer dos parámetros: 1) x_0, el umbral entre la declarativa y la interrogativa dónde la probabilidad perceptiva de cada una es 50%; 2) b, el parámetro relacionado con la pendiente de la curva de identificación. Los resultados obtenidos coinciden con los presentados en el anexo A, demostrando que la percepción de la modalidad oracional es un proceso dinámico en que los elementos acústicos de múltiples dimensiones se integran para conseguir una estimación precisa de la categoría entonativa. En concreto, se han extraído las siguientes conclusiones:

1) Los oyentes nativos de español se diferencian de los oyentes chinos por presentar un umbral entre declarativa e interrogativa significativamente más alto,

[④] El algoritmo Levenberg-Marquardt, conocido como LMA o simplemente LM en matemáticas o computación, combina dos métodos de minimización y se utiliza comúnmente para resolver problemas de mínimos cuadrados no lineales (ajuste de curva) y la estimación de parámetros de un modelo.

es decir, que necesitan un movimiento final de F0 más elevado para percibir una oración como interrogativa.

2) La curva de identificación más abrupta encontrada para los oyentes españoles en las diferentes condiciones acústicas sugiere, en conjunto, que estos tienen una mayor sensibilidad a las transiciones lineales de F0 percibidas como entonación que los oyentes chinos.

3) Los correlatos acústicos secundarios como la duración y la intensidad también tienen un efecto significativo en el reconocimiento de la modalidad oracional por parte de los oyentes españoles. En cambio, los oyentes chinos solo son sensibles a las modulaciones de la duración (aparte de la F0) y en menor medida que los oyentes españoles.

4) Los oyentes pueden necesitar un ascenso final de F0 más acusado para la identificación de la interrogatividad cuando el valor de la duración o la intensidad de la sílaba final reduce, y viceversa. La capacidad de los oyentes para compensar las variaciones acústicas de la entonación está relacionada con su sensibilidad a los correlatos acústicos que sean relevantes para el reconocimiento del contraste categorial.

Anexo A: ***Perception of Spanish statements and questions by native Chinese speakers: Effect of language experience and individual features.*** El artículo correspondiente al anexo A examina cómo los oyentes chinos y españoles perciben los patrones entonativos de las declarativas e interrogativas del español, y cómo las variables lingüísticas e individuales influyen en la evaluación de la categoría entonativa. Los estímulos se crean basando en el método de *gating*[⑤] (Face, 2007, 2011) que consiste en segmentar la secuencia melódica de cuatro pares de declarativas e interrogativas en varios fragmentos (o *gates*) que contienen distintos eventos entonativos. El test de percepción se realiza con una encuesta en línea para recopilar los datos sociolingüísticos y las respuestas

[⑤] El método de *gating* se emplea con frecuencia en los estudios de percepción para investigar en qué posición oracional los oyentes consiguen identificar o discriminar la modalidad oracional. La regla básica de este consiste en dividir, según el objetivo de investigación, el contorno melódico del enunciado en varios segmentos (conocidos como *gates* en inglés) que coinciden con los eventos entonativos de la lengua estudiada. En esta tesis, se mantienen los dos términos (*gating* y *gate*) en inglés porque no existe una terminología bien asentada en la literatura escrita en español.

auditivas, así como evaluar la habilidad pragmática de los informantes chinos usando el Cociente de Espectro Autista (*Autism Spectrum Quotient*, en adelante AQ por sus siglas en inglés) para los adultos (Baron-Cohen et al., 2001; Lau et al., 2013). El análisis de los resultados arroja las siguientes conclusiones:

1) Los oyentes nativos y no nativos se diferencian en la capacidad de usar los distintos eventos entonativos para discriminar la modalidad oracional del español. En concreto, los oyentes chinos se diferencian significativamente de los nativos de español en la percepción del pico inicial y el movimiento tonal en la sílaba nuclear de la oración.

2) El 67% de los oyentes chinos son capaces de identificar la modalidad oracional solo después del pico inicial, mientras que según Trimble (2013b) la mayoría de los aprendices estadounidenses de español no pueden distinguir la declarativa de la interrogativa hasta el movimiento final. El mejor rendimiento en este sentido de los sinohablantes se atribuiría a su haber estado expuestos a una lengua tonal que les ayuda a discriminar los eventos tonales contrastivos en una LE.

3) Además de la influencia de la L1, la percepción del contraste entonativo puede estar condicionada por las características individuales de los aprendices y por las propiedades lingüísticas de los estímulos. Los oyentes chinos con un nivel alto de competencia, alta habilidad pragmática y experiencia musical presentan mayor tasa de éxito en la identificación de la modalidad oracional. Asimismo, las oraciones formadas por palabras llanas son más fácilmente percibidas que las compuestas por palabras agudas.

Anexo B: ***Dynamic cue weighting in the perception of Spanish intonation contrasts: Differences between tonal and non-tonal language listeners.*** El objetivo de este artículo es investigar: 1) cómo los oyentes chinos y españoles usan cambios en las tres dimensiones acústicas (F0, duración e intensidad) para percibir el contraste entonativo entre declarativas e interrogativas del español; 2) cómo se integran e interactúan los distintos correlatos acústicos en el proceso perceptivo; y 3) hasta qué punto la percepción de los oyentes puede estar afectada por la L1, la edad y el género. Para estas finalidades, se realizan dos test perceptivos en que los estímulos se crean a través de la

manipulación de la curva de F0 junto con la duración o la intensidad de la sílaba final del enunciado. A partir de los resultados obtenidos, concluimos lo siguiente:

1) La percepción de declarativas e interrogativas por parte de los oyentes nativos de español está significativamente afectada por cambios en los tres correlatos acústicos citados, mientras que los oyentes chinos son sensibles solo a las modulaciones de la F0 y la duración. Además, se observa que los oyentes españoles tienen mayor sensibilidad a las variaciones de F0 y duración que los oyentes chinos. Sin embargo, los oyentes españoles necesitan un movimiento final de F0 más elevado que los oyentes chinos para la percepción de la interrogativa, posiblemente debido al triple contraste de los tonos de frontera (L%, !H%, H%) al final de las oraciones del español peninsular.

2) La duración e intensidad interactúan de forma significativa con la F0 en el proceso perceptivo. La reducción de uno de los correlatos acústicos secundarios (es decir, la duración y la intensidad) puede compensarse aumentando el valor del correlato acústico primario (F0) para que los oyentes puedan recuperar la percepción de la modalidad oracional deseada.

3) Los oyentes femeninos o de mayor edad son más propensos a percibir un enunciado como interrogativa total. Los enunciados formados por una palabra llana son más probables que se perciban como interrogativas que los compuestos por una palabra aguda en la misma condición acústica.

En el capítulo 6, dedicado a las conclusiones, se presenta en primer lugar una recapitulación de los resultados de las seis publicaciones, exponiendo de forma global las aportaciones e implicaciones para el campo del aprendizaje de lenguas extranjeras. Además, se desarrollan unas reflexiones sobre las limitaciones del estudio y las futuras líneas de investigación dentro del campo de la adquisición de la entonación de ELE.

Por último, en el anexo C, se hacen constar el documento de autorización, el cuestionario sociolingüístico, la encuesta perceptiva, el corpus y los contextos usados para la elicitación. También están incluidos en este apartado los datos de los informantes y la lista de los estímulos auditivos creados.

1.5 Referencias del Capítulo 1

Alexander, J. A., Wong, P. C. M., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. In *9th European Conference on Speech Communication and Technology* (pp. 397–400). Lisboa, Portugal.

Alexander, J., & Wang, D. Y. (2016). The perceptual assimilation model for suprasegmentals and cross-language lexical-tone identification. *The Journal of the Acoustical Society of America*, *139*(4), 2161.

Alharbi, A. (2020). *First language attrition and second language acquisition: exploring the role of phonetic aptitude and language use in highly proficient late Arabic-English and English-Arabic bilinguals* (Doctoral dissertation). Bangor University.

Archila-Suerte, P., Zevin, J., Bunta, F., & Hernandez, A. E. (2012). Age of acquisition and proficiency in a second language independently influence the perception of non-native speech. *Bilingualism: Language and Cognition*, *15*(1), 190–201.

Archila-Suerte, P., Zevin, J., & Hernandez, A. E. (2015). The effect of age of acquisition, socioeducational status, and proficiency on the neural processing of second language speech sounds. *Brain and Language*, *141*, 35–49.

Archila, P., Ramos, A. I., Zevin, J., & Hernandez, A. E. (2010). How age of acquisition and proficiency predicts sound detection in bilinguals. *Poster presented at Armadillo Conference, College Station, TX*.

Backman, N. (1979). Intonation Errors in Second-Language Pronunciation of Eight Spanish-Speaking Adults Learning English. *Interlanguage Studies Bulletin*, *4*(2), 239–265.

Baigorri, M., Campanelli, L., & Levy, E. S. (2019). Perception of American – English Vowels by Early and Late Spanish – English Bilinguals. *Language and Speech*, *62*(4), 681–700.

Baills, F., Suárez-González, N., González-Fuente, S., & Prieto, P. (2019). Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition*, *41*(1), 33–58.

Baills, F., Zhang, Y., Cheng, Y., Bu, Y., & Prieto, P. (2021). Listening to Songs and Singing Benefitted Initial Stages of Second Language Pronunciation but Not Recall of Word Meaning. *Language Learning*, *71*(2), 369–413.

Ballesteros, S. (2014). La atención selectiva modula el procesamiento de la información y la memoria implícita. *Acción Psicológica*, *11*(1), 7–20.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). Tone Language Speakers and Musicians Share Enhanced Perceptual and Cognitive Abilities for Musical Pitch: Evidence for Bidirectionality between the Domains of Language. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17.

Barto, K. A. (2015). *Mandarin Speakers' Intonation in their L2 English* (Doctoral dissertation). The University of Arizona.

Belchí, D. A. I. (1994). Estrategias comunicativas: la pregunta retórica en español. *Revista Española de Lingüística*, *24*(2), 329–344.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*(3), 305–330.

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception. In O. S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honour of James Emil Flege* (pp. 13–34). John Benjamins.

Bishop, J., Kuo, G., & Kim, B. (2020). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics*, *82*, 100977.

Bolinger, D. (1978). Intonation across languages. In J. Greenberg (Ed.), *Universals of Human Language Volume 2: Phonology* (pp. 471–524). Stanford University Press.

Brebner, J. (2003). Gender and emotions. *Personality and Individual Differences*, *34*(3), 387–394.

Chang, Y. S., Yao, Y., & Huang, B. H. (2017). Effects of linguistic experience on the perception of high-variability non-native tones. *The Journal of the Acoustical Society of America*, *141*(2), EL120–EL126.

Chen, Z., & Wen, X. (2001). On Echoic Utterances in Verbal Comunication. *Journal of PLA University of Foreign Language*, *24*(4), 24–17.

Chien, P. J., Friederici, A. D., Hartwigsen, G., & Sammler, D. (2020). Neural correlates of intonation and lexical tone in tonal and non-tonal language speakers. *Human Brain Mapping*, *41*(7), 1842–1858.

Choi, W. (2022). Theorizing positive transfer in cross-linguistic speech perception: The Acoustic-Attentional-Contextual hypothesis. *Journal of Phonetics*, *91*, 101135.

Consejo de Europa. (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Ministerio de Educación Cultura y Deporte.

Cortés Moreno, M. (2004). Análisis acústico de la producción de la entonación española por parte de sinohablantes. *Estudios de Fonética Experimental*, *13*, 80–110.

Defior, S., & Serrano, F. (2017). Learning to Read Spanish. In L. Verhoven & C. Perfetti (Eds.), *Learning to Read across Languages and Writing Systems*(pp. 243–269). Cambridge University Press.

Ding, H., Hoffmann, R., & Hirst, D. (2016). Prosodic transfer: A comparison study of f0 patterns in L2 english by chinese speakers. In *8th International Conference on Speech Prosody* (pp. 756–760). Boston, US.

Divenyi, P., Greenberg, S., & Meyer, G. (2006). *Dynamics of speech production and perception* (Vol. 374). Ios Press.

Eggins, S. (2004). *Introduction to systemic functional linguistics*. A&C Black.

Escandell-Vidal, M. V. (1999). Los enunciados interrogativos. Aspectos semánticos y pragmáticos. In V. Demonte & I. Bosque (Eds.), *Gramática Descriptiva de La Lengua Española: Entre la oración y el discurso. Morfología* (pp. 3929-3992). Espasa Libros, S.L.

Escandell-Vidal, M. V. (2002). Echo-syntax and metarepresentations. *Lingua*, *112*(11), 871–900.

Face, T. L. (2005). F0 Peak Height and the Perception of Sentence Type in Castilian Spanish. *Revista Internacional de Lingüística Iberoamericana*, *3*(2), 49–65.

Face, T. L. (2007). The role of intonational cues in the perception of declaratives and absolute interrogatives in Castilian Spanish. *Estudios de Fonética Experimental*, *16*, 186–225.

Face, T. L. (2011). *Perception of Castilian Spanish Intonation: Implications for Intonational Phonology*. Lincom Europa.

Fan, J. (1982). The Syntactic Structure of Yes-no Questions. *Studies of the Chinese Language*, *6*, 426–434.

Fawcett, R. (2008). *Invitation to systemic functional linguistics through the Cardiff grammar: An extension and simplification of Halliday's systemic functional grammar*. Equinox.

Feng, J., Tao, S., Wu, X., Alsbury, K., & Liu, C. (2019). The effects of amplitude and duration on the perception of English statements vs questions for native English and Chinese listeners. *The Journal of the Acoustical Society of America*, *145*(5), EL449–EL455.

Font-Rotchés, D. (2005). *L'entonació del català. Patrons melòdics, tonemes i marges de dispersió* (Tesis doctoral). Didàctica de la Llengua i la Literatura, Universitat de Barcelona.

Gabriel, C., & Kireva, E. (2014). Prosodic transfer in learner and contact varieties: Speech rhythm and intonation of Buenos Aires Spanish and L2 Castilian Spanish produced by Italian native speakers. *Studies in Second Language Acquisition*, *36*(2), 257–281.

Gass, S. M. (1988). Second language acquisition and linguistic theory: The role of language transfer. In S. Flynn & W. O'Neil (Eds.), *Linguistic theory in second language acquisition* (pp. 384–403). Springer.

Graham, C., & Post, B. (2018). Second language acquisition of intonation: Peak alignment in American English. *Journal of Phonetics*, *66*, 1–14.

Guo, R. (2000). The Certainty and Forms of Answer in "Ma" Questions. *Chinese Teaching in the World*, *2*, 13–23.

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421.

Halliday, M. A. K., Matthiessen, C., & Halliday, M. (2014). *An introduction to functional grammar* (3rd ed.). Routledge.

Halliday, M. A. K., & McDonald, E. (2004). Metafunctional profile of the grammar of Chinese. *Language Typology: A Functional Perspective*, *253*, 305–396.

Hanley, T. D., Snidecor, J. C., & Ringel, R. L. (1966). Some acoustic differences among languages. *Phonetica*, *14*(2), 97–107.

Haverkate, H. (2006). Aspectos pragmalingüísticos de la interrogación en español con atención especial a las secuencias de preguntas. *Cultura, Lenguaje y Representación*, *III*, 27–40.

Holt, L. L., Tierney, A. T., Guerra, G., Laffere, A., & Dick, F. (2018). Dimension-selective attention as a possible driver of dynamic, context-dependent re-weighting in speech processing. *Hearing Research*, *366*, 50–64.

Hong, W. (2012). *An experimental study of Chinese students' English intonation pattern*. Nankai University.

Huang, G. (1986). A Primary Study of the Particle "Ma". *Studies in Language and Linguistics]*, *2*, 131–135.

Ji, X., Wang, X., & Li, A. (2009). Intonation patterns of yes-no questions for Chinese EFL learners. In *2009 Oriental COCOSDA International Conference on Speech Database and Assessments* (pp. 88–93). Urumqi, China

Ji, X., Xia, W., Yuan, J., & Li, A. (2012). Intonation patterns of wh-questions from Chinese EFL learners. In *6th International Conference on Speech Prosody* (pp. 741–744). Shanghai, China.

Jilka, M. (2000). *The contribution of intonation to the perception of foreign accent* (Doctoral dissertation). Universitat Stuttgart.

Jin, Z. (2011). *Study of the Meaning of Modal Particles in Modern Chinese*. Fudan University.

Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *IRAL-International Review of Applied Linguistics in Language Teaching*, *28*(2), 99–118.

Keating, P., & Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *The Journal of the Acoustical Society of America*, *132*(2), 1050–1060.

Kuang, J., & Cui, A. (2018). Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics*, *71*, 194–214.

Lau, W. Y. P., Gau, S. S. F., Chiu, Y. N., Wu, Y. Y., Chou, W. J., Liu, S. K., & Chou, M. C. (2013). Psychometric properties of the Chinese version of the Autism Spectrum Quotient (AQ). *Research in Developmental Disabilities*, *34*(1), 294–305.

Lee, B., & Sidtis, D. V. L. (2017). The bilingual voice: Vocal characteristics when speaking two languages across speech tasks. *Speech, Language and Hearing*, *20*(3), 174–185.

Lü, S. (2002). *Zhong'guo Wenfa Yaolüe.* Liaoning Education Press.

Ma, J. K. Y., Ciocca, V., & Whitehill, T. L. (2008). Acoustic cues for the perception of intonation in Cantonese. In *9th Annual Conference of the International Speech Communication Association* (pp. 520–523). Brisbane, Australia.

Mennen, I. (2015). Beyond segments: towards an L2 intonation learning theory (LILt). In E. Delais-Roussarie, M. Avanzi, & S.Herment (Eds.), *Prosody and language in contact* (pp. 171–188). Springer.

Ng, M. L., Hsueh, G., Leung, S., & Shing, C. (2010). Voice pitch characteristics of Cantonese and English produced by Cantonese-English bilingual children. *International Journal of Speech-Language Pathology*, *12*(3), 230–236.

Nguyễn, T. A.-T., Ingram, C. L. J., & Pensalfini, J. R. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics*, *36*(1), 158–190.

Niebuhr, O. (2007). Categorical perception in intonation: A matter of signal dynamics? In *8th Annual Conference of the International Speech Communication Association*, *1* (pp. 642–645). Antwerp, Belgium.

Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism*. Cambridge University Press.

Ortega-Llebaria, M., Nemogá, M., & Presson, N. (2017). Long-term experience with a tonal language shapes the perception of intonation in English words: How Chinese-English bilinguals perceive "Rose?" vs. "Rose." *Bilingualism*, *20*(2), 367–383.

Patel, R., & Grigos, M. I. (2006). Acoustic characterization of the question–statement contrast in 4, 7 and 11 year-old children. *Speech Communication*, *48*(10), 1308–1318.

Peng, S. C., Chatterjee, M., & Lu, N. (2012). Acoustic Cue Integration in Speech Intonation Recognition With Cochlear Implants. *Trends in Amplification*, *16*(2), 67–82.

Peters, J. (2019). Fluency and speaking fundamental frequency in bilingual speakers of High and Low German. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 1–5). Melbourne, Australia.

Pierrehumbert, J. (1980). The phonology and phonetics of English intonation. *RELC Journal*, *15*(1), 117–118.

Pierrehumbert, J., & Beckman, M. (1988). Japanese tone structure. *Linguistic Inquiry Monographs*, *15*, 1–282.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*(2), 191–215.

Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux Cahiers de Linguistique Française*, *28*(2007), 41–66.

Real Academia Española. (2009). *Nueva Gramática de la lengua Española: Sintaxis II*. Espasa Libros, S.L.U.

Repp, B. H. (1982). Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*(1), 81–110.

Richards et al. (2018). Age of acquisition and proficiency in a second language independently influence the perception of non-native speech. *Physiology & Behavior*, *176*(5), 139–148.

Roca, I. (2019). Spanish Word Stress: an updated multidimensional account. In R., Heinz, J. Goedemans, & H. van der Hulst (Eds.), *The Study of Word Stress and Accent: Theories, Methods and Data* (pp. 256–292). Cambridge University Press.

Roseano, P., Fernández Planas, A. M., Elvira-García, W., Massó, R. C., & Celdrán, E. M. (2015). La entonación de las preguntas parciales en catalán. *Revista Española*

de Lingüística Aplicada, *28*(2), 511–554.

Ryalls, J., Dorze, G. Le, Lever, N., Ouellet, L., & Larfeuil, C. (1994). The effects of age and sex on speech intonation and duration for matched statements and questions in French. *Journal of the Acoustical Society of America*, *95*(4), 2274–2276.

Santa-Cecilia, Á. G. (2007). Plan curricular del instituto cervantes: niveles de referencia para el espanol. *MarcoELE: Revista de Didáctica*, *5*, 1–10.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge university press.

Sensui, H. (1995). Percepción de la entonación interrogativa del español: un estudio experimental. *Sophia Linguistica*, *38*, 1–23.

Sensui, H. (2003). A pilot case study on sentence pattern perception of Spanish. *PoLAR*, 1731–1734.

Shang, P., Elvira-García, W., & Roseao, P. (2021). La modalidad interrogativa en español y en chino: un enfoque funcionalista. *Círculo de Lingüística Aplicada a La Comunicació*, *88*, 235–254.

Shao, J. (2014). *The Study of Interrogatives in Chinese Mandairn*. The Commercial Press.

Shao, J. (2012). The comparison of the internal types of yes-no questions and the subdivision of incertitude. *Chinese Teaching in the World*, *26*(3), 347–356.

Shi, S., Zhang, J., & Xie, Y. (2014). Cross-language comparison of F0 range in speakers of native Chinese, native Japanese and Chinese L2 of Japanese: Preliminary results of a corpus-based analysis. In *Proceedings of the 9th International Symposium on Chinese Spoken Language Processing* (pp. 241–244). Singapore.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273–293.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Harvard University Press Cambridge.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*(3), 434–464.

Trimble, J. C. (2013a). *Acquiring Variable L2 Spanish Intonation in a Study Abroad Context* (Unpublished doctoral dissertation). University of Minnesota.

Trimble, J. C. (2013b). Perceiving Intonational Cues in a Foreign Language : Perception of Sentence Type in Two Dialects of Spanish. In C. Howe, S. E. Blackwell, & M. L. Quesada (Eds.), *15th Hispanic Linguistics Symposioum* (pp. 78–92).

Urbani, M. (2012). Pitch range in L1/L2 English. An analysis of f0 using LTD and linguistic measures. In M. G. Busà & S. Antonio (Eds.), *Methodological perspectives on L2 prosody: Papers from ML2P 2012* (pp. 79–83).

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, *38*(3), 253–265.

Van Dommelen, W. A., & Moxness, B. H. (1995). Acoustic parameters in speaker height and weight identification: sex-specific behaviour. *Language and Speech*, *38*(3), 267–287.

Wang, J. (2016). On the attribute, functions and the relation between “ma” and sentence intonation. *Chinese Language Learning*, *5*, 3–13.

Wiener, S., & Bradley, E. D. (2020). Harnessing the musician advantage: Short-term musical training affects non-native cue weighting of linguistic pitch. *Language Teaching Research*, 1362168820971791.

Wilson, D., & Sperber, D. (2002). *Relevance theory*. Blackwell.

Xing, F. (2002). *Three Hundred Questions of Chinese Mandarin Grammar*. The comercial press.

Xu, H. (2009). *A survey study of Chinese EFL learners' acquisition of English intonation: A functional perspective* (Unpublished MA thesis). Jiangsu University.

Yu, M. (2021). El español en china. *El español en el mundo Anuario del Instituto Cervantes 2021* (pp. 657–684). Retrived from https://cvc.cervantes.es/lengua/anuario/anuario_21/el_espanol_en_el_mundo_anuario_instituto_cervantes_2021.pdf

Yuan, C., González-Fuente, S., Baills, F., & Prieto, P. (2019). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, *41*(1), 5–32.

Yuan, J. (2006). Mechanisms of question intonation in Mandarin. In M. B. Huo Q., C. ES., & L. H. (Eds.), *Chinese Spoken Language Processing ISCSLP 2006. Lecture Notes in Computer Science* (Vol. 4274, pp. 19–30). Springer.

Yuan, J., Dong, Q., Wu, F., Luan, H., Yang, X., Lin, H., & Liu, Y. (2018). Pitch characteristics of L2 English speech by Chinese speakers: A large-scale study. In *Proceedings of the Annual Conference of the International Speech Communication Association* (pp. 2593–2597). Hyderabad, India.

Yuan, J., & Liberman, M. (2014). F0 declination in English and Mandarin broadcast news speech. *Speech Communication*, *65*, 67–74.

Zarate-Sandez, G. A., Morales-Front, A., & Sanz, C. (2015). *Perception and production of intonation among English-Spanish bilingual speakers at different proficiency levels* (Doctoral dissertation). Georgetown University.

Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B., & Trouvain, J. (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Proceedings of the 7th International Conference on Speech Prosody* (pp. 1037–1041). Dublin, Ireland.

Capítulo 2

La modalidad interrogativa en español y en chino: Un enfoque funcionalista

Citado como:

Shang, P.; Elvira-García, W. & Roseano, P. (2021). La modalidad interrogativa en español y en chino: un enfoque funcionalista. *Círculo de Lingüística Aplicada a la Comunicación*, *88*, 235-254. doi: https://dx.doi.org/10.5209/clac.78313.

Círculo de Lingüística Aplicada a la Comunicación
ISSN: 1576-4737

EDICIONES COMPLUTENSE

https://dx.doi.org/10.5209/clac.78313

La modalidad interrogativa en español y en chino: un enfoque funcionalista

Peizhu Shang 尚培珠[1]; Wendy Elvira-García[2]; Paolo Roseano[3]

Resumen. Este artículo presenta un análisis contrastivo de la modalidad interrogativa en español y en chino a partir de un marco teórico funcional-pragmático. A través del estudio, se muestra en primer lugar que la modalidad interrogativa no desempeña solo la función prototípica de solicitar información, sino que sirve para finalidades comunicativas diversas, tales como confirmar hipótesis, formular ruegos u órdenes, ofrecer objetos o sugerencias, expresar sorpresas, etc. Estas funciones pragmáticas se realizan mediante diferentes estrategias discursivas y gramaticales, caracterizadas por distintas combinaciones de recursos morfológicos, sintácticos y prosódicos. Además, tras la comparación entre las dos lenguas objeto de estudio se evidencia que la mayor diferencia entre ellas en lo que concierne a la modalidad interrogativa reside en los planos suprasegmentales, las partículas modales y el orden de palabras.
Palabras clave: Modalidad interrogativa; función pragmática; estrategia lingüística; comparación entre el español y el chino

[en] The interrogative modality in Spanish and Chinese: a functionalist focus

Abstract. This article presents a contrastive analysis of the interrogative modality in Spanish and Chinese, based on a functional-pragmatic theoretical framework. After research, firstly we have shown that the interrogative modality can not only perform the prototypical function of requesting information, but also serves for diverse communicative purposes, such as confirming hypotheses, formulating requests or orders, offering objects or suggestions, expressing surprises, etc. These pragmatic functions are encoded through different discursive and grammatical strategies, which are characterized by distinct combinations of morphological, syntactic and prosodic resources. Additionally, by comparing the two languages in this study, we demonstrated that the greatest difference between them in terms of the interrogative modality derives from the suprasegmental levels, modal particles and word orders.
Keywords: Interrogative modality; pragmatic function; linguistic strategy; comparison between Spanish and Chinese

[ch] 基于语用功能的西汉疑问语气对比研究

摘要：本文从系统功能语言学和语用学理论出发，对比分析了汉语和西班牙语中疑问语气的表达异同。经过研究，首先，本文表明了疑问语气不仅可以完成求取信息这一典型功能，还可以用来实现多种不同的交流目的，比如证实猜想、提出请求、发号施令、给予实物、提供建议、表达诧异等。为实现上述语用价值，说话人需要综合运用不同的话语和语法策略，即一系列不同的词法，句法和韵律手段来组合完成表达。此外，通过对本研究中两门语言的对比，我们证实了西汉疑问语气表达的最大差异主要来自于超音段，语气词，以及词序等方面。
关键词：疑问语气；语用功能；语言策略；西汉对比

Cómo citar: Shang, Peizhu 尚培珠; Elvira-García, Wendy; Roseano, Paolo. La modalidad interrogativa en español y en chino: un enfoque funcionalista. *Círculo de Lingüística Aplicada a la Comunicación* 88, 235-254, https://dx.doi.org/10.5209/clac.78313

[1] Universitat de Barcelona. Correo electrónico: pshangsh7@alumnes.ub.edu. ORCiD: 0000-0001-5407-4476.
[2] Universitat de Barcelona. Correo electrónico: wendyelvira@ub.edu. ORCiD: 0000-0001-7002-9851.
[3] Universidad Nacional de Educación a Distancia & University of South Africa. Correo electrónico: paolo.roseano@flog.uned.es. ORCiD: 0000-0002-9743-9735.

1. Introducción

La definición de la modalidad interrogativa y su clasificación interna son temas que han despertado interés tanto en la lengua española como en la china. Aun así, no existe un consenso entre los diferentes autores, que han fundamentado sus propuestas en múltiples marcos: desde la morfología hasta la pragmática, la sintaxis, la semántica o la prosodia. Tradicionalmente, se definía la modalidad interrogativa como aquella utilizada para preguntar y solicitar informaciones. Sin embargo, en las últimas décadas, con el auge de las teorías basadas en la pragmática, se ha comprobado que la función de la modalidad interrogativa no se circunscribe al uso prototípico de petición de información, sino que puede usarse para confirmar hipótesis, formular ruegos u órdenes, ofrecer objetos o sugerencias, hacer aserciones de ideas u opiniones o expresar sorpresas, entre otros (para el español ver a Escandell-Vidal, 1984, 1999, 2002; Escandell-Vidal y Prieto, n.d.; Guardado, 2017 y Haverkate, 2006, etc.; para el chino, a Chen y Wen, 2001; Gao y Zhang, 2009; Yuan y Yurie, 2019; Shao, 2014, 2012; Shao y Zhu, 2002; Shi, 2018; Yurie, 2014 y Zhang, 2010, etc.).

Sin embargo, los artículos que, desde la pragmática, constatan la existencia de estas funciones en las interrogativas, suelen hacer menos énfasis en tratar las estrategias lingüísticas que usar las lenguas para diferenciar esas funciones, sobre todo si se trata de estrategias entonativas. Y cuando lo hacen (mayormente para el caso del español como se podrá comprobar en secciones posteriores de este artículo), tratan solo una de las dos lenguas, pero no existe hasta la fecha un estudio contrastivo de las estrategias para la expresión de la modalidad interrogativa en chino y en español.

La expresión lingüística en la interacción conversacional de propósito comunicativo en las interrogativas se puede realizar en las dos lenguas (chino y español) mediante estrategias de nivel morfológico, sintáctico, prosódico o discursivo (Eggins, 2004; Fawcett, 2008; Halliday, Matthiessen y Halliday, 2014). Sin embargo, dado que el español y el chino son dos lenguas tipológicamente muy diferentes y filogenéticamente alejadas, se espera que las estrategias de expresión o los recursos lingüísticos utilizados para expresar una misma función pragmática sean distintos en las dos lenguas. Por este motivo, en este artículo, se pretende proporcionar un análisis contrastivo sistemático de las estrategias lingüísticas que se adoptan en cada una de las dos lenguas para la expresión de las diferentes fuerzas ilocutivas de modalidad interrogativa. La comparación de estas semejanzas y distinciones expresivas será útil para los estudios de tipología, así como para los estudios de lengua extranjera tanto ES>Chino como Chino>ES.

Para poder realizar dicho análisis, en este artículo se lleva a cabo, en primer lugar, una revisión de las principales perspectivas para abordar la modalidad interrogativa en español y en chino. Posteriormente, se presenta una clasificación basada en una visión funcional-pragmática como el punto de partida del este estudio contrastivo de la modalidad interrogativa. En los dos apartados siguientes, se analizan y comparan las estrategias lingüísticas utilizadas en las dos lenguas para la expresión de las distintas funciones pragmáticas de modalidad interrogativa. En los apartados finales, se presenta una sistematización de las diferencias expresivas que se han observado entre las dos lenguas a la hora de codificar una fuerza ilocutiva específica.

2. Interpretación y clasificación de la modalidad interrogativa en español y en chino

La modalidad interrogativa, entendida como una categoría lingüística distinta de las declarativas, exhortativas, exclamativas, etc., ha sido abordada tanto en español como en chino a partir de diferentes teorías lingüísticas. En los siguientes párrafos se resumen las principales contribuciones al respecto, organizándolas en cinco corrientes: (1) gramática escolástica, (2) teoría de la relevancia, (3) lingüística china tradicional, (4) gramática funcional y (5) la teoría de los actos de habla.

(1) Tanto la gramática escolástica del chino, como la del español, suele definir la modalidad interrogativa como aquella que sirve para pedir una información nueva (es decir, desconocida por el sujeto que formula la pregunta). En español, se suelen clasificar las interrogativas, a partir de una amalgama de criterios sintáctico-prosódicos, en preguntas parciales, preguntas totales y preguntas disyuntivas (RAE, 2009) (Real Academia Española, RAE). En chino, a estos tipos se añade la interrogativa positivo-negativa caracterizada por una construcción especial (llamada "X-no-X") de yuxtaposición de dos formas verbales diferenciadas únicamente por la negación (Lü, 2002; Xing, 2002).

(2) Dentro del marco de la teoría de relevancia de Sperber y Wilson, las interrogativas se han explicado como "interpretaciones de respuestas que el hablante consideraría relevantes si fueran verdaderas" (1986: 252). De acuerdo con este acercamiento, en las preguntas parciales, el hablante expresa una proposición incompleta e indica que la compleción sería la respuesta relevante, mientras en las preguntas totales, se ha optado por representar toda la proposición que el hablante ha considerado relevante a la respuesta deseable (Sperber y Wilson, 1986). A partir de esta teoría, diversos autores chinos y occidentales (como Blakemore, 1994; Chen y Wen, 2001; Escandell-Vidal, 2002; Shao 2014 y Noh; 1998) han realizado una diferenciación entre las preguntas auténticas y las preguntas eco que reproducen lingüísticamente o pragmáticamente toda o alguna parte de la proposición anterior. Las preguntas eco, según Escandell-Vidal (2002: 872), son ejemplos típicos del doble uso interpretativo del lenguaje, o en otras palabras, son preguntas metalingüísticas que representan otra interpretación atributiva.

(3) En la lingüística china tradicional se ha estudiado la modalidad interrogativa desde una perspectiva que comparte sus fundamentos con los estudios occidentales basados en el concepto de epistemicidad (Aikhenvald, 2003; Chafe y Nicolás, 1986; Palmer, 2001), es decir, el grado de compromiso con la verdad de un enunciado. En la lingüística china, el concepto utilizado para clasificar las preguntas es el grado de interrogación, que se define como el grado de incerteza del hablante hacia el contenido proposicional. El grado de interrogación de la gramática china es un concepto complementar y simétrico a la idea de "grado de compromiso epistémico" propuesto por la lingüística contemporánea occidental en el sentido de que si el grado de certeza aumenta, disminuye el grado de interrogación y viceversa (Shao, 2014: 71). La identificación del grado de interrogación depende de un conjunto de factores lingüísticos y extralingüísticos como el tipo interrogativo, la partícula final, el orden de palabras, el adverbio modal, la prosodia, la situación contextual, etc. Entre los factores en cuestión, Shao (2014) define el tipo de pregunta como el factor interno fundamental que determina el grado de interrogación del hablante y a partir de esto, se clasifican cinco niveles de interrogación para la modalidad interrogativa correspondientes con cinco tipos de pregunta en chino, como se observa en la figura 1. Además de Shao (2014), diversos autores chinos (Guo, 2000; Huang, 1986; Shao, 2014, 2012; Xu y Zhang, 1985) han realizado una clasificación de la modalidad interrogativa basándose en el grado de interrogación.

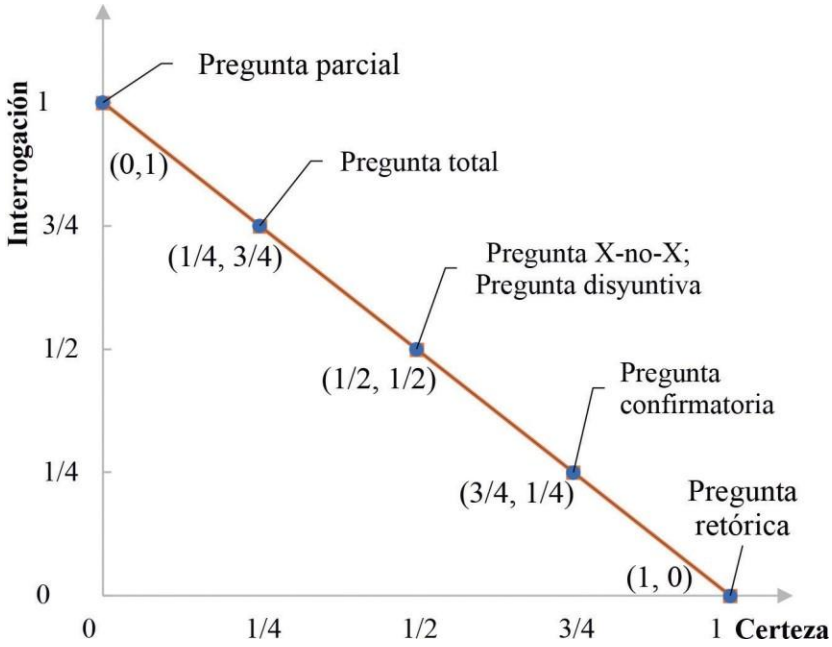


Figura 1. Clasificación de interrogativas en chino según el grado de interrogación realizada por Shao (2014).

Sin embargo, este modelo no es exento de críticas. En primer lugar, dado que la interrogatividad es un concepto complejo, su reducción a dos únicas variables de tipo numérico parece arbitraria. En segundo lugar, el resultado de análisis de Shao (2014) es incompleto en lo referente a las preguntas totales, puesto que no consigue diferenciar entre los tres tipos de preguntas totales que existen en esa lengua: i) preguntas totales sin la partícula *ma* marcadas únicamente por la entonación; ii) preguntas totales marcadas por la partícula ma^0, de tono neutro; iii) preguntas totales marcadas por la partícula ma^2, con un rango tonal más expandido y elevado (como su pronunciación es parecida al tono 2 del chino, se anota como ma^2 en el artículo para diferenciarla de la ma^0, de tono neutro). En tercer lugar, cabe destacar que la clasificación de las preguntas a partir del grado de interrogación es, en su esencia, circular: para definir los tipos de preguntas se utiliza el concepto de "grado de interrogación", que a su vez se define última instancia a partir del tipo de pregunta.

(4) La cuarta propuesta para abordar la modalidad interrogativa se realiza dentro del marco de la gramática funcional. A partir de las funciones metalingüísticas, Halliday y McDonald (2004: 305-396) clasifican las preguntas del chino en preguntas parciales y totales, y separan las últimas en totales sesgadas y totales neutras (ver la figura 2). La distinción entre las totales sesgadas y las neutras consiste en que en las primeras el hablante emite un enunciado afirmativo/negativo y pide una confirmación del oyente; mientras en las segundas el hablante es neutral y no formula ninguna hipótesis hacia el contenido proposicional (Halliday y McDonald, 2004). De acuerdo con esos autores, las totales neutras son formalmente las que presentan la estructura interrogativa *X-no-X*; mientras las totales segadas poseen dos formas expresivas (que se explican con detalle en el apartado 4.1.2. y 4.1.5.). Sin embargo, la clasificación funcional que se propone en (Halliday y McDonald, 2004) es demasiado general, ya que no se engloban todas las formas interrogativas posibles en chino (por ejemplo, no se hace referencia a las preguntas confirmatorias, eco, disyuntivas, etc.), ni se explica claramente la función pragmática de cado tipo de pregunta.

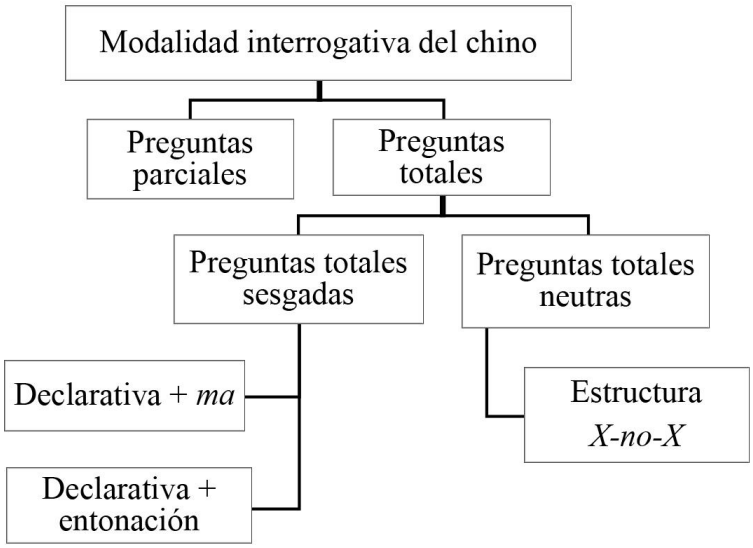


Figura 2. Clasificación de modalidad interrogativa en chino realizada por Halliday y McDonald (2004).

(5) Desde la teoría de los actos de habla de Searle (1969), Haverkate (2006) propone la existencia de cuatro actos ilocutivos que se pueden realizar con la modalidad interrogativa: acto exhortativo, asertivo, comisivo y expresivo. Sin embargo, en este estudio no se ha adoptado el esquema de Haverkate (2006) porque, aunque postula una taxonomía relativamente completa de las fuerzas ilocutivas de modalidad interrogativa, su trabajo se centra en la discusión del estatus pragmalingüístico de las secuencias de preguntas. Esto hace que no se detenga en explicar cómo se codifica y se diferencia lingüísticamente cada una de estas funciones ilocutivas. Además, para nuestro objetivo, la propuesta de cuatro funciones ilocutivas no es suficiente para una descripción pragmática exhaustiva de todas las clases de modalidad interrogativa dado que algunas preguntas, por ejemplo, las preguntas de función confirmatoria,no se han podido incluir en la taxonomía pragmática de Haverkate (2006).

Para completar esta revisión bibliográfica, es preciso hacer mención a otros estudios que no se pueden adscribir a ninguna de las perspectivas anteriores. Entre los más habituales, encontramos la clasificación de las preguntas a partir del tipo de respuesta esperada (Fan, 1982; Shao, 2014). Además, en el caso del español, se ha propuesto clasificar las preguntas a partir del patrón entonativo (Font-Rotchés, 2005; Roseano et al., 2015; Salcioli Guidi, 1987). En el caso del chino, por otra parte, un acercamiento muy habitual es el que consiste en clasificar las preguntas a partir de las partículas finales interrogativas (Jin, 2011; Wang, 2016).

En resumidas cuentas, queda evidente que no hay un consenso entre los lingüistas sobre cómo clasificar los diferentes tipos de preguntas, tanto en la lengua española como en la china. Este hecho, junto con las diferencias tipológicas en las estrategias de codificación entre las dos lenguas, hace que ninguno de los marcos teóricos existentes, ni desde la gramática del chino ni desde la del español, sea aplicable de manera perfecta a ninguna de las dos lenguas.

3. Una visión funcional-pragmática para el análisis contrastivo de modalidad interrogativa en español y en chino

En el apartado anterior, se han ilustrado las perspectivas más usadas para abordar la modalidad interrogativa en chino y en español. Sin embargo, en su forma actual ninguna de ellas es aplicable de manera perfecta a ninguna de las dos lenguas. Puesto que las funciones interactivas del lenguaje son interidiomáticas (es decir, son las mismas para el español y el chino) (Belchí, 1994: 329), para poder llegar a una clasificación de las interrogativas que sea válida para ambas lenguas, este artículo va a adoptar un punto de vista funcional-pragmático. Por lo tanto, en este artículo se establecerán relaciones entre la función comunicativa o los actos de habla realizados y las estrategias lingüísticas que se emplean en chino y en español, para conseguir determinados fines comunicativos.

Más en concreto, a partir del papel discursivo adoptado por el hablante y el tipo de producto que se intercambia, Eggins (2004), Fawcett (2008) y Halliday, Matthiessen y Halliday (2014) plantean la existencia de cuatro tipos de funciones comunicativas: i) solicitar informaciones, ii) ofrecer informaciones, iii) solicitar bienes y servicios, iv) ofrecer bienes y servicios. En las dos primeras funciones, se puede solicitar una información (sea nueva o algo que se ha presentado de alguna forma en el discurso anterior) o se pueden pedir una confirmación u ofrecer y trasmitir informaciones como opiniones y sentimientos mediante una interrogación retórica. De la misma manera, en los últimos dos tipos de funciones, el hablante puede pedir un objeto o una acción, u ofrecer algo material e inmaterial (como una sugerencia para influenciar el comportamiento del oyente).

La modalidad interrogativa constituye un tipo oracional especialmente interesante desde el punto de vista funcional-pragmático, ya que permiten la realización de una gran variedad de actos ilocutivos, que no se circunscriben al

acto prototípico informativo, sino que abarcan casi todos los actos ilocutivos registrados por Searle (1969), como el acto confirmativo, asertivo, imperativo, exhortativo, comisivo, etc.

El presente trabajo combina el planteamiento de intercambio conversacional de Eggins (2004), Fawcett (2008) y Halliday, Matthiessen y Halliday (2014) con la teoría de los actos de habla de Searle (1969) para proporcionar una visión sistemática de la modalidad interrogativa. Por ello, proponemos en primer lugar una clasificación funcional-pragmática de las interrogativas como el punto de partida del análisis de estrategias de expresión. Como se observa en la figura 3 (que contiene el esquema es el que se seguirá en los siguientes apartados), dentro de las dos clases de intercambio conversacional existen diferentes tipos interrogativos que se corresponden con unos actos de habla específicos, por ejemplo, preguntas informativas (acto informativo), preguntas confirmatorias (acto confirmativo), preguntas eco (acto citativo), preguntas retóricas (acto asertivo), preguntas imperativas (acto imperativo) y preguntas comisivas (acto comisivo).

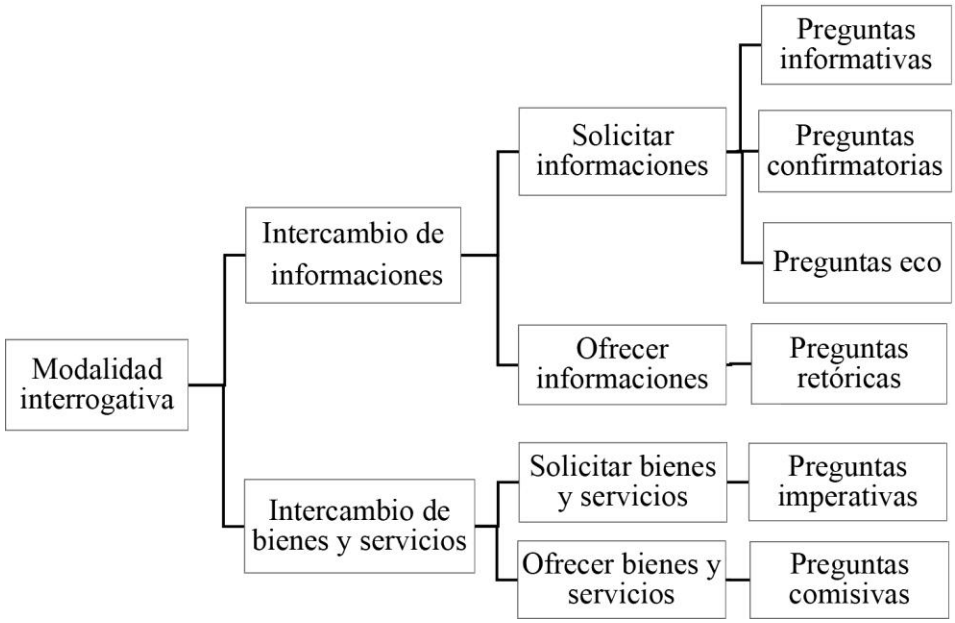


Figura 3. Clasificación funcional-pragmática de la modalidad interrogativa.

4. Estrategias para la expresión de las preguntas de intercambio de información en español y en chino

En la interacción conversacional, el hablante dispone de estrategias para que el oyente pueda identificar la función comunicativa del enunciado, es decir, la intención del mismo hablante al pronunciarlo. Estas estrategias pueden ser lingüísticas, gestuales o multimodales. Desde el punto de vista lingüístico, las estrategias son recursos comunicativos codificados en cada sistema gramatical y suelen manifestarse en el nivel morfológico, sintáctico y prosódico. En este sentido, el español y el chino, dos lenguas tipológicamente muy distintas, tienen diferentes estrategias lingüísticas para realizar preguntas con diferentes funciones pragmáticas.

4.1. Solicitar información: preguntas informativas, confirmatorias y eco

En primer lugar, se van a analizar las estrategias para la expresión de las preguntas de la clase de petición de información. En esta sección se explicarán los tres tipos de preguntas informativas originales que se pueden realizar en las dos lenguas: preguntas parciales (4.1.1.), totales (4.1.2.) y disyuntivas (4.1.3.). A continuación, se discutirán las preguntas confirmatorias (4.1.4.) y eco (4.1.5.).

4.1.1. Preguntas parciales informativas

La función pragmática de las preguntas parciales informativas es la de solicitar una información "que no está presente en el contenido proposicional de la misma proposición" (Roseano et.al, 2015: 5). El hablante normalmente no tiene ninguna certeza o conocimiento con respecto a la incógnita presentada y pregunta con el objetivo de pedir una respuesta específica, como en el ejemplo (1), en el que se presenta, además del enunciado de la pregunta en las dos lenguas, un contexto comunicativo para ayudar a clarificar la función de la interrogativa.

(1) Contexto: El hablante entra en una papelería y quiere comprar un nuevo lápiz, pero no sabe el precio y se lo pregunta al vendedor.

ESP: ¿Cuánto cuesta este lápiz?

CH:	zhe^{4}-zhi^{1}-qian^{1}bi^{3}	duo^{1}shao^{3}	qian^{2}?
	este-CLF-lápiz	cuánto	dinero
	¿Cuánto cuesta este lápiz?		

Tal como se observa en los ejemplos, en español y en chino, la característica más evidente de las preguntas parciales es que contienen un pronombre (como *qué*, *quién*, *cuál*, *cuánto*, etc.) o adverbio interrogativo (como *dónde*, *cómo*, *cuándo*, *por qué*, etc.) con los que se introduce una incógnita en el enunciado (Mora, 1999). La incógnita en cuestión no se puede despejar con un simple *sí* o *no*, sino proporcionando una información más específica.

Aunque las dos lenguas comparten una estrategia morfológica para formular las preguntas parciales (es decir, el uso de una palabra de tipo Qu-), existen diferencias significativas en la formación sintáctica de estos constituyentes. En español, las palabras interrogativas se caracterizan por presentar un movimiento sintáctico de tipo Qu- (conocido como *Wh-movement* en inglés): la palabra Qu- se desplaza desde la posición original del argumento hacia la periferia izquierda de la proposición y, como resultado, se produce una inversión sujeto-verbo, como en el ejemplo (1) (Escandell-Vidal y Prieto, n.d.). En chino, en cambio, la palabra Qu- aparece *in situ* (Cheng, 1997), por lo que, en principio, se mantiene la misma estructura sintáctica de las declarativas y el pronombre interrogativo puede aparecer en cualquier posición oracional dependiendo de la posición del argumento. Los ejemplos de (3) a (5) ejemplifican la ausencia de *wh-movement* en chino: las palabras Qu- *shui*^{2} (quién), *shen*^{2}*me*^{0}*shi*^{2}*hou*^{0} (cuándo) y *na*^{3}*li*^{3} (dónde) aparecen en su posición canónica de sujeto, modificador adverbial y objeto, respectivamente, que es la misma que ocupan los constituyentes correspondientes en la declarativa neutra que aparece en (2).

(2) Declarativa:

Lin^{2}yi^{1}	ming^{2}tian^{1}	qu^{4}	bei^{3}jing^{1}
Linyi	mañana	ir	beijing
Linyi irá mañana a Beijing.			

(3) Interrogativo inicial:

shui^{2}	ming^{2}tian^{1}	qu^{4}	bei^{3}jing^{1}
quién	mañana	ir	beijing
¿Quién irá mañana a Beijing?			

(4) Interrogativo medio:

Lin^{2}yi^{1}	shen^{2}me^{0}shi^{2}hou^{0}	qu^{4}	bei^{3}jing^{1}
Linyi	cuándo	ir	beijing
¿Cuándo irá Linyi a Beijing?			

(5) Interrogativo final:

Lin^{2}yi^{1}	ming^{2}tian^{1}	qu^{4}	na^{3}li^{3}
Linyi	mañana	ir	dónde
¿Adónde irá Linyi mañana?			

Desde un punto de vista prosódico, puesto que el español es una lengua entonativa, la prosodia tiene una gran importancia como indicador de la modalidad interrogativa. Este marcaje se manifiesta a través de la modulación de F0 (sobre todo en la parte final del contorno entonativo, que en la gramática tradicional española se ha llamado tonema y que abarca la última sílaba tónica del enunciado y sus postónicas). Hay un consenso entre los diferentes autores a la hora de afirmar que las preguntas parciales desde el punto de vista entonativo están caracterizadas por un acento prenuclear alto H* en la palabra Qu- y un tonema descendente L* L% (Estebas-Vilaplana y Prieto, 2010: 35). En general, se considera que esta solución es la más común y neutra para expresar una pregunta parcial informativa en español.

Por el contrario, en una lengua tonal como el chino, puesto que el F0 codifica principalmente información léxica (es decir, sirve para diferenciar palabras) (Cai, 2019), nos esperaríamos que ese parámetro pudiera vehicular información acerca de la modalidad oracional. Lo que se observa es que, efectivamente, no hay un patrón entonativo que caracterice las preguntas parciales, puesto que una pregunta parcial puede tener tanto una melodía descendente como una ascendente, dependiendo del tono léxico de la última sílaba. Sin embargo, el F0 contribuye –en parte– a vehicular información acerca de la modalidad oracional mediante el rango tonal (Shen, 1990). En chino , de hecho, las preguntas parciales informativas tienen un rango tonal más amplio que las declarativas y, además, en el foco de la pregunta (que suele ser la palabra Qu-) presentan un movimiento más amplio de F0, una mayor duración y una mayor intensidad (Liu y Xu, 2005: 81; Yan, Wang y Shi, 2016).

La última distinción entre las dos lenguas es de tipo léxico y consiste en el uso de partículas. En chino, los recursos léxicos son habituales para la expresión de la modalidad y de la función pragmática. En el caso de las preguntas

parciales (y también de las declarativas) es habitual el uso de la partícula *ne*^{0} (呢), sobre cuya posición sintáctica y función pragmática ha habido un debate intenso. Cheng (1991) en su *Clausal Typing Hypothesis* plantea que la partícula *ne*^{0} es una *typing particle* de las preguntas parciales del chino, es decir, una marca léxica que pone de manifiesto el tipo de la pregunta. Sin embargo, esta afirmación se ha comprobado que es empíricamente problemática de acuerdo con los estudios de Shi y Zhang (1995), Qu (2006) y Jin (2011), puesto que en las preguntas parciales sin *ne*^{0} se puede expresar también el significado interrogativo completo, como en el ejemplo (6), en que la presencia de la partícula *ne*^{0} *no* es obligatoria.

(6) CH: Hu^{2}fei^{1} chi^{1}-le^{0} shen^{2}me^{0} (ne^{0})
Hu^{2}fei^{1} comer-PRF qué (Q_{NE})
¿Qué comió Hufei?

Estudios más recientes retoman la cuestión de la función pragmática de la partícula *ne*^{0}. Según Shi (2004, 2006) y Jin (2011), *ne*^{0} se utiliza cuando el hablante posee algún conocimiento previo sobre la respuesta que su pregunta va a recibir, aunque no conozca su contenido exacto, como en el ejemplo (7).

(7) Contexto: Dos hermanos están al lado del sujeto en el sofá. Miran al sujeto de reojo y después siguen hablando y riendo en voz baja. El sujeto se da cuenta de qué están hablando de él y les pregunta:

CH: ni^{3}men^{0} liao^{2} shen^{2}me^{0} ne^{0}
vosotros hablar qué Q_{NE}
¿De qué estáis hablando?

En la situación (7) el hablante ya sabe que los dos hermanos están hablando de él, pero no conoce exactamente el contenido de la conversación y quiere descubrirlo. En los ejemplos de (3) a (5), en los que no aparece la partícula *ne*^{0}, el hablante no tiene ningún tipo de conocimiento o expectativa acerca de la respuesta que va a obtener. En conclusión, las preguntas parciales con la partícula *ne*^{0} se diferencian pragmáticamente de las parciales sin *ne*^{0} por un mayor grado de conocimiento del hablante hacia la respuesta.

4.1.2. Preguntas totales informativas

La función pragmática de las preguntas totales informativas consiste en pedir una información nueva cuando el hablante no tiene expectativas sobre la respuesta que obtendrá, como en (8). Las totales informativas son también denominadas "preguntas sí-no" porque, tanto en chino como en español, se responden normalmente con un "sí" o con un "no", aunque también sea posible el uso de respuestas como "tal vez", etc. (Guo, 2000; Li, 1990).

(8) Contexto: El sujeto entra en una frutería donde no ha estado nunca. Mira en las estanterías y no ve que haya sandía. Le pregunta al empleado de la tienda si tienen sandía.

ESP: ¿Tenéis sandía?

CH: ni^{3}men^{0} you^{3} xi^{1}gua^{1} ma^{0}
vosotros tener sandía Q_{MA}
¿Tenéis (vosotros) sandía?

Desde una perspectiva sintáctica, las preguntas totales informativas en chino mantienen el mismo orden que las enunciativas (es decir, SVO). En español, sin embargo, hay un cambio sintáctico: mientras las declarativas neutras presentan un orden SVO, en las interrogativas totales en las que aparece el sujeto se observa una permutación de sujeto y predicado y, por tanto, el orden sintáctico es VSO (Escandell-Vidal, 1999; Haverkate, 2006). Además, en español las totales informativas se distinguen de las declarativas por una serie de rasgos entonativos, entre los cuales el más importante es la configuración nuclear bajo-ascendente L* HH% (Estebas-Vilaplana y Prieto, 2010: 27). El chino, por otra parte, al ser una lengua tonal no diferencia la pregunta total mediante un contorno entonativo distinto.

En chino, el marcaje de las preguntas totales se observa fundamentalmente en el nivel léxico. En concreto, según Yuan y Yurie (2019), Jin (2011) y Wang (2016), las preguntas totales informativas neutras se realizan en chino por una partícula final *ma*^{0}, de tono fonológico neutro. Sin embargo, cabe destacar que en chino hay otra partícula modal *ma*^{2}, homógrafa pero fonológicamente distinta, que se pronuncia con un rango tonal más expandido y elevado en la parte final. Según Yuan y Yurie (2019), Jin (2011), Guo (2000), las preguntas totales marcadas por la partícula *ma*^{2} no son neutras, en el sentido de que expresan un sesgo epistémico de sorpresa o sospecha hacia el contenido proposicional (sobre el que se volverá en el apartado 4.1.5.).

Además de las totales informativas marcadas por la partícula *ma*^{0}, existe en chino otra forma de pregunta con la misma función informativa. Ese segundo tipo de preguntas informativas se diferencia de las preguntas con la partícu-

la *ma*^{0} en varios aspectos, tanto formales como pragmáticos. Sintácticamente, de hecho, presentan una construcción del tipo X-no-X: a diferencia de las preguntas totales marcadas por *ma*^{0}, en las que se presenta una sola opción de respuesta en forma explícita, en las preguntas X-no-X las dos respuestas antagónicas aparecen en la misma proposición y están conectadas por la negación *bu*^{4}, como en el ejemplo (9). Las diferencias pragmáticas con las preguntas marcadas por *ma*^{0} son dos: i) en primer lugar, las preguntas X-no-X exigen, normalmente, una respuesta específica, es decir, no se pueden responder con "sí" o "no"; por ejemplo, en la situación anterior (9) la repuesta no podría ser *shi*^{4} ('sí') o *bu*^{4}*shi*^{4} ('no') sino que tendría que ser *qu*^{4} ('ir') o *bu*^{4}*qu*^{4} ('no ir'); ii) en segundo lugar, las preguntas marcadas con *ma*^{0} expresan un mayor grado de formalidad o cortesía, mientras las preguntas X-no X se suelen utilizar en los contextos en los que hay cierta familiaridad entre los interlocutores (Yuan y Yurie, 2019; Wang, 2016; Yurie, 2014).

(9) Contexto: Llega el período de exámenes finales de curso y el hablante quiere ir a la biblioteca para repasar una asignatura. Le pregunta a su amigo si quiere ir con él a la biblioteca.

CH:	ni^{3}	qu^{4}-bu^{4}-qu^{4}	tu^{2}shu^{1}guan^{3}
	tú	ir-NEG-ir	biblioteca
	¿Vas a la biblioteca?		

Desde el punto de vista léxico, las preguntas X-no-X pueden llevar opcionalmente una partícula *ne*^{0}, de tono neutro, al final del enunciado. La misma pregunta del ejemplo (9) puede aparecer como en (10). La presencia de *ne*^{0} no ejerce influencia en el significado pragmático interrogativo y su función es sencillamente la de intensificar la pregunta (Jin, 2011: 164).

(10) CH:	ni^{3}	qu^{4}-bu^{4}-qu^{4}	tu^{2}shu^{1}guan^{3} ne^{0}	
	tú	ir-NEG-ir	biblioteca Q_{NE}	
	¿Vas a la biblioteca?			

Además, cabe destacar que estudios recientes (Yan, Wang y Shi, 2014a; Yuan y Yurie, 2013; Yurie, 2014), afirman que las preguntas X-no-X tienen, además, una característica prosódica. Las preguntas en cuestión, de hecho, en comparación con otras interrogativas neutras presentan una reducción del rango de los movimientos tonales en la última palabra del enunciado.

Por último, cabe destacar que en español existe una construcción de forma cercana a las preguntas X-no-X del chino, que se caracteriza por finalizar con una interrogativa del tipo *¿... o no?*, como por ejemplo en *¿Vienes a cenar o no?* A primera vista, las preguntas de esta forma presentan cierto paralelismo sintáctico con las preguntas X-no-X del chino. Sin embargo, la fórmula *¿... o no?* del español vehicula una variedad de valores pragmáticos distintos y, sobre todo, no neutros. Estas preguntas, de hecho, pueden contener un matiz de "insistencia, presión, recriminación o vehemencia, pero también duda o inseguridad en ciertos casos" (RAE, 2009: 159). Por ejemplo, en el caso de la pregunta *¿Vienes a cenar o no?* el hablante puede estar dejando entender que espera que la respuesta sea positiva, o bien está subrayando que quiere obtener una respuesta de una vez (en este último caso es posible que la pregunta vaya precedida por el marcador *Pero*).

4.1.3. Preguntas disyuntivas informativas

La función de las preguntas disyuntivas consiste en presentar dos o más alternativas al oyente y pedirle que elija entre ellas. Por un lado, se diferencian de las preguntas parciales en que el emisor de estas preguntas explicita las respuestas posibles, mientras que en las parciales el hablante no tiene ningún conocimiento sobre la incógnita. Por otro, difieren de las preguntas totales informativas en el tipo de respuesta, en las preguntas disyuntivas el enunciado presenta todas las posibilidades alternativas y la respuesta debe ser una opción concreta de todas ellas, como en (11), mientras en las preguntas totales se responden con una respuesta afirmativa o negativa.

(11) Contexto: El hablante está hablando con un amigo sobre las carreras universitarias; quieres saber cuál es su disciplina favorita entre la física y la química y se lo pregunta.

ESP:	¿Te gusta la física o la química?					
CH:	(a)	ni^{3}	xi^{3}huan^{1}	wu^{4}li^{3} (ne^{0})	hai^{2}shi^{4}	hua^{4}xue^{2} (ne^{0})
		tú	gustar	física (Q_{NE})	o.CONJ	química (Q_{NE})
	(b)	ni^{3}	xi^{3}huan^{1}	wu^{4}li^{3} ne^{0}		hua^{4}xue^{2} ne^{0}
		tú	gustar	física Q_{NE}		química Q_{NE}
	¿Te gusta la física o la química?					

Sintácticamente, como se observa en los ejemplos, la última opción de la pregunta disyuntiva necesita estar separada de las opciones anteriores por un nexo léxico. En español este nexo es la conjunción *o*, mientras en chino,

el significado disyuntivo está marcado con la conjunción *hai²shi⁴* o por la partícula modal *ne⁰*. La posición y el uso de la partícula *ne⁰* es muy flexible en las disyuntivas: esta partícula puede aparecer junto con la conjunción *hai²shi⁴* y aparecer después de cada alternativa o aparecer solamente en la última alternativa (11a). La partícula *ne⁰* también puede constituir el único recurso morfosintáctico para codificar el significado alternativo, como en (11b).

En chino también se han descrito fenómenos prosódicos en las preguntas disyuntivas. En concreto, según Yan, Wang y Shi (2014b) en las preguntas disyuntivas el rango tonal del primer constituyente es más elevado y amplio frente al último. Además, se ha defendido que en chino la velocidad de habla de las preguntas disyuntivas es más lenta que la de las preguntas totales (tanto en las que están marcadas por la partícula *ma⁰* como en las del tipo X-no-X) y de las parciales (Shi, 1980: 76).

En español la pregunta disyuntiva se diferencia de la declarativa disyuntiva -y de las demás interrogativas- gracias al marcaje prosódico. A diferencia de la interrogativa total, que acaba con un tono alto, la interrogativa disyuntiva termina con un tono bajo. Al mismo tiempo, los dos elementos de una pregunta disyuntiva como *¿Quieres mandarinas o limones?* están separados por un tono de frontera interna de tipo HH- que aparece al final del primer elemento (*mandarinas*), mientras en una declarativa disyuntiva como *No sabe si quieres mandarinas o limones* el tono al final del primer elemento (*mandarinas*) es H- (Estabas-Vilaplana y Prieto, 2010: 27). En otras palabras, la altura del tono de frase intermedia es distintiva.

4.1.4. Preguntas confirmatorias

La función de las preguntas que, tanto en la tradición china como en la española, se han llamado "confirmatorias" es pedir una confirmación del contenido proposicional. Estas preguntas se diferencian de las preguntas informativas porque el hablante prevé o espera que la respuesta a su pregunta será afirmativa. De acuerdo con el grado de seguridad o expectativa que muestra el hablante ante esa hipótesis, se van a clasificar estas preguntas en confirmatorias débiles y confirmatorias fuertes. Una pregunta es confirmatoria "débil" cuando el grado de compromiso con la veracidad del enunciado (en inglés *commitment to the truth-content of the content of the proposition*) por parte hablante es bajo (González, Roseano, Borràs-Comes y Prieto, 2017). Es "fuerte" si el mismo grado de compromiso con la veracidad del enunciado es alto. En otros términos, en una pregunta confirmatoria débil el hablante está moderadamente seguro de recibir una respuesta afirmativa, mientas que en una confirmatoria "fuerte" está muy seguro de ello.

i) Confirmatorias débiles

En español, las preguntas confirmatorias débiles tienen las mismas características sintácticas de las interrogativas totales informativas (v. 4.1.2.), pero se diferencian de ellas en el nivel prosódico. Las preguntas confirmatorias débiles, de hecho, presentan un contorno melódico claramente diferenciado de las totales informativas, que se caracteriza por una configuración nuclear descendente del tipo H+L* L% (Estebas-Vilaplana y Prieto, 2010: 34). En chino, la misma función pragmática se puede vehicular de diferentes maneras, tal como aparece en el ejemplo (12).

(12) Contexto: El hablante está mirando la televisión con un compañero de piso. De repente, el compañero de piso estornuda tres veces seguidas y el hablante le pregunta si está resfriado.

ESP: ¿Estás resfriado?

CH:	(a)	ni³	shi⁴	gan³mao⁴-le⁰	ma⁰
		tú	sí.ADV	resfriar-PRF	Q_{MA}
		¿Es verdad que estás resfriado?			
	(b)	ni³	shi⁴-bu⁴-shi⁴	gan³mao⁴-le⁰	
		tú	estar-NEG-estar	resfriado-PRF	
		¿Estás o no estás resfriado?			

La primera se caracteriza por la presencia de un adverbio de afirmación *shi⁴* ('sí') delante del verbo o adjetivo, como en (12a). En ese tipo de interrogativa, el adverbio *shi⁴* sirve de *focus marking* para enfatizar y confirmar el contenido proposicional posterior mientras la partícula final *ma⁰* es el marcador interrogativo que se utiliza para formular una pregunta total (Guo, 2000; Jin, 2011: 185).

La segunda manera para formular una pregunta confirmatoria débil en chino consiste en utilizar la construcción *shi-bu-shi* VP ('ser no ser VP'). Las preguntas *shi-bu-shi* son un tipo especial de la pregunta X-no-X. Como se expone en la sección 4.1.2., la estructura X-no-X constituye una pregunta informativa epistémica neutra, en el sentido de que el hablante no se espera una respuesta afirmativa. Esa neutralidad epistémica se da si el elemento X es un verbo de contenido (como por ejemplo *ir*, *ver* o *comer*). Sin embargo, si la posición X es ocupada por el verbo *shi⁴* ('ser') como en (12b), la pregunta tiene una función confirmatoria y es cortés (Liu y Pan, 2001; Shao y Zhu, 2002; Tao, 1998). Finalmente, cabe destacar que igual que las preguntas neutras X-no-X, la pregunta confirmatoria *shi-bu-shi VP* está caracterizada por una compresión del rango tonal del último elemento.

ii) Confirmatorias fuertes

En español, cuando el hablante tiene un elevado grado de compromiso con la veracidad del enunciado, tiene a su disposición una estrategia que consiste en proponer el contenido proposicional P en forma de oración declarativa, seguida por una "pregunta coletilla" o "pregunta de apéndice" (en inglés *tag question*). El segundo elemento (es decir, la pregunta coletilla), es el marcador pragmático que expresa la fuerza ilocutiva de toda la construcción, que suele ser la petición de una confirmación (uso confirmatorio) o a veces la petición de una acción (uso directivo) (Cuenca, 1997). En esta parte, debido a la delimitación temática, nos centramos en el análisis del uso confirmatorio.

El apéndice, como marcador del significado confirmatorio, presenta formas sintácticas diferentes dependiendo de la lengua y la situación comunicativa situada. Aun así, Cuenca (1997) considera que la mayoría de las lengua del mundo, por lo menos en su análisis, las lenguas románicas y las lenguas germánicas disponen de una serie de apéndices confirmativos convencionales y similares. En el caso del español, los apéndices se pueden dividir en tres tipos (Cuenca, 1997): el primero (13a) contiene un elemento léxico perteneciente al área semántica de la "verdad" (como *verdad*, *cierto*…) y, al ser una pregunta, presenta el patrón ascendente típico de las interrogativas neutras (Uribe, 1996: 8; Estebas-Vilaplana y Prieto, 2010: 29); el segundo (13b) también presenta un esquema entonativo interrogativo (Uribe, 1996: 8; Estebas-Vilaplana y Prieto, 2010: 29) y constituye un marcador de modalidad, generalmente en forma negativa y relacionado con el valor de verdad; y el último tipo (13c), aunque lexicalmente pertenece a la clase de las interjecciones (como *eh* o *ah*), se pronuncia con una entonación de tonema ascendente típicamente de las interrogativas y se interpreta como una "apelación al oyente solicitando de manera reforzada que confirme, ratifique o acepte lo dicho" por el hablante (Hidalgo Navarro, 2015: 91).

(13) Contexto: Los resultados del examen de fonética se acaban de publicar en la página web de la Facultad. El hablante los está consultando desde su móvil, mientras su compañero de piso hace los mismo desde el propio. Al ver la cara feliz de su compañero de piso, el hablante infiere que ha aprobado y le pregunta si es así.

ESP: (a) Has aprobado el examen, ¿verdad?/¿no es cierto?
(b) Has aprobado el examen, ¿no?
(c) Has aprobado el examen, ¿eh?/¿ah?

Para las preguntas confirmatorias fuertes del chino se ha descrito un conjunto de estrategias comparables. Gao y Zhang (2009), han propuesto dos grupos de apéndices confirmatorios para el chino . El primer grupo está formado por elementos léxicos relacionados con el concepto de "verdad" (como *shi*^{4} ('sí') o *dui*^{4} ('correcto')), a los que sigue una partícula modal (que puede ser *ma*^{0} o *ba*^{0}), como en (14a y 14b), o bien la yuxtaposición de un adverbio epistémico en forma afirmativa y negativa, como en (14c). El segundo conjunto de estrategias consiste en utilizar una interjección *ha*^{0} (哈) como en (14d). Prosódicamente, no se han encontrado hasta la actualidad descripciones acústicas relativas a estas preguntas con apéndices confirmativos en chino, por lo que no es posible caracterizarlas.

(14) Contexto: El hablante llega a casa a las nueve de la noche y ve que hay platos sucios en la mesa. Supone que su pareja debe de haber cenado, y se lo pregunta para confirmarlo.

CH: (a) ni^{3} chi^{1}guo^{4}-le^{0} shi^{4} ma^{0}/ba^{0}
tú comer-PRF sí $Q_{MA/}Q_{BA}$
Has comido, ¿verdad?/¿no?/¿no es cierto?
(b) ni^{3} chi^{1}guo^{4}-le^{0} dui^{4} ma^{0}/ba^{0}
tú comer-PRF correcto $Q_{MA/}Q_{BA}$
[Has comido, ¿verdad?/¿no?/¿no es cierto?]
(c) ni^{3} chi^{1}guo^{4}-le^{0} shi^{4}-bu^{4}-shi^{4}/dui^{4}-bu^{4}- dui^{4}
tú comer-PRF sí-NEG-sí/correcto-NEG-correcto
Has comido, ¿no?/¿no es cierto?
(d) ni^{3} chi^{1}guo^{4}-le^{0} ha^{0}
tú comer-PRF INTERJ.
Has comido, ¿eh?

Además de las construcciones con apéndice que se acaban de describir, el chino dispone de otra estrategia para expresar una pregunta confirmatoria fuerte que se caracteriza por la presencia de la partícula *ba*^{0} (吧) al final de una pregunta total y un rango tonal más reducido que las totales informativas *ma*^{0} (Wang y Ruan, 2005: 348), como en (15).

(15) Contexto: El hablante sabe que su compañero de piso ha suspendido el examen final de matemáticas. Ve que tiene cara triste y los ojos rojos y le pregunta si ha llorado.

CH: ni^{3} ku^{1}-le^{0} ba^{0}
tú llorar-PRF Q_{BA}
¿Has llorado?

En esta situación, basándose en las evidencias directas que tiene, el hablante formula una inferencia sobre el hecho de que ha llorado su compañero de piso, pero no puede estar complemente seguro de que su inferencia sea correcta, de manera que utiliza la partícula *ba^{0}* (que tiene una mayor carga de certeza en comparación con *ma^{0}*) para pedir una confirmación por parte del oyente.

Finalmente, en cuanto a los aspectos léxicos cabe destacar que en las preguntas del tipo (14) y (15) y en las marcadas por un apéndice confirmativo, suelen aparecer delante de los elementos predicativos adverbios de significado afirmativo, como *ken^{3}ding^{4}* (definitivamente), *hen^{3}ke^{3}neng^{2}* (seguramente) y *ba^{1}cheng^{2}* (probablemente), que según (Zhang, 2010: 113) contribuyen a vehicular la fuerza ilocutiva confirmatoria fuerte.

4.1.5. Preguntas eco

En la tradición de estudios sobre la lengua española y otras lenguas indoeuropeas (Bolinger y Bolinger, 1989, entre otros) se han descrito y analizado un conjunto de tipos de preguntas que comparten un rasgo de tipo formal, a pesar de tener funciones pragmáticas distintas. El rasgo formal que comparten es el de repetir, de manera más o menos literal, un enunciado que se acaba de presentar en la conversación. Por el hecho de repetir un enunciado, se han denominado preguntas eco o reiterativas o de retoma. A pesar de compartir el rasgo formal que se acaba de mencionar, las preguntas eco tienen dos funciones pragmáticas claramente diferenciables y, por esa razón, se van a presentar por separado en los próximos dos subapartados. La primera función pragmática es la de reparar un error en la comprensión del enunciado anterior, mientras la segunda es la de expresar sorpresa frente a un contenido proposicional que va en contra de las expectativas del hablante.

i) Eco de comprensión

Las preguntas eco de comprensión se emplean cuando el hablante no ha acabado de percibir o entender bien de todo lo que ha dicho el interlocutor y pregunta repitiendo las informaciones ya conocidas en la conversación, para pedir una confirmación o repetición de la parte desconocida, como se ejemplifica en (16) y en (17).

(16) Contexto: El hablante (B) no sabe qué hora es y se lo pregunta a su compañero de clase (A). Ese le contesta, pero, como hace mucho ruido cuando está hablando, el hablante no consigues entender bien todas sus palabras. Le parece que le haya contestado que son las once, y le pregunta si es eso lo que le ha dicho.

ESP: A: Son las on…
B: ¿(Que) Son las once?

CH: A: shi^{2}…dian^{3}
on…hora
Son las on…
B: (ni^{3}-shuo^{1}) shi^{2}yi^{1} dian^{3}
(tú-decir) once hora
¿(Dices que) Son las once?

(17) Contexto: Tú y tu amigo han quedado esta tarde para ir al cine. Tu amigo te pregunta para saber cuándo sales de la casa. Pero como la señal del teléfono no es buena, no estás seguro de lo que se ha percibido y le preguntas para confirmar.

ESP: A: ¿Cuándo sales?
B: ¿(Que) Cuándo salgo?

CH: A: ni^{3} shen^{2}me^{0}shi^{2}hou^{0} chu^{1}fa^{2}
tú cuándo salir
¿Cuándo sales?
B: (ni^{3}-wen^{4}-wo^{3}) shen^{2}me^{0}shi^{2}hou^{0} chu^{1}fa^{2}
(tú-preguntar-yo.OBJ) cuándo salir
¿(Me preguntas que) Cuándo salgo?

Las preguntas eco de comprensión tanto en español como en chino pueden replicar la estructura sintáctica de la frase que repiten, sea esa una declarativa como en (16) o de una pregunta parcial como en (17): en español la pregunta

eco mantiene el movimiento Qu- de la pregunta parcial original, mientras que en chino se mantiene la palabra Qu-*in situ*. Desde el punto de vista del marcaje léxico, cabe destacar que las preguntas eco en ambas lenguas suelen ir encabezadas por una partícula introductora opcional, que en español es *que* mientras en chino pueden ser las fórmulas *ni*^{3}*wen*^{4}*wo*^{3} (‘me preguntas que…’) o *ni*^{3}*shuo*^{1} (‘dices que…’). Para el español, esos elementos han sido analizados pragmáticamente como marcadores del discurso (Ballesteros, 2000) o como marcas de “interpretación citativa” relacionada al discurso (Gras, 2011: 272).

Desde un punto de vista entonativo, en español, en general las preguntas eco de comprensión como la en (16) se caracterizan por un acento nuclear del tipo L+¡H* seguido por un tono de frontera bajo L% (Escandell-Vidal, 1999). Por otra parte, según Estebas-Vilaplana y Prieto (2010: 35), las preguntas eco que repiten una pregunta parcial como en (17) presentan un acento nuclear diferente, es decir ¡H*, seguido por un tono de frontera bajo. En cuanto al chino, estudios recientes han destacado que F0 desempeña cierto papel en algunos casos. Según Liu (2009), por ejemplo, si las preguntas eco presentan un elemento focalizado, el elemento en cuestión está caracterizado por un movimiento tonal más amplio en comparación al que presenta en las declarativas neutras.

i) Eco de sorpresa

Las preguntas eco de sorpresa se utilizan cuando a alguien le sorprende de lo que ha oído (o incluso, visto o inferido a partir del contexto o una acción del otro interlocutor). La función pragmática principal de estas preguntas no es la de pedir que el interlocutor repita o confirme lo que ha dicho, sino la de expresar sorpresa o incluso incredulidad por parte del hablante, como en el caso (18a).

(18) Contexto: Un buen amigo le dice al hablante que ayer en un restaurante italiano comió grillos. El hablante está muy sorprendido y le pregunta que si comió grillos.

(a)	ESP:	¿(Que) Comiste grillos?			
	CH:	(ni^{3}-shuo^{1})	ni^{3}	chi^{1}-le^{0}	xi^{1}shuai^{4} (ma^{2})
		(tú-decir)	tú	comer-PRF	grillo (Q_{MA})
		¿(Dices que) Comiste grillos?			
(b)	ESP:	¿(Que) Comiste qué?			
	CH:	(ni^{3}-shuo^{1})	ni^{3}	chi^{1}-le^{0}	shen^{2}me^{0}
		(tú-decir)	tú	comer-PRF	qué
		¿(Dices que) Comiste qué?			

Desde el punto de vista sintáctico, en español las totales eco de sorpresa en general repiten la estructura del enunciado que retoman. Sin embargo, si lo que destaca no es el contenido del enunciado entero, sino sólo una parte (como por ejemplo en 18b, en que el interlocutor no se sorprende de que su amigo hubiese comido sino de que hubiese comido grillos), cabe la posibilidad de formular una pregunta con un elemento Qu-*in situ*, lo que difiere del orden típico de Qu-*ex situ* en las preguntas parciales informativas del español. Entonativamente, a diferencia de las eco de comprensión, las eco de sorpresa en español se caracterizan por una configuración nuclear ascendente (del tipo L+H* HH%) (Estebas-Vilaplana y Prieto, 2010: 28–35).

En chino, no se han encontrado descripciones entonativas satisfactorias de las preguntas eco de sorpresa. Sin embargo, dos tipos de preguntas tradicionalmente descritos como preguntas totales informativas neutras (total sin la partícula *ma*; total con *ma*^{2} realizada por un rango tonal expandido y elevado) en realidad, tienen una función de preguntas eco de sorpresa. Al igual que las preguntas eco prototípicas descritas en el contexto (18), la producción de las totales sin *ma* requiere una información contextual o discursiva previa, es decir, no se puede utilizar esta pregunta para iniciar un nuevo tema de conversación (Shao, 2012). Las preguntas sin *ma* (que también conllevan una mayor altura tonal) expresan un matiz de sorpresa o contraexpectativa del hablante (Jin, 2011; Shao, 2012; Wang, 2016). También el tipo de preguntas descritas en (18a), las preguntas con *ma*^{2} realizadas por un rango amplio y superior expresan la sorpresa del hablante hacia el contenido proposicional anterior (Hu, 2005; Lee, 2005; Yuan y Yurie, 2019).

4.2. Ofrecer información: preguntas retóricas

Las preguntas retóricas presentan la fuerza ilocutiva de aserciones. Sin embargo, a pesar de que en estas preguntas no se solicita una información ni se espera una respuesta, desde el punto de vista formal no se niega su estatus interrogativo (Escandell-Vidal, 1984). Las preguntas retóricas, en concreto, tienen la función de ofrecer informaciones y pueden servir en ciertos contextos para transmitir una opinión, una idea, una creencia del hablante, e incluso pueden enfatizar un hecho ya compartido entre los interlocutores (Borillo, 1981). Además, la expresión de una interrogación retórica suele ir acompañada de una fuerte emotividad por parte del emisor (Belchí, 1994; Escandell-Vidal, 1999), como crítica, ironía, desacuerdo, etc.

Una de las características más destacadas de las preguntas retóricas en español reside en su argumentación inversa de la presuposición formulada, es decir, "cuando son totales y se enuncian en forma negativa esperen o admiten una respuesta afirmativa, y viceversa" (Belchí, 1994: 9), como se ejemplifica en (19). Cuando son parciales, como en (20), el mecanismo semántico es el mismo, en el sentido que si la pregunta contiene una negación espera una respuesta positiva (del tipo "Todo el mundo hace X") o doble negativa (del tipo "Nadie no hace X").

(19) Contexto: Los amigos del hablante han planeado pinchar un clavo en el neumático del coche de su profesor de matemática y quieren que el hablante se sume a ellos. El hablante no tiene la más mínima intención de hacerlo y les pregunta si se creen que está loco

ESP:	¿Acaso creéis que estoy loco?				
CH:	nan^{2}dao^{4}	ni^{3}men^{0}	ren^{4}wei^{2}	wo^{3}	feng^{1}-le^{0} ma^{0}
	acaso	vosotros	creer	yo.OBJ	loco-PRF Q_{MA}
	¿Acaso creéis que estoy loco?				

(20) Contexto: El hablante está discutiendo con un buen amigo sobre el tema de la muerte. Su amigo le pregunta si tiene miedo a la muerte y el hablante le pregunta quién no.

ESP:	¿Quién no teme la muerte?		
CH:	shui^{2}	bu^{2}pa^{4}	si^{3} ne^{0}
	quién	NEG-temer	muerte Q_{NE}
	¿Quién no teme la muerte?		

En términos convencionales, algunos autores consideran que las preguntas retóricas se diferencian de otros tipos de pregunta únicamente en los aspectos pragmáticos y que no difieren de las preguntas informativas desde el punto de vista sintáctico y en cuanto al uso de las partículas (Belchí, 1994; Lü, 2002). A pesar de ser cierto lo que esos autores afirman, también hay que destacar que otros autores han puesto de manifiesto que existen algunas estrategias de tipo sintáctico y prosódico que caracterizan las preguntas retóricas frente a las informativas. En concreto, en ambas lenguas, las preguntas retóricas suelen caracterizarse por la utilización de algunos adverbios negativos (acaso '*nan^{2}dao^{4}*'), verbos modales (poder '*neng^{2}*'), términos de polaridad negativa (nadie/ninguno '*mei^{2}you^{3}ren^{2}*'), formulaciones con negaciones externas (¿no es cierto que…? '*bu^{4}shi^{4}...ma^{0}?*'; ¿no crees que…? '*ni^{3}bu^{4}ren^{4}wei^{2}...?*') e incluso unos adverbios o pronombres interrogativos en el enunciado interrogativo (Escandell-Vidal, 1999; Shi, 2018; Yin, 2007).

Desde el punto de vista prosódico, en español, las preguntas retóricas con una connotación emotiva suelen pronunciarse en una tesitura vocal más alta y con un rango tonal expandido (Escandell-Vidal y Prieto, n.d.). En cuanto al chino, además de las mismas estrategias prosódicas descritas en español, se ha descrito la presencia de movimientos tonales más amplios de lo habitual en las marcas formales de negación (Shi, 1980: 76).

5. Estrategias para la expresión de las preguntas de intercambio de bienes y servicios en español y en chino

Una serie de autores (entre los cuales Eggins (2004), Fawcett (2008) y Halliday, Matthiessen y Halliday (2014)) reconocen que en las interacciones entre los hablantes es común ofrecer o solicitar bienes y servicios. Cuando un hablante ofrece bienes o servicios, realiza un acto de habla denominado comisivo; cuando los solicita, su acto de habla es de tipo imperativo. De acuerdo con los actos de habla de Searle (1969), la realización de estas fuerzas ilocutivas mediante una forma interrogativa (como *¿Te puedo ofrecer un café?* o *¿Me podrías prestar 5 euros?*) es un acto de habla indirecto. El hecho de utilizar un acto de habla indirecto, según la teoría de Brown y Levinson (1987) es una estrategia de cortesía negativa, puesto que tanto el acto de ofrecer como el de pedir amenaza la imagen negativa del oyente (Haverkate, 2006; Orozco, 2008). En otras palabras, las preguntas imperativas constituyen una forma atenuada de imperativo (Álvarez y Blondet, 2003: 5).

5.1. Solicitar bienes y servicios: preguntas imperativas

La función de las preguntas imperativas es dar una orden para que el oyente realice una determinada acción. Las preguntas imperativas, que pueden ser tanto totales como parciales, pueden presentar grados diferentes de imperatividad y cortesía; en este apartado, nos ocuparemos en primer lugar de las preguntas que se configuran como corteses y, a continuación, trataremos las descorteses.

Para formular un imperativo cortés, es posible utilizar tanto preguntas totales (21a) como parciales (21b), que en español no difieren sintácticamente de las preguntas informativas, excepto la presencia habitual de la perífrasis verbal *poder+infimitivo*, como en (21a), y algunas expresiones de cortesía, como *por favor* en (21b). Sin embargo, la dife-

rencia en cuanto a la función pragmática es importante: a diferencia de lo que pasaría en una pregunta total informativa, en (21a) el emisor no espera recibir una respuesta que le aclare si el interlocutor sería capaz de pasarle el agua, sino que le está pidiendo que lo haga. De la misma manera, en (21b) el hablante no quiere saber por qué razón el interlocutor se niega a pasarle el agua (circunstancia que no sería cierta), sino que le ordena, de forma atenuada, que lo haga.

Por otra parte, en chino, las preguntas imperativas corteses, tanto totales como parciales, se caracterizan por el uso habitual de la perífrasis verbal *neng*^{2} ('poder')+*verbo*, las partículas *ma*^{0}/*ne*^{0}, así como una construcción especial de orden sintáctico dislocado que se marca con una palabra funcional obligatoria, *ba*^{3} (把), como se observa en (21a) y (21b) en las que el orden es SOV: sujeto-*ba*^{3}-objeto-verbo (Ye, Zhan y Zhou, 2007). La utilización del patrón sintáctico *ba*^{3} tiene ciertas restricciones. Solo se pueden utilizar si el verbo es transitivo y acepta construcciones de doble acusativo, es decir, puede llevar dos complementos directos. Además, el objeto directo de la construcción *ba*^{3} suele ser algo material específico y, por añadidura, debe ser información conocida por los dos interlocutores (Li y Thompson, 1989: 463–468; Ye, Luo, Friederici y Zhou, 2006).

(21) Contexto: El hablante está comiendo con un buen amigo en su piso y, como la comida está muy salada, le entra sed. Como el agua no está a su alcance (está al lado de su amigo) le pregunta si se la puede pasar.

(a)

ESP: ¿Me puedes pasar el agua?

CH:	ni^{3}	neng^{2}-ba^{3}	shui^{3}	di^{4}-gei^{3}	wo^{3} ma^{0}
	tú	poder-ba	agua	pasar-PREP	yo.OBJ Q_{MA}
	¿Me puedes pasar el agua?				

(b)

ESP: ¿Por qué no me pasas el agua, por favor?

CH:	ni^{3}	zen^{3}me^{0}	bu^{4}ba^{3}	shui^{3}	di^{4}-gei^{3}	wo^{3} ne^{0}
	tú	por qué	NEG-ba	agua	pasar-PREP	yo.OBJ Q_{NE}
	¿Por qué no me pasas el agua?					

Desde el punto de vista prosódico, se solía considerar que las peticiones corteses en forma interrogativa (es decir, las preguntas exhortativas) se expresaban en español mediante el mismo contorno entonativo de las preguntas auténticas informativas (Haverkate, 1994; Quilis, 1993; Sosa, 1999; Tomás, 1990). Sin embargo, estudios recientes (Álvarez y Blondet, 2003; Devís Herraiz, 2011; Orozco, 2008; Roldán, 2000) han comprobado que las preguntas imperativas corteses poseen un conjunto de estrategias prosódicas específicas: según Estebas-Vilaplana y Prieto (2010), por ejemplo, en español peninsular central las preguntas totales corteses tienen una configuración nuclear característica que se transcribe con H* M%; para la misma variedad no se ha descrito específicamente la entonación de las parciales imperativas corteses, pero sí la de las parciales corteses de invitación, que presentan un acento nuclear L+¡H* seguido de un tono de frontera HL% (Estebas-Vilaplana y Prieto, 2010: 38). En cuanto al chino, no se han podido encontrar descripciones que las caractericen desde la perspectiva suprasegmental.

Por otro lado, con respecto a las preguntas imperativas descorteses, cabe destacar en primer lugar que en las dos lenguas no se diferencian de las preguntas informativas (parciales o totales, respectivamente), como se observa en los ejemplos (22a) y (22b), extraídos de (Prieto y Roseano, 2009-2013). Sin embargo, tal y como se ha puesto de manifiesto en el caso de la interrogativas imperativas corteses, en (22) el emisor no espera recibir una respuesta que le aclare si los interlocutores tienen la intención de callar o que le explique por qué razón hablan, sino que solicita una acción por parte de sus nietos.

(22) Contexto: Los nietos del hablante hacen mucho ruido y no le dejan oír las noticias en la televisión/radio. El hablante le pide que se callen.

(a)	ESP:	¿Queréis callar?		
	CH:	ni^{3}men^{0}	neng^{2}	bi^{4}zui^{3} ma^{0}
		vosotros	poder	callar Q_{MA}
		¿Podéis callar?		

(b)	ESP:	¿Por qué no os calláis?		
	CH:	ni^{3}men^{0}	zen^{3}me^{0}	bu^{4}bi^{4}zui^{3}
		vosotros	por qué	NEG-callar
		¿Por qué no os calláis?		

Desde un punto de vista entonativo, las preguntas imperativas descorteses del español se caracterizan por una configuración nuclear H+L* L% y una velocidad de habla más rápida y un rango tonal relativamente reducido que las preguntas informativas (Estebas-Vilaplana y Prieto, 2010: 29-38). Con respecto al chino, debido a la carencia de estudios relativos, no se ha podido caracterizar la propiedad entonativa de este tipo de interrogativas.

5.2. Ofrecer bienes y servicios: preguntas comisivas

La función principal de las preguntas comisivas consiste en ofrecer al oyente algo material como en el ejemplo (23), o algo inmaterial como una invitación o sugerencia como en (24). En ambas lenguas, una pregunta comisiva se puede contestar con *Sí/No* seguidos de una fórmula de agradecimiento (como *gracias*), o incluso un solo agradecimiento (RAE, 2009: 3162). Sintácticamente, tanto en chino como en español, las preguntas totales de ofrecimiento no se diferencian de las totales informativas.

(23) Contexto: Un amigo ha ido de visita a casa del hablante. Cuando llega, el hablante le ofrece une cerveza.

ESP: ¿Te apetece una cerveza?

CH:	ni^{3}	xiang^{3}lai^{2}-bei^{1}	pi^{2}jiu^{3} ma^{0}
	tú	querer-vaso.CLF	cerveza Q_{MA}
	¿Quieres un vaso de cerveza?		

Desde el punto de vista léxico, en ambas lenguas pueden aparecer verbos que indican la volición: en español aparecen a menudo los verbos *querer*, *apetecer* o *gustar* (como en el ejemplo 24), mientras en chino se utiliza, además, con frecuencia una construcción del tipo X-no-X que se fundamenta en los verbos de volición, como *xiang^{3}-bu^{4}-xiang^{3}* (‘desear no desear’) en (24).

(24) Contexto: Después de cenar con un amigo en un restaurante, el hablante se siente muy lleno y le preguntas a su amigo si quiere dar un paseo con él.

ESP: ¿Quieres dar un paseo?

CH:	ni^{3}	xiang^{3}-bu^{4}-xiang^{3}	qu^{4}	san^{4}bu^{4}
	tú	desear-NEG-desear	ir	pasear
	¿Quieres dar un paseo?			

Con respecto a la caracterización entonativa de las preguntas totales comisivas, las descripciones tradicionales del español afirman que se distinguen de las preguntas totales informativas y las de petición por una curva melódica de final bajo-descendente (RAE, 2011: 471). En cuanto al chino, no se han documentado hasta la fecha diferencias entonativas entre ellas, y la diferenciación se realiza a menudo a través del significado semántico y la situación contextual en que se genera el enunciado.

6. Resumen y discusión

En los apartados anteriores, se han presentado las estrategias lingüísticas que se usan en español y en chino para codificar las distintas funciones pragmáticas de la modalidad interrogativa. En esta sección, siguiendo los resultados resumidos en la tabla 1, se pretende discutir las similitudes y diferencias entre las estrategias adoptadas por las dos lenguas para expresar una misma función pragmática.

i) Preguntas informativas. En primer lugar, para la realización de una función pragmática informativa, es decir, preguntas que piden información nueva, existen tres tipos de interrogativas informativas en las dos lenguas que se pueden distinguir por a) el tipo de la información nueva que se quiere solicitar y b) la forma de contestarlas (sí/no: totales o con un elemento: parciales).

Las preguntas parciales informativas se pide una información específica de una incógnita. En español este tipo de preguntas están caracterizas por una configuración nuclear L* L% y un desplazamiento sintáctico de tipo Q-; mientras el chino mantiene la palabra Qu-*in situ*. Además, en chino, de manera opcional, estas preguntas pueden ir marcadas por una partícula *ne^{0}*. La presencia de la partícula *ne^{0}* da énfasis al tono inquisitivo y pone en manifestado que el hablante tiene cierto conocimiento hacia lo que se pregunta (Jin, 2011; Shi, 2004, 2006), por lo que se podría defender que en realidad no son informativas, sino que tienen un ligero sesgo confirmatorio.

Tabla 1. Estrategias de expresión y funciones pragmáticas de modalidad interrogativa en español y en chino.

	Español	**Chino**
Tipo	**Estrategia**	**Estrategia**
Preguntas de intercambio de informaciones		
Parciales Informativas	Palabra Qu- + Qu-*ex situ* + inversión sujeto-verbo + L* L%	Palabra Qu- + Qu-*in situ* + (partícula ne^0) + rango tonal más amplio que la declarativa
Totales informativas	VSO + L* HH%	1) SVO + ma^0 + entonación declarativa; 2) construcción X-no-X + L%
Disyuntivas informativas	Conjunción *o* + L+H* HH- L*L%	(hai^2shi^4) + (ne^0/a^0) + rango elevado y amplio en el primer constituyente; bajo y reducido en el último
Confirmatorias débiles	Pregunta total + H+L* L%	1) Pregunta total + adverbio afirmativo *shi* + ma^0 + entonación declarativa; 2) construcción *shi-bu-shi VP* + L%
Confirmatorias fuertes	Apéndice confirmativo + L* H%	1) Apéndice confirmativo + entonación no descrita; 2) pregunta total + ba^0 + rango tonal más bajo que las totales ma^0
Eco de comprensión	(Partícula introductora *que*) + réplica estructural + L+¡H* L%	(Palabra introductora ni^3shuo^1/ni^3wen^4) + réplica estructural + entonación no descrita
Eco de sorpresa	1) Total: (Que) + L+ H* HH%; 2) Parcial: (Que) + Qu-*ex situ* + L+¡H* HH%	1) (ni^3shuo^1/ni^3wen^4) + réplica estructural + entonación no descrita; 2) total sin *ma* + altura tonal más alta; 3) total con ma^2 + rango expandido y superior
Retóricas	*(Acaso, nadie/ninguno, ¿no es cierto que…?/ ¿no crees que…?/ ¿no es verdad que…?, poder, etc).* + (supuesto compartido) + entonación alta y rango expandido	(nan^2dao^4, $shui^2bu^4$…?, $neng^2$, $bu^4shi^4...ma^0$?, $ni^3bu^4ren^4wei^2$…?, *etc.*) + (supuesto compartido) + rango alto y expandido + focalización en las marcas de negación
Preguntas de intercambio de bienes y servicios		
Imperativas corteses	*Poder* + inf. + (por favor) + H%	$Neng^2$ + construcción ba^3 + SOV + ma^0/ne^0 + entonación no descrita
Imperativas descorteses	H+L* L% + velocidad de habla rápida y rango tonal reducido	No se encuentran diferencias morfosintácticas + entonación no descrita
Comisivas	(*Desear, querer, gustar, apetecer, etc.*) + L+¡H* HL% para las preguntas parciales de ofrecimiento de invitación	Verbo/construcción de volición $xiang^3$ *o* $xiang^3bu^4xinag^3$ + entonación no descrita

Las preguntas totales informativas requieren una respuesta sí/no. Estas se caracterizan en español por un orden sintáctico VSO y un patrón entonativo final ascendente L* HH%; mientras el chino, dispone de dos grupos de estrategias para conseguir esta función pragmática. La primera es la realización de la pregunta total mediante la partícula ma^0, de tono neutro; la segunda se realiza mediante una estructura sintáctica *X-no-X* y un tono de frontera bajo-descendente. Los dos tipos de preguntas totales informativas son intercambiables cuando el hablante no se inclina hacia una respuesta determinada (Yuan y Yurie, 2019). La diferencia entre ellas reside en que la partícula ma^0 expresa un matiz de cortesía/formalidad (Wang, 2016).

Las preguntas disyuntivas informativas, comparten una realización prosódica general en las dos lenguas. Pero a diferencia del español, la presencia de la conjunción hai^2shi^4 ("o" en español) no es obligatoria en chino si existe una partícula modal para conectar las diferentes alternativas.

ii) Preguntas confirmatorias. Según el grado de creencia que el hablante posee ante el predicado, las preguntas se describen como confirmatorias. Estas se pueden subdividir a su vez en confirmatorias fuertes o débiles, según el grado de certeza del hablante.

Las preguntas confirmatorias débiles están codificadas en español por un tonema descendente L* L% típico de las declarativas. Mientras en chino, las estrategias principales para realizar la misma función consisten en la construcción *shi-bu-shi VP* y el tono de frontera bajo L%.

Las confirmatorias fuertes se realizan en las dos lenguas mediante unos apéndices confirmativos e interjecciones (por ejemplo, *¿no?, ¿verdad?, ¿eh?,* etc.) que contribuyen a enfatizar las expectativas del hablante por obtener una confirmación de lo que se emite en la parte anterior del enunciado, el cual tiene normalmente una forma de declarativa. Prosódicamente, en español estos apéndices se pronuncian con un tonema ascendente L* H% de los interrogativos prototípicos, mientras en chino no se han encontrado datos relativos a la caracterización entonativa de los mismos. Aparte de eso, en chino existe una segunda manera para realizar las confirmatorias fuertes, que está

caracterizada por el uso de la partícula final *ba*^{0} y un rango tonal más bajo que las totales informativas *ma*^{0} (Wang y Ruan, 2005).

iii) Preguntas eco. Las preguntas eco de comprensión son aquellas en las que se pide una repetición o confirmación de la información anterior. Se expresan, en las dos lenguas, mediante una réplica sintáctica de las preguntas originales y, de manera opcional, una marca introductora que encabeza la pregunta.

En español, las preguntas eco presentan diferencias según si son de comprensión o de sorpresa. Las eco de comprensión se caracterizan por un acento nuclear ascendente y un tono de frontera final bajo (L+¡H* L%). En cambio, las eco de sorpresa se caracterizan por un tono de frontera final alto-ascendente HH%. Además, las eco de sorpresa del español pueden presentar una estructura Qu-*in situ* imposible en las parciales originales.

En chino, las preguntas eco siempre se mantienen una forma de Qu-*in situ*. Por lo demás, se expuesto que los dos tipos que tradicionalmente se consideran en chino como interrogativas totales informativas (total sin la partícula *ma*; total con *ma*^{2}+rango expandido y superior) se acercan más a una pregunta eco, ya que expresan una sorpresa del hablante hacia lo que ha visto, oído e incluso inferido del contexto.

iv) Preguntas retóricas. La función asertiva de la modalidad interrogativa se realiza por medio de las preguntas retóricas, que en las dos lenguas se codifica de forma muy similar. Básicamente, la argumentación afirmativa se consigue en español y en chino a través de un conjunto alternativo de estrategias morfosintácticas tales como el uso de un adverbio negativo (por ejemplo, *acaso),* término de polaridad negativa (como *nadie*, *ninguno*, etc.), formulación con negación externa (como *¿No es verdad que...?*), el verbo modal *poder*, así como algunas palaras interrogativas. Con respecto a las estrategias prosódicas, en general, las preguntas retóricas se caracterizan por una altura tonal más alta y un rango expandido en las dos lenguas.

v) Preguntas imperativas y comisivas. Por último, la función pragmática imperativa y comisiva se realiza, morfosintácticamente, de manera similar a las preguntas informativas en chino y en español.

En concreto, para la realización de un imperativo cortés, el español usa sobre todo la perífrasis verbal *poder+inf.*, algunas expresiones corteses (por ejemplo, *por favor*), así como un final entonativo mayormente ascendente. En chino, tanto los imperativos corteses como los descorteses se distinguen de las interrogativas neutras por la presencia de una construcción *ba*^{3} de orden sintáctico SOV.

Por otro, el español diferencia los imperativos descorteses de las preguntas neutras entonativamente mediante: una configuración nuclear H+L* L%, una velocidad de habla más rápida y un rango general reducido. En chino, debido a la carencia de estudios, no se ha podido describir este tipo desde un punto de vista entonativo.

Por último, la función pragmática de ofrecimiento se expresa mediante las preguntas comisivas y se caracteriza en español y en chino por la utilización de unos verbos de volición (por ejemplo, *querer*, *apetecer*, *gustar*, etc.). En cuanto a la caracterización prosódica de este tipo interrogativo, aunque no está claramente descrita ni en cualquiera de las dos lenguas, en español, tradicionalmente, se ha considerado que estas preguntas presentan una curva melódica de final descendente (RAE, 2011: 471), específicamente, un patrón entonativo L+¡H* HL% en caso de que sea un ofrecimiento de invitación (Estebas-Vilaplana y Prieto, 2010: 38).

7. Conclusiones

Los resultados que hemos discutido en este trabajo suponen el primer intento de análisis contrastivo español-chino sobre las estrategias de expresión de modalidad interrogativa a partir de un punto de vista funcional-pragmático.
En concreto, a través del estudio comparativo, se ha evidenciado el carácter general e interidiomático de los valores funcional-pragmáticos de la modalidad interrogativa y las diferencias en las estrategias lingüísticas adoptadas por cada lengua.

Más específicamente, se ha podido confirmar que, en general, la modalidad interrogativa se asemeja en las dos lenguas en las funciones pragmáticas que se pueden realizar dentro de las dos dimensiones fundamentales del intercambio comunicativo: intercambio de informaciones; intercambio de bienes y servicios. En la primera categoría, la manera de pedir y ofrecer información está vinculada al grado de certeza que tiene el hablante hacia el contenido proposicional y, a veces, a la necesidad de reproducir lingüísticamente o pragmáticamente el discurso anterior (por ejemplo, en las preguntas eco). En la segunda clase, el acto de pedir y ofrecer bienes y servicios (en concreto, actos imperativos y comisivos) también puede conllevar una atenuación o intensificación de acuerdo con la voluntad del enunciador, la cortesía, la relación social entre los participantes o el efecto comunicativo que se quiere conseguir en la interacción conversacional.

Desde el punto de vista de su realización lingüística, nuestro análisis reafirma el carácter sintético y analítico del español, donde además del modo verbal, la entonación y el orden sintáctico tienen un peso expresivo relevante en la expresión de un acto de habla de las interrogativas. Mientras en chino, en cambio, debido al carácter analítico, la codificación y la diferenciación de las funciones pragmáticas se realizan en su mayoría a través de las partículas modales y algunas construcciones sintácticas. Además, en comparación con el español cuya entonación se describe

en términos de ascensos y descensos en el tonema, el chino, al ser una lengua tonal, utiliza los parámetros de rango tonal y de duración para caracterizar las propiedades entonativas. Pero en general, se observa que el chino presenta un menor grado de dependencia de la entonación debido al marcaje morfológico de partículas modales.

Este hecho pone de relieve la necesidad de diseñar métodos educativos para practicar y mejorar la producción entonativa del español por parte de los sino-hablantes, a la vez que permite determinar en qué casos las diferencias entre las funciones expresivas de entonación y un proceso entonativo distinto pueden causar una transferencia prosódica negativa entre el español y el chino.

Agradecimiento

Esta investigación constituye parte de un proyecto doctoral financiado por China Scholarship Council (No. 201808390059).

Apéndice: Lista de abreviaturas

ADV	Adverbio
CLF	Clasificador
CONJ	Conjunción
INF	Infinitivo
INTERJ	Interjección
NEG	Negación, negativo
OBJ	Objeto
PREP	Preposición
PRF	Perfecto
Q	Partícula interrogativa
VP	Sintagma verbal

Bibliografía

Aikhenvald, Alexandra Yurievna. (2003). Evidentiality in typological perspective. *Typological Studies in Language*, *54*, 1–32. https://doi.org/10.1075/tsl.54.04aik.

Álvarez, Alexandra, y Blondet, María Alejandra. (2003). Cortesía y prosodia: un estudio de la frase cortés en el español de Mérida (Venezuela). *La Tonía. Dimensiones Fonéticas y Fonológicas*, 319–330. http://dx.doi.org/10.4067/S0718-48832013000200007.

Belchí, Dolores Anunciación Igualada. (1994). Estrategias comunicativas: la pregunta retórica en español. *Revista Española de Lingüística*, *24*(2), 329–344. http://www.sel.edu.es/pdf/jul-dic-94/24-2-Igualada.pdf.

Blakemore, Diane. (1994). Echo questions: A pragmatic account. *Lingua*, *94*(4), 197–211. https://doi.org/10.1016/0024-3841(94)90009-4.

Bolinger, Dwight, y Bolinger, Dwight Le Merton. (1989). *Intonation and its uses: Melody in grammar and discourse*. Stanford university press.

Borillo, Andrée. (1981). Quelques aspects de la question rhétorique en français. *DRLAV. Documentation et Recherche En Linguistique Allemande Vincennes*, *25*(1), 1–33. https://www.persee.fr/doc/drlav_0754-9296_1981_num_25_1_969.

Chafe, Wallace, y Nicolás, Johanna. (Eds.). (1986). *Evidentiality: The linguistic coding of epistemology*. Norwood, N.J.: Ablex Pub. Corp. https://doi.org/10.1525/aa.1988.90.1.02a00770.

Chen, Zhi'an 陈治安, y Wen, Xu 文旭. (2001). 论言语交际中的回声话语 [On Echoic Utterances in Verbal Comunication]. 解放军外国语学院学报 *[Journal of PLA University of Foreign Language]*, *24*(4), 24–17.

Cheng, Lisa Lai Shen. (1997). *On the typology of wh-questions*. Taylor & Francis.

Uribe, Miriam Elizabeth Cid. (1996). La entonación de las preguntas ratificadas en español y de sus equivalentes "Question tags" en inglés: una comparación. *Onomázein: Revista de Lingüística, Filología y Traducción de La Pontificia Universidad Católica de Chile*, (1), 56–70.

Cuenca, Maria Josep. (1997). Form-use mappings for tag questions. *AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES 4*, 3–20. https://doi.org/10.1075/cilt.151.04cue.

Devís Herraiz, Empar. (2011). Rasgos melódicos de la cortesía atenuadora en el español coloquial. *Moenia: Revista Lucense de Lingüistica & Literatura*, (17), 475–490.

Eggins, Suzanne. (2004). *Introduction to systemic functional linguistics*. A&C Black.

Escandell-Vidal, María Victoria. (1984). La interrogación retórica. *Dicenda. Estudios de Lengua y Literatura Españolas*, (3), 9.

Escandell-Vidal, María Victoria. (1999). Los enunciados interrogativos. Aspectos semánticos y pragmáticos. *BOSQUE, I.; DEMONTE, V. Gramática Descriptiva de La Lengua Española: Entre La Oración y El Discurso–Morfología*, *3*. https://doi.org/10.5817/ERB2016-2-15.

Escandell-Vidal, María Victoria, y Prieto, Pilar. (n.d.). Pragmatics and Prosody. In C. Félix-Brasdefer & D. A. Koike (Eds.), *The Routledge Handbook of Spanish Pragmatics*.

Escandell-Vidal, María Victoria. (2002). Echo-syntax and metarepresentations. *Lingua*, *112*(11), 871–900. https://doi.org/10.1016/S0024-3841(02)00051-7.

Real Academia Española. (2009). *Nueva Gramática de la lengua Española. Sintaxis II*. Baecelona: Espasa Libros, S.L.U.

Real Academia Española. (2011). Nueva gramática de la lengua española. Fonética y Fonología. Barcelona: Espasa Libros, SLU.

Estebas-Vilaplana, Eva, y Prieto, Pilar. (2010). Castilian Spanish intonation. *Transcription of Intonation of the Spanish Language*, 17–48.

Fan, Jiyan 范继淹. (1982). 是非问句的句法形式 [The Syntactic Structure of Yes-no Questions]. 中国语文 *[Studies of the Chinese Language]*, *6*, 426–434.

Fawcett, Robin. (2008). *Invitation to systemic functional linguistics through the Cardiff grammar: An extension and simplification of Halliday's systemic functional grammar*. United Kingdom: Equinox. https://doi.org/10.1075/fol.17.2.04bac.

Cai, Yazhi 蔡雅芝. (2019). Estudio comparativo del eufemismo en chino y español. *Círculo de Lingüística Aplicada a la Comunicación*, (77), 3-20. https://doi.org/10.5209/CLAC.63267.

Gao, Hua 高华, y Zhang, Wei 张惟. (2009). 汉语附加问句的互动功能研究 [The Functions of Tag Questions in Chinese: An Interactional Analysis]. 语言教学与研究 *[Language Teaching and Linguistic Studies]*, *5*, 45–52.

González, Monserrat, Roseano, Paolo, Borràs-Comes, Joan, y Prieto, Pilar. (2017). Epistemic and evidential marking in discourse: Effects of register and debatability. *Lingua*, *186*, 68–87. http://dx.doi.org/10.1016/j.lingua.2014.11.008.

Gras, Pedro. (2011). *Gramática de construcciones en interacción. Propuesta de un modelo y aplicación al análisis de estructuras independientes con marcas de subordinación en español*. Universitat de Barcelona: Tesis doctoral.

Guardado, Amalia Lorena Camacho. (2017). Mecanismos de modalidad interna y externa para expresar probabilidad en el aula de E/LE. *Normas: Revista de Estudios Lingüísticos Hispánicos*, *7*(1), 107–125. http://dx.doi.org/10.7203/Normas.7.10427.

Guo, Jimao 郭继懋. (1997). 反问句的语义语用特点 [The Semantic and Pragmatic Properties of Rhetorical Questions]. 中国语文 *[Studies of the Chinese Language]*, *2*, 111–121.

Guo, Rui 郭锐. (2000). "吗"问句的确信度和回答方式 [The Certainty and Forms of Answer in "Ma" Questions]. 世界汉语教学 *[Chinese Teaching in the World]*, (2), 13–23.

Halliday, Michael Alexander Kirkwood, Matthiessen, Christian, y Halliday, Michael. (2014). *An introduction to functional grammar* (3rd ed.). Routledge.

Halliday, Michael Alexander Kirkwoo, y McDonald, Edawrd. (2004). Metafunctional profile of the grammar of Chinese. *Language Typology: A Functional Perspective*, *253*, 305–396. https://doi.org/10.1075/cilt.253.08hal.

Haverkate, Henk. (1994). *La cortesía verbal. Estudio pragmalingüístico*. Madrid: Gredos. https://revistas.ucm.es/index.php/DIDA/article/view/DIDA9595110177A.

Haverkate, Henk. (2006). Aspectos pragmalingüísticos de la interrogación en español con atención especial a las secuencias de preguntas. *Cultura, Lenguaje y Representación*, *III*, 27–40. http://dx.doi.org/10.6035/CLR.

Hidalgo Navarro, Antonio. (2015). Prosodia y partículas discursivas: sobre las funciones de atenuación, intensificación como valores (des)corteses en los marcadores conversacionales. *Círculo de Lingüística Aplicada a La Comunicación*, (62), 76–104. https://doi.org/10.5209/rev_CLAC.2015.v62.49499.

Hu, Fang 胡方. (2005). 普通话疑问词韵律的语音学分析 [A phonetic study of prosody of Wh-words in Standard Chinese]. 中国语文 *[Studies of the Chinese Language]*, (3), 269–278.

Huang, Guoying 黄国营. (1986). "吗"字用法初探 [A Primary Study of the Particle "Ma"]. 语言研究 *[Studies in Language and Linguistics]*, (2), 131–135.

Jin, Zhiyan 金智研. (2011). 现代汉语句末语气词意义研究 *[Study of the Meaning of Modal Particles in Modern Chinese]*. 复旦大学博士论文 [Doctoral dissertation of Fudan University].

Lee, Ok Joo. (2005). *The prosody of questions in Beijing Mandarin*. The Ohio State University: Doctoral dissertation.

Li, Charles, y Thompson, Sandra. (1989). *Mandarin Chinese: A functional reference grammar*. University of California Press. https://doi.org/10.2307/2055534.

Li, Xiaofan 李小凡. (1990). 也谈反复问句 [On the A-not-A Questions]. In 语言学与汉语教学 *[Linguistics and Chinese Language Teaching]* (pp. 179–188). 北京语言学院出版社 [Beijing Language and Culture College Press].

Liu, Fang, y Xu, Yi. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica*, *62*(2–4), 70–87. https://doi.org/10.1159/0000 90090.

Liu, Yanli 刘妍李. (2009). 汉语焦点重音位置对回声疑问句句尾语调的影响初探 *[On the intonation of Chinese echo question influenced by the stressed position of focus]*. 暨南大学硕士论文 [Master's thesis of Jinan University].

Liu, Yuehua 刘月华, y Pan, Wenyu 潘文娱. (2001). *Shiyong Xiandai Hanyu Yufa (*实用现代汉语语法*)*. Beijing: 商务印书馆 [The Commercial Press].

Lü, Shuxiang 吕叔湘. (2002). *Zhong'guo Wenfa Yaolüe(*中国文法要略*)*. 辽宁教育出版社[Liaoning Education Press].

Yuan, Mengxi 袁梦溪, y Yurie, Hara. (2019). 现代汉语"吗"问句与"A 不 A"问句的语义差异 [The Semantic Distinction between Chinese ma Questions and A-not-A Questions]. 当代语言学 [*Contemporary Linguistics*], *21*(1), 43–67.

Vidal, Isabel Molina. (2014). Entonación, Intención y Relevancia. La importancia de la entonación y su enseñanza en el aula de ELE. Algunas propuestas didácticas. *Revista de Didáctica ELE*, *19*(1885–2211), 1–21.

Mora, Hang Ferrer. (1999). Los modos oracionales en la gramática española: rasgos morfológicos, sintácticos, semánticos y pragmáticos para una taxonomía. *Linguistik Online*, *4*, 3.

Navarro Tomás, Tomás. (1990). *Manual de pronunciación española*. Madrid: CSIC Press.

Noh, Eun-Ju. (1998). Echo questions: Metarepresentation and pragmatic enrichment. *Linguistics and Philosophy*, *21*(6), 603–628.

Orozco, Leonor. (2008). Peticiones corteses y factores prosódicos. *Fonología Instrumental. Patrones Fónicos y Variación*, 335–355.

Palmer, Frank Robert. (2001). *Mood and modality* (2nd ed.). New York: Cambridge University Press.
Ballesteros, Margarita Porroche. (2000). Algunos aspectos del uso de que en el español conversacional (“que” como introductor de oraciones “independientes”). *Círculo de Lingüística Aplicada a La Comunicación*, (3), 110–116.
Prieto, Pilar, y Roseano, Paolo. (2009-2013). Atlas interactivo de la entonación del español. Retrieved from http://prosodia.upf.edu/atlasentonacion/.
Quilis, Antonio. (1993). *Tratado de fonología y fonética españolas* (Vol. 2). Gredos Madrid.
Roldán, Yasna. (2000). Correlatos acústicos de actos de habla atenuados del español de Chile. *Onomázein: Revista de Lingüística, Filología y Traducción de La Pontificia Universidad Católica de Chile*, (5), 107–118.
Roseano, Paolo., Fernández Planas, Ana María, Elvira-García, Wendy, Massó, Ramón Cerda, y Martínez Celdrán, Eugenio. (2015). La entonación de las preguntas parciales en catalán. *Revista Española de Lingüística Aplicada*, *28*(2), 511–554. https://doi.org/10.1075/resla.28.2.07ros.
Searle, John Rogers. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge university press.
Shao, Jingmin 邵敬敏. (2014). 现代汉语疑问句研究 *[The Study of Interrogatives in Chinese Mandairn]*. Beijin: 商务印书馆 [The Commercial Press].
Shao, Jingmin 邵敬敏, y Zhu, Yan 朱彦. (2002). “是不是 VP” 问句的肯定性倾向及其类型学意义 [The affirmative inclination of the shì-bù-shì + VP question and its typological significance]. 世界汉语教学 *[Chinese Teaching in the World]*, (3), 23–36.
Shao, Jingmin 邵静敏. (2012). 是非问句内部类型的比较以及“疑惑”的细化 [The comparison of the internal types of yes-no questions and the subdivision of incertitude]. 世界汉语教学 *[Chinese Teaching in the World]*, *26*(3), 347–356.
Shen, Xiao-nan Susan. (1990). *The prosody of Mandarin Chinese* (Vol. 118). Universiy of California Press.
Shi, Fenru 史芬茹. (2018). 言语行为视角下的现代汉语反问句研究 [A Study of Chinese Rhetorical Questions from the Perspective of Speech Acts]. 语言文字运用 *[Applied Linguistics]*, (3), 143. https://doi.org/10.16499/j.cnki.1003-5397.2018.03.019.
Shi, Peiwen 石佩雯. (1980). 四种句子的语调变化 [The Intonational Variation in Four Sentence Types]. 语言教学与研究 *[Language Teaching and Linguistic Studies]*, *2*, 71–81.
Shi, Yuzhi 石毓智. (2004). 汉语研究的类型学视野 *[Linguistic Typology and Chinese Language]*. 江西教育出版社 [Jiangxi Education Press].
Shi, Yuzhi 石毓智. (2006). 现代汉语疑问标记的感叹用法 [The Exclamative Use of Question Markers in Chinese Mandarin]. 汉语学报 *[Chinese Linguistics]*, *4*, 14–26.
Sosa, Juan Manuel. (1999). *La entonación del español: Su estructura fónica, variabilidad y dialectología*. Madrid: Cátedra.
Sperber, Dan, y Wilson, Deirdre. (1986). *Relevance: Communication and cognition* (Vol. 142). Cambridge MA: Harvard University Press.
Tao, Lian 陶炼. (1998). “是不是” 问句说略 [A Study of “Shi-bu-shi” Questions]. 中国语文 *[Studies of The Chinese Language]*, (2), 105–107.
Wang, Jue 王珏. (2016). 再论“吗”的属性、功能及其与语调的关系 [On the attribute, functions and the relation between “ma” and sentence intonation]. 汉语学习 *[Chinese Language Learning]*, (5), 3–13.
Wang, Yunjia 王韫佳, y Ruan, Lüna 阮吕娜. (2005). 普通话疑问句语调的实验研究 [The Experimental Investigation of the Intonation in Chinese Mandarin’s Interrogatives]. In 第八届全国人机语音通讯学术会议论文集 *[The 8th National Conference on Man-Machine. Speech Communication]* (pp. 344–349).
Wilson, Deirde, y Sperber, Dan. (2002). Relevance theory. Blackwell.
Xing, Fuyi 刑福义. (2002). 汉语语法三百问 *[Three Hundred Questions towards Chinese Mandarin Grammar]*. Beijing: 商务印书馆 [The Commercial Press].
Xu, Jie 徐杰, y Zhang, Linlin 张林林. (1985). 疑问程度和疑问句式 [The Interrogativity and Interrogative Constructions]. 江西师范大学学报 *[Journal of Jiangxi Normal University]*, *2*, 71–79.
Yan, Jinting 阎锦婷, Wang, Ping 王萍, y Shi, Feng 石锋. (2014a). 普通话疑问标记复用的声学实验——以正反疑问句为例 [Acoustic Analysis on Mark Multiplexing in Interroagtive Sentences: for the Positive and Negative Interroagitve Sentences]. 语言教学与研究 *[Language Teaching and Linguistic Studies]*, (5), 97–102.
Yan, Jinting 阎锦婷, Wang, Ping 王萍, y Shi, Feng 石锋. (2014b). 普通话选择问句的语调格局 [The Intonation of Disyunctive Questions in Chinese Mandarin]. 语言文字运用 *[Applied Linguistics]*, (1), 53–59.
Yan, Jinting 阎锦婷, Wang, Ping 王萍, y Shi, Feng 石锋. (2016). 普通话特指问句语调的声学实验分析 [An Acoustic Analysis of Specific Interrogation of Mandarin]. 汉藏语学报 *[Journal of Sino-Tibetan Linguistics]*, (1), 121–129.
Ye, Zheng, Luo, Yue-jia, Friederici, Angela D., y Zhou, Xiaolin. (2006). Semantic and syntactic processing in Chinese sentence comprehension: Evidence from event-related potentials. *Brain Research*, *1071*(1), 186–196. https://doi.org/10.1016/j.brainres.2005.11.085.
Ye, Zheng, Zhan, Weidong, y Zhou, Xiaolin. (2007). The semantic processing of syntactic structure in sentence comprehension: An ERP study. *Brain Research*, *1142*, 135–145. https://doi.org/10.1016/j.brainres.2007.01.030.
Yin, Shulin 殷树林. (2007). 现代汉语反问句特有的句法结构 [The Special Syntactic Structures of Rhetorical Questions in Modern Chinese]. 湖南科技大学学报 *(社会科学版) [Journal of Hunan University of Science & Technology (Social Sicence Edition)]*, *10*(3), 101–105.
Yuan, Mengxi, y Yurie, Hara. (2013). Questioning and asserting at the same time: the L% tone in A-not-A questions. In *Proceedings of the 19th Amsterdam Colloquium* (pp. 265–272).
Yurie, Hara. (2014). Semantics and Pragmatics of Cantonese Polar Questions: an inquisitive approach. In *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing* (pp. 605–614).
Zhang, Hongxin 张洪欣. (2010). 疑问句系统的疑问程度与语用功能的关系之研究 [A Study on the Relationship between Interrogativity Levels and Pragmatic Functions in Interrogative System]. 济南职业学院学报 *[Joural of Jinan Vocational College]*, (3), 112–114.

Capítulo 3

Second language acquisition of Spanish prosody by Chinese speakers: Nuclear contours and pitch characteristics

Citado como:

Shang, P. & Elvira-García, W. Second language acquisition of Spanish prosody by Chinese speakers: Nuclear contours and pitch characteristics. *Vigo International Journal of Applied Linguistics*, (19), 129-176. doi: https://doi.org/10.35869/vial.v0i19.376

Second language acquisition of Spanish prosody by Chinese speakers: Nuclear contours and pitch characteristics

Peizhu Shang
Department of Catalan Philology and General Linguistics
University of Barcelona, Spain
pshangsh7@alumnes.ub.edu

Wendy Elvira-García
Department of Spanish Language and General Linguistics
UNED, Spain
welvira@flog.uned.es

Abstract

Despite the increasing number of studies in L2 prosody, little research has been carried out on the Chinese-Spanish language pair. This article sets out to examine the L2 acquisition of nuclear contours and pitch implementation details of Spanish spoken by Chinese speakers. To this end, 555 utterances (produced by 37 informants) were analyzed within an autosegmental-metrical framework, and pitch values were evaluated using long-term distributional (LTD) and pitch dynamism quotient (PDQ) measures. The results suggest a hierarchy of difficulties in acquiring the prosodic features of different sentence types. The most salient intonational error made by the Chinese learners was the tendency to replace low nuclear accents with high/rising tones. Furthermore, the higher pitch level, narrower span, and lower F0 variance found for Chinese speakers lend support to previous hypotheses which proposed a general pitch compression pattern for L2 speech. Nevertheless, with increasing proficiency in Spanish, learners appear to develop more target-like intonation contours and pitch profiles. Finally, gender and stress effects as well as other interactions prove that L2 prosody learning is more complex than previously stated, and is influenced not only by the L1 system and oral competence but is also correlated with some psychological and sociocultural factors.

Key words: Nuclear contour; pitch characteristics; L2 speech acquisition; prosodic transfer; Peninsular Spanish by L1 Chinese speakers.

Resumen

A pesar del creciente número de estudios sobre L2, el par de lenguas chino-español ha sido abordado en escasos estudios. Este artículo examina la adquisición de los contornos entonativos y los detalles de implementación tonal del español por parte de los sinohablantes. Para hacerlo, se han analizado 555 oraciones producidas por 37 informantes usando la teoría métrico-autosegmental, y se han evaluado acústicamente sus valores tonales a través de las medidas de distribución a largo plazo (LTD) y el cociente del dinamismo de tono (PDQ). Los resultados sugieren una jerarquía de dificultad en la adquisición de los patrones entonativos relacionada con los diferentes tipos oracionales. Además, el error entonativo más destacado para los sinohablantes es la tendencia a reemplazar los acentos nucleares bajos por tonos altos/ascendentes. Además, nuestros resultados apoyan las hipótesis que proponen una tendencia general de compresión tonal en la L2, ya que se ha documentado un nivel tonal más alto, un rango más reducido y una variación menor de F0 en los sinohablantes que en el grupo control. Sin embargo, los aprendices parecen desarrollar contornos entonativos y perfiles tonales más parecidos al objetivo a medida que aumenta la competencia en L2. Finalmente, el efecto del género y acento, así como otras interacciones, demuestran que la adquisición de la prosodia es más compleja de lo expuesto anteriormente, y que se ve influenciada no solo por el sistema de la L1 y las competencias orales, sino también por factores psicológicos y socioculturales.

Palabras clave: Contorno nuclear; característica tonal; adquisición del habla de L2; transferencia prosódica; español peninsular por los sinohablantes.

1. Introduction

Prosody in second language acquisition (SLA) research has long been promoted through a monolingual linguistic ideology. The vast majority of SLA studies on phonetics and phonology, to date, overtly or covertly tend to treat native speakers as a yardstick for second language (L2) learning. In this regard, L2 learners are defined as failed and deficient monolinguals that will not achieve a native proficiency because of their unchangeable conditions at birth (Cook, 2012; Ortega, 2010, 2011, 2014). However, since the mid-1990s, there has been continual criticism against this dominant paradigm and the concept of monolingualism as a default norm for SLA. For instance, Ortega pointed that the major deleterious consequence of the monolingual view is that it unethically turns bilinguality into an invisible reality by "erasing bilinguals' other language competence from analysis" (2010: 56). To overcome this pervasive monolingual bias, a new trend of bi/multilingual turn has been developed in recent years (May, 2013; Ortega, 2010, 2011, 2014; Valdés, Poza & Brooks, 2015). The creativity

of this new paradigm is that it conceives SLA as a dynamic and constantly interactive state between the first (L1) and the second language, rather than the sum of two clearly separate monolinguals (May, 2013). Besides, instead of examining SLA development based on native speakers, it proposes to investigate learners' total language repertories, and utilize fully-developed bilinguals or successful L2 users as a new empirical baseline for L2 learning (Cook, 2016; Ortega, 2010, 2014).

Despite the advantage of a bilingual approach portraying a more complete picture for SLA research, it has not been sufficiently worked out to replace the predominant SLA theories, at least in the domain of phonetics and phonology. The main causes of this, according to May (2013), have been correlated with the ideological roots of monolingualism as the normative condition of the discipline. Another possible reason lies in the ambiguity and complexity of correctly defining the mature and experienced L2 users as the reference for SLA (Cook, 2016). Moreover, the lack of integration between theories and practices in different SLA contexts and fields also have stunted the epistemological change toward bilingual turn in SLA (Pandey, 2016). Finally, as mentioned by Pandey (2016), cross-continental examples or typologically distinct language pairs are still needed to explore the global appeal of current frameworks on the bilingual turn. Given these remaining uncertainties in the bilingual turn and the typological distance between Chinese and Spanish prosody (see the discussion in section 2.1), we decided in this study to investigate SLA from the classical view of native speakers. Unlike Spanish and most Romance languages that are intonation languages, Chinese is widely known as a tonal language, with each tone having a different internal pattern of rising and falling pitch contours (Yip, 2002). Arguably, problems would arise when comparing the intonation of Chinese with Spanish, given that the pitch contours of Chinese are highly dependent on the tone type of the sentence-final syllable. Therefore, rather than direct comparisons across learners' language pairs, it seems to be more feasible in our research to interpret the L1 influence on L2 acquisition based on previous empirical studies on Chinese prosody. The L2 transfer to L1 prosody from the opposite direction will be considered for further research.

Cross-language transfer effects have been discussed over the last decades in SLA research and most studies have focused on segmental differences across languages. However, several intonational aspects, for instance, F0 register, pitch range, and intonational pitch patterns are, in essence, more vulnerable to cross-language influences, and therefore more difficult to learn, to interpret, and to investigate in SLA (Mackay, 2000; Atoye, 2005; Mennen, 2014, 2015). These difficulties in L2 prosody are not merely caused by the typological distance between the phonological and phonetic systems of the L1 and the L2, but are also closely linked to the "complexity and multidimensionality" of intonation (Mennen & Leeuw, 2014: 187). Intonation

can interact with other prosodic structural elements (e.g. duration, rhythm, intensity and lexical stress) and signal multiple pragmatic functions in the speech. Therefore, it is difficult to determine whether some intonational variations in the phonetic implementation are either categorical or gradient to the phonological representations (Mennen, 2015; Nolan, 2006). Besides, the divergence of different types of frameworks on prosodic typology (e.g. see the discussion between Beckman & Venditti, 2011; Hyman, 2006, 2012; Jun, 2006; Ladd, 2001) adds to the existing difficulty of describing and characterizing intonational features. This methodological issue remained unsolved in the prosodic field until the advent of the AM (autosegmental-metrical) theory that has allowed researchers to analyze and uniformly compare the intonation systems of many languages.

The broadly recognized autosegmental-metrical framework was developed around the core idea of isolating the categorial phonological elements from its surface phonetic realization (Gussenhoven, 2004; Ladd, 1996; Pierrehumbert, 1980). This separation between the phonological and phonetic dimensions is of great importance, not only for the prosodic investigation of a wide variety of languages (e.g. Hualde, 2003; Face & Prieto, 2007; Estebas-Vilaplana & Prieto, 2008, 2010 for Spanish; Tseng, Huang, & Beckman, 2011 for Mandarin Chinese), but also for cross-language or cross-dialectal comparisons of intonation (e.g. English and Spanish comparison by Bowen, 1956 and Vilaplana, 2008; Mandarin and English by Crosswhite & McDonough, 2000; Spanish and Catalan by Vilaplana, 2008; Majorcan and Minorcan Catalan by Payà & Vanrell, 2005). According to the AM theory, intonational pitch contours are phonologically comprised of two types of tonal units: pitch accents and boundary tones, which are realized as either a high or low tone attached to metrically prominent syllables or the edges of prosodic phrases. Different languages may differ typologically in the inventory of pitch accents, boundary tones and combinations, using strategies of the structural elements according to the text structure (Graham & Post, 2018; Ladd, 1996; Mennen, 2015). Beyond this, the surface phonetic shapes (e.g. pitch height, pitch span, alignment, speech rhythm, etc.) of the intonation primitives (basic components) may also vary to different degrees because of different conventions of pitch implementation across language communities.

Furthermore, in the recently developed L2 Intonation Learning Theory (LILt), Mennen (2015) suggested that there can be some cross-language intonation differences in the semantic and frequency dimensions, not merely in the systematic (or ‘phonological’) and realizational (or ‘phonetic’) aspects. Concerning deviations in the semantic dimension, Chinese learners of L2 English were found to have difficulties in signaling new information (Juffs, 1990), marking contrastive stress (Wennerstrom, 1998) and prominence relationships (McGory, 1997) in a native-like way, even

having reached a high level of proficiency in the L2. Besides, significant contrasts were observed in the frequency of usage of phonological tones in the L1 and the L2, probably due to the transfer of the first language (see examples in Backman, 1979; Hewings, 1995; Jilka, 2000; Mennen et al., 2010). Based on this multi-dimension modeling, L2 intonation research is expected not only to examine learners' phonetic and phonological deviations from the target language but also to shed new light on their ability to use appropriately different categorical types and their ability to use intonation to signal different functional meanings in the cross-language speech.

Overall, the current study was undertaken based on the deep phonological understanding of intonation (AM) and previous evidence developed in the L2 prosody learning model (LILt). It was of interest to investigate the systematic acquisition of the intonation contours and pitch implementation details in five question types with different linguistic meanings in Spanish by L1 speakers of Mandarin Chinese and Peninsular Spanish. Furthermore, this study examined the correlation between different pitch variables and compares prosodic performance across different proficiency levels, question types, stress positions, and gender. The rest of the paper is structured as follows: Section 2 describes the intonation system of Peninsular Spanish and Mandarin Chinese and reviews the existing literature on pitch range implementation. Beyond this, the cross-language differences between L1 and L2 prosody will be introduced in this section. Next, Sections 3 and 4 describe the present study with the specific research questions and the methodology. Sections 5 and 6 report the experimental results and discuss the main findings of the work. Finally, in Section 7, we present the conclusions, the pedagogical implications, and some potential limitations of the current study.

2. Literature review

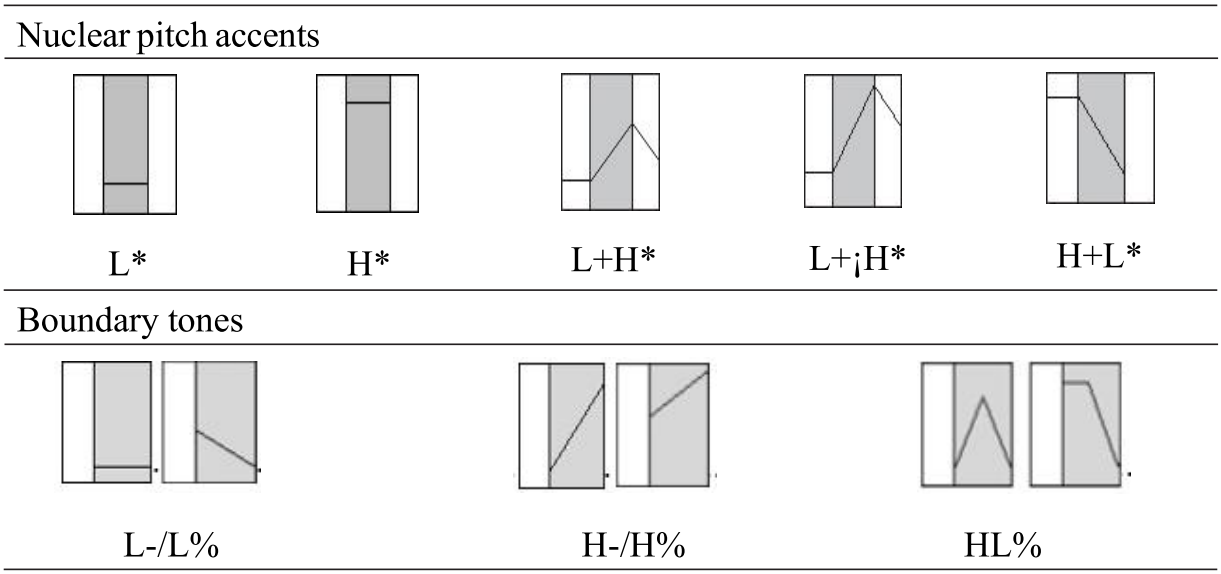
2.1. Intonation contours in L1 and L2

As Spanish is an intonation language, there is a general consensus on the use of ToBI (Tones and Break Indices) labeling systems for the description of intonational phonology. ToBI conventions distinguish two types of tonal events (pitch accents and boundary tones) and describe the intonational patterns by means of two tones that indicate prosodic level (e.g. low is L, high is H). If the tone is associated with a prominent or metrically strong syllable, the letter is followed by a star (*), for example, L* stands for a low stressed syllable. In Spanish, if the prominent syllable has more than one tone attached, the tones are linked with a plus sign and the star indicates the position of the stress. For example, L+H* signals a rising stressed syllable, whereas

L*+H signals a low stressed syllable and a high poststressed syllable. The group of tones associated to a strong syllable is called a pitch accent. If the tone is associated with the edge of an intonation group, it is marked with a percentage sign (%) when the intonation group is final (signaling an intonation phrase), or with a dash (-) when it is non-final (signaling an intermediate phrase).

Based on the ToBI framework, 2 monotonal pitch accents (L*, H*), 3 bitonal pitch accents (L+H*, L+¡H*, H+L*), 2 monotonal boundary tones (L-/L%, H-/H%) and 2 bitonal boundary tones (LH%, HL%) (see the schematic representations in Figure 1) have been proposed for the nuclear configuration of questions in Peninsular Spanish (Estebas-Vilaplana & Prieto, 2010; Estebas-Villaplana, 2008; Face & Prieto, 2007; Hualde & Prieto, 2015; Prieto & Roseano, 2019). While the monotonal accent L* is commonly found in the nuclear position of interrogative modality (e.g. in the nuclear location of information-seeking *yes-no* questions, information-seeking *wh*-questions, confirmation-seeking tag questions as well as the final constituent of disjunctive questions), the high pitch accent H* often appears in the nuclear position of echo and rhetorical questions (Estebas-Vilaplana & Prieto, 2010). With regard to the bitonal accents L+H* and L+¡H*, these are frequently linked to obviousness meanings or some nuance of surprise and exclamation, and normally appear in statements of the obvious, exclamative statements and various counterexpectational questions of Peninsular Spanish. The phonetic distinction between the two accent types typically lies in the F0 scaling, with L+¡H* exhibiting a step rise into the highest peak of the utterance. In contrast with the two previous rising accents, in Peninsular Spanish a falling pattern (e.g. the H+L* L% found for confirmation-seeking *yes-no* question) is frequently used to express "the degree of commitment that a speaker has in the truth-value of a proposition" (which technically has been called the epistemicity) (Roseano, González, Borràs-Comes, & Prieto, 2016: 3). Nevertheless, the epistemicity encoded by the intonation strategy is generally considered to be weaker than that marked with confirmative tags, for example, *¿no?* (no?), *¿verdad?* (is it?), *¿no es cierto?* (isn´t it?). In Spanish, these syntactic structures are typically pronounced with a low nuclear accent L* followed by a high-rising boundary tone H%. Aside from tag questions, the H% or H- edge tone is also attested in information-seeking *yes-no* questions and the non-final constituents of disjunctive questions. Conversely, the low-falling boundary tone L% is usually observed at the end of information-seeking *wh*-questions and disjunctive questions as well as confirmation-seeking *yes-no* questions.

Figure 1. Schematic representations of the pitch accents and boundary tones of Peninsular Spanish questions presented in our study (adapted from Estebas-Vilaplana & Prieto, 2010: 19-21)



In Mandarin Chinese, however, the realization of intonation types differs strongly from intonation languages such as English and Spanish, where pitch contours are used to convey linguistic meanings and communicative purposes (Prieto & Roseano, 2019) although it is worth noting that pitch variation is typically accompanied by changes in other prosodic features, like duration, amplitude, or voice quality. As a tone language, on the one hand, the key issue in Chinese prosody research is concerned with the interaction between tone and intonation. It has long been established that Mandarin Chinese has four lexical tones and one neutral tone in its inventory, namely, tone1 (high), tone2 (rising), tone3 (falling-rising), tone4 (falling), and tone0 (phonological neutral). Though both tone and intonation use F0 as a primary cue, the functional load of F0 in Chinese lies mostly in word distinction (Chen & Gussenhoven, 2008; Yuan & Shih, 2004). Thus, it is expected that intonational meanings in Mandarin Chinese might be recognized through the modification of other prosodic indicators (normally, pitch range, pitch level, and duration), due to the restriction in manipulating the F0 curve (Chen & Gussenhoven, 2008). On the other hand, despite the growing body of knowledge on tone-intonation interaction (Chao, 1968; Chen & Gussenhoven, 2008; Girding, Zhang, & Svantesson, 1983; Kochanski & Shih, 2003; Wu, 1982; Xu, 2005, 2015; Xu & Wang, 2001; Yuan, 2004, 2006; Yuan & Shih, 2004), there seems to be no consensus as to the formulation of a general framework for Chinese prosody. Various models have been proposed to investigate the Chinese intonation mechanism, the most influential ones being the following: the PENTA model (Xu, 2005; Xu & Wang, 2001), the STEM-ML model (Kochanski & Shih, 2003) and the

Static Target model (Shih, 1988). Although models differ in the way they interpret and analyze intonation and its interaction with lexical tones, it is generally accepted that the intonation of Chinese questions and statements diverges increasingly at the sentence-final location. These observations are generally in line with the ToBI-based intonation school who tends to differentiate intonation types through the final nuclear configuration. Nevertheless, the high (H-/H%) and low (L-/L%) boundary labels used in ToBI systems were found to be problematic for the prosodic annotation of Mandarin Chinese (e.g. Pan-ToBI or C-ToBI system), particularly in the intonation of questions, due to the fact that the surface F0 contour of Chinese questions could have either a rising or falling end depending on the tonal identity of the last syllable.

The typological differences of prosodic structure and intonation mechanism between the tone languages and the intonation languages could pose significant challenges for L2 learners of tone languages over the course of learning an intonation language. In the last decades, with the growing body of literature on L2 prosody acquisition, various intonation errors have been reported for Chinese learners of L2 English. For instance, it has been found that Chinese speakers of English tend to employ systematically high-level tones (H*) in target nuclear accents where the phonetic realization is characterized by a low plateau (henceforth L*) during the stressed syllable (Barto, 2015; Hong, 2012; Ji, 2010; Ji et al., 2009, 2012; Shao et al., 2011; Xu, 2009). The tendency to use high-level tones in the L2 nuclear position has been mostly correlated with the lack of a steady low tone (L*) in the tonal inventory of Chinese (Ji et al., 2009; Yuan et al., 2019). Additionally, the intonation stress in Chinese is achieved by mainly raising the H target tone rather than changing the low tone. Therefore, as suggested by Ling (2003) and Jia, Xiong & Li (2005), Chinese learners may unconsciously transfer the high tone strategy from their L1 to the L2 target intonation patterns when speaking an intonation language like English. Compared to the numerous findings regarding L2 English intonation produced by Chinese speakers, little research has been carried out in the case of language contact between Spanish and Chinese. To our knowledge, the only empirical investigations carried out have been those by Cortés Moreno (1997, 1999, 2001, 2004, 2005) and Liu (2005) on the L2 acquisition of Spanish prosody by Mandarin-speaking Taiwanese students. Specifically, Cortés Moreno (1997, 2004) found that Taiwanese students had considerably more problems producing the L2 Spanish intonation than perceiving it, even advanced learners who had shown native-like performance in perceiving the target intonation patterns. Moreover, studies by Cortés Moreno (2001, 2005) seem to suggest a hierarchy of difficulties in the acquisition of L2 Spanish intonation types whereby *yes-no* questions were the most difficult pattern to learn in comparison to statements and *wh*-questions. This ranking of learning difficulties holds essential implications for L2 prosody teaching. However, it is interesting to note that the said

proposal was the opposite of the sequence put forth by Yuan et al. (2019). In this latest study, statements were reported to be the most difficult L2 patterns for Chinese students due to their less perceptually salient target pitch movements compared to *yes-no* questions and requests, which involve more H-L contrasts.

2.2. Cross-language research of pitch range variation

Aside from the phonological system, languages may also differ typologically in the way they use pitch to phonetically implement the categorical elements. According to Ladd (1996), Cruttenden (1997) and Gussenhoven (2004), there are two main types of pitch range variation, namely, level and span. Whereas pitch level, or "register" has been referred to as the overall height of a speaker's voice (Cruttenden, 1997), span, or "F0 range" has been described as a speaker's "excursion size" of range of frequencies in the speech ('t Hart, Collier & Cohen, 1990: 75). A positive correlation exists between the two dimensions of pitch range; the higher the F0 level, the wider the pitch span (Urbani, 2012). In Gussenhoven's view (2004), the phonetic implementation of pitch values has generalizability in the paralinguistic intonational meanings and can be explained using a number of biological codes. For instance, the Frequency Code, which is based on the correlation between the size of the larynx and the rate of vocal fold vibration, suggested that a higher pitch frequently signals a smaller larynx and expresses speaker uncertainty, whilst a lower F0 tends to be associated with a larger organ of production and an assertive interpretation (Ohala, 1984, 1994).

Based on the universality of biological codes, however, some languages were found to go against the general form-function patterns and show some language-specificities in the use of pitch codes. Evidence for this tendency of cross-language differences can be found in a growing number of recent studies in which different languages or language varieties have been reported to have different levels of pitch range and F0 variability in the speech. A crucial difference was found, for example, between tone languages such as Mandarin Chinese and stress languages like English. Compared to English, Mandarin Chinese shows generally higher pitch level, greater F0 fluctuation, and faster pitch change rate in the prose passage (Keating & Kuo, 2012; Eady, 1982) as well as wider F0 span in broadcast news speech (Yuan & Liberman, 2014). Similar F0 features of higher register and wider span were also reported in the Chinese dialect Min when compared with American English (Chen, 2005). Nevertheless, in Cantonese-English bilingual children's speech, significantly lower values of speaking fundamental frequency and F0 range were found for Cantonese speakers than for English-speaking children (Ng, Hsueh, Leung, & Shing, 2010). This difference in pitch implementation might be correlated with the tonal structure of Cantonese, although language proficiency and sociocultural conventions may also play a role in the

divergence of voice pitch characteristics. Furthermore, a recent study on cross-language comparison suggested a wider F0 range and higher F0 register in native Chinese speech than in native Japanese speech (Shi, Zhang, & Xie, 2014). However, when Japanese was compared to American English, Spanish and Dutch, it was found to have the highest values of mean F0 in read speech (Hanley, Snidecor, & Ringel, 1966; Van Bezooijen, 1995). The higher pitch values found for Japanese speakers, particularly for Japanese women, are not due to mere physiological or anatomical differences across gender and language communities (Rendall, Vokey, & Nemeth, 2007; Van Dommelen & Moxness, 1995), but are more closely linked to the social-cultural behaviors and relative powerlessness image of Japanese women (Van Bezooijen, 1995). Aside from the above-mentioned cases, further support for cross-language pitch value differences can be found in a large number of studies on distinct language pairs (see Mennen et al., 2014 and Ordin & Mennen., 2017 for a short review).

Compared to the accomplishments of cross-language research, few studies have shed light on L2 speech deviations from the target language. Because the acquisition of L2 intonation always entails some degree of interaction between the L1 and L2 systems, it is expected that the cross-language differences in the F0 register and span may also appear in the L2 production. Generally, it is suggested that L2 learners have a compressed F0 span and less variable pitch when compared to native speakers of the target language. For instance, Chen (1972) and Juffs (1990) reported that Chinese L2 learners had a narrower F0 span than native speakers of English. Recently, using a large-scale dataset extracted from a language learning app, Yuan et al. (2018) reconfirmed that, compared to that of native English speakers, the speech of Chinese L2 learners was characterized by a narrower pitch span, slower pitch change rate and more small “ripples” on the F0 contour. In addition, it was found that Chinese learners of L2 English and L2 German had higher values than native speakers in F0 span on the phoneme level, and in pitch change amount on the utterance level, due to the negative influence of L1 mandarin prosody (Ding, Hoffmann, & Hirst, 2016; Ding, Jokisch, & Hoffmann, 2012). This general trend of compressed range and less F0 variability in L2 speech can be observed in many L1-L2 combinations, for example, in Spanish learners of L2 English (Backman, 1979), in Chinese learners of L2 Japanese (Shi et al., 2014), among many others. The consistency seen in L2 pitch implementation patterns is probably influenced by the L1 prosody, but more frequently it has been correlated with learners’ lack of confidence or cautiousness when speaking a non-native language (Mennen, 1998; Shi et al., 2014; Volín, Poesová & Weingartová, 2015). Another plausible reason for those L2 speech deviations could be that learners are too focused on the segmental pronunciation and stress emphasis, thus, there might be a lack of attention given to extending and varying the F0 pitch in a native-like way (Zimmerer et al., 2014).

In light of all these findings, it seems that there is a universal trend of pitch range compression when speaking an L2. However, in previous studies, most experiments were conducted with a small number of subjects. As a result, the conclusions drawn are potentially less convincing due to the high variability of the F0 range within speakers (Ladd, 1996). Another problem concerning previous studies is that in some cases the pitch range has been treated as a unitary concept without distinguishing the level from the span (Mennen et al., 2014), and have analyzed the data using different quantification measures. Most importantly of all, to date, none of the studies have explored the L2 pitch implementation characteristics of Spanish by L1 Chinese learners, thus, this research would be interesting evidence for the investigation of a general pitch compression pattern in the L2.

3. The present study

As can be seen in the literature review, prior studies in L2 prosody learning give inconsistent evidence for the cross-language differences in the phonological and phonetic dimensions, either due to the different speaking materials and quantification methods used in the research or because of the different language pairs under investigation. Of these studies, few have shed light on the production of Spanish intonation by L1 Chinese speakers, probably owing to the typologically substantial differences between the two language systems. Therefore, with the present study we intend to fill the existing gap in second language research and investigate the acquisition of intonation contours and pitch implementation details in L2 Spanish, by taking into account proficiency level (or L1), question type, gender and stress position. Specifically, the current study addresses the following questions:

1. Is the L1 prosodic system transferred (either positively or negatively) to the L2 Spanish intonation and therefore can the L1 account for some L2 deviation errors? If so, does the acquisition of L2 intonation patterns reflect different levels of proficiency?

2. Does the acquisition of pitch implementation details (as measured by six pitch variables based on the F0 distribution: mean F0, max F0, min F0, 100% span, 80% span and Pitch dynamism quotient -PDQ-) in an L2 reflect different levels of proficiency, and do they differ among different question types or stress positions as well as between male and female speakers?

3. Do our findings of L2 pitch range variation point towards a universal developmental trajectory (narrower span and less variable pitch) during the

L2 learning process, or are they highly dependent on the L1-L2 language pairs under study?

4. Does the acquisition of L2 Spanish intonation show different levels of difficulty depending on pragmatically different question types, and, if so, does this difficulty ranking exist only in the phonological dimension, or it can also appear in the phonetic dimension?

4. Methodology

4.1. Participants

The subjects of the present study were 5 native speakers of Peninsular Spanish and 32 learners of Spanish with Mandarin Chinese as their first language. The age of the participants ranged from 18 to 31 years (mean age: 23.97; SD=2.872). None of the individuals reported any speech, hearing or communicative impairments. The native control group consisted of 5 women who were born and/or lived for more than 20 years in Barcelona, and who had a comparable level of education (mean age: 23.2; SD=4.87). Although some of these participants were Catalan-Spanish bilinguals, they reported that Spanish was their dominant language.

As for the Chinese speakers (26 females and 6 males), they were all students and lived in Barcelona at the time of the recordings (mean age: 24.09; SD=2.53). Peninsular Spanish was the language variety (dialect) to which they had been predominantly exposed both during their leaning period in China and their immersion period in Spain. The L2 participants were divided into 2 groups according to proficiency level in line with the Common European Framework of Reference for Languages (CEFR): intermediate level (B1-B2) and advanced level (C1-C2). The Spanish language proficiency of the Chinese speakers was judged using the official language qualification DELE (Diploma of Spanish as a Foreign Language), with the exception of those learners who did not have this certificate. In the latter case (roughly 15% of the L2 learners), participants were required to state their self-evaluated L2 proficiency on the basis of Spanish language courses they had completed. In order to ensure that the learners were aware of the criteria of self-assessment, explicit descriptions of the six levels of European language proficiency were explained to those speakers at an early stage of this process.

In this study, we did not specifically control for Chinese learners' origin, age of L2 acquisition, or length of exposure to the target-language environment, due

to the dramatic reduction in subject pool which would result from including these selection criteria. However, as these variables were reported in previous literature to exert certain effects in the L2 speech (Cadierno et al., 2020; Juan-Garau & Pérez-Vidal, 2007; Kharkhurin, 2008; Pfenninger & Singleton, 2016), we decided to include this additional information for the non-native groups (see Appendix A for more details). The L2 participants were native speakers of Mandarin Chinese and declared that this was their dominant language, despite having different places of origin within China. The majority of the learners acquired Spanish in adulthood (mean age: 18.81; SD=2.08), only two female participants reported being teenage learners (they started to learn Spanish at 12 and 17 years old). Although the number of months of exposure to the target language was quite different between individual learners, the mean exposure time of advanced speakers (mean length: 22.8; SD=18.) was generally longer in comparison to the intermediate group (mean length: 19.13; SD=9.51).

4.2. Materials and recording procedures

In order to collect natural speech, we used the DCT (Discourse Completion Task) (Billmyer & Varghese, 2000; Félix-Brasdefer, 2010; Golato, 2003) method to elicit the corpus. Specifically, 15 daily scenarios were designed to elicit five question types with different linguistic meanings in the target language, namely, information-seeking *yes-no* questions ('YN'), information-seeking *wh*-questions ('WH'), disjunctive questions ('DJ'), confirmation-seeking *yes-no* questions ('CYN') and confirmation-seeking tag questions ('TAG'). Each question type varied in nuclear stress position (two positions: final and penultimate stressed syllable). Test items were mostly comprised of words with high familiarity ratings or high frequency (Tanaka & Terada, 2011), for the benefit of non-natives' comprehension during the task activity (see Appendix B for more details). The average syllable number in the study was 5.8 per utterance.

Situational contexts were presented by an interlocutor with whom the participants were somewhat familiar, and speakers were asked to produce the target sentence used in that situation. The task was performed only once except in cases where there was a problem with the speaker's first realization. Subjects were allowed to reproduce the test item if they made a mistake. All recordings took place in a soundproof room with a head-mounted microphone. Speech files were digitized at a sampling rate of 44.1 kHz and with a quantization precision of 16 bits. Each utterance was saved separately as a *wav* format file and annotated to a *TextGrid* object using a *Praat* script.

4.3. Data collection and analysis

4.3.1. Intonation contours

For the benefit of intonation labeling, unvoiced segments were interpolated through and F0 trace was smoothed using a *Praat* script with bandwidth set to 10 Hz. In our study, intonation transcriptions were realized by combining the visual representation of the F0 curve with the auditory perception of pitch accents. To facilitate the transcription work, all test items were initially annotated using a prosodic tool (Eti-ToBI) which automatically labels intonational events in Spanish utterances (Elvira-García et al., 2016). However, as this script was developed based on Sp_ToBI and Cat_ToBI conventions for native speakers, many unexpected pitch movements produced by Chinese learners could not be appropriately assigned (mainly due to differences in pitch alignment). Thus, manual correction of all labeling was conducted by the first author of the paper, according to the guidelines for Castilian Spanish intonation (Vilaplana, 2008; Vilaplana & Prieto, 2010). Furthermore, the annotation results were checked and revised by the second author of the paper and a third expert in Sp_ToBI labeling.

To conduct a cross-language comparison of intonation contours, the proportion of occurrence of pitch accents, boundary tones and nuclear configurations was calculated separately within the 5 question types. In this study, we were particularly interested in the nuclear configuration realization which has been referred to as the most salient part of an intonation contour (Prieto & Roseano, 2019) although it is worth noting that pitch variation is typically accompanied by changes in other prosodic features, like duration, amplitude, or voice quality. Regarding the set of linguistic functions that intonation (together with other prosodic factures, other parts of the F0 contour were excluded from further analysis.

4.3.2. F0 pitch extraction

Pitch tracking was performed automatically in *Praat* using the ESPS algorithm (‘get F0’) (Talkin, 1995), with pitch floor set to 70 Hz and pitch ceiling 600 Hz. A time step of 10 ms was used for the computation of F0. The original F0 data was refined by a manual correction based on *Pitch* objects. Specifically, cases with octave jumps and measurement errors (e.g. mistakes caused by creaky voice or laryngealizaiton, or false voicing in silent fragments) were marked as “unvoiced” and therefore excluded from further analysis. The speakers’ pitch characteristics were analyzed along three dimensions: (a) *pitch level*, which has been referred to as a “reference line” (or a constant value) averaged by the rising and falling movements in the overall pitch contour

(Gussenhoven, 2004: 76); (b) *pitch span*, that is the distance between the highest and lowest values of the F0 contour, and finally, (c) *pitch variation*, which describes the degree of F0 variability in the speech (J't Hart, Collier & Cohen, 1990).

To quantify pitch characteristics in Spanish L1 and L2, two types of measures were included in the analysis. Firstly, for pitch level and span, we used the long-term distributional (LTD) measures developed by Mennen (2014). Specifically, 3 parameters have been extracted for level, and 2 parameters for span including both the absolute pitch excursion (100% span: max F0 - min F0) and the 80% span (the 90th and 10th percentile span) which has been reported to be the best F0 distributional measure (Mennen, Schaeffler, & Docherty, 2009; Niebuhr & Skarnitzl, 2019). Furthermore, because female and male speakers differed greatly in group size, we decided to introduce another pitch dynamism quotient measure (PDQ) to normalize the F0 variation data and to make the LTD results more robust. PDQ values were calculated by dividing the standard deviation by the mean frequency of each utterance. This metric gives an account of pitch variability and generally, the higher the PDQ, the greater the F0 variance and the emotional turmoil in subjects' speech (Shi, Zhang, & Xie, 2014; Wang & Qian, 2018; Zimmerer et al., 2014). To sum up, the following six dependent variables were extracted for the comparison of L1-L2 pitch implementation characteristics:

- Pitch level: *min F0*, *max F0* and *mean F0*
- Pitch span: *100% span*, *80% span*
- Pitch variability: *PDQ*

To assess the effect of scales, pitch measures of the span were also transformed into another two psycho-acoustic scales. Compared to the corresponding linear scale (Hz), the logarithmic (semitones) or the near-logarithmic (ERB-rate) scales have been reported to be the best measures for modeling intonational equivalence between females and males, and for capturing the frequency differences across speakers and languages (Nolan, 2003; Patterson & Ladd, 1999). Besides, correlation coefficients were calculated using the Pearson method in order to examine whether, and to what extent, F0 span is correlated with other pitch variables.

4.4. Statistical analysis

During the first stage, the between-group differences with regard to the production of pitch accents and boundary tones were tested in the R environment (R Core Team, 2016) using the chi-square statistic with continuity correction when all the cells of

the contingency tables for expected values had a value greater than 5. In the case that this condition was not met, we applied Fisher's exact probability test to examine the significance. At the second stage, statistical analysis of pitch results was performed using the linear mixed effects model. The model was run separately for six dependent pitch variables (namely, min F0, max F0, mean F0, 100% span, 80% span and PDQ) with Proficiency, Gender, Question type and Stress position, as well as all their possible interactions as fixed effects. Subjects were included as random effects with all possible random intercepts. We used the Anova function to test the significance of main effects, and *p* values were fitted by eliminating non-significant effects of the initial model with the *lmerTest* package and evaluated using Satterthwaite's approximation (Kuznetsova, Brockhoff, & Christensen, 2017). The post-hoc comparisons were conducted using the single-step function of the *multcomp* package (Hothorn et al., 2016) supported by the *emmeans* function (Lenth, Singmann, & Love, 2019). Finally, the correlation coefficients of various F0 distributional measures were obtained using Pearson's method (Benesty et al., 2009).

5. Results

5.1. Results of overall pitch accents and boundary tones

The first analysis examines the proportion of pitch accents and boundary tones of the five question types within each of the three language groups, namely, L2 Chinese intermediate learners (hereafter CI), L2 Chinese advanced learners (hereafter CA), and L1 Native Spanish speakers (hereafter SN). Of the 555 speech items, 119 items were realized with a low-rising accent L+H* and 61 items, with a high-level tone H* in the nuclear location, regardless of final pitch movements. Interestingly, both accent types were produced by Chinese L2 speakers, with a relatively higher proportion for the CI group than for the CA group (see Table 1). However, the probability test revealed no statistical significance between the two learner groups with regard to the production of the two pitch accents H* (χ^2=1.20, p>0.1) and L+H* (χ^2=0.72, p>0.1). The vast number of high plateaux and rising patterns in L2 Spanish corroborate previous findings for L2 English and L2 German spoken by Chinese learners (Ding et al., 2012; Ji et al., 2009, 2012). Nevertheless, concerning the nuclear accent L*, the SN group used a significantly higher number of steady low tones compared to the CI (χ^2=8.94, p<0.01) and CA learners (χ^2=5.12, p<0.05). Similar results have also been observed in the realization of the pitch accent H+L*, whereby Spanish L1 speakers presented a significantly higher proportion of falling contours than the two learner groups (CI-SN: p<0.001; CA-SN: p<0.01).

Table 1. Proportion of pitch accents produced by the three language groups over the five question types

	H*	**L+H***	**L+¡H***	**L***	**H+L***	**Total**
CI	14.58%	26.67%	0%	58.75%	0%	240
CA	10.83%	22.92%	0%	63.75%	2.50%	240
SN	0%	0%	10.67%	78.67%	10.67%	75

In regard to the boundary tones, Table 2 shows that, compared to the Spanish native group, Chinese L2 learners, particularly the CI group, produced significantly more rising boundaries (H%: χ^2=4.39, p<0.05) and less falling boundaries (L%: χ^2=5.99, p<0.05) over the five question types. In addition, a small proportion of rising-falling boundaries (HL%) was exclusively observed in the L2 intonation (see Table 2). These results appear to be contradictory, to some extent, to our initial expectation that L2 learners tend to use a final fall more frequently than a final rise for Spanish questions, given that in their L1 Chinese questions are marked mostly by interrogative particles. A possible explanation for this phenomenon may be that Chinese learners have overproduced the typical final rise for all question types, due to a lack of knowledge about the intonational phonology of the target language (both its possible patterns and their communicative functions). Furthermore, our study seems to suggest a general trend of improvement during the L2 acquisition of phonological tones. More precisely, it was found that compared to Chinese intermediate learners, the advanced group showed systematically a more native-like performance in producing the target pitch accents and boundary tones. However, these between-group differences are not statistically significant (p>0.1 in all cases except for the nuclear accent H+L* where the p value was less than 0.05). This is likely because the oral proficiency levels of Chinese intermediate (B2 level) and advanced learners (mostly C1 level) were very similar in our research.

Table 2. Proportion of boundary tones produced by the three language groups over the question types

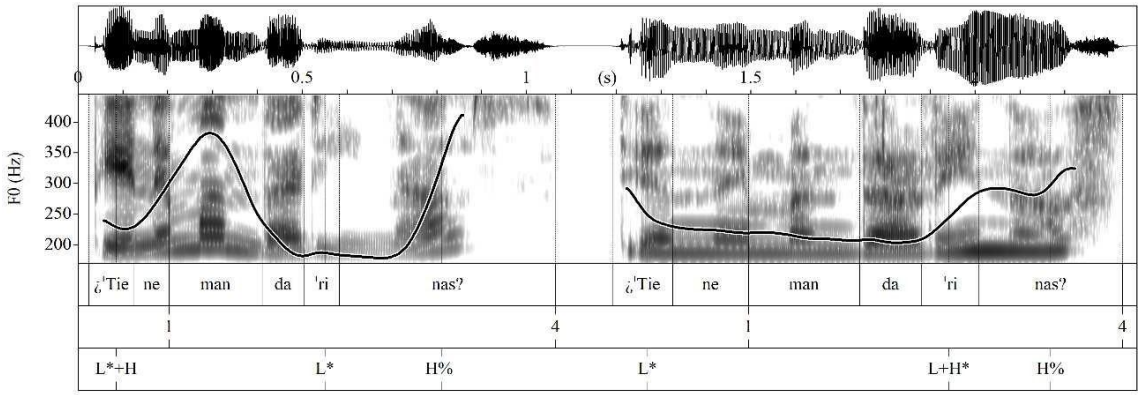
	H%	**L%**	**HL%**	**Total**
CI	67.50%	30.42%	2.08%	240
CA	64.17%	34.17%	1.67%	240
SN	53.33%	46.67%	0%	75

5.2. *Results of nuclear configurations*

5.2.1. Information-seeking yes-no questions

Since pitch contours did not differ significantly between the two Chinese groups (see the statistical results in section 5.1), we treated the non-native subjects as a single group when comparing them to L1 native speakers in the nuclear configuration analysis. Specifically, for information-seeking *yes-no* questions, only around 40% of the Chinese learners successfully acquired the native-like pattern (L* H%), which was significantly less than the Spanish L1 group who consistently used the typical low-rising contours for *yes-no* questions (χ^2=16.64, p<0.001). The majority of L2 speakers were found to apply an early rising accent (L+H*: 35%) or use high-level accents (H*: 18%) with a rising end (H%) on the nuclear position (see Figure 2). The large number of deviated accents produced by Chinese learners may be explained as a negative transfer from their L1. Mandarin Chinese does not have a steady low tone (L*) in its tonal inventory and intonational stress is realized mainly through the pitch range expansion of high tone, therefore, Chinese learners may unconsciously transfer the high-level tone (namely Tone1) or the rising tone (namely Tone2) from their L1 into the L2 Spanish prosody. Moreover, it was noted that most of the Chinese speakers were capable of producing the final rising boundaries (H%) of *yes-no* questions, with only 7% of the L2 speakers failing to achieve the high pitch targets in sentence-final locations.

Figure 2. Waveform, spectrogram and F0 trace of the information-seeking yes-no question "¿Tiene mandarinas?" 'Do you have Tangerines?' produced by a native speaker (L* H%) (left) and a Chinese learner (L+H* H%) (right)

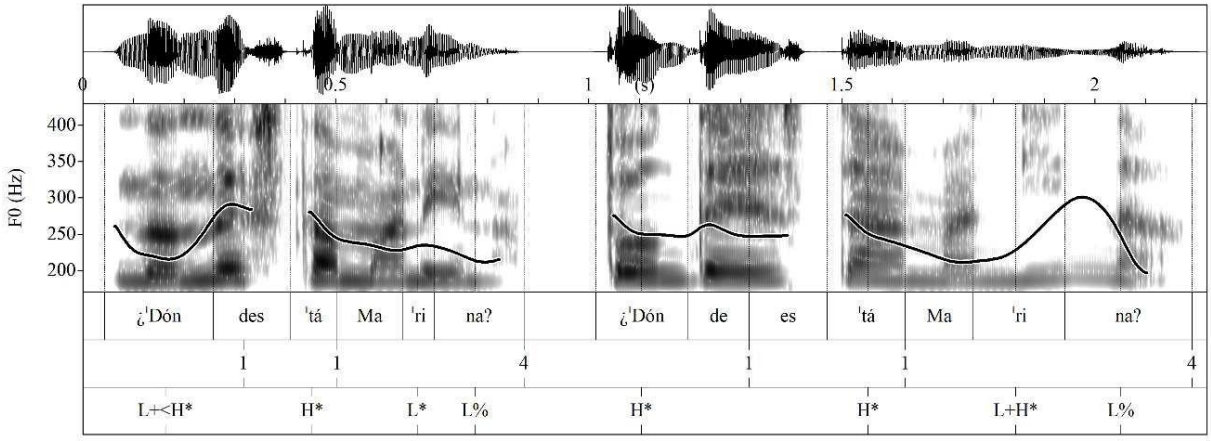


5.2.2 Information-seeking wh-questions

With regard to information-seeking *wh*-questions, the Spanish L1 speakers produced three possible pitch contours in the nuclear position: L* L% (30%), L* H% (18%) and L+¡H* L% (53%). Of these patterns, the first falling pattern (L* L%) has been interpreted as the most neutral and default realization of *wh*-questions, while the second rising contour (L* H%) expresses a nuance of politeness and the speaker's interest or greater involvement in the informative speech act (Estebas-Vilaplana & Prieto, 2010: 35). The third pattern L+¡H* L% is not an information-seeking *wh*-question, but has instead been described as a form of the reintroduction of a previously stated topic. The high proportion of this complex contour in our corpus can likely be attributed to different interpretations of the situational contexts by the L1 speakers.

As for the L2 intonation results, the Chinese learners showed a great probability of success in producing the two typical falling (L* L%: 30%) and rising contours (L* H%: 18%) used for the target *wh*-questions, and no statistically significant difference was found between the Spanish L1 and L2 groups (p>0.1). Aside from the native-like production, the rest of the L2 speakers were found to show deviation problems similar to those of information-seeking *yes-no* questions, that is, a tendency to apply rising tones (L+H*: 44%) or high-level (H*: 8%) tones instead of low accents (L*) in the nuclear location (see Figure 3).

Figure 3. Waveform, spectrogram and F0 trace of the information-seeking wh-question "¿Dónde está Marina?" 'Where is Marina?' produced by a native speaker (L* L%) (left) and a Chinese learner (L+H* L%) (right)

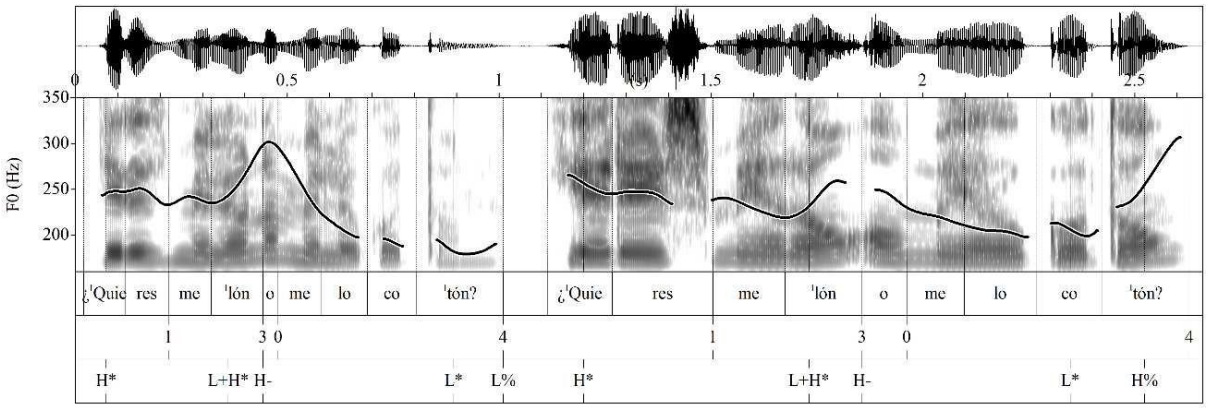


5.2.3 Information-seeking disjunctive questions

Disjunctive questions in L1 Peninsular Spanish were realized mainly with a rising contour (L+H* H-: 93%) in the first prosodic unit and a falling movement (L* L%: 100%) at the end of the utterance (see Figure 4). Among Chinese learners, in the

first prosodic unit, 81% of the participants successfully produced the target rising pitch pattern (L+H* H-), statistically they showed no significant difference compared to the L1 native performance (χ^2= 0.62, p>0.1). For the second prosodic group, more than half of the L2 learners (roughly 66%) were able to produce the final falling pitch contours (L* L%) found in disjunctive questions. Nevertheless, according to Fisher's exact test, there is a significant difference (p<0.01) between the L1 and L2 speakers with regard to their intonational performance. The high level of accessibility of target intonation contours of disjunctive questions by Chinese speakers could be explained as a positive transfer of L1 intonation strategies. In Mandarin Chinese, disjunctive questions are realized in a similar way to those of Peninsular Spanish, that is, by the expansion of pitch range in the first prosodic unit and compression in the last constituent of the sentence. This similarity in the phonetic dimension between the source and target languages appears to benefit or accelerate learners' rate of acquisition of target intonation patterns. In addition to the low tone (L*) deviations, it is also interesting to note that some Chinese learners might make mistakes when producing the intermediate and final boundaries of disjunctive questions. For instance, roughly 24% of Chinese speakers were found to use a high-rising boundary (H%) instead of a low-falling boundary (L%) in the final constituent of disjunctive questions (see Figure 4). This may happen because of their lack of intonational knowledge and the cognitive bias in marking questions with final rising pitch movements.

Figure 4. Waveform, spectrogram and F0 trace of the disjunctive question "¿Quieres melón o melocotón?" 'Do you want melon or peach?' by a native speaker (L+H* H- L* L%) (left) and a Chinese learner (L+H* H- L* H%) (right)



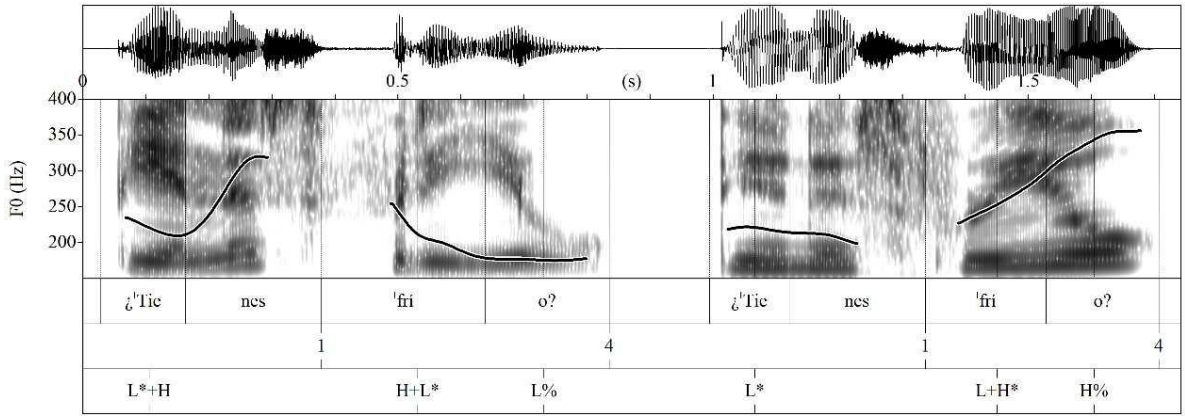
5.2.4 Confirmation-seeking yes-no questions

In addition to the prototypical informative speech act, questions in Peninsular Spanish can also be used to convey the illocutionary force of confirmation. Unlike information-seeking questions in which speakers do not have any expectations

regarding the answer, confirmation-seeking questions are biased questions where the information is previously given or shared through the context or presumptions, and the speaker usually expects a positive answer from the other interlocutor (Frota & Prieto, 2015; Vanrell et al., 2010). In our corpus, more than half of the L1 speakers applied a falling contour (H+L* L%: 53%) to express a relatively higher certainty on the proposition, whilst the rest (47%) maintained the canonical rising pitch pattern (L* H%) used for information-seeking *yes-no* questions.

As for the Chinese learners, only a small number (6%) had acquired the native-like falling pitch accent (H+L*) combined with a low boundary tone (L%) for this type of question. By contrast, 35% of the L2 speakers borrowed the nuclear contours of information-seeking *yes-no* questions, while the rest exhibited several different patterns which are congruent with a deviated form of L* H% with the low nuclear accent (L*) misproduced (see Figure 5). These results proved significantly different (χ^2=16.01, p<0.001) from those produced by L1 native speakers.

Figure 5. Waveform, spectrogram and F0 trace of the confirmation-seeking yes-no question "¿Tienes frío?" 'Are you cold?' produced by a native speaker (H+L* L%) (left) and a Chinese learner (L+H* H%) (right)

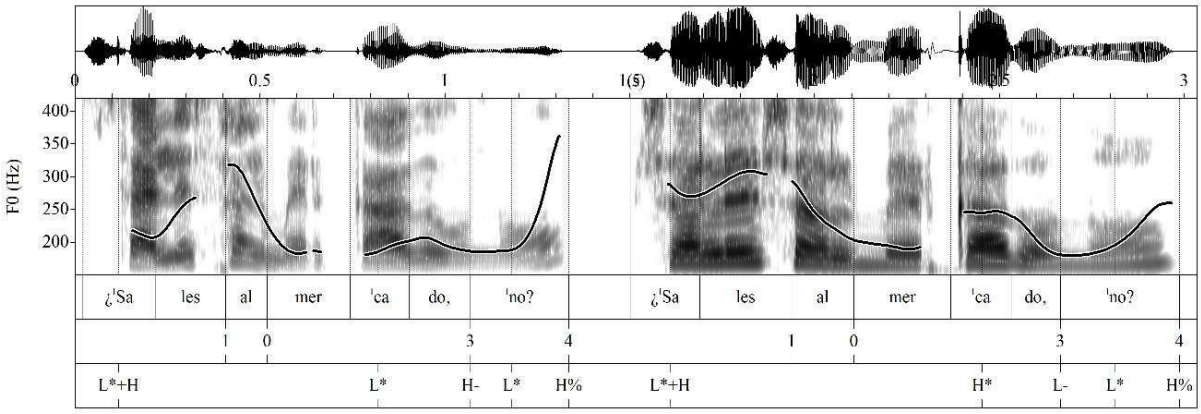


5.2.5 Confirmation-seeking tag questions

Confirmation-seeking questions can be produced, in addition to the typical falling pattern presented in the section above, with various confirmative tags (i.e. *¿no?* (no?), *¿verdad?* (is it?), *¿no es cierto?* (isn't it?)). This syntactic resource is valid both in Mandarin Chinese and in Peninsular Spanish, as it expresses a stronger commitment by the subject to the truth value of the proposition. In our experiment, all question tags were consistently produced with a low-rising pitch contour (L* H%) by Spanish L1 and L2 speakers. A possible explanation for this phenomenon has been correlated with the (alleged) universality in marking question tags with various alternative strategies

such as final rise, final pitch range expansion, late prominence and some emphasis gestures (Cuenca, 1997; Gussenhoven & Chen, 2000). Besides, the simple structure (mostly 1 or 2 syllables) of those tags has also been proposed as a plausible factor which accounts for the high intonational accuracy in L2 tag questions. Note, however, that the production of target phonological patterns does not necessarily mean that L2 speakers have acquired a 100% native-like prosodic performance. It has been observed, for example, that the question tags produced by Chinese learners differed phonetically from those produced by the L1 native speakers in the F0 differences of the final rising movements (see Figure 6). Nevertheless, this comparison is outside the scope of this paper.

Figure 6. Waveform, spectrogram and F0 trace of the confirmation-seeking tag question "¿Sales al mercado, no?" 'You are going to the market, aren't you?' produced by a native speaker (L* H%) (left) and a Chinese learner (L* H%) (right)



5.3. *Pitch implementation results*

5.3.1. Pitch level results

This section aims to show the differences in pitch level (min, max and mean F0) between Spanish L1 and L2 speakers. The statistical analysis revealed a tendency among female learners of the CI and CA groups to have systematically higher pitch levels than female speakers of the SN group (see Table 3), due to the influence of L1 tonal structure. Nevertheless, results on main effects (Table 4) show no significance of Proficiency on any of the three measures of pitch level. On the contrary, the effects of Gender, Question as well as the interaction of Proficiency and Question on the three pitch measures. Furthermore, Table 4 shows that there is a mild effect of Stress on the mean F0 and the factor Stress strongly interacted with Proficiency on the variable of mean F0. No significant interactions were found between Gender and Question, Gender and Stress on the three measures of pitch level.

Table 3. Descriptive statistics of pitch level results depending on language group and gender

	Mean F0	**Min F0**	**Max F0**
CI_M	143 Hz	108 Hz	214 Hz
CI_F	236 Hz	189 Hz	329 Hz
CA_F	232 Hz	175 Hz	335 Hz
SN_F	220 Hz	167 Hz	327 Hz

Table 4. Effects (F values) of Proficiency, Gender, Question type, Stress position and its interactions on the three measures of pitch level (Hz) (N=555, '***' p<0.001; '**' p<0.01; '*' p<0.05; ' •' p< 0.1)

	Mean F0	**Min F0**	**Max F0**
Proficiency	1.06	2.50•	0.17
Gender	95.01***	63.17***	37.90***
Question	8.49***	5.73***	13.38***
Stress	5.80*	0.37	3.70•
Proficiency *Question	7.80***	3.65***	6.48***
Proficiency*Stress	8.12***	0.45	0.90

On the one hand, the lack of a significant effect of Proficiency, but its strong interaction with Question type on the three F0 measures, further suggests that pitch level may differ within each of the five question types produced by learners of different levels of proficiency. Specifically, the post-hoc analysis indicates that the CI group has a significantly higher minimum (t(2)=2.59, p<0.05) and mean F0 values (t(2)=3.09, p<0.01) in CYN questions than the L1 native speakers. Also, Figure 7 shows that speakers in the CI group tend to use a higher mean pitch in WH questions than the native participants (t(2)=2.56, p<0.05). By contrast, the SN group was found to have a significantly higher maximum F0 in WH questions when compared to the CA group (t(2)=2.47, p<0.05). On the other hand, Figure 8 indicates that the two Chinese groups tend to have significantly lower mean F0 (CI: t(2)=-3.383, p<0.01; CA: t(2)=-4.359, p<0.001) in utterances with final-syllable stressed words compared to those with stress on the penultimate syllable. Moreover, the CI group was found to use a statistically lower maximum F0 in final-syllable stressed questions (t=-2.45, p<0.05). These results differ from those of the SN group, which exhibited higher pitch values in final-syllable stressed sentences (oxytone words), as described in Figure 8. A possible explanation for

this deviation in L2 pitch has been correlated with the extra difficulty of combining intonational patterns and those required by lexical stress, since some pitch contours require adaptations when they are applied to oxytone words (e.g. L* H% becomes L+H* L%) (Prieto & Roseano, 2019) although it is worth noting that pitch variation is typically accompanied by changes in other prosodic features, like duration, amplitude, or voice quality. Regarding the set of linguistic functions that intonation (together with other prosodic factures.

Figure 7. Mean pitch of the three language groups depending on proficiency and question type. Upper and lower levels indicate maximum and minimum F0

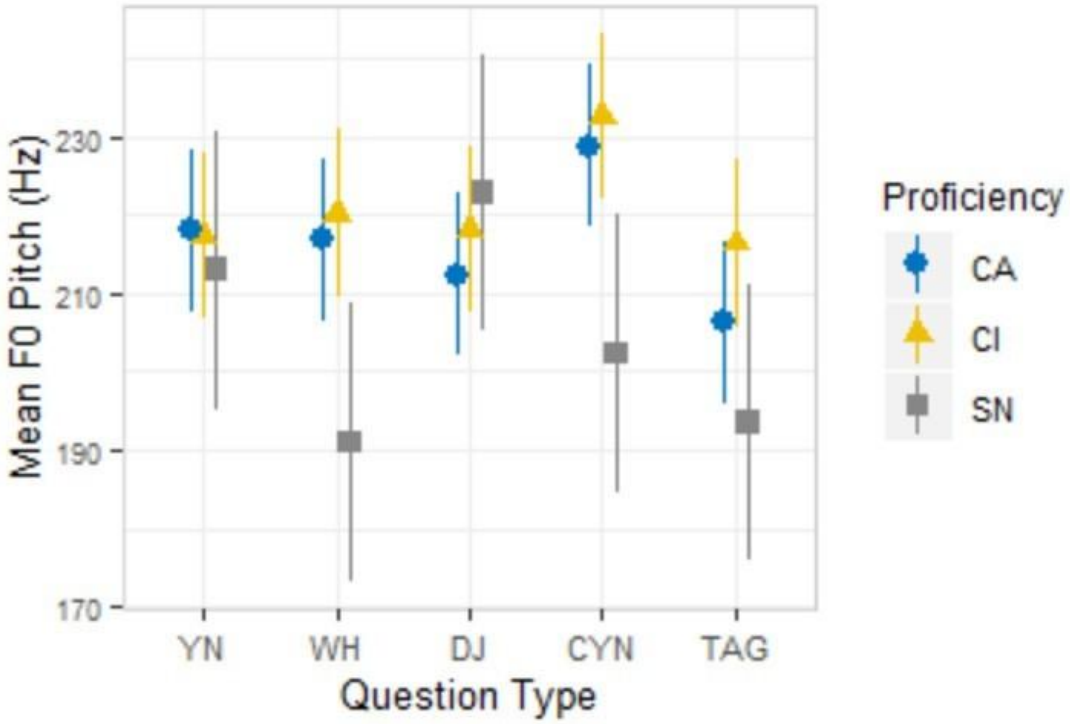
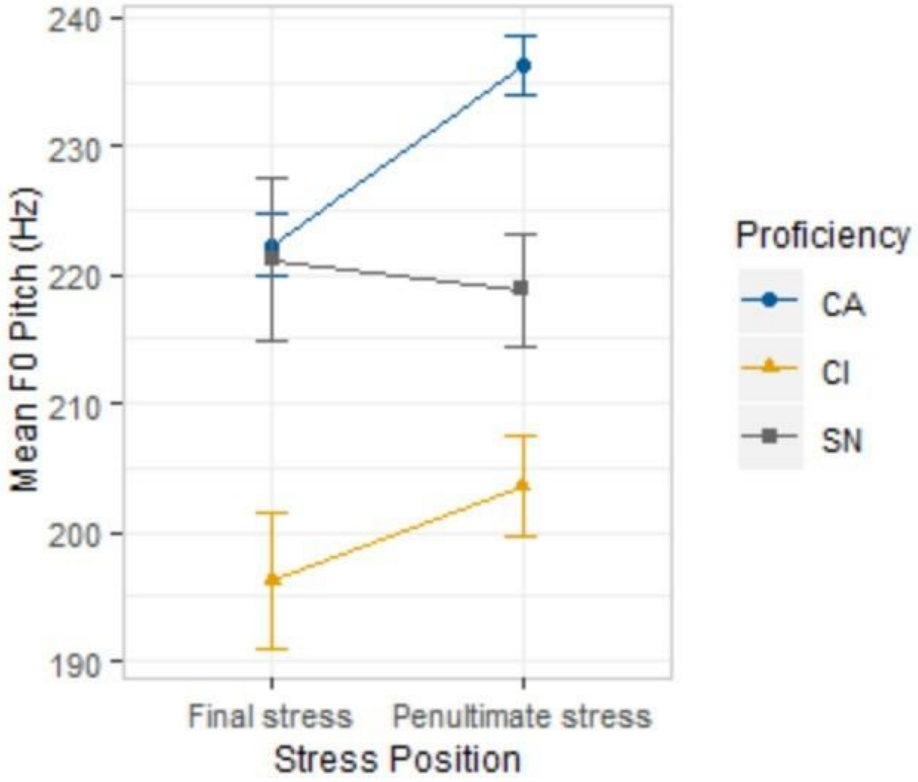


Figure 8. Mean pitch of the three language groups depending on proficiency and stress position. Error bars indicate ±1SE



5.3.2. Pitch span results

Of the two F0 span measures, only the 80-percentile span measured in St showed a significant main effect of Proficiency ($F(2, 38)=3.47$, $p<0.05$). As expected, the two learner groups showed a more reduced span than the L1 native speakers (80% span: CI:7.11 St; CA: 8.34 St; SN: 9.45 St). Nevertheless, the F0 differences were only statistically significant when comparing the CI and SN groups ($t(2)=-2.53$, $p<0.04$). Besides this, there was a strong main effect of Question type (80% span: $F(4, 518)=9.16$, $p<0.001$; 100% span: $F(4, 518)=12.30$, $p<0.001$) as well as its interaction with Proficiency (80% span: $F(8, 518)=8.78$, $p<0.001$; 100% span: $F(8, 518)=7.01$, $p<0.001$) on the two span measures. Since statistical results among scales of Hz, St and ERB were similar on Question type and on the interaction between Proficiency and Question type, only the St version is reported in this section.

Furthermore, a post-hoc test which looked for the interaction between Proficiency and Question on the 80% span showed that within the five question types, the CI and CA groups tended to use a significantly narrower span than the L1 speakers in YN questions (CI-SN: $t(2)=-3.38$, $p<0.01$; CA-SN: $t(2)=-2.82$, $p<0.05$), DJ questions (CI-SN:

t(2)=-3.70, p<0.01; CA-SN: t(2)=-3.53, p<0.01) and CYN questions (CI-SN: t(2)=-2.53, p<0.05). By contrast, regarding WH questions, Figure 9 indicates that the two Chinese groups had a wider span than the native speakers, although this trend did not reach statistical difference in our study. Besides, the factor Proficiency was also found to interact significantly with Stress on the 80% span measured in ERB (F(2, 518)=3.54, p<0.05) and in Hz (F(2, 518)=4.33, p<0.05). More precisely, the post-hoc tests on all scales indicated that the two learner groups, particularly the CA group (e.g. St: t(2)=-2.70, p<0.01), tended to compress the span in questions with final-syllable stressed words more than those with stress on the penultimate syllable. By contrast, results in Figure 10 seem to suggest an opposite trend for the SN group regarding the span performance on the two stress positions. Finally, unlike the mean F0, no significant main effect of Gender or Stress was found on the span measures on any of the scales.

Figure 9. 80% F0 span of the three language groups depending on proficiency and question type

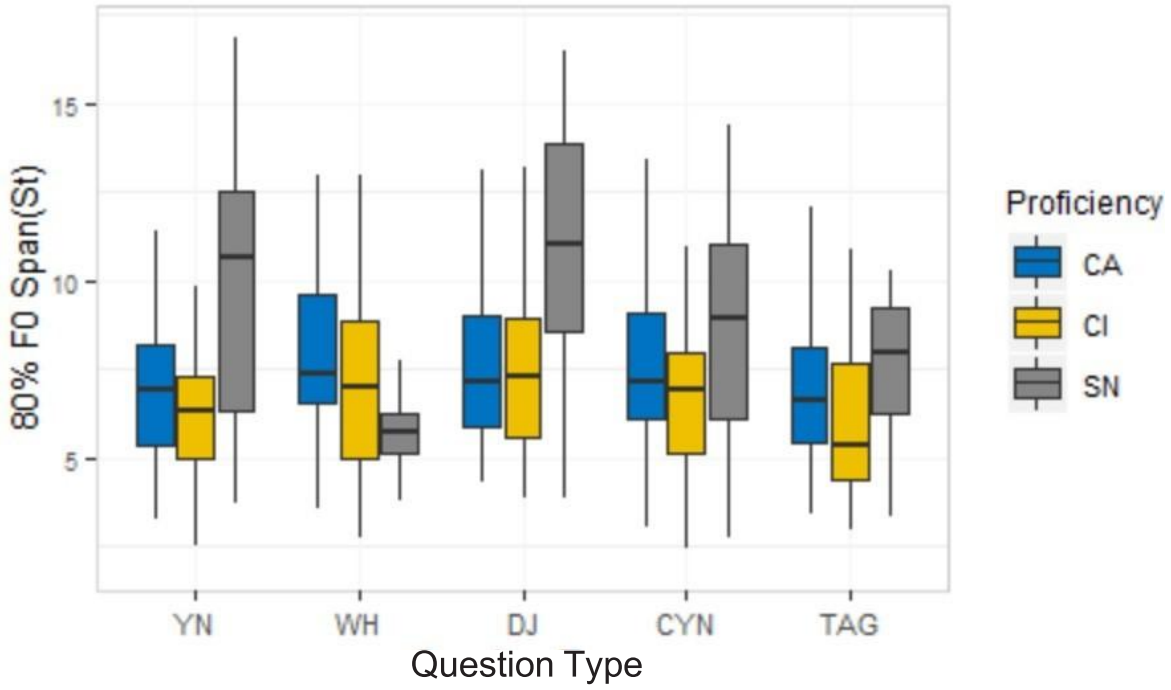
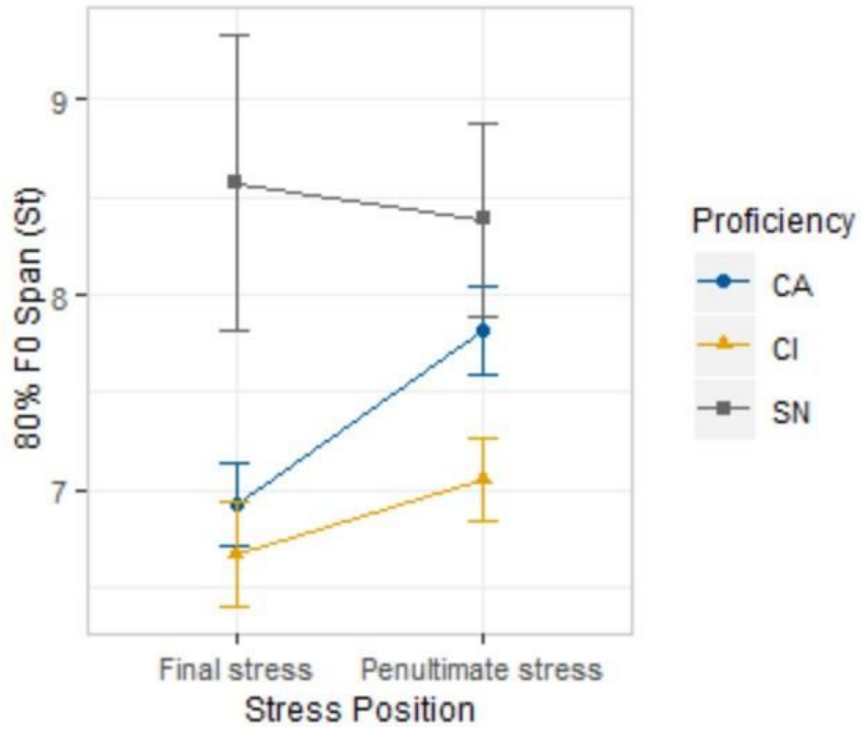


Figure 10. 80% F0 span of the three language groups depending on proficiency and stress position. Error bars indicate ±1SE



5.3.3. Pitch variation results

With regard to the pitch variation, the statistical analysis showed a significant main effect of Gender ($F(1, 38)=4.56$, $p<0.05$), Question type ($F(4, 518)=10.07$, $p<0.001$) and Stress position ($F(1, 518)=3.98$, $p<0.05$). To be specific, our results showed that female speakers were significantly more variable than male speakers in their use of F0, and as a whole the participants had higher F0 variance (or higher pitch dynamism quotient -PDQ-) in DJ questions (mean: 0.214), YN questions (mean: 0.211) and CYN questions (mean: 0.200) than in WH (mean: 0.173) and TAG questions (mean: 0.171). In addition, Figure 12 indicates that speakers of all three proficiency levels had consistently higher F0 variability in questions with penultimate syllable stressed words (mean: 0.200) than those with stress on the final syllable (mean: 0.188). In addition to those features, it is interesting to note that the two Chinese groups produced an overall lower PDQ (CI: 0.168; CA: 0.198) than the L1 native speakers (SN: 0.216), although the factor Proficiency did not show significant main effect in this analysis. Nevertheless, the statistical model revealed a strong interaction between Question type and Proficiency ($F(8, 518)=8.30$, $p<0.001$). As can be seen in Figure 11, compared to the two Chinese groups, the Spanish L1 speakers were higher in PDQ for all question types, in contrast to WH questions in which native speakers produced the lowest PDQ. More precisely,

the post-hoc test indicated that the native group had statistically higher PDQ in DJ questions than the CI (t(2)=-3.317, p<0.01) and CA groups (t(2)=-2.593, p<0.05), as well as in YN questions when compared to the CI group (t(2)=-3.218, p<0.01). No other factors or interactions reached significance on the variable of PDQ.

Finally, in order to examine whether, and to what degree, the F0 span is interdependent on other pitch variables (i.e. mean F0, max F0, min F0 and PDQ), we analyzed the correlation between the span and the rest of the variables that we have taken into account in the pitch range analysis. The statistical results indicated that the 100% and the 80% span were more closely correlated with the maximum F0 (100% span: r=0.84, p<0.001; 80% span: r=0.66, p<0.001) than with the mean F0 (100% span: r=0.38, p<0.001; 80% span: r=0.41, p<0.001) or the minimum F0. These results appear to suggest that the more a speaker is able to approximate to the high pitch targets, the wider the entire F0 range of his/her speech. Additionally, there was a strong positive correlation of PDQ with the 100% span (r=0.78, p<0.001) and the 80% span (r=0.80, p<0.001), as illustrated in the right panel of Figure 13. The high consistency between the two span measures and the PDQ measure also consolidated the results of pitch range variation in our study, despite the unbalanced group size across gender and language group. In general, our findings seem to suggest that the wider the F0 span, the more variable the speech.

Figure 11. Mean PDQ of the three language groups depending on proficiency and question type

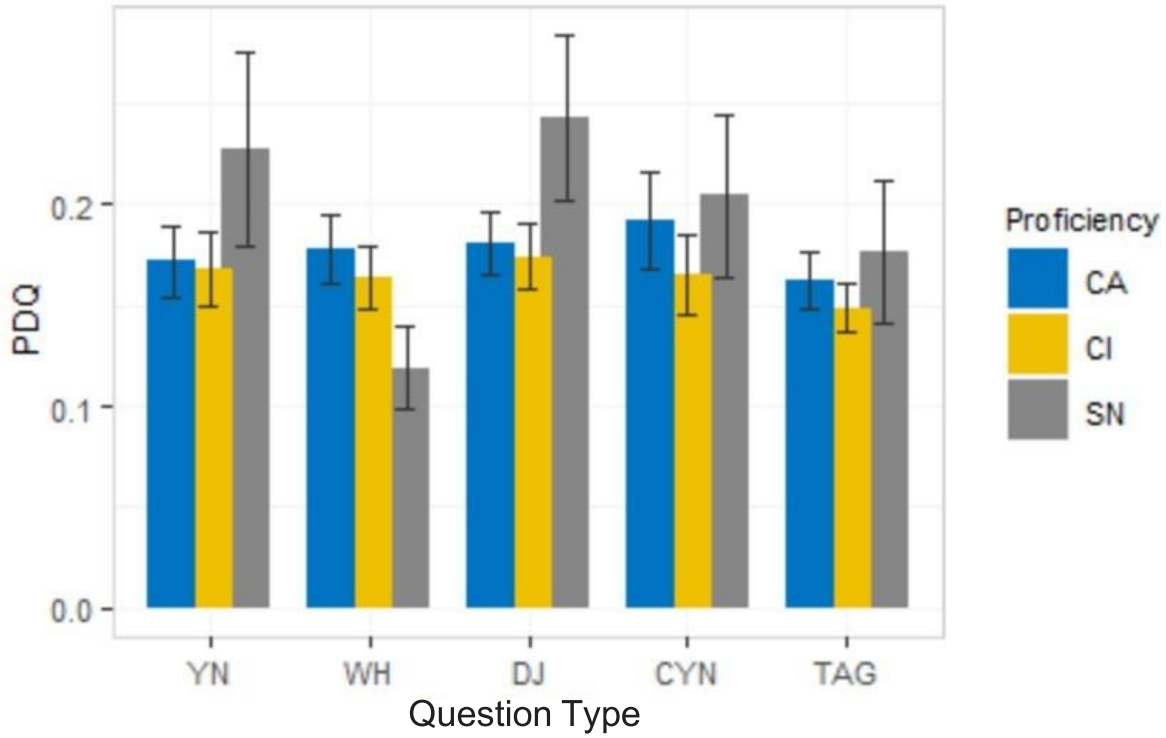


Figure 12. Mean PDQ of the language three groups depending on proficiency and stress position

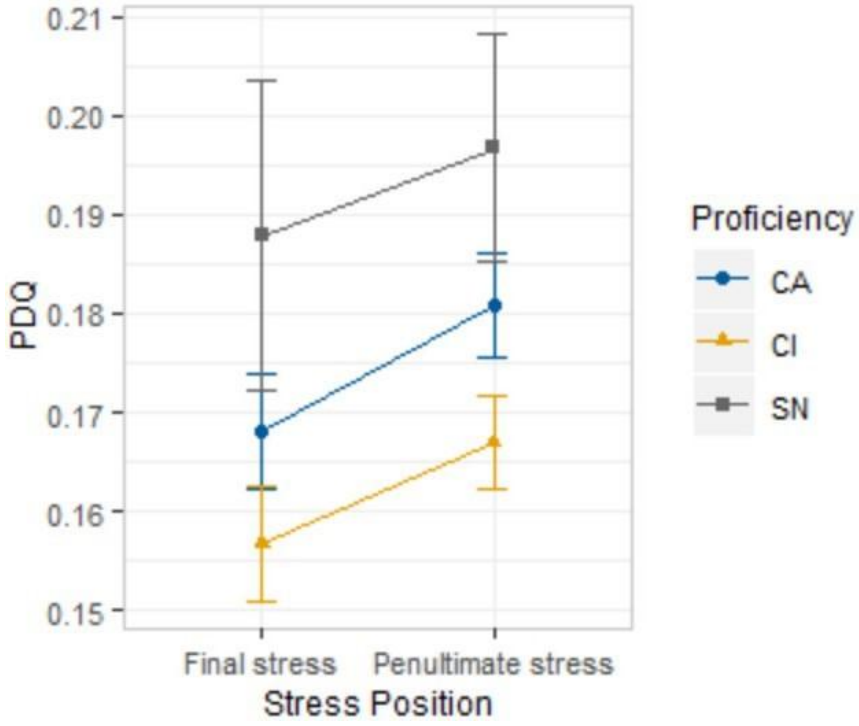
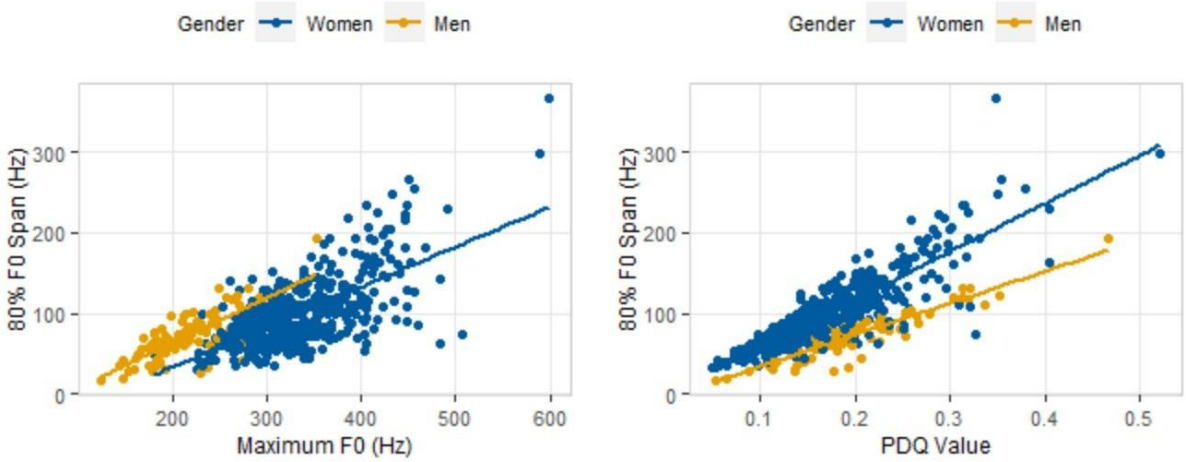


Figure 13. Correlation of the 80% F0 span with the maximum F0 (left) and with the PDQ (right)



6. Discussion

The present study sets out to investigate the acquisition of L2 Spanish prosody by Chinese L1 speakers. For this purpose, percentages of occurrence of the pitch accents,

boundary tones and nuclear configurations related to different speech acts, as well as various pitch implementation variables have been examined across the three language groups.

The first research question enquired whether the acquisition of intonation contours in an L2 reflects different levels of proficiency, and whether prosodic knowledge from the L1 can be transferred onto L2 pitch movements. In general, it was found that the L2 intonation performance of the phonological contours is dependent on proficiency level, and most of the intonation deviations and success in the L2 could be accounted for under an L1 transfer hypothesis; therefore the answer to the first research question is affirmative. Specifically, the descriptive statistics suggested that compared to the L2 intermediate speakers, L2 advanced learners generally acquired more native-like patterns in producing pitch accents and boundary tones. However, these differences between the two learner groups did not reach statistical significance in our study, probably due to the modest gap of oral proficiency between the Chinese intermediate (B2 level) and advanced group (mostly C1 level). Furthermore, our study found that both the L2 intermediate and the L2 advanced groups produced a significantly higher proportion of high plateaux (H*) and rising accents (L+H*) in the nuclear location than the Spanish L1 speakers, who did not produce any of the aforementioned accent types but showed a statistically higher number of low tones (L*) or falling movements (H+L*) instead. These deviation errors produced by Chinese speakers have been mostly correlated with the negative transfer of their L1. As described before, the lack of a steady low tone in Mandarin's lexical tone inventory and its different stress realization mechanisms (through the expansion of high tone) may be contributing factors in the frequent use of high/rising accents in L2 Spanish intonation. Generally, our findings regarding high/rising tone strategies are consistent with previous studies in L2 speech, which also reported a similar tendency for Chinese learners of other intonation languages such as English and German (Ding et al., 2012; Ji et al., 2009, 2012; Yuan et al., 2019). Moreover, it was surprising that the two Chinese groups, particularly the intermediate group, produced a significantly larger number of high boundaries (H%) and fewer low prosodic tones (L%) than L1 native speakers at the end of the five question types. Such results were contrary to our initial expectations that Chinese learners would use more final falls than rises in L2 Spanish intonation, due to the negative influence of L1 question marking strategies using a particle. A possible explanation for this phenomenon might be that Chinese speakers had at their disposal few intonation contours of the target language and therefore could not assign the correct pattern of different sentence types with specific functional meanings, thus, they tended to overproduce all the questions with the most prototypical final rising intonation. Nevertheless, in general, our results seem to suggest that the performance of L2 speakers in the production of pitch accents and boundary tones can progress

towards native-like shapes by increasing their proficiency level and phonological awareness in the L2.

As for the nuclear configurations, it was found that, excluding the low-rising contours (L* H%), more than half of the learners deviated from the L1 native speakers in the overproduction of high *plateaux* (H* H%) or early rising accents (L+H* H%) in information-seeking *yes-no* questions. Moreover, unlike Spanish L1 speakers who consistently used a high boundary tone to mark the information-seeking *yes-no* question, the final boundaries of Chinese learners could be either high (H%) or low (L%), although low boundaries constituted only a small proportion in their L2 intonation. These results corroborate previous studies on L2 intonation (Ding et al., 2012; Ji et al., 2009, 2012) that have reported similar problems for Chinese L1 speakers of L2 English regarding *yes-no* question patterns. The prevalence of high/rising accents was also observed in the nuclear accents of information-seeking *wh*-questions and confirmation-seeking *yes-no* questions, as a result of the systematic negative transfer from L1 prosody explained above. But in the case of confirmation-seeking *yes-no* questions, most L2 learners have borrowed the intonation inventory of information-seeking *yes-no* questions and only a few of them were able to produce the most salient falling pattern (H+L* L%) for the confirmative speech act. The high comparability in the realization of the two question types might be attributed to the fact that they share the same syntactic structure in Spanish. However, it is possible that learners faced difficulties in correlating different functional meanings with the specific pitch contour shapes in the L2. Besides, we found that the nuclear patterns of *wh*-questions were variable even in L1 native speech, likely due to individual differences in the interpretation of the situational contexts. Whilst the falling (L* L%) and rising contours (L* H%, with a nuance of politeness and interest) tend to be interpreted as neutral contours for *wh*-questions, the complex rising-falling patterns (L+¡H* L%) in our dataset were biased forms that focused on the reintroduction of topics instead of requesting information. Overall, it seems that the rate of acquisition of target intonation contours is dependent, to some extent, on the systematic similarities between the first prosody and the target prosody, and in this way, we can reasonably explain the great success of Chinese learners in producing L2 disjunctive questions and tag questions as a positive transfer of L1 intonation.

The second research question asked whether the L2 acquisition of pitch implementation details reflects different levels of proficiency, and whether the pitch range values differ among question type, stress position, and gender. The overall results seem to give an affirmative answer to the first part of the question. More precisely, our study found a better pitch performance for the L2 advanced learners than for the L2 intermediate speakers, although the factor "proficiency" did not reach

significance on any of the three pitch measures (except the 80% span measured in St). In addition, the significant interactions of proficiency with other dependent variables show that the L2 prosody acquisition is more complex than previously stated and the proficiency effect was modulated by question type and stress position. Furthermore, the significant main effect of question type for all the pitch measures appears to give a positive answer to the second research question, showing that different sentence types are encoded with different pitch profiles which serve different pragmatic purposes. In other words, the language-specific pitch implementation values were highly dependent on question type. Specifically, we found that the intermediate learners' pitch increased significantly in information-seeking *wh*-questions and confirmation-seeking *yes-no* questions. In addition, it was observed that the two learner groups compressed the span and the F0 variability in all of the question types except for the *wh*-questions in which Chinese informants used wider span and greater F0 variation compared to the L1 native speakers.

As for the pitch differences on stress position, it was interesting to observe that the Chinese learners tended to increase the F0 register, expand the pitch span and exhibit more frequent F0 contour variation in questions where the stress falls on the penultimate-syllable of the word than those with stress on the final syllable of the word, whereas the native speakers (on the contrary) produced a wider span and higher register for the oxytone words. The factor which accounts for these differences between L1 and L2 is mainly correlated with learners' unfamiliarity and lack of knowledge of rules in producing the final-syllable stressed words, particularly when they appear in the sentence-final position that requires compressed pitch movements. In such circumstances, learners may pay more attention to the pronunciation of words and reduce the use of F0 strategies in the L2 (Mennen, Schaeffler, & Dickie, 2014; Peters, 2019; Yuan et al., 2018; Zimmerer et al., 2014). Finally, with regard to gender effects, our results are congruent with previous findings which report that women use significantly higher pitch level and greater pitch variation than men. Hence, the answer to the latter part of the second research question is also affirmative. According to Ohala (1994), Van Bezooijen (1995), Urbani (2012), Mennen et al. (2014) and Peters (2019) pitch differences across gender are not merely driven by physiological and anatomical factors, but are also closely linked to some sociocultural aspects such as social status and gender roles. Moreover, it is assumed that the greater F0 variation of female speakers found in our study might be correlated with the finding that women tend to express many emotions more frequently than men, except for pride and power (Brebner, 2003). However, further research on the current topic is required to provide direct and strong evidence for our speculation. Finally, in the data obtained for our study, we also found a strong positive correlation between the F0 span and the maximum F0 as well as the F0 variability. More precisely, our study

appears to suggest that the wider the entire F0 span, the higher the F0 pitchtargets and the more variable the F0 speech.

The third research question enquired whether our pitch implementation findings point towards a universal developmental path during the L2 learning process, regardless of the language combinations under study. In the present study, we found that compared to Spanish L1 speakers, Chinese learners tended to use higher pitch level (specifically higher minimum and mean F0), narrower span and less variable pitch in the L2. These results are in agreement with other preliminary studies on L2 speech which also documented a compressed pitch pattern for L2 speakers of different language backgrounds, and reported a similar trend towards improvement as their L2 experience increased (Busà & Urbani, 2011; Mennen, Schaeffler, & Docherty, 2009; Mennen, 1998; Peters, 2019; Shi, Zhang, & Xie, 2014; Ullakonoja, 2007; Urbani, 2012; Yuan et al., 2018; Zimmerer et al., 2014). As for the differences of F0 height, the higher pitch level observed in L2 intonation has been explained as a result of an increased cognitive effort when speaking a non-dominant language (Zimmerer et al., 2014). However, in our study, we speculate that the L1 pitch characteristics may also play a role in the F0 rise of L2. To test this hypothesis, further investigations comparing L1 Chinese and L1 Spanish are needed.

The final research question examined whether the acquisition of the phonological and phonetic patterns of an L2 reflects different levels of difficulty corresponding to pragmatically different question types. On the one hand, it is interesting to note that the probability of success with regard to the production of L2 nuclear contours was different across the five question types. Therefore, the answer to the last research question must be affirmative in the phonological dimension. Specifically, our results seem to suggest a hierarchy of difficulty in the intonation learning of different question patterns of Spanish by Chinese L1 speakers, whereby the confirmation-seeking *yes-no* question was the most difficult pattern, followed by the information-seeking *yes-no/wh* question, then the disjunctive question and finally the confirmation-seeking tag question. Along the same lines, Cortés Moreno (1999, 2004), Liu (2005), Mennen (2015) and Yuan et al. (2019) have also reported that different sentence types may imply different degrees of difficulty during the learning process of L2 phonological contours. On the other hand, the varied pitch performance among the five question types appears to extend previous hypotheses on L2 intonation (Cortés Moreno, 1999, 2004; Liu, 2005; Yuan et al., 2019) from the phonological to the phonetic dimension, suggesting that there is also a ranking of difficulty in implementing the pitch values of different question patterns in L2. The relatively faster rate of acquisition of the two intonational aspects (phonological and phonetic) has been partially correlated with the typological closeness between the first language and the target language.

Moreover, this has also been explained as a result of the "perceptual salience" of some intonational movements in the target language (Yuan et al., 2019).

7. Conclusion

All things considered, the present study has shown that the L2 intonation learning process is more complex than previously stated, whereby the phonological and phonetic dimensions may develop in a non-parallel way. Specifically, it is worth noting that learners who have successfully acquired the target-like intonation contours may still deviate from native speakers in the pitch implementation details of the L2 prosodic system. However, as their L2 proficiency increases, they are capable of approximating to the target language settings both in the systematic dimension and in some realizational aspects. This finding of non-uniform development paths for L2 phonetic and phonological acquisition has also been observed in other L1-L2 language pairs and seems to be a universal feature that occurs throughout the course of foreign language learning (Ding et al., 2012; Graham & Post, 2018; Mennen & Leeuw, 2014; Mennen, 2015). In addition, the distinct pitch performance in L2 appears to suggest a differing degree of difficulty in acquiring Spanish intonation depending on sentence type, stress position and gender. Beyond this, our study extends previous hypotheses by proposing a progression in difficulty levels from the phonological to the phonetic dimension, suggesting that this difficulty ranking exists not merely when acquiring the L2 phonological contours, but also when implementing the pitch values of different sentence types.

From a teaching perspective, our study sheds light on L2 prosody learning, particularly on the L2 acquisition of Spanish intonation by Chinese learners which, despite its crucial importance, has not been addressed in many publications. Overall, it is proposed that there should be special training methods based on specific tasks to help reduce learners' foreign accents, as they showed distinct performances for L2 intonation patterns, according to the similarities and dissimilarities between the first and target language. Furthermore, our results suggest that the training program should not only include the phonological knowledge of target intonation contours but also, and perhaps more importantly, should allow learners to interact with language-appropriate contexts and to produce pitch implementation details in a native-like way. At this point, a growing number of recent investigations are devoted to the development of intonation teaching techniques. For instance, pitch gestures have been reported to benefit L2 prosody learning (Baills, 2016; Bernardis & Gentilucci, 2006; Gullberg, 2006; Jia & Wang, 2013a, 2013b; Morett & Chang, 2015), particularly in acquiring the low nuclear accents, which constitute the most difficult patterns for Chinese L2 learners. Other approaches, such as music training activities and speech visualization

tools, can also help learners to progress in L2 speech production. Nevertheless, considering that Spanish intonation has different levels of difficulty depending on sentence type, any pedagogical proposal should include scaffolding techniques in order to ease the way into the phonetic implementation.

Finally, some limitations should be noted in the current study. One such limitation is the lack of more precise criteria on the sample profile. Specifically, due to the dramatic reduction of possible samples, we did not specifically control for Chinese subjects' origin, age of L2 acquisition, or length of exposure to the target-language environment. Moreover, the sample sizes of male and female speakers were unequal across the three language groups. However, the effect of gender imbalance on the pitch range values was minimized by introducing the PDQ measure, which can effectively normalize the F0 variation data and make the LTD results more robust. Another limitation is related to the elicitation of the corpus. Due to the ambiguity of the situational contexts designed for the information-seeking *wh*-question, there were some unexpected nuclear pitch patterns in the L1 production, for instance, L+¡H* L%, which has been described as a form of the reintroduction of a previous topic rather than a request for new information. Finally, although our study found that there were some general pitch range deviations for L2 speakers, we did not discover exactly how the F0 range was realized depending on the syntactic position of the phrase and in which positions the learners deviated most from the L1 pitch patterns. Therefore, further research could take into account the effect of syntactic position to examine whether the increased pitch level and the compressed pitch span occur along with the overall L2 utterances or whether they are position-sensitive (e.g. increasing only on the low pitch targets while high tones remain basically unchanged).

8. Acknowledgments

This research was supported by a Ph.D. scholarship provided by China Scholarship Council (No.201808390059) to the first author, and by the postdoctoral program Juan de la Cierva awarded to the second author. Special thanks to Paolo Roseano who kindly revised the Sp_ToBI notation of the corpus. Also, we want to thank all Chinese and Spanish participants that allowed us to record their voices for this study.

9. References

Abu-Al-Makarem, A. & Petrosino, L. (2007) Reading and spontaneous speaking fundamental frequency of young Arabic men for Arabic and English languages: A comparative study. *Perceptual and Motor Skills*, 105(2), 572–580.

Andreeva, B., Demenko, G., Wolska, M., Möbius, B., Zimmerer, F., Jügler, J., Magdalena, O.P. & Trouvain, J. (2014) Comparison of pitch range and pitch variation in Slavic and Germanic languages. In N. Campbell, D. Gibbon & D. Hirst (eds) *Proceedings of the 7th International Conference on Speech Prosody*. (pp. 776–780). Dublin, Ireland: Trinity College Dublin.

Atoye, R. O. (2005) Non-native perception and interpretation of English intonation. *Nordic Journal of African Studies*, 14(1), 26-42.

Backman, N. (1979) Intonation Errors in Second-Language Pronunciation of Eight Spanish-Speaking Adults Learning English. *Interlanguage Studies Bulletin*, 4(2), 239–265.

Baills, F. (2016). *Producing pitch gestures enhances the acquisition of Mandarin Chinese tones*. Unpublished MA thesis. Departament of Translation and Language Science, Universitat de Ponpeu Fabra.

Barto, K. A. (2015). *Mandarin Speakers' Intonation in their L2 English*. Unpublished doctoral dissertation. Program in Second Language Acquisition and Teaching, University of Arizona.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.

Beckman, M. E. & Venditti, J. J. (2011) Intonation. In J. Goldsmith, J. Riggle & A. C. L. Yu (eds) *The handbook of phonological theory*. (pp. 485–532). Oxford, UK: Blackwell.

Benesty, J., Chen, J., Huang, Y. & Cohen, I. (2009) Pearson correlation coefficient. In I. Cohen, Y. Huang, J. Chen & J. Benesty (eds) *Noise reduction in speech processing*. (pp. 1–4) Heidelberg, Germany: Springer.

Bernardis, P. & Gentilucci, M. (2006) Speech and gesture share the same communication system. *Neuropsychologia*, 44(2), 178–190.

Billmyer, K. & Varghese, M. (2000) Investigating instrument-based pragmatic variability: Effects of enhancing discourse completion tests. *Applied Linguistics*, 21(4), 517–552.

Bowen, J. D. (1956) A comparison of the intonation patterns of English and Spanish. *Hispania*, 39(1), 30–35.

Brebner, J. (2003) Gender and emotions. *Personality and Individual Differences*, 34(3), 387–394.

Busà, M. G. & Urbani, M. (2011) A Cross Linguistic Analysis of Pitch Range in English L1 and L2. In W.-S. Lee & E. Zee (eds) *Proceedings of the 17th International*

Congress of Phonetic Sciences. (pp. 380–383). Hong Kong, China: City University of Hong Kong.

Cadierno, T., Hansen, M., Lauridsen, J. T., Eskildsen, S. W., Fenyvesi, K., Jensen, S. H. & aus der Wieschen, M. V. (2020) Does younger mean better? Age of onset, learning rate and short-term L2 proficiency in young Danish learners of English. *Vigo International Journal of Applied Linguistics*, 17, 57–86.

Chao, Y. R. (1968) *A Grammar of Spoken Chinese*. Beijing: The Commercial Press.

Chen, G. T. (1972). *A comparative study of pitch range of native speakers of Midwestern English and Mandarin Chinese: An acoustic study*. Unpublished doctoral dissertation. Department of Philosophy, University of Wisconsin-Madison.

Chen, S. H. (2005) The effects of tones on speaking frequency and intensity ranges in Mandarin and Min dialects. *The Journal of the Acoustical Society of America*, 117(5), 3225–3230.

Chen, Y. & Gussenhoven, C. (2008) Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, 36(4), 724–746.

Cook, V. (2012) The native speaker and Multicompetence. In P. Robinson (ed) *Routledge Encyclopedia of Second Language Acquisition*. (pp. 447-450). London, UK: Routledge.

___. (2016) Premises of multicompetence. In V. Cook & W. Li (eds) *The Cambridge Handbook of Linguistic Multicompetence*. (pp. 1-25). Cambridge: Cambridge University Press.

Cortés Moreno, M (1997). Sobre la percepción y adquisición de la entonación española por parte de hablantes nativos de chino. *Estudios de Fonética Experimental*, 9, 67–134.

___. (1999) *Adquisición de la entonación española por parte de hablantes nativos de chino*. Unpublished doctoral dissertation. Departamento de Didáctica de la Lengua y la Literatura, Universitat de Barcelona.

___. (2001) Percepción y adquisición de la entonación española en enunciados de habla espontánea: el caso de los estudiantes taiwaneses. *Estudios de Fonética Experimental*, 11, 90–119.

___. (2004) Análisis acústico de la producción de la entonación española por parte de sinohablantes. *Estudios de Fonética Experimental*, 13, 80–110.

___. (2005) Análisis experimental del aprendizaje de la acentuación y la entonación españolas por parte de hablantes nativos de chino. *Phonica*, 1, 1–25.

Cruttenden, A. (1997) *Intonation*. Cambridge: Cambridge University Press.

Cuenca, M. J. (1997) Form-use mappings for tag questions. *Amsterdam Studies in the Theory and History of Linguistic Science*, 4, 3–20.

Ding, H., Hoffmann, R. & Hirst, D. (2016) Prosodic transfer: A comparison study of f0 patterns in L2 English by Chinese speakers. *Proceedings of the 8th International Conference on Speech Prosody* (pp. 756–760). Boston, USA: Boston University.

Ding, H., Jokisch, O. & Hoffmann, R. (2012) A phonetic investigation of intonational foreign accent in Mandarin Chinese learners of German. In Q. Ma, H. Ding & D. Hirst (eds) *Proceedings of the 6th International Conference on Speech Prosody.* (pp. 374–377). Shanghai, China: Tongji University Press.

Elvira-García, W., Roseano, P., Fernández-Planas, A. M. & Martínez-Celdrán, E. (2016) A tool for automatic transcription of intonation: Eti_ToBI a ToBI transcriber for Spanish and Catalan. *Language Resources and Evaluation*, 50(4), 767–792.

Estebas-Vilaplana, E. & Prieto, P. (2010) Castilian Spanish intonation. In P. Prieto & P. Roseano (eds) *Transcription of intonation of the Spanish language*. (pp. 17–48). München, Germany: Lincom Europa.

Estebas-Villaplana, E. (2008) The evaluation of intonation: pitch range differences in English and in Spanish. In G. Thompson & L. Alba-Juez (eds) *Evaluation in Context*. (pp. 285–301) Amsterdam, The Netherlands: John Benjamins.

Face, T. L. & Prieto, P. (2007) Rising accents in Castilian Spanish: A revision of Sp_ToBI. *Journal of Portuguese Linguistics*, 6(1), 117–146.

Félix-Brasdefer, J. C. (2010) Data collection methods in speech act performance: DCTs, role plays, and verbal reports. *Speech Act Performance: Theoretical, Empirical and Methodological Issues*, 26, 41-56.

Font-Rotchés, D. (2005). *L'entonació del català. Patrons melòdics, tonemes i marges de dispersió.* Unpublished doctoral dissertation. Department of Didactics and Literature, Universitat de Barcelona.

Frota, S. & Prieto, P. (2015) *Intonation in Romance*. Oxford: Oxford University Press.

Girding, E., Zhang, J. & Svantesson, J.-O. (1983) A generative model for tone and intonation in Standard Chinese. *Working Papers (Lund University: Department of Linguistics)*, 25, 53-65.

Graham, C. & Post, B. (2018) Second language acquisition of intonation: Peak alignment in American English. *Journal of Phonetics*, 66, 1–14.

Gullberg, M. (2006) Some reasons for studying gesture and second language acquisition (Hommage à Adam Kendon). *IRAL-International Review of Applied Linguistics in Language Teaching*, 44(2), 103–124.

Gussenhoven, C. (2004) *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Gussenhoven, C. & Chen, A. (2000) Universal and language-specific effects in the perception of question intonation. In B. Yuan, T. Huang & X. Tang (eds) *Proceedings of the 6th International Conference on Spoken Language Processing*. (pp. 91–94). Beijing, China: China Military Friendship Publish.

Hanley, T. D., Snidecor, J. C. & Ringel, R. L. (1966) Some acoustic differences among languages. *Phonetica*, 14(2), 97–107.

Hewings, M. (1995) The English Intonation of Native Speakers and Indonesian Learners: a Coniparative Study. *RELC Journal*, 26(1), 27–46.

Hong, W. (2012) *An experimental study of Chinese students' English intonation pattern*. Unpublished doctoral dissertation. Department of Engllish Philosophy, Nankai University.

Hothorn, T., Bretz, F., Westfall, P., Heiberger, R. M., Schuetzenmeister, A., Scheibe, S. & Hothorn, M. T. (2016) Package 'multcomp': Simultaneous Inference in General Parametric Models. *Project for Statistical Computing.* Vienna, Austria.

Hualde, J. I. (2003) El modelo métrico y autosegmental. In P. Prieto (coord) *Teorías de La Entonación*. (pp. 155–184). Barcelona, Spain: Ariel.

Hualde, J. I. & Prieto, P. (2015) Intonational variation in Spanish: European and American varieties. In S. Frota & P. Prieto (eds) *Intonation in Romance*. (pp. 350–391). Oxford, UK: Oxford University Press.

Hyman, L. M. (2006) Word-prosodic typology. *Phonology*, 23(2), 225–257.

___. (2012) In defense of prosodic typology: A response to Beckman and Venditti. *Linguistic Typology*, 16(3), 341–385.

Ji, X. (2010) *Acquisition of intonation by Chinese EFL learners—An empirical study based on experimental phonetics*. Unpublished MA thesis. Department of English Linguistics, Zhejiang University.

Ji, X., Wang, X. & Li, A. (2009) Intonation patterns of yes-no questions for Chinese EFL learners. *Proceedings of the 2009 Oriental COCOSDA International Conference on Speech Database and Assessments*. (pp. 88–93). Urumqi, China: IEEE.

Ji, X., Xia, W., Yuan, J. & Li, A. (2012) Intonation patterns of *wh*-questions from Chinese EFL learners. In Q. Ma, H. Ding & D. Hirst (eds.) *Proceedings of the*

6th International Conference on Speech Prosody. (pp. 741–744). Shanghai, China: Tongji University Press.

Jia, L. & Wang, J. Q. (2013a). On the effects of visual processing on tone production by English-speaking learners of Chinese. *TCSOL Studies*, 52(4), 30–34.

___. (2013b.) The effects of visual processing on tone perception by native English-speaking learners of Chinese. *Chinese Teaching in the World*, 27, 548–557.

Jia, Y., Xiong, Z. & Li, A. (2006) The Effect of Focal Accents upon Sentential Pitch in Standard Chinese. *Report of Phonetic Research*, 61-68. Beijing, China.

Jilka, M. (2000) Testing the contribution of prosody to the perception of foreign accent. *New Sounds*, 4, 199–207.

Juan-Garau, M. & Pérez-Vidal, C. (2007) The effect of context and contact on oral performance in students who go on a stay abroad. *Vigo International Journal of Applied Linguistics*, 4, 117–134.

Juffs, A. (1990) Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *IRAL-International Review of Applied Linguistics in Language Teaching*, 28(2), 99–118.

Jun, S.-A. (ed) (2006) *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Keating, P. & Kuo, G. (2012) Comparison of speaking fundamental frequency in English and Mandarin. *The Journal of the Acoustical Society of America*, 132(2), 1050–1060.

Kharkhurin, A. V. (2008) The effect of linguistic proficiency, age of second language acquisition, and length of exposure to a new cultural environment on bilinguals' divergent thinking. *Bilingualism: Language and Cognition*, 11(2), 225–243.

Kochanski, G. & Shih, C. (2003) Prosody modeling with soft templates. *Speech Communication*, 39(3–4), 311–352.

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017) lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1-27.

Ladd, D. R. (1996) *Intonational phonology*. Cambridge: Cambridge University Press.

___. (2001) Intonational universals and intonational typology. In M. Haspelmath, E. König, W. Oesterreicheran & W. Raible (eds) *Language Typology and Language Universals*: *An International Handbook.* (pp. 1-18). Berlin, Germany: De Gruyter.

Lenth, R., Singmann, H. & Love, J. (2019) Emmeans: Estimated marginal means, aka least-squares means. *R Package Version 1.3.4*.

Ling, F. (2003). *A Study on the acoustic correlates of emphasized Tone 3 in Mandarin Chinese*. Unpublished doctoral dissertation. Deparment of linguistics, Peking University.

Liu, F. & Xu, Y. (2005) Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica*, 62(2–4), 70–87.

Liu, Y. H. (2005) La entonación del español hablado por taiwaneses. *Biblioteca Phonica*, 2, 1–807.

Mackey, W.F. (2000) The description of bilingualism. In L. Wei (ed) *The Bilingualism Reader*. (pp.26-54). Oxford, New York: Routledge.

Majewski, W., Hollien, H. & Zalewski, J. (1972) Speaking fundamental frequency of Polish adult males. *Phonetica*, 25(2), 119–125.

May, S. (ed) (2013) *The multilingual turn: Implications for SLA, TESOL, and bilingual education*. New York: Routledge.

McGory, J. T. (1997). *Acquisition of intonational prominence in English by Seoul Korean and Mandarin Chinese speakers*. Unpublished doctoral dissertation. Deparment of Speech and Hearing Sciences,The Ohio State University.

Mennen, I., Schaeffler, F. & Docherty, G. (2007) Pitching it differently: a comparison of the pitch ranges of German and English speakers. In J. Trouvain & W. J. Barry (eds) *Proceedings of the 16th International Congress of Phonetic Sciences*. (pp. 1769–1772). Saarbrücken, Germany: Pirrot GmbH.

Mennen, I. & Leeuw, E. De. (2014) Beyond segments. *Studies in Second Language Acquisition*, 36(2), 183–194.

Mennen, I. (1998) Can language learners ever acquire the intonation of a second language? In *ESCA Workshop on Speech Technology in Language Learning (STiLL 98)*. (pp. 17–20). Marholmen, Sweden: Department of Speech, Music and Hearing, KTH.

____(2015) Beyond segments: towards an L2 intonation learning theory (LILt). In E. Delais-Roussarie, M. Avanzi & S. Herment (eds) *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Multilingual Situations*. (pp. 171–188). Berlin, Germany: Springer.

Mennen, I., Chen, A. & Karlsson, F. (2010) Characterising the internal structure of learner intonation and its development over time. In K. Dziubalska-Kołaczyk, M. Wrembel & M. Kul (eds) *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech*. (pp. 319–324). Poznań, Poland: Adam Mickiewicz University.

Mennen, I., Schaeffler, F. & Dickie, C. (2014) Second language acquisition of pitch range in German learners of English. *Studies in Second Language Acquisition*, 36(2), 303–329.

Mennen, I., Schaeffler, F. & Docherty, G. (2012) Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America*, 131(3), 2249–2260.

Mennen, I. (2007) Phonological and phonetic influences in non-native intonation. In J. Trouvain & U. Gut (eds) *Non-Native Prosody: Phonetic Description and Teaching Practice*. (pp. 53–76). Berlin, Germany: De Gruyter Mouton.

Morett, L. M. & Chang, L.-Y. (2015) Emphasising sound and meaning: Pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, 30(3), 347–353.

Ng, M. L., Hsueh, G., Leung, S. & Shing, C. (2010) Voice pitch characteristics of Cantonese and English produced by Cantonese-English bilingual children. *International Journal of Speech-Language Pathology*, 12(3), 230–236.

Niebuhr, O. & Skarnitzl, R. (2019) Measuring a speaker's acoustic correlates of pitch–but which? A contrastive analysis based on perceived speaker charisma. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (eds) *Proceedings of the 19th International Congress of Phonetic Sciences*. (pp. 1774-1778). Melbourne, Australia: Australasian Speech Science and Technology Association Inc.

Nolan, F. (2003) Intonational equivalence: an experimental evaluation of pitch scales. In M. J. Solé, D. Recasens & J. Romero (eds) *Proceedings of the 15th International Congress of Phonetic Sciences*. (Vol. 771, pp. 2–5). Barcelona, Spain.

___. (2006) Intonation. In B. Aarts & A. McMahon (eds) *The Handbook of English Linguistics*. (pp. 433–457). Hoboken, USA: Blackwell Publishing Ltd.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, 41(1), 1–16.

___. (1994) The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols & J. J. Ohala (eds) *Sound symbolism*. (pp. 1337-1340). Cambridge, UK: Cambridge University Press.

Ortega, L. (2010) The Bilingual Turn in SLA. Plenary delivered at the *Annual Conference of the American Association for Applied Linguistics*. Atlanta, GA, March 6-9.

___. (2011) New trends in SLA research: Theories, methods, ethics. *Invited lecture at National Tsing Hua University*. Taiwan, June 8.

___. (2014) Ways forward for a bi/multilingual turn in SLA. In Stephen M. (ed) *The multilingual turn: Implications for SLA, TESOL and bilingual education*. (pp. 32-53). New York: Routledge.

Pandey, A. (2016). The multilingual turn: Implications for SLA, TESOL, and bilingual education. Stephen M. (ed). 2014. New York: Routledge. *World Englishes*, 35(3), 475-477.

Payà, M. & Vanrell, M. del M. (2005) Yes-no questions and echo questions intonation in Majorcan and Minorcan Catalan: A cross-dialectal comparison. In M. Zdrojkowski (ed) *Phonetics and Phonology in Iberia (PaPI)*. (pp. 20–21). Barcelona, Spain.

Peters, J. (2019) Fluency and speaking fundamental frequency in bilingual speakers of High and Low German. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (eds) *Proceedings of the 19th International Congress of Phonetic Sciences*. (pp. 1-5). Melbourne, Australia: Australasian Speech Science and Technology Association Inc..

Pfenninger, S. E. & Singleton, D. (2016). Age of onset, socio-affect and cross-linguistic influence: a long-term classroom study. *Vigo International Journal of Applied Linguistics*, 3, 147–179.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation. Department of Linguistics and Philosophy, Massachusetts Institute of Technology.

Prieto, P. & Roseano, P. (2019) Current issues and challenges in Spanish intonational research. In S. Colina & M.-G. Fernando (eds) *The Routledge Handbook of Spanish Phonology*. (pp. 222–233). UK: Taylor & Francis.

Rendall, D., Vokey, J. R. & Nemeth, C. (2007) Lifting the curtain on the Wizard of Oz: Biased voice-based impressions of speaker size. *Journal of Experimental Psychology: Human Perception and Performance*, 33(5), 1208–1219.

Roseano, P., González, M., Borràs-Comes, J. & Prieto, P. (2016) Communicating epistemic stance: How speech and gesture patterns reflect epistemicity and evidentiality. *Discourse Processes*, 53(3), 135–174.

Schack, K. (2000) Comparison of intonation patterns in Mandarin and English for a particular Speaker. In K. M. Crosswhite & J. McDonough (eds) *Working Papers in the Language Science*. (pp. 24-55). Rochester, USA: University of Rochester.

S.J.Eady. (1982) Differences in the f0 patterns of speech: Tone language versus stress language. *Language and Speech*, 25, 29–42.

Shao, P., Jia, Y. & Li, A. (2011) Intonation patterns of exclamations of Chinese EFL learners from Zhejiang. *Proceedings of the 2011 International Conference on Speech Database and Assessments (Oriental COCOSDA).* (pp. 74–78). Hsinchu City, Taiwan: IEEE.

Shi, S., Zhang, J. & Xie, Y. (2014) Cross-language comparison of F0 range in speakers of native Chinese, native Japanese and Chinese L2 of Japanese: Preliminary results of a corpus-based analysis. In M. Dong, J. Tao, H. Li, T. F. Zheng & Y. Lu (eds) *Proceedings of the 9th International Symposium on Chinese Spoken Language Processing*. (pp. 241–244). Singapore: IEEE.

Shih, C. (1988) Tone and intonation in Mandarin. *Working Papers, Cornell Phonetics Laboratory*, 3, 83–109.

J't Hart, J., Collier, R. & Cohen, A. (1990) *A Perceptual Study of Intonation: An Experimental Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.

Talkin, D. (1995) A robust algorithm for pitch tracking (RAPT). In W. B. Klejin & K. K. Paliwal (eds) *Speech coding and synthesis*. (pp. 495–518). Amsterdam, The Netherlands: Elsevier Science B.V.

Tanaka-Ishii, K. & Terada, H. (2011) Word familiarity and frequency. *Studia Linguistica*, 65(1), 96–116.

Team, R. C. (2016) R: *A language and environment for statistical computing* [Computer software manual]. Vienna, Austria.

Todaka, Y. (1990). *An error analysis of Japanese students' intonation and its pedagogical applications*. Unpublished MA thesis. Department of Lingsuitics, University of California.

Ullakonoja, R. (2007) Comparison of pitch range in Finnish (L1) and Russian (L2). In J. Trouvain & W. J. Barry (eds) *Proceedings of the 16th International Congress of Phonetic Sciences*. (pp. 1701–1704). Saarbrücken, Germany: Pirrot GmbH.

Urbani, M. (2012) Pitch range in L1/L2 English. An analysis of f0 using LTD and linguistic measures. In M. G. Busà & S. Antonio (eds) *Methodological perspectives on Second Language prosody*. (pp. 79–83). Padova, Italia: Università di Padova

Valdés, G., Poza, L. & Brooks, M. D. (2015) Language acquisition in bilingual education. In Wayne E. W., Sovicheth B. & Ofelia G. (eds) *The handbook of bilingual and multilingual education*. (pp. 56-74). Malden, MA: Wiley-Blackwell.

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, 38(3), 253–265.

Van Dommelen, W. A. & Moxness, B. H. (1995) Acoustic parameters in speaker height and weight identification: sex-specific behaviour. *Language and Speech*, 38(3), 267–287.

Vanrell, M. del M., Mascaró, I., Torres-Tamarit, F. & Prieto, P. (2010) When intonation plays the main character: information-vs.confirmation-seeking questions in Majorcan Catalan. *Proceedings of the 15th International Conference on Speech Prosody*. (pp. 4–7). Chicago, USA: S. Karger AG.

Vilaplana, E. E. (2008) Modelling final declarative intonation in English and Spanish. In T. Gibert & L. Alba (eds) *Estudios de Filología Inglesa: Homenaje a La Dra. Asunción Alba Pelayo*. (pp. 109–129). Madrid, Spain: Ediciones UNED.

Volín, J., Poesová, K. & Weingartová, L. (2015) Speech melody properties in English, Czech and Czech English: Reference and interference. *Research in Language*, *13*(1), 107–123.

Wennerstrom, A. (1998) Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition*, 20(1), 1–25.

Wu, Z. (1982) Rules of intonation in Standard Chinese. *Preperation for Working Group on Intonation, 8th International Congress of Linguists*. (pp. 95–108). Tokyo, Japan.

Xu, H. (2009). *A survey study of Chinese EFL learners' acquisition of English intonation: A functional perspective*. Unpublished MA thesis. Department of English Philosophy, Jiangsu University.

___. (2005) Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 46(3–4), 220–251.

___. (2017) Intonation in Chinese. In W. Behr, Y. Gu, Z. Handel, C.-T. J. Huang & J. Myers (eds) *Encyclopedia of Chinese language and linguistics*. (pp. 458–466). Boston: Brill.

Xu, Y. & Wang, Q. E. (2001) Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, 33(4), 319–337.

Yamazawa, H. & Hollien, H. (1992) Speaking fundamental frequency patterns of Japanese women. *Phonetica*, 49(2), 128–140.

Yip, M. (2002) *Tone.* Cambridge: Cambridge University Press.

Yuan, C., González-Fuente, S., Baills, F. & Prieto, P. (2019) Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, 41(1), 5–32.

Yuan, J. (2004). *Intonation in Mandarin Chinese: Acoustics, perception, and computational modeling*. Unpublished doctoral dissertation. Department of lingusitics, Cornell University.

___. (2006) Mechanisms of question intonation in Mandarin. In B. Ma, Huo Q., E.-S. Cheng & H. LI (eds) *Proceedings of the 5th International Symposium on Chinese Spoken Languag.* (Vol. 4274, pp. 19–30). Berlin, Germany: Springer.

Yuan, J., Dong, Q., Wu, F., Luan, H., Yang, X., Lin, H. & Liu, Y. (2018) Pitch characteristics of L2 English speech by Chinese speakers: A large-scale study. *Proceedings of the 19th Annual Conference of the International Speech Communication Association*. (pp. 2593–2597). Hyderabad, India: Curran Associates, Inc.

Yuan, J. & Liberman, M. (2014) F0 declination in English and Mandarin broadcast news speech. *Speech Communication*, 65, 67–74.

Yuan, J. & Shih, C. (2004) Confusability of Chinese Intonation. In B. Bernardl & M. Isabelle (eds) *Proceedings of the 2nd International Conference on Speech Prosody*. (pp. 1–4). Nara, Japan.

Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B. & Trouvain, J. (2014) Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. In N. Campbell, D. Gibbon & D. Hirst (eds) *Proceedings of the 7th International Conference on Speech Prosody*. (pp. 1037–1041). Dublin, Ireland: Trinity College Dublin.

Appendix A

Biographical information for L2 learners in the study, with Mandarin Chinese as the L1.

Code	Gender	Age at test	Age of acquisition	Month of residence in Spain	L2 proficiency	China origin	L2 dialect
ci01	F	22	18	10	B2	Henan	PS
ci02	F	22	18	10	B2	Sichuan	PS
ci03	M	26	21	12	B2 (SJ)	Shandong	PS
ci04	F	22	18	12	B2	Liaoning	PS
ci05	M	23	19	13	B2	Jiangsu	PS
ci06	F	23	19	14	B2	Jiangsu	PS
ci07	F	24	21	20	B2 (SJ)	Liaoning	PS
ci08	F	22	18	22	B2	Jiangsu	PS
ci09	M	31	24	22	B2	Shandong	PS
ci10	F	22	18	22	B2	Zhejiang	PS
ci11	F	23	17	25	B2	Anhui	PS
ci12	M	24	18	25	B2	Jiangxi	PS
ci13	M	24	18	25	B2	Shanxi	PS
ci14	M	21	19	32	B2 (SJ)	Guangdong	PS
ci15	F	23	19	40	B2 (SJ)	Ningxia	PS
ci16	F	23	18	60	B2 (SJ)	Shanghai	PS
ca17	F	24	18	1	C1	Shandong	PS

ca18	F	22	19	2	C1	Gansu	PS
ca19	F	24	20	2	C1	Jilin	PS
ca20	F	23	19	3	C1	Neimenggu	PS
ca21	F	22	18	7	C1	Beijing	PS
ca22	F	22	18	8	C1	Shanghai	PS
ca23	F	24	19	10	C1	Shanxi	PS
ca24	F	22	18	11	C1	Heilongjiang	PS
ca25	F	26	24	24	C1	Guangxi	PS
ca26	F	25	18	25	C1	Heilongjiang	PS
ca27	F	30	19	25	C2	Liaoning	PS
ca28	F	24	22	25	C1	Zhejiang	PS
ca29	F	29	18	36	C1	Henan	PS
ca30	F	24	18	40	C1	Heilongjiang	PS
ca31	F	29	19	40	C1	Anhui	PS
ca32	F	26	12	48	C2	Tianjin	PS

Note: Proficiency (SJ = self-judgement by L2 learners); L2 dialect (PS = Peninsular Spanish).

Appendix B

Test items

(1)	Sample contexts for information-seeking *yes-no* question: Entras en una frutería donde no has estado nunca y le preguntas al dependiente si tiene mandarinas. *¿Tiene mandarinas?*
(2)	Sample contexts for information-seeking *wh*-question: Has quedado con dos buenas amigas para hacer compras esta tarde. Pero al llegar la hora, sólo se presenta una chica y no ves a la otra amiga, Marina. Pregúntale dónde está Marina. *¿Dónde está Marina?*
(3)	Sample contexts for disjuntive question: Has invitado a buen amigo a tu piso para una cena. Después de acabar los platos principales, le preguntas si quiere tarta o helado de postre. *¿Quieres tarta o helado?*
(4)	Sample contexts for confirmation-seeking *yes-no* question: Estás hablando con tu compañero de piso y ves que está cubierto con dos mantas. Tú infieres que tu compañero de piso debe de tener frío y le preguntas si es así. *¿Tienes frío?*
(5)	Sample contexts for confirmation-seeking tag question: Vas a una discoteca con un amigo y ves que baila muy bien, así que supones que practica baile a menudo (mucho) y le preguntas si es así. *Bailas a menudo, ¿verdad?*

Other test items

Stress position	Penultimate stressed syllable	Final stressed syllable
	¿Tiene cambio?	¿Tiene melón?
	¿Cuándo viene?	¿Dónde está Manuel?
	¿Viene el lunes o el martes?	¿Quieres melón o melocotón?
	¿Tienes hambre?	¿Tienes calor?
	Sales al mercado, ¿no?	No te encuentras bien, ¿verdad?

Capítulo 4

Cross-linguistic comparison of the pitch and temporal profiles between L1 and Chinese L2 speakers of Spanish

Citado como:

Shang, P. (en prensa). Cross-Linguistic Comparison of the Pitch and Temporal Profiles between L1 and Chinese L2 Speakers of Spanish. *Loquens*.

Cross-Linguistic Comparison of the Pitch and Temporal Profiles between L1 and Chinese L2 Speakers of Spanish

Peizhu Shang

Universitat de Barcelona
Pshangsh7@alumnes.ub.edu

ABSTRACT

Cross-linguistic studies between intonational languages suggest that there is a universal trend during the L2 learning process regarding pitch and temporal characteristics. We extend these hypotheses to Chinese learners of Peninsular Spanish—a new pairing of tone and non-tone languages. Using six pitch and temporal metrics, we examine how Chinese learners' pitch and temporal profiles deviated from those of L1 native speakers and explore the factors that may contribute to L2 speech deviations. The Discourse Completion Task was conducted to elicit five question types produced by 37 participants, who were divided into three language groups. Consistent with previous literature, our study shows that Chinese L2 learners had a compression of pitch span (at both the utterance and syllable levels) and pitch variability, as well as a strong reduction of pitch change rate, speech rate, and articulation rate compared to L1 Spanish speakers. Most pitch and temporal deviations in L2 Spanish intonation are closely linked to psychological and cognitive attributes rather than being determined by physiological factors or L1 tonal transfer. Moreover, the lack of prosodic knowledge of the target intonation patterns concerning the different question types may also hinder L2 learners from approaching a native-like pitch and temporal profile.

Keywords: pitch span, pitch variability, temporal features, Chinese speakers of L2 Spanish.

RESUMEN

Estudios interlingüísticos entre lenguas entonativas sugieren que puede haber una tendencia universal durante el proceso de aprendizaje de la L2 con respecto a las características tonales y temporales. Extendemos estas hipótesis a los aprendices chinos de español peninsular—una nueva combinación lingüística entre lenguas tonales y entonativas. Usando seis métricas tonales y temporales, pretendemos examinar cómo los aprendices chinos se desvían de los hablantes nativos en los perfiles tonales y temporales, y explorar los factores que contribuyen a las desviaciones en el habla de la L2. Se ha realizado la Tarea de Finalización del Discurso para elicitar cinco tipos de preguntas producidas por los 37 participantes divididos en tres grupos lingüísticos. En línea con la literatura anterior, nuestro estudio muestra que los aprendices chinos tenían una compresión del rango tonal (tanto a nivel oracional como a nivel silábico) y la variedad tonal, así como una reducción significativa en la tasa del cambio tonal, la velocidad del habla, y la tasa de articulación en comparación con los hablantes nativos de español. La mayoría de las desviaciones tonales y temporales en la entonación de la L2 están estrechamente relacionadas con atributos psicológicos y cognitivos más que con factores fisiológicos o con la transferencia tonal de la L1. Además, la falta de conocimiento prosódico de los patrones entonativos relativos a los diferentes tipos de pregunta en la lengua meta también impide que los aprendices de L2 se asimilen a un perfil tonal y temporal similar a los nativos.

Palabras clave: rango tonal, variedad tonal, características temporales, español como L2 producido por sinohablantes.

1. INTRODUCTION

In the last decades, while there has been a growing body of work on the acquisition of non-native Spanish segments (i.e., Chen, 2007; Cobb & Simonet, 2015; Liu, 2019; Morrison, 2003), stress (i.e., Chen, 2007b; Cortés

Moreno, 2005; Kim, 2015; Kimura, Sensui, & Takasawa, 2015), prominence (i.e., Kim, 2016; Van Maastricht, Krahmer, & Swerts, 2016), and intonation contours (i.e., Gabriel & Kireva, 2014; Henriksen, Geeslin, & Willis, 2010; Silva & Barbosa, 2017; Trimble, 2013; Yuan et al., 2019), little is known about the acoustic-phonetic realization of pitch and temporal patterns in L2 Spanish, particularly in environments of language contact between tone and non-tone languages such as Chinese and Spanish. Therefore, the goal of the present study is to fill in the gap by examining cross-linguistic differences of pitch and temporal profiles between first- (L1) and second-language (L2) speakers of Peninsular Spanish.

Pitch profiles consist of the oscillations of fundamental frequency (F0) and are claimed to have quasi-universal and language-specific characteristics in human communication (Chen, Gussenhoven, & Rietveld, 2004; Gussenhoven & Chen, 2000). The generalizability in the use of pitch to convey certain paralinguistic meanings is often explained with biologically determined codes. For example, the frequency code proposes that high pitch is related to a small larynx and often serves as a marker of uncertainty, whilst low pitch is associated with a larger organ of production and is used to signal assertiveness (Gussenhoven, 2002; Ohala, 1983). However, despite this commonality, it is broadly recognized that language communities differ from each other in the specific phonetic implementation of pitch patterns, such as register and range. For instance, by combining the linguistic and the long-term distributional (LTD) measures, Mennen et al. (2012) found that English female speakers had a significantly higher F0 register and a larger F0 span than their German counterparts. Similar cross-linguistic differences in pitch profiles have also been observed for Polish vs. English (Majewski et al., 1972), Russian vs. German (Nebert, 2013), Mandarin vs. English (Keating & Kuo, 2012), Mandarin vs. Japanese (Shi et al., 2014), Slavic and Germanic languages (Andreeva et al., 2014), and many others (see Mennen et al., 2012 and Ordin & Mennen., 2017 for a review). Apart from the influence of the L1 prosodic system and some physiological factors such as vocal tract length, gender, and age, the language-specific pitch properties are possibly more closely linked to some social-cultural attributes. Unmistakable evidence for this is that Japanese speakers, particularly women, have a higher F0 register and F0 span than native speakers of Chinese (Shi et al., 2014), Dutch (Van Bezooijen, 1995), American English, and Spanish (Hanley et al., 1966). The preference for high pitches shown by Japanese women is explained in the context of their relative powerlessness in social status and the gender roles they are expected to play according to cultural conventions.

Furthermore, since the speech of a foreign language often entails some degree of interaction, the cross-language differences between the first and the second language can be expected to impact the target speech patterns. Studies have shown that most L2 segmental and suprasegmental errors could be attributed to a prosodic transfer from the L1 system into the phonetic and phonological knowledge of the L2 (Graham & Post, 2018; Mennen, 2015). However, importantly, several studies have found that some deviated use of pitch is common in L2 speech, revealing itself as a consistent development trajectory during the L2 speech-learning process. For example, the results in previous literature (i.e., Busà & Urbani, 2011; Chen, 1972; Mennen, Schaeffler, & Dickie, 2014; Shi et al., 2014; Ullakonoja, 2007; Yuan et al., 2018) suggest that foreign speakers, regardless of their L1–L2 backgrounds, are often characterized by a narrower F0 range and less variable pitch when producing the L2 speech on the utterance level. In contrast, on the phonemic level, Chinese L2 speakers were reported to have a wider pitch span and smaller F0 fluctuations than native English speakers, mostly due to the negative attachment of L1 lexical tones to stressed syllables in the L2 (Ding et al., 2016; Yuan et al., 2018).

The difficulty of accurately implementing the target pitch profiles has been mainly correlated with the L2 learners' lack of confidence and insecurity when speaking a

foreign language (Ding et al., 2016; Shi et al., 2014; Yuan et al., 2018), and not merely due to the language specificities and the different socio-cultural identities. Another plausible factor that may constrain the pitch variance is the learners' increased cognitive efforts in producing segments and stress (Zimmerer et al., 2014). Nevertheless, fortunately, studies showed that, with the aid of speech technology or with developing their proficiency in L2, learners were able to fine-tune the production of the L2 pitch and finally approach native-like pitch patterns (Hincks & Edlund, 2009; Ullakonoja, 2007).

On the other hand, L2 speech is also found to be characterized by a decrease in oral fluency (Peters, 2019). The differences in fluency between the L1 and the L2 are frequently measured by various temporal metrics. For example, Ding et al. (2016) showed that, in comparison with native English speakers, Chinese learners tend to have a lower speech rate and articulation rate in their L2 English. Lee and Sidtis (2017) and Peters (2019) made similar observations. The decrease in speech fluency in the non-native language has been explained with reference to the same psychological and cognitive factors as L2 pitch compression—cautiousness and increased cognitive efforts when speaking a foreign language. However, unlike the two variables of speech rate and articulation rate, the temporal assumption of pitch change rate is controversial, especially when it is examined in a stress language such as English compared to a tone language like Chinese. For instance, Yuan et al. (2018) reported a faster pitch change rate for L1 English speakers than for L2 Chinese learners, while in Ding et al. (2016), there was no significant difference between the two language groups with regards to the speed of pitch changes.

Despite the large body of cross-linguistic analyses of pitch and temporal differences, it is somewhat difficult to compare the results of these findings. This is partly because the F0 estimation methods and the fluency measures used for evaluating the pitch and temporal properties differed across studies. Another aspect is that the distinct discourse conditions designed to elicit the speech may also cause inconsistent results. For instance, Yuan and Liberman (2014) reported that Chinese native speakers have a wider pitch range and greater F0 fluctuations in broadcast news speech than native English speakers. However, regarding prose passages (Keating & Kuo, 2012), there was no significant difference in pitch range on the utterance level between Chinese and English speech.

Given the inconsistency of prior results and the typological differences between Chinese and Spanish, it is of great importance to examine the pitch and temporal characteristics in the CH-ES language pair, which has received little attention in the prosodic field to date. Of particular interest to us is to investigate (1) whether the pitch and temporal profiles produced by L2 Chinese learners are highly dependent on their L1 properties or if they support the L2 general trend hypothesis, (2) whether speakers' pitch and temporal implementations are influenced by the gender and the level of proficiency in Spanish, and finally (3) whether the production of L2 pitch and temporal features reflects different levels of difficulty depending on question type and stress position. For these purposes, we extend the previous studies by accounting for proficiency level, gender, question type, and stress position, which allows us to examine the interaction between proficiency and other fixed factors concerning various pitch and temporal metrics.

2. METHODOLOGY

2.1. Participants

The participants of this study included: 5 female native speakers of Peninsular Spanish and 32 learners of Spanish (26 females and 6 males) whose first language is Mandarin Chinese. The ages of Chinese learners ranged from 21 to 31 (mean age: 24.09; *SD* = 2.53), while those of L1 Spanish speakers ranged from 18 to 24, with a mean age of 23.2 years (*SD* = 4.87). All subjects were divided into three language groups according to their proficiency level in Spanish: intermediate (B1-B2 level), advanced (C1-C2 level), and native. The Spanish proficiency of most Chinese

speakers was judged using the information from their most recent official language qualification DELE (Diploma of Spanish as a Foreign Language). Chinese learners who did not have this certificate (approximately 15%) were asked to self-evaluate their L2 proficiency based on the Spanish language courses they had completed. The criteria for the six levels of European language proficiency were explained to those participants to help them to reach a reliable self-assessment.

Although the age of acquisition and the length of exposure to the target language are reported to influence L2 speech (Cadierno et al., 2020; Kharkhurin, 2008; Pfenninger & Singleton, 2016), we did not control for these variables, as this would have significantly reduced the number of L2 Chinese participants. However, most of the Chinese learners in this study acquired Spanish in adulthood (mean age: 18.81; *SD* = 2.08). Only one subject reported starting to learn Spanish at 12 years of age. All the Chinese participants were in an immersion situation at the time of recording.

2.2. Task and materials

The corpus was elicited by utilizing the DCT (Discourse Completion Task) technique (Billmyer & Varghese, 2000; Félix-Brasdefer, 2010). Specifically, we designed 15 brief dialogues structured as situational contexts to elicit five question types with different functional meanings in Spanish, namely, information-seeking *yes-no* question ('YN'), information-seeking *wh*-question ('WH'), *disjunctive* question ('DJ'), confirmation-seeking *yes-no* question ('CYN'), and confirmation-seeking *tag* question ('TAG'). The conversational interaction was initiated by an interlocutor with whom the participant was familiar so that politeness-related effects (e.g., power, and social distance) could be minimized (Borràs-Comes, Sichel-Bazin, & Prieto, 2015; Roseano et al., 2015). A sample context for eliciting the disjunctive question is as follows:

- Interlocutor: *Has invitado a un buen amigo a tu piso para una cena. Después de acabar los platos principales, le preguntas si quiere tarta o helado de postre.* (You have invited a good friend to your apartment for dinner. After finishing the main courses, you ask her if she wants cake or ice cream for dessert.)
- Participant: *¿Quieres tarta o helado?* (Do you want cake or ice cream?)

Each of the five question types varied in the nuclear stress position (two positions: penultimate syllable stress—paroxytone; final syllable stress—oxytone). To facilitate L2 speakers' comprehension during the task, all test items consisted of words with high frequency for L1 and L2 Spanish speakers (Tanaka & Terada, 2011).

The recordings took place in a soundproof room with a head-mounted microphone. Speech files were digitalized at a sampling rate of 44.1 kHz and with a quantization precision of 16 bits. Each utterance was saved separately and annotated to a *TextGrid* object in *Praat* (Boersma & Weenink, 2020).

2.3. Data extraction

For the purposes of this paper, two types of measurements were conducted: (a) pitch and (b) temporal measures. In order to extract the pitch information from the utterances, firstly, the ESPS algorithm ('get F0') (Talkin, 1995) was automatically conducted in *Praat* with the pitch floor and ceiling set to 70 Hz and 600 Hz, respectively. A time step of 10 ms was used for the computation of F0. After the automatic extraction, the raw F0 data were corrected manually, unvoicing those pitch points with octave jumps or measurement errors, such as false voicing in silent fragments, creaky voice, and laryngealization. The linear results in Hz were then transformed into the near-logarithmic scale (ERB-rate), which is one of the best psycho-acoustic measures for modelling the intonational equivalence between men and women, and for capturing the F0 differences across languages (Nolan, 2003).

In specific, pitch characteristics in this study were evaluated by means of the three F0 variables: (1) 80% pitch span on the utterance

level (the 90th and 10th percentile span), (2) absolute span on the syllable level (the 100th percentile span), and (3) pitch dynamism quotient (abbreviated as PDQ). The PDQ metric was included as a normalization of the F0 variation data since it can minimize the effects caused by gender and different group size. The PDQ value gives an account of the pitch variability in the utterance, and it is calculated by dividing the standard deviation by the F0 mean. In general, the previous literature indicates that the higher the PDQ, the more variable the speech (Shi, Zhang, & Xie, 2014; Wang & Qian, 2018; Zimmerer et al., 2014).

Further, considering the temporal traits, three variables were examined between L1 and L2 speech: (1) pitch change rate (the average of the absolute pitch differences in every 10-ms interval), (2) speech rate (number of syllables / total duration of the utterance), and (3) articulation rate (number of syllables / (total duration–internal pauses). The minimum pause length calculated for fluency judgments was set to 0.05 s instead of the larger values of 0.25 s adopted in the study of Peters (2019). The underlying reason is that the speech materials used in our experiment were single utterances with an average syllable number of 5.8—unlike the passages in Peters (2019) that frequently required the use of long pauses as a linguistic cue for narrative segmentation (Oliveira, 2002).

2.4. Statistical analysis

The data analysis was conducted in the R environment (R Core Team, 2020). A linear mixed-effects analysis was carried out using the *lmerTest* package for R (Kuznetsova et al., 2017). The six pitch and temporal parameters (80% span on the utterance level, PDQ, 100% span on the syllable level, pitch change rate, speech rate, and articulation rate) were entered into the model successively as dependent variables, with *Proficiency Level* in Spanish (intermediate < advanced < native), *Gender* (female vs. male), *Question Type* (i.e., YN, WH, DJ, CYN, and TAG), *Stress Type* (Oxytone vs. Paroxytone), and their possible interactions as fixed effects. Participants were included as random effects with all possible random intercepts. The significance of the main effects was tested using the *ANOVA* function. *P*-values were fitted by eliminating the non-significant effects of the initial model and calculated with Satterthwaite's method (Kuznetsova et al., 2017). The post-hoc analysis was performed using the single-step function of the *multcomp* package (Hothorn et al., 2016) supported by the *emmeans* algorithm (Lenth et al., 2019).

3. RESULTS

The following two sections present the results of the three pitch variables measured on the utterance (80% F0 span, and PDQ) and syllable level (100% F0 span), and the results of the three temporal parameters (pitch change rate, speech rate, and articulation rate).

3.1. Pitch results

First, we considered the differences in the use of pitch across the three language groups. The analysis of variance indicated that *Proficiency Level* was not a significant factor for the three pitch variables (see Table 1). However, Figures 1, 2, and 3 indicate that Chinese intermediate (hereafter CI) and advanced learners (hereafter CA) tend to produce a less variable pitch and narrower span on the utterance and syllable levels compared to L1 Spanish speakers (hereafter SN). These findings generally are consistent with previous studies that reported a reduced pitch for non-native speakers (Busà & Urbani, 2011; Mennen, Schaeffler, & Docherty, 2007; Shi et al., 2014; Yuan et al., 2018; Zimmerer et al., 2014), suggesting that there may be a universal trend of pitch range compression in L2 speech. Additionally, the results in Figures 1, 2, and 3 indicated that, in comparison with the lower proficiency group (i.e., the CI group), highly proficient learners of the CA group were closer to SN speakers in the implementation the F0 pitch, although this trend was not strong enough to be statistically significant (see Table 2).

Table 1: Effects (*F*-values) of Proficiency level, Question type, Gender, Stress position, and their interactions on the three pitch variables ('***' $p < 0.001$; '**' $p < 0.01$; '*' $p < 0.05$; '.' $p < 0.1$).

	80% utterance span	PDQ	100% syllable span
Proficiency	2.99.	2.80.	2.53.
QuestionType	10.53***	8.99***	22.26***
Gender	0.00	8.76**	1.33
StressType	3.12.	4.00*	0.42
Proficiency*QuestionType	9.98***	8.42***	8.58***
Proficiency* StressType	3.54*	0.22	0.37

Table 2: Pairwise comparisons between the three language groups regarding the three pitch variables.

	80% utterance span	100% syllable span	PDQ
CI-CA	$t = -1.802$, $p = 0.179$	$t = -1.783$, $p = 0.185$	$t = -1.791$, $p = 0.182$
SN-CA	$t = 1.029$, $p = 0.559$	$t = 0.766$, $p = 0.723$	$t = 0.932$, $p = 0.620$
SN-CI	$t = 2.289$, $p = 0.067$	$t = 2.029$, $p = 0.116$	$t = 2.190$, $p = 0.083$

Next, as with the *Question Type* factor, it is apparent in Table 1 that there is a significant main effect on the three pitch variables. In contrast, the factors *Gender* and *Stress Type* were found to be significant only for the variable of PDQ. In particular, our results indicated that female speakers (mean PDQ: 0.175) had significantly more F0 variability than males (mean PDQ: 0.127) in speech [$t(70) = 2.14$, $p < 0.05$). We also observed a significant effect of *Stress Type* on the variable of PDQ. Specifically, it is noteworthy in Figure 2 (see the right panel) that participants of the three language groups consistently had a more variable pitch in questions with a paroxytone than those with an oxytone in the final word.

As with the 80% utterance span, Figure 1 shows that the two Chinese groups had a wider pitch span in questions ending with a paroxytone word, but this tendency was statistically significant only for the CA group [$t(539) = 3.07$, $p < 0.01$]. Regarding the SN group, we did not find a statistically significant difference in realizing the pitch between the two stress types [$t(539) = 0.04$, $p = 0.76$], although SN speakers were more likely to compress the F0 span in questions ending with a paroxytone word (see the right panel of Figure 1). The pitch performance exhibited by the CI and CA groups may be because the paroxytone is the most frequent and unmarked stress pattern in Spanish and, therefore, the most familiar one for L2 speakers (Defior & Serrano, 2017; Roca, 2019). This means that Chinese learners may experience the least cognitive difficulties when producing such stressed words in Spanish, which allows more planning time to fine-tune the corresponding pitch profiles in a native-like way. In contrast, it is unclear why SN speakers had an opposite trend for implementing the F0 span between the two stress types. Since we only had five Spanish subjects in this work, future investigations with a larger sample size are needed to validate this finding.

The results of the linear mixed model also revealed a strong interaction effect between *Proficiency Level* and *Question Type* on the three pitch variables (see Table 1). The post-hoc analysis indicated that the pitch performance of CI and CA learners was highly dependent on the question type in which they were engaged. More precisely, we found that, in comparison with the SN group, the CI and CA group had a particularly narrower span and less pitch variability in DJ [e.g. 80% span: CI-SN: $t(2) = -4.04$, $p < 0.001$; CA-SN: $t(2) = -3.79$, $p < 0.001$] and YN questions [e.g. PDQ: CI-SN: ($t(2) = -3.35$, $p < 0.01$); CA-SN: ($t(2) = -2.47$, $p < 0.05$)]. By contrast, in WH questions, it is noteworthy that the two

Chinese groups had a higher PDQ and a wider pitch span on both utterance and syllable levels than the SN group (see Figures 1, 2, and 3). This finding can be explained by the overproduction of WH questions by Chinese learners. Specifically, we notice that some L2 learners, irrespective of their level of proficiency, tend to produce a high-rising nuclear pitch accent or a final rising boundary tone in WH questions. Although the final rising contour can also be used in WH questions, it is not frequently found in the L1 native speech (i.e., all the SN speakers in our study produced the WH questions with a final-falling pitch movement) since the interrogative particles in Spanish (e.g., *qué, dónde, quién, cuál*) are clear enough for signalling this type of question.

Figure 1: 80% pitch span on the utterance level of the three language groups depending on Question Type (left) and Stress Type (right). Error bars indicate ± 1SE.

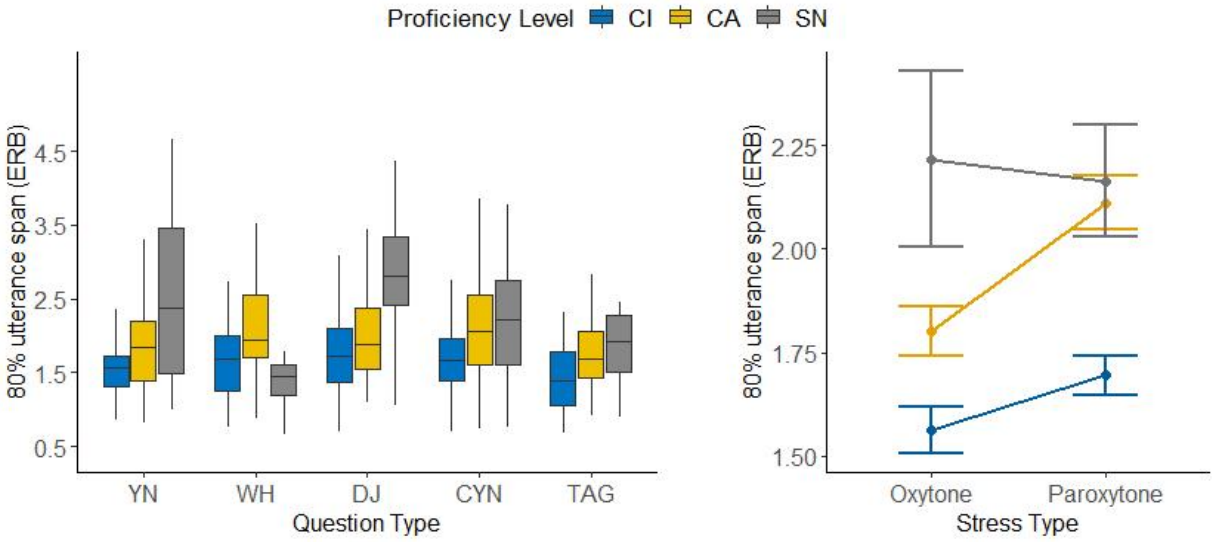


Figure 2: Mean PDQ of the three language groups depending on Question Type (left) and Stress Type (right). Error bars indicate ± 1SE.

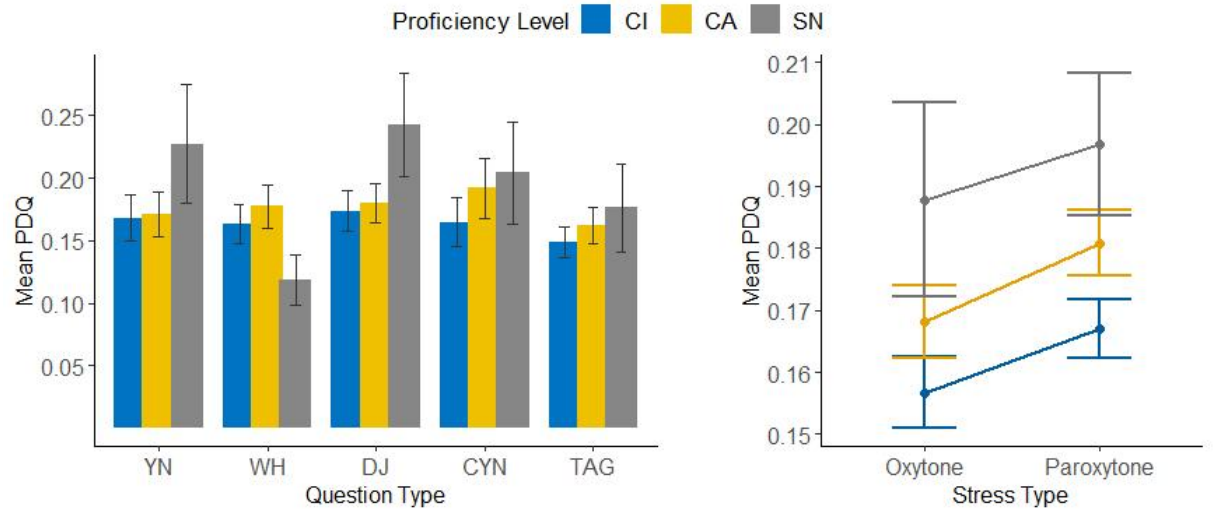
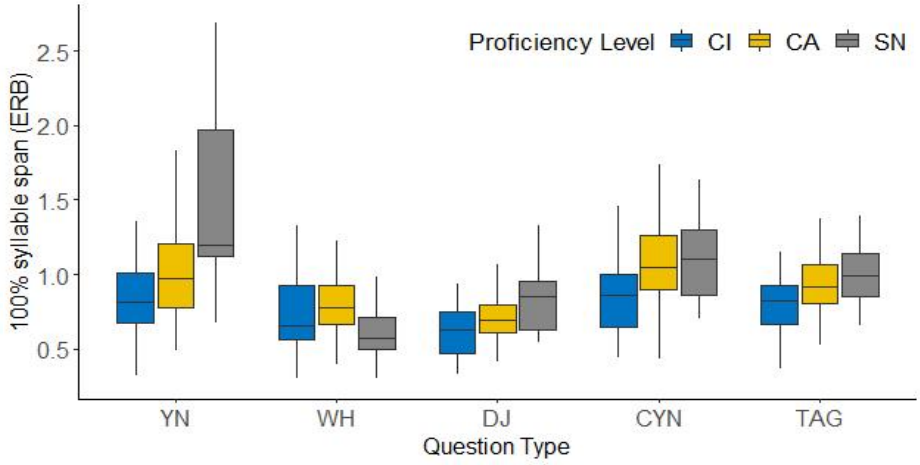


Figure 3: 100% pitch span on the syllable level depending on Proficiency Level and Question Type.



3.2. Temporal results

The main effects of the linear mixed models fitted for the three temporal variables are shown in Table 3. For ease of exposition, we discuss these results by referring to Figures 4 and 5, which display the specific temporal values produced by the three language groups in the five question types. First, considering individual effects, the output in Table 3 revealed that there was a significant main effect of *Proficiency* and *Question Type* on the outcome variables of pitch change rate, speech rate, and articulation rate. By contrast, *Stress Type* and *Gender* were insignificant factors for the three temporal variables. Moreover, the pairwise comparisons of *Proficiency Level* showed that, in comparison with the SN group, the two Chinese groups had a significantly lower pitch change rate [CI-SN: $t(2) = -4.71$, $p < 0.001$; CA-SN: $t(2) = -3.75$, $p < 0.01$], speech rate [CI-SN: $t(2) = -5.71$, $p < 0.001$; CA-SN: $t(2) = -5.62$, $p < 0.001$], and articulation rate [CI-SN: $t(2) = -5.58$, $p < 0.001$; CA-SN: $t(2) = -5.44$, $p < 0.001$] in their speech. These findings corroborate previous studies that reported a reduced oral fluency for L2 speakers in the non-native language (Ding et al., 2016; Peters, 2019). Nevertheless, unlike our previous findings—which showed that high-proficiency Chinese learners achieved a target-like pitch performance—(see Section 3.1), we did not observe any significant improvement in speech rate and articulation rate between the CI and CA groups.

Further, the results in Table 3 indicated a strong interaction between *Proficiency* and *Question Type* on the three temporal variables. Particularly, as shown in Figure 5, SN speakers had higher values of pitch change rate than the CI and CA learners in all questions except for WH questions. As discussed above, the faster pitch change in L2 WH questions may be attributed to the fact that most Chinese learners excessively varied their F0 contours by producing either a high pitch accent or a final rising boundary in the nuclear position. In addition, although each question type was realized with a specific temporal value, the two Chinese groups were consistently lower than the SN speakers regarding the speech and articulation rates (see Figure 5). Finally, it is interesting that the results of speech rate and articulation rate were similar in this work. This is perhaps because the speech stimuli used in this work consisted of short utterances produced with low frequency and short pauses.

4. DISCUSSION

The aim of the present study was to investigate the L2 production of Spanish questions by Chinese speakers with regards to pitch and temporal characteristics and to explore the factors that may contribute to the pitch and temporal deviations in L2 speech. Six pitch and temporal metrics of L1 and L2

Table 3: Effects (*F*-values) of Proficiency Level, Question type, Stress Type, and their interactions on the three temporal variables ('***' $p < 0.001$; '**' $p < 0.01$; '*' $p < 0.05$; '.' $p < 0.1$).

	Pitch change rate	Speech rate	Articulation rate
Proficiency	11.23***	18.75***	17.75***
QuestionType	14.95***	10.56***	4.35**
Gender	0.03	2.71	3.26.
StressType	0.01	0.22	0.00
Proficiency*QuestionType	11.46***	3.80***	2.64**

Figure 4: Pitch change rate of the three language groups depending on Proficiency Level and Question Type. Error bars indicate ± 1SE.

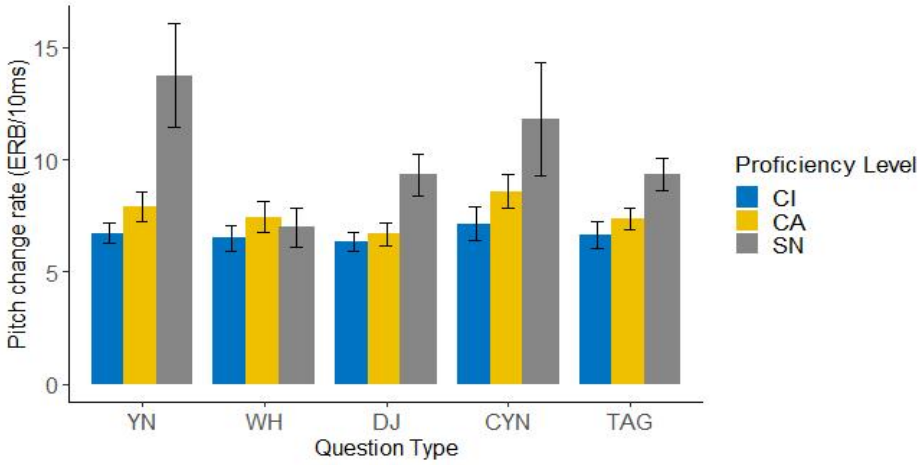
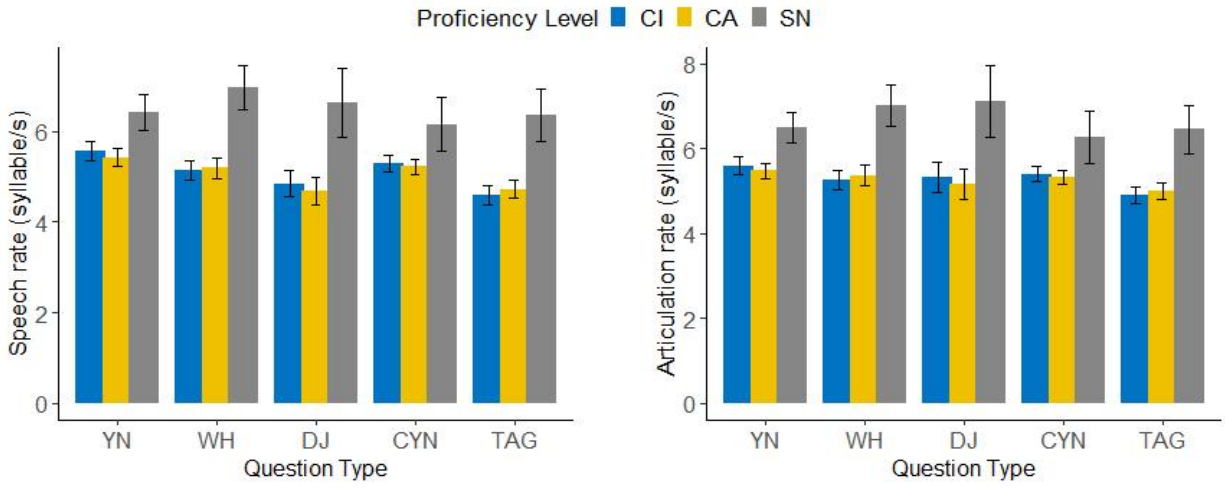


Figure 5: Speech rate (left) and articulation rate (right) of the three language groups depending on Proficiency Level and Question Type. Error bars indicate ± 1SE.



Spanish speakers were examined and compared using a linear mixed-effects analysis. The findings of our study are discussed below.

First, our results confirm that there are indeed some cross-linguistic differences between Spanish L1 and L2 regarding pitch performance. The evidence in support of this is that the L2 Spanish in this study was produced

with a narrower span (on both utterance and syllable levels) and less variable pitch than that of L1 native speakers. This supports previous studies that reported a pitch range compression effect for L2 speakers with typologically different L1 backgrounds (e.g., Busà & Urbani, 2011; Liu, 2005; Mennen et al., 2007, 2012, 2014; Peters, 2019; Shi et al., 2014; Ullakonoja, 2007; Urbani, 2012; Yuan et al., 2018; Zimmerer et al., 2014). The consistency of the findings for L2 pitch and temporal production suggests that non-native learners may have universal developmental pathways for acquiring specific aspects of L2 speech, independent of the specificity of their L1 system. We cannot provide a definitive explanation for this quasi-universal effect in L2 speech. However, rather than being shaped by the L1 phonetic system, the compressed pitch patterns in L2 have previously been attributed to the lack of confidence and insecurity of L2 learners when speaking a non-native language (Peters, 2019; Volín, Poesová, & Weingartová, 2015; Yuan et al., 2018). Additionally, the increased cognitive efforts in producing the non-native segmental or suprasegmental features (i.e., vowels and consonants, stress, and prominence) are also plausible factors that may lead to a lower pitch variability in L2 utterances. For instance, Zimmerer et al. (2014) pointed out that L2 learners can frequently overlook the variation of F0 pitch in a native-like way because they are too focused on the correct production of words and stress in the non-native language.

Another noteworthy point in the pitch results is the F0 span at the syllable level. As a typical tone language, Chinese makes use of the F0 information for encoding lexical tone meanings (Yuan, 2011). Therefore, it is expected that Chinese learners would show greater F0 variations on the syllable level because of L1 tonal transfer. However, unlike Ding et al. (2016), we did not find a wider pitch span on the syllable level for Chinese learners of Spanish. This seems to imply that the production of the L2 syllable span was not necessarily affected by the learners' long-term experience with a tone language. The discrepancy between the results could be justified by the distinct language pairs examined in the experiment: In Ding et al. (2016), English was the Chinese learners' L2, whereas in our study, it was Spanish. Future studies regarding the pitch range differences between English and Spanish at the syllable level would help us elucidate whether this is the primary cause of the discordances found. On the other side, based on our observed data, another possible explanation for the reduced syllable span in L2 Spanish might be that Chinese learners were too cautious to vary the pitch due to a lack of intonational skills and language experience, thereby exhibiting a flat F0 contour without many fluctuations until they reached the great F0 changes in the nuclear location. Further investigations of L2 phonetic performance are required to test this hypothesis, considering the position sensitivity of pitch changes in the utterance.

Further, although the factor proficiency statistically failed to reach significance in the three pitch variables, the results seem to suggest that Chinese learners of L2 Spanish can progressively fine-tune their production of F0 values and approach a target-like pitch pattern with increasing proficiency in their L2. Moreover, results of the three pitch parameters revealed a strong interaction between proficiency level and question type, illustrating that the L2 learning of pitch implementation details is susceptible to pragmatically different question types. For instance, we found that Chinese intermediate and advanced learners consistently had a reduced pitch span and lower PDQ in all utterances except for WH questions. As is clear from the above discussion, the opposite performance of Chinese speakers on WH-questions can account for their overproduction of a high pitch accent or a final-rising boundary in the nuclear position. Or, in a more general way, it can be attributed to the fact that learners were unfamiliar with the target intonation contours of WH-questions due to the typological distance between the L1 and the L2. Thus, most would simply assume that Spanish WH-questions are produced with a high pitch in the utterance-final location based on their knowledge of the typical use of the F0 cue.

As with other question types (e.g., the information-seeking *yes-no* question and the disjunctive question), we found that most F0 targets in the utterance-final position could be accurately achieved by Chinese learners, while those in the prenuclear position were deviated and produced with a less variable contour. In this regard, our findings suggest that the compressed pitch in L2, rather than being solely determined by psychological and cognitive factors (i.e., uncertainty, cautiousness, and increased efforts when speaking the L2), is also constrained by the learners' knowledge of the target intonation categories. Overall, the different pitch performance of the L2 speakers in the five question types gives support to previous findings which proposed a scaffolding from the phonological to phonetic dimensions (Cortés Moreno, 2004; Yuan et al., 2019), suggesting that there is a hierarchy of difficulties in implementing the L2 pitch patterns depending on the prosodic similarities and dissimilarities between the first and the target language.

Considering the gender effect, our study revealed that men and women differ significantly only in the variable of PDQ. Congruent with previous works (Ordin & Mennen., 2017), female speakers in our study varied their F0 contours more frequently than male speakers. The gender differences in pitch variability are more closely linked to the speakers' willingness to express emotions in communication rather than physiological factors. Research has shown that humans express a range of emotions by readily modulating their F0 pitch, and female speakers tend to express most emotions more frequently than males in speech—except for pride and power (Brebner, 2003; Pisanski et al., 2020). In this sense, the greater pitch variance observed in the data of female speakers could be attributed to their greater emotional involvement in speech than male participants. Nevertheless, because the number of male and female speakers differed strongly in this task, this research needs to be replicated with a well-balanced design to consolidate the results presented here.

Further interesting findings related to pitch are that the F0 variation was highly modulated by the stress type, whereby all speakers produced more variable pitch in questions with a final-paroxytone word than in those with a final-oxytone word. Similarly, for the 80% F0 span, Chinese learners (particularly those of the advanced group) showed a significantly wider pitch span in questions ending with a paroxytone word. We speculate that this could be related to the relative cognitive efforts required to process the two stress types for L2 learners. Since the paroxytone is the most frequent and unmarked stress pattern in Spanish (hence the most familiar one for L2 learners), Chinese speakers may show fewer difficulties when producing it in questions and have more planning time, allowing them to better approach a target-like pitch profile. Although L1 Spanish speakers had a reduced pitch span in sentences with a final-paroxytone word, this effect did not reach statistical significance, and their average pitch span was still higher than that of Chinese learners with such stimuli. So far, we have no clear explanation for the behaviour of Spanish speakers. Since there were only five native subjects in the control group, future investigations with a larger sample size are required to test whether there is a difference of pitch span for L1 Spanish speakers in questions ending with different stress patterns.

Regarding the temporal characteristics, our study revealed significantly lower pitch change rate, speech rate, and articulation rate in L2 Spanish. These results are consistent with previous studies that reported a similar reduction of oral fluency (Ding et al., 2016; Peters, 2019) and slower pitch rises and falls in L2 speech than L1 speech (Yuan et al., 2018). Moreover, it has been noted that although Chinese is a lexical tone language with F0 peaks or valleys in every syllable, the speed of F0 changes is not significantly faster than in stress languages such as English (Xu & Sun, 2002). If this is the case, we speculate that there is no negative transfer of L1 Chinese in terms of the pitch change rate in this study. The lower values of Chinese L2 learners on the three temporal metrics might also be

attributed to their increased cognitive efforts in producing the segments or their lack of experience in the target speech.

Additionally, the interaction effects found for the three temporal variables indicate that the proficiency effect was strongly modulated by question type. Whereas the speech rate and articulation rate were lower in all question types for L2, the average pitch change rate showed an exception for the WH questions in which the F0 directions varied more frequently in L2 than in L1. Since there is no indication that the L2 deviation on WH questions was caused by the systematic differences between the two languages, we speculate that the higher values of pitch change rate and F0 span in WH questions reflected overproduction by Chinese speakers due to a lack of target intonational knowledge. Finally, the main effect of proficiency seems to suggest a trend of pitch improvement with learners' increasing L2 proficiency. In particular, our study replicates previous findings (i.e., Ullakonoja, 2007; Yuan et al., 2018; Zimmerer et al., 2014) that highly proficient learners were closer to L1 native speakers in the realization of pitch change rate, pitch span on the utterance and syllable level, and pitch variability. Further, as suggested by neurobehavioral research, the advantage of high-proficiency speakers in the L2 can be attributed to their enhanced ability to use higher-level cognition (i.e., attention) to process non-native speech components (Archila-Suerte et al., 2012, 2015).

5. CONCLUSION

The study presented here was intended to explore the pitch and temporal characteristics of native and Chinese L2 speakers of Spanish. Using six different metrics, we examined the pitch and temporal implementation in five question types of Peninsular Spanish and obtained several important findings regarding the cross-linguistic differences in the speech. First, congruent with previous literature on L2 speech, the results of this study suggest that Chinese speakers of L2 Spanish deviate from L1 native speakers mainly in the compression of pitch span (both on the utterance and syllable levels) and pitch variability, and the strong reduction of pitch change rate, speech rate, and articulation rate. Second, these pitch and temporal deviations in L2 speech are attributed to psychological-cognitive factors and the learners' lack of knowledge and intonation skills in the target language rather than physiological factors or the L1 effect.

From the pedagogical perspective, our findings hold important implications for understanding the cross-linguistic differences between L1 and the speech, underlining the importance of preparing special training methods with varied materials and contexts to reduce learners' foreign accents and improve their phonetic knowledge of the L2. Further research on native Chinese and native Spanish will be conducted to explore more cross-linguistic differences that may account for the L2 speech deviations. It is also interesting to consider how pitch span and pitch variability are realized depending on the syntactic and phonological positions of the phrase and in which locations L2 learners mostly deviate from the L1 native speakers.

6. ACKNOWLEDGEMENTS

The author is extremely grateful to Dr. Wendy Elvira-García for her technical support. An earlier version of this paper was presented at the 4th Phonetics and Phonology in Europe conference (Barcelona, June 21-23, 2021), and I want to thank the audience there for their feedback. This work has been funded by a Ph.D. research grant provided by the China Scholarship Council (No. 201808390059).

7. REFERENCES

Andreeva, B., Demenko, G., Wolska, M., Möbius, B., Zimmerer, F., Jügler, J., Oleskowicz-Popiel, M. & Trouvain, J. (2014). Comparison of pitch range and pitch variation in Slavic and Germanic languages. In *Proceedings to the 7th Speech Prosody Conference* (pp. 776–780). Dublin, Irland.

Archila-Suerte, P., Zevin, J., Bunta, F., & Hernandez, A. E. (2012). Age of acquisition and proficiency in a second language independently influence the perception of non-native speech. *Bilingualism*, *15*(1), 190.

Archila-Suerte, P., Zevin, J., & Hernandez, A. E. (2015). The effect of age of acquisition, socioeducational status, and proficiency on the neural processing of second language speech sounds. *Brain and Language*, *141*, 35–49.

Billmyer, K., & Varghese, M. (2000). Investigating instrument-based pragmatic variability: Effects of enhancing discourse completion tests. *Applied Linguistics*, *21*(4), 517–552.

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 5.3.82. http://www.praat.org/

Borràs-Comes, J., Sichel-Bazin, R., & Prieto, P. (2015). Vocative intonation preferences are sensitive to politeness factors. *Language and Speech*, *58*(1), 68–83.

Brebner, J. (2003). Gender and emotions. *Personality and Individual Differences*, *34*(3), 387–394.

Busà, M. G., & Urbani, M. (2011). A Cross Linguistic Analysis of Pitch Range in English L1 and L2. In *XVII International Congress of Phonetic Sciences* (pp. 380–383). Hong Kong, China.

Cadierno, T., Hansen, M., Lauridsen, J. T., Eskildsen, S. W., Fenyvesi, K., Jensen, S. H., & aus der Wieschen, M. V. (2020). Does younger mean better? Age of onset, learning rate and shortterm L2 proficiency in young Danish learners of English. *Vigo International Journal of Applied Linguistics*, *17*, 57–86.

Chen, A., Gussenhoven, C., & Rietveld, T. (2004). Language-specificity in the perception of paralinguistic intonational meaning. *Language and Speech*, *47*(4), 311–349.

Chen, G. T. (1972). *A comparative study of pitch range of native speakers of Midwestern English and Mandarin Chinese: An acoustic study* [Doctoral dissertation. University of Wisconsin.

Chen, Y. (2007a). *A comparison of Spanish produced by Chinese L2 learners and native speakers—an acoustic phonetics approach* [Doctoral dissertation]. Department of Philosophy, University of Illinois at Urbana-Champaign.

Chen, Y. (2007b). From tone to accent: the tonal transfer strategy for Chinese L2 learners of Spanish. In *Proceedings of 16th International Congress of Phonetic Sciences* (pp. 6–10). Saarbrücken, Germany.

Cobb, K., & Simonet, M. (2015). Adult second language learning of Spanish vowels. *Hispania*, 47–60.

Cortés Moreno, M. (2004). Análisis acústico de la producción de la entonación española por parte de sinohablantes. *Estudios de Fonética Experimental*, *13*, 80–110.

Cortés Moreno, M. (2005). Análisis experimental del aprendizaje de la acentuación y la entonación españolas por parte de hablantes nativos de chino. *Phonica*, *1*, 1–25.

Defior, S., & Serrano, F. (2017). Learning to Read Spanish. In Verhoeven, L., & Perfetti, C. (Eds.), *Learning to Read across Languages and Writing Systems* (pp. 243–269). Cambridge University Press.

Ding, H., Hoffmann, R., & Hirst, D. (2016). Prosodic transfer: A comparison study of f0 patterns in L2 english by chinese speakers. In *8th International Conference on Speech Prosody* (pp. 756–760). Boston, US.

Félix-Brasdefer, J. C. (2010). Data collection methods in speech act performance. *Speech Act Performance: Theoretical, Empirical and Methodological Issues*, *26*(41), 69–82.

Gabriel, C., & Kireva, E. (2014). Prosodic transfer in learner and contact varieties: Speech rhythm and intonation of Buenos Aires Spanish and L2 Castilian Spanish produced by Italian native speakers. *Studies in Second Language Acquisition*, *36*(2), 257–281.

Graham, C., & Post, B. (2018). Second language acquisition of intonation: Peak alignment in American English. *Journal of Phonetics*, *66*, 1–14.

Gussenhoven, C. (2002). Intonation and Interpretation: Phonetics and Phonology. In *International Conference on Speech Prosody 2002* (pp. 47–57). Aix-en-Provence, France.

Gussenhoven, C., & Chen, A. (2000). Universal and language-specific effects in the perception of question intonation. In *6th International Conference on Spoken Language Processing* (pp. 91–94). Beijing, China.

Hanley, T. D., Snidecor, J. C., & Ringel, R. L. (1966). Some acoustic differences among languages. *Phonetica*, *14*(2), 97–107.

Henriksen, N. C., Geeslin, K. L., & Willis, E. W. (2010). The development of L2 Spanish intonation during a study abroad immersion program in León, Spain: Global contours and final boundary movements. *Studies in Hispanic and Lusophone Linguistics*, *3*(1), 113–162.

Hincks, R., & Edlund, J. (2009). Using speech technology to promote increased pitch variation in oral presentations. In *International Workshop on Speech and Language Technology in Education* (pp. 1–4). Wroxall, UK.

Hothorn, T., Bretz, F., Westfall, P., Heiberger, R. M., Schuetzenmeister, A., Scheibe, S., &

Hothorn, M. T. (2016). Package 'multcomp.' *Simultaneous Inference in General Parametric Models. Project for Statistical Computing*, Vienna, Austria.

Keating, P., & Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *Journal of the Acoustical Society of America*, *132*(2), 1050–1060.

Kharkhurin, A. V. (2008). The effect of linguistic proficiency, age of second language acquisition, and length of exposure to a new cultural environment on bilinguals' divergent thinking. *Bilingualism: Language and Cognition*, *11*(2), 225–243.

Kim, J.-Y. (2015). Perception and production of Spanish lexical stress by Spanish heritage speakers and English L2 learners of Spanish. In *Proceedings of the 6th Conference on Laboratory Approaches to Romance Phonology* (pp. 106–128). Cascadilla, Somerville, MA.

Kim, J. Y. (2016). *The perception and production of prominence in Spanish by heritage speakers and L2 learners* [Doctoral dissertation]. Department of Spanish and Portuguese, University of Illinois at Urbana-Champaign.

Kimura, T., Sensui, H., & Takasawa, M. (2010). A Pilot Study on Perception of Spanish Stress by Japanese Learners of Spanish. In *Second Language Studies: Acquisition, Learning, Education and Technology* (pp. 1–4). Tokyo, Japan.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13).

Lee, B., & Sidtis, D. V. L. (2017). The bilingual voice: Vocal characteristics when speaking two languages across speech tasks. Speech, *Language and Hearing*, *20*(3), 174–185.

Lenth, R., Singmann, H., & Love, J. (2019). Emmeans: Estimated marginal means, aka least-squares means. *R Package Version 1.3.4*.

Liu, Y. H. (2005). La entonación del español hablado por taiwaneses. *Phonica*, 1–807.

Liu, Z. (2019). *Análisis de las obstruyentes en chino y en español como L3: estudio acústico y perceptivo para la categorización de errores* [Tesis doctoral]. Departamento de Filología Española, Universitat Autònoma de Barcelona.

Majewski, W., Hollien, H., & Zalewski, J. (1972). Speaking fundamental frequency of Polish adult males. *Phonetica*, *25*(2), 119–125.

Mennen, I, Schaeffler, F., & Docherty, G. (2007). Pitching it differently: a comparison of the pitch ranges of German and English speakers. In *16th International Congress of Phonetic Sciences* (pp. 1769–1772). Saarbrücken, Germany.

Mennen, Ineke. (2015). Beyond segments: towards an L2 intonation learning theory (LILt). In *Prosody and language in contact* (pp. 171–188). Berlin, Heidelberg: Springer.

Mennen, Ineke, Schaeffler, F., & Dickie, C. (2014). Second language acquisition of pitch range in german learners of English. *Studies in Second Language Acquisition*, *36*(2), 303–329.

Mennen, Ineke, Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *Journal of the Acoustical Society of America*, *131*(3), 2249–2260.

Morrison, G. S. (2003). Perception and production of Spanish vowels by English speakers. In *Proceedings of the 15th international congress of phonetic sciences* (pp. 1533–1536). Barcelona, Spain.

Nebert, A. U. (2013). *Der Tonhöhenumfang der deutschen und russischen Sprechstimme: Vergleichende Untersuchung zur Sprechstimmlage* (Hallesche Schriften zur Sprechwissenschaft und Phonetik 46). Frankfurt Am Main: Lang.

Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. In *Proceedings of the 15th international congress of phonetic sciences* (Vol. 771, pp. 2–5). Barcelona, Spain.

Ohala, J. J. (1983). Cross-language use of pitch: an ethological view. *Phonetica*, *40*(1), 1–18.

Oliveira, M. (2002). The role of pause occurrence and pause duration in the signaling of narrative structure. In *International Conference for Natural Language Processing in Portugal* (pp. 43–51). Springer.

Ordin, M., & Ineke Mennen. (2017). Cross-Linguistic Differences in Bilinguals' Fundamental Frequency Ranges. *Journal of Speech, Language, and Hearing Research*, *60*(6), 1493–1506.

Peters, J. (2019). Fluency and speaking fundamental frequency in bilingual speakers of High and Low German. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 1–5). Melbourne, Australia.

Pfenninger, S. E., & Singleton, D. (2016). Age of onset, socio-affect and cross-linguistic influence: a long-term classroom study. Vigo International *Journal of Applied Linguistics*, *13*, 147–179.

Pisanski, K., Raine, J., & Reby, D. (2020). Individual differences in human voice pitch are

preserved from speech to screams, roars and pain cries. *Royal Society Open Science*, *7*(2), 191642.

R Core Team (2020). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria.

Reeder, J. T. (1998). English Speakers' Acquisition of Voiceless Stops and Trills in L2 Spanish. *Texas Papers in Foreign Language Education*, *3*(3), 101–118.

Roca, I. (2019). Spanish Word Stress: an updated multidimensional account. In Goedemans, R., Heinz, J., & van der Hulst, H. (Eds.), *The Study of Word Stress and Accent: Theories, Methods and Data* (pp. 256–292). Cambridge University Press.

Roseano, P., Fernández Planas, A. M., Elvira-García, W., Massó, R. C., & Celdrán, E. M. (2015). La entonación de las preguntas parciales en catalán. *Revista Española de Lingüística Aplicada*, *28*(2), 511–554.

Shi, S., Zhang, J., & Xie, Y. (2014). Cross-language comparison of F0 range in speakers of native Chinese, native Japanese and Chinese L2 of Japanese: Preliminary results of a corpus-based analysis. In *Proceedings of the 9th International Symposium on Chinese Spoken Language Processing* (pp. 241–244). Singapore.

Silva, C. C., & Barbosa, P. A. (2017). The contribution of prosody to foreign accent: A study of Spanish as a foreign language. *Loquens*, *4*(2), e041–e041.

Talkin, D. (1995). A robust algorithm for pitch tracking (RAPT). In W. B. Klejin & K. K. Paliwal (Eds.), *Speech coding and synthesis* (pp. 495–518). Elsevier Science B.V.

Tanaka-Ishii, K., & Terada, H. (2011). Word familiarity and frequency. *Studia Linguistica*, *65*(1), 96–116.

Trimble, J. C. (2013). Perceiving Intonational Cues in a Foreign Language: Perception of Sentence Type in Two Dialects of Spanish. In C. Howe, S. E. Blackwell, & M. L. Quesada (Eds.), *Proceedings of the 15th Hispanic Linguistics Symposioum* (pp. 78–92). University of Georgia, USA.

Ullakonoja, R. (2007). Comparison of pitch range in Finnish (L1) and Russian (L2). In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1701–1704). Saarbrücken, Germany.

Urbani, M. (2012). Pitch range in L1/L2 English. An analysis of f0 using LTD and linguistic measures. In Maria Grazia Busà & S. Antonio (Eds.), *Methodological perspectives on L2 prosody: Papers from ML2P 2012* (pp. 79–83).

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, *38*(3), 253–265.

Van Maastricht, L., Krahmer, E., & Swerts, M. (2016). Prominence Patterns in a Second Language: Intonational Transfer From Dutch to Spanish and Vice Versa. *Language Learning*, *66*(1), 124–158.

Volín, J., Poesová, K., & Weingartová, L. (2015). Speech melody properties in English, Czech and Czech English: Reference and interference. *Research in Language*, *13*(1), 107–123.

Xu, Y., & Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America*, *111*(3), 1399–1413.

Yuan, C., González-Fuente, S., Baills, F., & Prieto, P. (2019). Observing pitch gestures favors the learning of Spanish intonation by mandarin speakers. *Studies in Second Language Acquisition*, *41*(1), 5–32.

Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *Journal of the Acoustical Society of America*, *130*(6), 4063–4069.

Yuan, J., Dong, Q., Wu, F., Luan, H., Yang, X., Lin, H., & Liu, Y. (2018). Pitch characteristics of L2 English speech by Chinese speakers: A large-scale study. In *Proceedings of the Annual Conference of the International Speech Communication Association* (pp. 2593–2597). Hyderabad.

Yuan, J., & Liberman, M. (2014). F0 declination in English and Mandarin broadcast news speech. *Speech Communication*, *65*, 67–74.

Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B., & Trouvain, J. (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Proceedings to the 7th Speech Prosody Conference* (pp. 1037–1041). Dublin, Ireland.

Capítulo 5

Cue weighting differences in perception of Spanish sentence types between native listeners of Chinese and Spanish

Citado como:

Shang, P.; Elvira-García, W. & Li, X. (en prensa). Cue weighting differences in perception of Spanish sentence types between native listeners of Chinese and Spanish. *Proceedings of the 11th International Conference on Speech Prosody*. Lisboa, Portugal.

Cue weighting differences in perception of Spanish sentence types between native listeners of Chinese and Spanish

Peizhu Shang[1], *Wendy Elvira-García*[1], *Xinyi Li*[2]

[1]Department of Catalan Philology and General Linguistics, Universitat de Barcelona, Spain
[2]Department of Informatics, Technical University of Munich, Germany
pshangsh7@alumnes.ub.edu, wendyelvira@ub.edu, super.xinyi@tum.de

Abstract

This study examined the acoustic cue weighting in Spanish question-statement identification between native listeners of a tonal (Chinese) and a non-tonal language (Spanish). Listeners' performance was evaluated using two identification tests, whereby the stimuli were generated by manipulating the F0 contour, duration, and amplitude of the word-final syllable. A logistic sigmoid model was used to fit the question-statement identification function, and the model parameters were further estimated using a linear mixed-effect analysis. Results showed that Spanish listeners were more sensitive to the F0 linear transitions perceived as intonation, suggesting that the tonal language benefit in pitch perception may be limited to specific dimensions of pitch events. Nevertheless, Spanish listeners needed a higher terminal pitch to identify questions than Chinese speakers, especially when the duration or amplitude was decreased. Besides, our study revealed that native Spanish speakers gave more weight to the secondary cues of duration and amplitude and made greater compensations for the acoustic perturbations than Chinese listeners. These cross-linguistic differences of cue weighting between the two groups can be attributed to several factors, including listeners' previous linguistic experience with the first and the target language and the perceptual compensatory capacity in speech.

Index Terms: Acoustic cue weighting, intonation perception, pitch sensitivity, secondary cues, Chinese speakers of Spanish

1. Introduction

Speech perception is characterized by combining information of multidimensional acoustic space. One classic example of this is voicing contrasts in English obstruents, which have more than 15 cues, including vowel length, voice onset time (VOT), fundamental frequency (F0) contour, and intensity of the glottal signal [1]. The goal of speech perception is to get the best estimate of a phonological category from the continuous values along each acoustic dimension. While listeners may rely on multiple sources of acoustic cues to recognize a phonological contrast accurately, the contribution or the weight of these cues is different to listeners' judgments [2], [3]. For instance, VOT has been found to be the strongest cue for perceiving onset voicing contrasts in English [4].

Although the acoustic correlates of intonation have been researched less, the F0 pitch has been broadly recognized as the most reliable cue for perceiving the question-statement contrast in most languages, such as Chinese [5], Cantonese [6], English [7], and Spanish [8]. Moreover, Chinese listeners are often expected to outperform speakers of non-tonal languages in pitch perception because of their long-term experience with a tonal language. For instance, [9] reported that Chinese listeners were more accurate than English speakers in identifying Cantonese-level tones. Similarly, [10]–[12] showed that Chinese listeners were more sensitive to the F0 direction and detected F0 mismatches in their native or non-native speech more rapidly than listeners of an intonational language. Consequently, a general hypothesis of tonal language benefit in pitch perception has been proposed, although current findings on this hypothesis are highly divergent. For instance, [13]–[15] did not find that tonal language listeners were superior to speakers of a non-tonal language in perceiving non-native tones or native intonation patterns. One possible reason for the inconsistent results is that the finer F0 perception exhibited by tonal language listeners depends on the specific pitch representations of the stimuli. For instance, [13] found that Chinese listeners had higher F0 sensitivity than English speakers, but only for pitch profiles specific to their native language (L1), such as F0 contrastive patterns that phonetically shift like their L1 lexical tones. This finding of a tonal language advantage under specific pitch conditions has been confirmed by recent neurobehavioral evidence. For example, [14] demonstrated that the processing of F0 as lexical tone engages a semantic brain area limited to tonal language speakers. In contrast, the processing of F0 as intonation involves bilateral brain regions and is generalized to tonal and non-tonal language listeners, irrespective of language-specific realization of intonation [14].

While F0 contour is the primary cue for intonation perception, other secondary cues such as duration and amplitude can also influence listeners' decisions by conveying important but less reliable information [15]. Furthermore, [16] pointed out that primary and secondary cues may constantly interact in perception due to the dynamics of speech processing. That is, a shift in the identification function for one cue correlates with an opposing change in another cue, a phenomenon called *phonetic trading relation* [10]. One instance of this trade-off relationship is that English L1 listeners were more likely than non-native Chinese speakers to use a high pitch to perceive questions when the duration or amplitude of the final word was decreased [7]. The capacity to compensate for changes during speech relates to listeners' perceptual sensitivity to acoustic cues and their previous linguistic experience with the target speech [18].

Overall, given the debate on tonal language benefit in pitch perception and potential differences of cue weighing between listeners from different language backgrounds, this study aimed to examine the perceptual weighting of the three acoustic cues—F0, duration, and amplitude—in perceiving Spanish statements and yes-no questions by native listeners of Peninsular Spanish (hereafter, SN) and non-native Mandarin Chinese learners of the Spanish language (hereafter, CN). In particular, we are interested in (1) which listener group is

more sensitive to F0 linear changes perceived as intonation, and (2) the extent to which the secondary cues of duration and amplitude can interact with F0 pitch and contribute to the question-statement recognition by either group. To this end, we designed two perceptual experiments in which the auditory stimuli were synthesized by manipulating the values of the three acoustic parameters—F0 pitch, duration, and amplitude.

2. Method

2.1 Participants

In total, 78 CN listeners and 39 SN listeners participated in Experiment 1 (mean age: 28.27; SD = 8.43), while 77 CN listeners and 33 SN listeners took part in Experiment 2 (mean age: 28.02; SD = 8.46). All subjects in the two listener groups had normal hearing sensitivity within a range of 0–20 dB HL (decibels Hearing Level).

2.2 Stimuli synthesis

Since the intonation of Spanish broad focus statements and yes-no questions differ in both the prenucleus and the nucleus [19], we used a one-word sentence in the study to exclude the effect of prenuclear accents. The trisyllabic word "Alcalá" (a Spanish city name) was used as the target item, and it was produced in a broad focus statement form by an adult female native speaker of Peninsular Spanish. The target stimuli were generated by manipulating the three most salient acoustic cues (F0, duration, and amplitude) in the word-final syllable.

The pitch contour of the last stressed syllable was replaced by a multi-step F0 continuum using the *To Manipulation* function in *Praat* [20]. Specifically, the pitch curve of the last syllable was stylized into two pitch points. The start point was located at the beginning of the last syllable's vowel, with a pitch height of 196 Hz. The endpoint was anchored at the last regular glottal pulse detected in the spectrogram, and its F0 value was nearly equal to that of the starting point (3 Hz differences). The endpoint was manipulated nine times upward and once downward, with a step size of 20 Hz. Thus, 11 F0 contours spanning over 200 Hz were generated for each experimental condition (i.e., duration and amplitude). The amplitude of the carrier segment "Alca" was normalized at 70 dB sound pressure level and further used as a reference for manipulating the amplitude of the last syllable.

In Experiment 1, the 11 pitch contours with different F0 offsets were manipulated at three duration levels: short, medium (i.e., the original duration of the last syllable's vowel, which was 185 ms), and long. The long duration was generated by adding in *Praat* 50 ms from the center of the nucleus of the original vowel in the last syllable. We used a +50 ms value following previous research, which showed that the final vowel of yes-no questions is 40–70 ms longer than that of statements [21]. The short duration was created by removing 40 ms from the same vowel's nucleus. The reduced value was defined by referring to the shortest production of a statement by the same speaker. Since each duration set had 11 different F0 contours, the total number of stimuli for Experiment 1 was 33.

In Experiment 2, the amplitude of the 11 F0 contours was manipulated at three levels using the *Constant Amplification* function in *Cool Edit Pro 2.1* [22]. The low, medium and high amplitudes were created by applying a change of -7 dB, 0 dB, and +7 dB, respectively, to the final syllable's vowel. The manipulated values took as reference the carrier segment's amplitude, which was normalized at 70 dB. Thus, a total of 33 stimuli were used in Experiment 2. Each amplitude level had 11 stimuli with different F0 offsets. All stimuli were digitized at a sampling rate of 44.1 kHz in this study.

2.3 Procedure

The tests were conducted using the web-based survey software *Alchemer*, which allows to upload audio files onto the platform. The study survey comprised three sections. The first was intended for gathering information on the participants' linguistic backgrounds. The second involved the 33 stimuli of Experiment 1, and the third included the 33 stimuli of Experiment 2. The entire survey was administered in the language preferred by the listeners (Chinese or Spanish). Subjects could participate in either one or both auditory experiments. The stimulus order of each experiment was randomized, and every stimulus was presented once. The text of the stimuli was displayed on the screen without punctuation marks. Participants were instructed to listen to the audio using earphones in a quiet room. The perceptual answers of the two experiments were arranged on a five-point Likert scale to capture listeners' perceptual changes more accurately. The listeners had to identify which of the five descriptions of sentence type was closer to the stimulus they had just heard by selecting one of the following options: "Statement," "More statement than question," "Either statement or question," "More question than statement," "Question."

2.4 Data analysis

The five-response scales were recoded as 0, 0.25, 0.5, 0.75, and 1 based on the degree of interrogativity of each sentence type. The recoded values are interpreted as the probability from 0 to 1 that a given stimulus is perceived as a question. Thus, given the probabilistic estimation of question identification and the nearly S-shaped curve of the classification results, a logistic sigmoid model was used to fit the question-statement identification function for each listener. The equation of the sigmoid model is shown in (1),

$$p = \frac{1}{1 + e^{(-(x - x_0)/b)}} \tag{1}$$

where p refers to the probability of perceiving a sentence as a question, x is the value of pitch manipulation applied to the synthesized F0 contour (ranging from -20 Hz to 180 Hz, with an interval of 20 Hz) compared to the original terminal pitch of the statement, x_0 is a parameter where the question-statement boundary was at 50% identification, and b is a steepness factor that is inversely correlated with the slope of the identification function. The sigmoid model was fitted in MATLAB [23] using the Levenberg-Marquardt algorithm [24], a hybrid technique used to extract the model parameters and is particularly effective in solving nonlinear least-squares curve-fitting problems [25]. Specifically, an iterative process was executed to calculate the values of x_0 and b that minimized the objective function $F(x_0, b)$ (the sum of the squares of the errors between the N pairs of real data points and the fitted values), as shown in equation (2),

$$[x_0, b]^* = \underset{x_0, b}{\operatorname{argmin}} F(x_0, b) = \underset{x_0, b}{\operatorname{argmin}} \sum_{i=1}^{N} \left\| p - \frac{1}{1 + e^{(-(x - x_0)/b)}} \right\|^2 \tag{2}$$

3. Results

The general results of the two experiments are presented in Figure 1, in which the average probability of question identification between CN and SN listeners was fitted and plotted using MATLAB for the three duration (short, medium, and long) and amplitude (63 dB, 70 dB, 77 dB) levels.

Further, to examine the specific cue weighting paradigm between CN and SN listeners, a linear mixed-effect analysis was performed using the *lme4* package for R [26]. The question-statement boundary (x_0) and steepness factor (b) were alternatively entered as dependent variables, with the experimental condition (i.e., duration or amplitude), listener group (CN vs. SN), and their interactions as fixed effects. The speakers were included as random effects. Type III ANOVA was used to examine the significance of the main effects. Then, multiple comparisons of the interaction were performed using the *emmeans* package with the *Tukey* method [27].

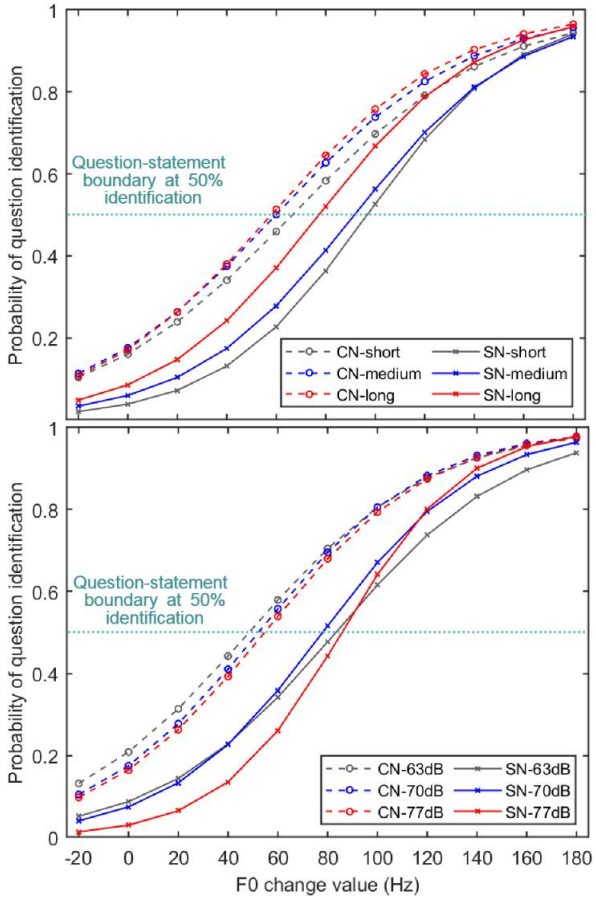


Figure 1: *Question identification as a function of the final F0 change at the three duration (upper) and amplitude (lower) levels for CN and SN listeners.*

3.1 Identification results of Experiment 1

For the question-statement boundary (x_0), the analysis of variance revealed a significant main effect of listener group [$\chi^2(1) = 24.88$, $p < .0001$], duration [$\chi^2(2) = 20.23$, $p < .0001$], and a significant interaction between listener group and duration [$\chi^2(2) = 8.38$, $p < .05$]. Moreover, pairwise comparisons showed that SN listeners had a significantly higher question-statement boundary than CN listeners at the three duration levels (see Table 1). However, the difference in intonation boundary between the two listener groups decreased when listening to sentences with a longer final syllable. This result is mainly due to SN listeners' behavior; that is, when listening to stimuli with a longer duration, SN listeners strongly decreased the pitch threshold (x_0) for hearing questions (see Table 2). A similar decrease of the question-statement boundary was observed for CN listeners over a longer duration. Nonetheless, this change was less significant in the CN group if we consider the overall magnitude of changes on the intonation boundary caused by the duration modulation (see also Figure 1), as revealed by the coefficients of the contrast between the short and long duration in Table 2. The different performances of the two listener groups indicate that SN listeners weighted the duration cue more heavily than CN listeners in perceiving Spanish sentence types.

In contrast, the results of the linear mixed model with the steepness indicator (b) as the dependent variable showed a significant main effect of the listener group [$\chi^2(1) = 4.60$, $p < .05$], while the effect of duration and its interaction with the L1 group was not significant. Moreover, the post-hoc analysis indicated that SN listeners had a steeper identification curve (i.e., smaller b) than CN listeners in the three duration types (see Figure 1), suggesting that SN listeners were more sensitive than CN speakers to the F0 linear transitions perceived as intonation. Notably, however, the difference in identification function was significant only for stimuli with a short final syllable [$t(246) = 2.13$, $p < .05$].

Table 1: *Pairwise comparisons of intonation boundaries (x_0) of the two groups in each duration.*

Duration	Contrast	Estimate	df	t	*p*
Short	CN – SN	-26.5	187	-4.95	< .0001
Medium	CN – SN	-30.7	187	-5.71	< .0001
Long	CN – SN	-18.5	187	-3.44	< .001

Table 2: *Pairwise comparisons of intonation boundaries (x_0) at three duration types for each group.*

Group	Contrast	Estimate	df	t	*P*
CN	1 – 2	7.37	238	2.94	< .05
	2 – 3	3.59	238	1.43	> .1
	1 – 3	10.96	238	4.37	< .001
SN	1 – 2	3.24	238	0.91	> .1
	2 – 3	15.83	238	4.47	< .0001
	1 – 3	19.06	238	5.38	< .0001

1, 2, 3 signify short, medium, and long duration, respectively.

3.2 Identification results of Experiment 2

The linear mixed model with the question-statement boundary (x_0) as the dependent variable showed a significant main effect of the listener group [$\chi^2(1) = 41.22$, $p < .0001$], while the effects of amplitude and the interaction of amplitude by the listener group were statistically nonsignificant. Specifically, in line with the results of Experiment 1, listeners in the SN group exhibited a significantly higher intonation boundary than those in the CN group at the three amplitude levels (all ps < .0001).

Regarding the steepness indicator (b), although there was no significant main effect or interaction for the listener group and amplitude, some clear tendencies should be noted for the prosodic evaluation of acoustic cues in intonation perception.

For instance, the identification curve of the SN group was steeper than for the CN group at the three amplitude levels, and this effect was marginally significant when they heard the stimuli with 77 dB [$t(182) = 6.34$, $p = 0.098$]. In addition, the pairwise comparisons of amplitude levels by the listener group showed that the sharpness of the identification curve increased to a different extent for the CN and SN groups when the amplitude of the last syllable was changed from 63 dB to 77 dB (see Figure 1). However, this amplitude effect was marginally significant and only present in the SN group [$t(210) = 2.29$, $p = 0.059$], suggesting that SN listeners were more sensitive to amplitude changes and had better compensatory behavior than CN listeners in perceiving Spanish intonation.

4. Discussion and conclusions

This study examined how CN and SN listeners use the three acoustic cues—F0 pitch, duration, and amplitude—to perceive question-statement contrasts in Spanish. In line with previous literature ([7], [18], [29]), the identification curves in Figure 1 revealed a reliable positive relationship between the F0 pitch and the probability of question recognition within a certain pitch interval. That is, the probability of question identification for the two listener groups rises robustly with increasing the F0 pitch. However, SN listeners differed from CN listeners by exhibiting a significantly higher question-statement boundary (i.e., higher x_0) in the three duration and amplitude levels. The possible differences between the linear pitch contours synthesized in this study and the natural production of Spanish yes-no questions (which has a nonlinear F0 trajectory at the end of the sentence) may explain the higher pitch threshold needed by SN listeners to hear questions [7]. Specifically, we speculate that the higher question boundary shown by SN listeners was because they had made greater compensations for the F0 loss in synthesized stimuli and the attenuation of other secondary cues in their L1.

Another possible reason for this finding relates to the categorical distinctions of intonation contours in Spanish phonology. Unlike Mandarin Chinese, which does not rely on the final rising or falling tone to discriminate sentence types, Peninsular Spanish has a three-way contrast of boundary tones at the end of sentences. In addition to low (L%) and high (H%) boundary tones, Spanish involves some mid-level tones that are used for signaling uncertainty statements (!H%) and statements of obvious (L!H%) [19]. Thus, the higher question-statement boundary exhibited by SN listeners may be due to the fact they have categorized both low- and mid-level tones as statements. In contrast, CN listeners tended to perceive any tone that was not strictly low as high, and thus related it to a question category.

Further, the steeper identification function (i.e., smaller b) consistently found for the SN group in the three duration and amplitude levels revealed that SN listeners were more sensitive than CN speakers to F0 linear transitions perceived as intonation. This finding seems to support the claim that the F0 advantage for CN listeners may be limited to specific dimensions of pitch processing, for example, to naturally occurring local tonal events that vary phonetically like Chinese lexical tones [29]. However, regarding the perception of F0 pitch as intonation, our study showed that tonal language listeners may have a lower F0 sensitivity, especially in a non-native language context, than native listeners of an intonational language. Previous research supporting this result can be found in [14], which showed that perceiving F0 pitch as intonation involved bilateral brain regions and was common to tonal and non-tonal language speakers. In contrast, processing F0 pitch as lexical tone, besides the left-hemispheric region shared between tonal and non-tonal language users, involved some semantic areas and subcortical regions limited to tonal language listeners [14]. Based on this neural analysis, CN and SN listeners were expected to have a comparable F0 sensitivity in the perception of intonation contours. Nevertheless, our study found that SN listeners had a steeper identification function (thus, higher F0 sensitivity) than CN listeners, partially because of their broader experience with Spanish prosody. Additionally, the potential F0 conflicts of stress and intonation bearing on the last syllable of the oxytone word (the target item) may increase CN listeners' cognitive efforts in processing F0 pitch, thus decreasing their sensitivity to the F0 perceived as intonation.

Although F0 contour plays a crucial role in Spanish intonation perception, other secondary cues, such as duration and amplitude, also significantly affect the performance of SN listeners. Experimental evidence for this is that SN listeners used a significantly lower question-statement boundary in the long duration condition and needed fewer F0 cues (thus, steeper identification function) to hear questions when the amplitude of the final syllable increased. The results of SN listeners resonate with the phonetic trading relation [7], [17], showing that attenuating one acoustic cue (i.e., duration) can be compensated by enhancing another (i.e., F0) so that the phonological category can be preserved. In contrast, CN listeners were sensitive only to duration variations (besides the F0 cue), and their reliance on this secondary cue was lower than SN listeners during question-statement identification.

The most frequent explanation for these cue weighting differences between the two listener groups relates to their compensatory capacity in speech processing. In general, the more sensitive listeners are to the acoustic changes, the greater the compensations for the perturbations during speech [21], [33]. Thus, SN listeners may have displayed higher sensitivity to duration and amplitude cues than CN listeners because they had superior compensatory behavior in their native language. In contrast, the limited experience of CN listeners with Spanish intonation and the overwhelming weight of F0 cues in their native prosody may plausibly cause their reduced sensitivity to the less prominent cues of duration and amplitude in Spanish sentence types [7].

In conclusion, the present study shows that, while the perception of Spanish intonation contrasts by SN listeners was influenced by the three acoustic cues of F0, duration, and amplitude, CN listeners were sensitive only to the F0 modulation and duration changes. Moreover, the degree of sensitivity of SN listeners to F0 and duration cues was generally greater than that of the CN group. The different cue weighing paradigm between CN and SN listeners in question-statement identification can be attributed to several reasons, including the linguistic experience with the first and the target language and the perceptual compensatory ability of native and non-native listeners. Finally, the trade-off relationship between the acoustic cues demonstrates that the perception of intonation contrasts involves the dynamic interplay of multiple cue sources rather than a single F0 pitch variation.

5. Acknowledgements

This work was supported by a research grant provided by China Scholarship Council (No. 201808390059) to the first author. The second author is a Serra Húnter Fellow.

6. References

[1] L. Lisker, "'Voicing' in English: A catalogue of acoustic features signaling/b/versus/p/in trochees," *Language and Speech*, vol. 29, no. 1, pp. 3–11, 1986.

[2] J. Kuang and A. Cui, "Relative cue weighting in production and perception of an ongoing sound change in Southern Yi," *Journal of Phonetics.*, vol. 71, pp. 194–214, 2018.

[3] G. Tillman, T. Benders, S. D. Brown, and D. van Ravenzwaaij, "An evidence accumulation model of acoustic cue weighting in vowel perception," *Journal of Phonetics.*, vol. 61, pp. 1–12, 2017.

[4] J. Schertz, T. Cho, A. Lotto, and N. Warner, "Individual differences in phonetic cue use in production and perception of a non-native sound contrast," *Journal of Phonetics.*, vol. 52, pp. 183–204, 2015.

[5] J. Yuan, "Perception of intonation in Mandarin Chinese," *Journal of the Acoustical Society of America*, vol. 130, no. 6, pp. 4063–4069, 2011.

[6] J. K.-Y. Ma, V. Ciocca, and T. L. Whitehill, "The perception of intonation questions and statements in Cantonese," *Journal of the Acoustical Society of America.*, vol. 129, no. 2, pp. 1012–1023, 2011.

[7] J. Feng, S. Tao, X. Wu, K. Alsbury, and C. Liu, "The effects of amplitude and duration on the perception of English statements vs. questions for native English and Chinese listeners," *Journal of the Acoustical Society of America*, vol. 145, no. 5, pp. EL449–EL455, 2019.

[8] T. L. Face, "The role of intonational cues in the perception of declaratives and absolute interrogatives in Castilian Spanish," *Estudios de Fonética Experimental*, vol. 16, pp. 186–225, 2007.

[9] Y. S. Chang, Y. Yao, and B. H. Huang, "Effects of linguistic experience on the perception of high-variability non-native tones," *Journal of the Acoustical Society of America*, vol. 141, no. 2, pp. EL120–EL126, 2017.

[10] Y. Xu, J. T. Gandour, and A. L. Francis, "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," *Journal of the Acoustical Society of America*, vol. 120, no. 2, pp. 1063–1074, 2006.

[11] M. Ortega-Llebaria, M. Nemogá, and N. Presson, "Long-term experience with a tonal language shapes the perception of intonation in English words: How Chinese-English bilinguals perceive 'Rose?' vs. 'Rose,'" *Bilingualism*, vol. 20, no. 2, pp. 367–383, 2017.

[12] M. L. D. Deroche *et al.*, "A tonal-language benefit for pitch in normally-hearing and cochlear-implanted children," *Scientific Reports*, vol. 9, no. 1, pp. 1–12, 2019.

[13] A. Krishnan, J. T. Gandour, G. M. Bidelman, and J. Swaminathan, "Experience-dependent neural representation of dynamic pitch in the brainstem," *Neuroreport*, vol. 20, no. 4, p. 408, 2009.

[14] P. J. Chien, A. D. Friederici, G. Hartwigsen, and D. Sammler, "Neural correlates of intonation and lexical tone in tonal and non-tonal language speakers," *Human Brain Mapping*, vol. 41, no. 7, pp. 1842–1858, 2020.

[15] S. C. Peng, M. Chatterjee, and N. Lu, "Acoustic Cue Integration in Speech Intonation Recognition With Cochlear Implants," *Trends in Amplification*, vol. 16, no. 2, pp. 67–82, 2012.

[16] P. Divenyi, S. Greenberg, and G. Meyer, *Dynamics of speech production and perception*, vol. 374. Amsterdam: Ios Press, 2006.

[17] B. H. Repp, "Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception," *Psychological Bulletin*, vol. 92, no. 1, pp. 81–110, 1982.

[18] D. R. Nault and K. G. Munhall, "Individual variability in auditory feedback processing: Responses to real-time formant perturbations capacity and their relation to perceptual acuity," *Journal of the Acoustical Society of America*, vol. 148, no. 6, pp. 3709–3721, 2020.

[19] E. Estebas-Vilaplana and P. Prieto, "Castilian Spanish intonation," in *Transcription of intonation of the Spanish language*, P. Prieto and P. Roseano, Eds. 2010, pp. 17–48.

[20] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer," *Version 6.1.05*. 2019 [Computer program]. Available from http://www.praat.org/

[21] L. Romera Barrios, A. M. Fernández-Planas, V. Salcioli Guidi, J. Carrera Sabaté, and D. R. Montes de Oca, "Una muestra del español de Barcelona en el marco AMPER," *Estudios de Fonética Experimental*, vol. 16, no. 16, pp. 147–184, 2007.

[22] Syntrillium Software Corporation, "Cool Edit Pro," *Version 2.1.* 2003 [Computer program]. Available from https://cool-edit-pro.soft32.es/

[23] The Mathworks, "MATLAB and Statistics Toolbox Release," *Version 2021b*. 2021 [Computer program]. Available from https://es.mathworks.com/products/matlab.html

[24] J. J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory," in *Numerical analysis*, Berlin: Springer, 1978, pp. 105–116.

[25] H. P. Gavin, "The Levenberg-Marquardt Algorithm For Nonlinear Least Squares Curve-Fitting Problems," *Department of Civil and Environmental Engineering, Duke University*, pp. 1–19, 2019.

[26] D. Bates *et al.*, "Package 'lme4," *R package version 1.17*. 2018.

[27] R. Lenth, H. Singmann, and J. Love, "Emmeans: Estimated marginal means, aka least-squares means," *R package version 1.3.4*. 2019.

[28] C. Liu and A. Rodriguez, "Categorical perception of intonation contrasts: Effects of listeners' language background," *Journal of the Acoustical Society of America*, vol. 131, no. 6, pp. EL427–EL433, 2012.

[29] B. Chandrasekaran, A. Krishnan, and J. T. Gandour, "Experience-dependent neural plasticity is sensitive to shape of pitch contours," *Neuroreport*, vol. 18, no. 18, pp. 1963–1967, 2007.

[30] V. M. Villacorta, J. S. Perkell, and F. H. Guenther, "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," *Journal of the Acoustical Society of America*, vol. 122, no. 4, pp. 2306–2319, 2007.

Capítulo 6

Conclusiones, limitaciones y líneas futuras de investigación

Esta tesis explora la adquisición de la entonación del español como LE por parte de sinohablantes y, más en concreto, se centra en las oraciones interrogativas del español. A la luz de los resultados obtenidos de las seis publicaciones realizadas, consideramos que se han cumplido los cinco objetivos previstos de este estudio: el objetivo relacionado con la definición del marco teórico, el objetivo de llevar a cabo una comparación lingüística, el objetivo de contrastar la validez de las teorías científicas en el campo de la adquisición de una LE, el objetivo descriptivo y el objetivo explicativo.

Desde el punto de vista general del marco teórico, se ha conseguido plantear un sistema interidiomático para la clasificación de la modalidad interrogativa a partir de un enfoque funcional-pragmático, que permite la realización del análisis contrastivo sobre una base teórica común (§2).

Desde el punto de vista de la comparación tipológica, se han especificado los recursos lingüísticos de distintos niveles que se emplean en chino y en español para expresar una serie de preguntas con distintas funciones pragmáticas. A partir de ese estudio comparativo, se han extraído las principales diferencias entre las dos lenguas en cuanto a las estrategias de codificación de la interrogatividad que podrían causar dificultades a los sinohablantes a la hora de adquirir las oraciones interrogativas del español. Este objetivo se ha cumplido en el capítulo dos (§2).

Desde el punto de vista de la contrastación empírica, los datos obtenidos en este trabajo suponen una importante contribución a las teorías de la adquisición de la pronunciación de lenguas extranjeras, como el modelo de LILt diseñado para predecir las características de la entonación de una L2 (Mennen, 2015), el modelo de asimilación perceptiva de los elementos suprasegmentales (*Perceptual Assimilation Model for Suprasegmentals*) (Alexander & Wang, 2016; So & Best, 2010, 2011, 2014) y las teorías de transferencia interlingüística (Choi, 2022; Gabriel & Kireva, 2014) y de compensación acústica en el proceso perceptivo (Nault & Munhall, 2020; Repp, 1982).

Desde el punto de vista de la descripción entonativa, se han descrito los patrones fonológicos y las propiedades fonético-acústicas de la prosodia de las producciones de los sinohablantes en comparación con la entonación de las oraciones interrogativas de hablantes nativos de español. Asimismo, se han identificado los fallos y los éxitos mostrados por el colectivo chino en el campo de la producción (§3 y §4) y la percepción (§5 y anexos A y B), lo que proporciona importantes conocimientos sobre la adquisición de la entonación de ELE por parte de hablantes de una lengua tonal. Este objetivo se ha conseguido a lo largo de varios capítulos de la tesis.

Desde el punto de vista explicativo, se han examinado una serie de factores lingüísticos y extralingüísticos que pueden incidir en el aprendizaje de la entonación de ELE y se han distinguido las causas que contribuyen a explicar el éxito o el fracaso en la adquisición de los patrones entonativos. Este objetivo también se ha conseguido de manera transversal en distintos apartados de este trabajo.

6.1 Conclusiones generales de la tesis

Las conclusiones que se derivan del presente trabajo de investigación enlazan entre sí los temas tratados en los capítulos principales (§2, §3, §4 y §5) y los anexos A y B, y son las que se exponen a continuación.

En el capítulo 2, se ha planteado una taxonomía de la modalidad interrogativa que permite clasificar un conjunto de preguntas con funciones pragmáticas distintas que existen tanto en chino como en español. En base a este marco teórico, se han analizado las estrategias lingüísticas usadas en cada lengua para expresar cada tipo de pregunta.

Como resultado de este análisis, se ha confirmado que las oraciones interrogativas del chino y del español pueden desempeñar las mismas funciones pragmáticas, quedando así confirmada la universalidad de las funciones pragmáticas del lenguaje. Por otra parte, se ha comprobado que, para la codificación de las preguntas, el español recurre sobre todo al modo verbal, a la entonación y al orden sintáctico, mientras que el chino, al ser una lengua analítica y tonal, utiliza con el mismo fin partículas modales y construcciones sintácticas. Puesto que en la expresión de la interrogatividad en chino la entonación desempeña un papel secundario, es lógico que la entonación sea uno de los aspectos más difíciles de aprender para los estudiantes sinohablantes de español. Las conclusiones extraídas de este capítulo soportan la Hipótesis 1 (§1.3.1).

A partir de las observaciones anteriores, se ha estudiado la adquisición de la entonación de las oraciones interrogativas del español por parte de aprendices chinos. En los capítulos 3 y 4, se ha analizado la producción de la entonación de cinco tipos de preguntas (preguntas totales informativas, preguntas parciales informativas, preguntas disyuntivas informativas, preguntas totales confirmatorias y preguntas confirmatorias de apéndice) desde diferentes puntos de vista.

En la dimensión fonológica, como se ha podido comprobar en el capítulo 3, la producción de los patrones entonativos por parte de los sinohablantes puede desviarse o aproximarse a la de los hablantes nativos, dependiendo en buena medida de las similitudes o disimilitudes sistemáticas del inventario tonal y realizaciones entonativas del chino y del español. Sin embargo, algunos fallos (p.ej., la sobreproducción de H%) de los aprendices, en vez de estar relacionado con la L1, se pueden atribuir a la falta de conocimiento de la fonología entonativa de la lengua meta, en concreto con la falsa idea según la cual todas las preguntas en español tienen un ascenso final de F0. En la dimensión pragmática, se ha comprobado que la mayoría de los aprendices chinos tuvieron problemas a la hora de utilizar diferentes patrones entonativos para distinguir entre la función informativa y confirmatoria de las preguntas totales. En la dimensión fonética, el análisis de los parámetros tonales y temporales llevó a la conclusión de que los aprendices chinos, además de un nivel tonal más alto, tenían una compresión del rango tonal (en el nivel oracional y silábico) y la variabilidad tonal (*pitch variation*), así como una reducción de la velocidad de habla, de la tasa de articulación y de la

velocidad del cambio de F0 (*pitch change rate*) en comparación con los hablantes nativos de español. Por último, se ha confirmado que la producción entonativa de los aprendices de ELE, además de estar determinada por el sistema fonético y fonológico de la L1, se ve influenciada por las características lingüísticas de los materiales de habla y las propiedades individuales de los aprendices. Estas conclusiones se relacionan con la Hipótesis 2 (§1.3.2) y la Hipótesis 3 (§1.3.3).

El artículo incluido en el anexo A ha analizado la percepción de los patrones entonativos de las declarativas neutras y de las interrogativas totales del español. Se ha comprobado que los aprendices chinos fueron capaces de reconocer la modalidad oracional mediante el contorno final y se diferenciaron significativamente de los oyentes nativos de español en la percepción del pico inicial y el movimiento tonal de la última sílaba acentuada. En comparación con los oyentes estadounidenses del español (Trimble, 2013b), los mejores resultados obtenidos por los aprendices chinos parecen confirmar que los hablantes nativos de una lengua tonal tienen ventajas a la hora de asimilar los eventos tonales contrastivos de una LE. Por otra parte, también se ha estudiado la influencia de los factores lingüísticos e individuales en la percepción de la entonación de ELE. Los resultados demuestran que los sinohablantes con un alto nivel de competencia en español, alta habilidad pragmática y experiencia musical obtuvieron mejores resultados en la tarea perceptiva. Además, se ha comprobado que las preguntas totales formadas por una palabra llana tuvieron mayor probabilidad de ser identificadas correctamente que las compuestas por una palabra aguda. Por consiguiente, se ha corroborado la Hipótesis 4 (§1.3.4) de esta tesis.

En el capítulo 5 y el anexo B, se ha contrastado la Hipótesis 5 (§1.3.5) a través del análisis del uso de tres correlatos acústicos (F0, duración e intensidad) para percibir la entonación declarativa e interrogativa total del español. Utilizando dos métodos estadísticos distintos, se ha confirmado que la identificación de la categoría entonativa por parte de los oyentes nativos de español está basada en la modulación de la F0, la duración y la intensidad, mientras que la percepción por parte de los oyentes chinos se fundamenta solo en la F0 y la duración. Además, se ha encontrado que los oyentes nativos de español necesitan un ascenso final de F0 más elevado para percibir la interrogatividad y son más sensibles no solo al cambio de la duración, sino también al

cambio de la F0 que los oyentes chinos. Éste último hallazgo contradice la segunda suposición que se deriva de la quinta hipótesis general (§1.3.5).

Otra cuestión importante que se ha investigado en estos dos apartados (§5 y anexo B) es la relación entre los diferentes correlatos acústicos en el proceso perceptivo. En relación con este aspecto, se ha podido comprobar la teoría de la *phonetic trading relation* (Repp, 1982) al demostrar que la reducción de los valores de la duración o la intensidad podía compensarse con el aumento de la F0 para asegurar la percepción del oyente de la categoría entonativa deseada, y viceversa.

El último aspecto tratado en el anexo B es la influencia de las variables lingüísticas y extralingüísticas en la percepción de los contrastes entonativos del español. En cuanto a las variables extralingüísticas, en nuestros datos las participantes de sexo femenino y aquellos de mayor edad muestran un sesgo hacia la identificación de las oraciones como interrogativas. En cuanto al efecto de las variables lingüísticas, las oraciones formadas por una palabra llana tienen mayor probabilidad de ser percibidas como interrogativas que las acabadas en palabra aguda.

En conjunto, los estudios realizados en este trabajo han cotejado todas las hipótesis propuestas y han alcanzado una serie de conclusiones coherentes con los objetivos iniciales de la tesis.

6.2 Aportaciones de la tesis

Como se ha expuesto anteriormente (§6.1), la realización de estudios comparativos entre la lengua materna y la lengua meta proporciona una valiosa base para investigar la adquisición de las lenguas extranjeras. Así pues, en uno de los primeros capítulos (§2) de esta tesis, se ha realizado un análisis contrastivo de la expresión de las oraciones interrogativas del chino y del español, lo que ha llevado al desarrollo de una nueva taxonomía de la modalidad interrogativa. Es de destacar que este sistema, al estar fundamentado en un criterio funcional-pragmático, permite clasificar un conjunto de preguntas que se encuentran en ambas lenguas. En este sentido, el estudio podría suponer una importante contribución a las teorías de la modalidad interrogativa y servir como base para futuros estudios interlingüísticos. Gracias a este sistema, se han podido

comparar las estrategias lingüísticas usadas en las dos lenguas para cada tipo de oración interrogativa. Los resultados comparativos constituyen una gran oportunidad para contribuir al desarrollo del conocimientos sobre las diferencias entre el chino y el español y también para impulsar el avance en el campo de los estudios contrastivos entre lenguas sino-tibetanas y lenguas romances.

Seguidamente, en los capítulos 3 y 4, se han aportado nuevos datos empíricos acerca de la producción de la entonación del español por parte de aprendices sinohablantes. A través de la descripción prosódica en diferentes dimensiones (fonológica, fonética, pragmática, individual, etc.), se ha podido ofrecer una interpretación detallada del aprendizaje de la entonación de ELE y ampliar nuestro conocimiento sobre los fallos y los logros entonativos de los sinohablantes en cinco tipos de oraciones interrogativas del español. Los resultados alcanzados coinciden, en general, con los estudios previos, sugiriendo la existencia de una jerarquía de dificultades para producir los patrones entonativos (Cortés Moreno, 2001, 2005; Yuan et al., 2019) y confirmando algunas tendencias generales de la adquisición de la pronunciación de lenguas extranjeras en cuanto a la producción de los rasgos tonales y temporales (Shi et al., 2014; Yuan et al., 2018; Zimsmerer et al., 2014 para el rango tonal de la L2; Ding et al., 2016; Peters, 2019 para la fluidez oral en la L2). En estos dos capítulos también se ha intentado explicar cómo las propiedades individuales de los aprendices y las características lingüísticas del corpus son responsables, juntamente con los fenómenos de transferencia de la L1, de la variabilidad de la producción entonativa del español como LE.

En el capítulo 5, junto con los anexos A y B, se ha realizado un estudio completo de la percepción de las declarativas e interrogativas totales del español desde el punto de vista fonológico y fonético-acústico. A partir de los resultados de los test de percepción, se ha podido observar que los aprendices chinos difieren de los hablantes nativos no solo en la capacidad de percibir los patrones entonativos, sino también en el uso de los correlatos acústicos para identificar las categorías entonativas. Los resultados obtenidos arrojan nueva luz sobre la caracterización perceptiva de la entonación de ELE y llaman la atención sobre una serie de dificultades que los sinohablantes suelen encontrar en la percepción de los eventos tonales contrastivos que no están presentes en el inventario de su L1.

Además, en estos tres apartados, el haber considerado como variables el ser hablante de una lengua tonal, así como otras características lingüísticas e individuales, ha posibilitado alcanzar una visión completa de la percepción entonativa de lenguas extranjeras y contribuye en manera determinante al debate sobre las diferencias de sensibilidad a la hora de percibir las informaciones postléxicas vehiculadas mediante la F0 entre los oyentes de lenguas tonales y no tonales. Por último, los hallazgos relacionados con los fenómenos de compensación perceptiva entre diferentes correlatos acústicos en el proceso perceptivo también suponen una importante aportación a los estudios que defienden el carácter multidimensional y dinámico de la percepción del habla (Peng et al., 2012; Toscano & McMurray, 2010).

En definitiva, los resultados de esta tesis mejoran nuestro entendimiento de la adquisición de la entonación del español como LE, confirmando así también la importancia de desarrollar métodos y técnicas didácticas eficaces para promover que los sinohablantes logren una entonación que se aproxime lo máximo posible a la de los hablantes nativos.

6.3 Limitaciones y futuras líneas de investigación

Tras la exposición de las aportaciones principales de esta tesis, es necesario reconocer una serie de limitaciones y proponer, en la medida de lo posible, fórmulas para superarlas, lo que permite trazar futuras líneas de investigación.

En primer lugar, una limitación metodológica tiene que ver con el reclutamiento de los informantes. Los aprendices de español de origen chino son necesariamente heterogéneos. La gran extensión del país hace que, por ejemplo, los estudiantes de español muchas veces tengan, además del mandarín, una L1 diferente (los que una parte de la literatura científica llama “dialectos chinos”). También suele ser heterogénea la edad de adquisición del español y el tiempo de exposición a la lengua en inmersión. En este estudio, atendiendo a un criterio de ecología (que consiste en reflejar lo mejor posible la situación de las aulas de español en España u otros países hispanohablantes), se ha optado por tener informantes de diferentes perfiles. Sin embargo, para trabajos posteriores y, sobre todo, si estos se centran en la enseñanza del español en China, todas estas variables se podrían controlar para explorar el impacto de los antecedentes

lingüísticos concretos en la adquisición de la entonación.

En segundo lugar, como se ha expuesto anteriormente, en nuestro estudio de producción solo se han recogido datos sobre cinco tipos de oraciones interrogativas, a pesar de que en español se han documentado más tipos de preguntas (Prieto & Roseano, 2010). De manera parecida, los test de percepción se han centrado solo en dos modalidades oracionales (concretamente, en las declarativas neutras y en las interrogativas totales no marcadas) dentro de las muchas posibles. Por consiguiente, si se quiere dibujar un panorama completo de la situación del aprendizaje de la entonación del español por parte de los sinohablantes, cabría ampliar el objeto de estudio para incluir más modalidades oracionales tanto desde el punto de vista de la producción como desde el punto de vista de la percepción.

Además, en esta tesis una parte del comportamiento de los aprendices en cuanto a la adquisición de la entonación se ha atribuido a la transferencia de la lengua materna (el chino mandarín) a la lengua meta, lo cual supondría una potencial limitación ya que no se ha considerado la influencia de otras lenguas habladas por los sinohablantes, tanto como L1 que como LE. Como se ha expuesto en una de las secciones iniciales (§1.2), el inglés es una lengua extranjera que se aprende antes que el español para la mayoría de los sinohablantes. Por tanto, es posible que los resultados obtenidos por los aprendices chinos de español también pudieran estar influenciados hasta cierto punto por sus conocimientos fonológicos y fonéticos del inglés, lo que según De Angelis (2007) constituiría un caso de influencia interlingüística combinada. En vista de esto, futuras investigaciones tendrían que abordar el aprendizaje de la entonación del español a partir de un enfoque multilingüe, lo que implicaría también establecer una relación entre todas las lenguas ya adquiridas por los aprendices y la lengua meta, para ampliar nuestra comprensión de los problemas que pueden surgir en el proceso de adquisición del español como lengua extranjera.

Por otro lado, el presente estudio se ha destinado investigar la adquisición de la entonación tanto en el ámbito de la producción como en la percepción, pero de manera separada. De esta manera, el trabajo que aquí se presenta se limita a aportar datos a una cuestión candente en el aprendizaje de la LE: la de si las desviaciones de los aprendices en la producción están relacionadas con su falta de competencia para percibir las

categorías fonéticas y fonológicas de la lengua meta (Strange, 2007). Es decir, en esta tesis no se trabaja la relación que se establece entre la habilidad para percibir contrastes entonativos y la habilidad para producirlos. Para incrementar nuestra comprensión sobre esta cuestión, sería recomendable evaluar resultados de producción y de percepción en un mismo grupo de aprendices, lo que daría la posibilidad de establecer una relación más directa entre los dos aspectos de la adquisición del habla de una lengua extranjera.

Finalmente, los resultados aquí presentados sirven para dar a conocer la adquisición de la entonación del español en el contexto chino y llamar la atención de los profesores y alumnos sinohablantes sobre este fenómeno lingüístico que a menudo no recibe suficiente atención en el proceso de enseñanza-aprendizaje. Como continuación natural de esta tesis, una importante línea de trabajo en el futuro podría ser la elaboración de métodos y técnicas didácticas que ayuden a mejorar la capacidad de los aprendices chinos para producir y percibir la entonación del español, como por ejemplo el uso de gestos de tonalidad (p.ej., Baills et al., 2019; Kelly et al., 2017; Yuan et al., 2019), programas de análisis audiovisual del habla (p.ej., Bengrait, 2018; Levis & Pickering, 2004; Olson, 2014) y actividades de entrenamiento musical (p.ej., Liao & Davidson, 2016; Moradi & Shahrokhi, 2014), entre otras.

6.4 Referencias del Capítulo 6

Baills, F., Suárez-González, N., González-Fuente, S., & Prieto, P. (2019). Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition*, *41*(1), 33–58.

Bengrait, N. (2018). *The Use of Speech Analysis Software for the Teaching of English Intonation* (Doctoral dissertation). University of Badji-Mokhtar Annaba.

Cortés Moreno, M. (2001). Percepción y adquisición de la entonación española en enunciados de habla espontánea: el caso de los estudiantes taiwaneses. *Estudios de Fonética Experimental*, *11*, 90–119.

Cortés Moreno, M. (2005). Análisis experimental del aprendizaje de la acentuación y la entonación españolas por parte de hablantes nativos de chino. *Phonica*, *1*, 1–25.

De Angelis, G. (2007). *Third or additional language acquisition*. Multilingual Matters.

Kelly, S., Bailey, A., & Hirata, Y. (2017). Metaphoric gestures facilitate perception of intonation more than length in auditory judgments of non-native phonemic contrasts. *Collabra: Psychology*, *3*(1).

Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System*, *32*(4), 505–524.

Liao, M.-Y., & Davidson, J. W. (2016). The effects of gesture and movement training on the intonation of children's singing in vocal warm-up sessions. *International Journal of Music Education*, *34*(1), 4–18.

Mennen, I. (2015). Beyond segments: towards an L2 intonation learning theory (LILt). In E. Delais-Roussarie, M. Avanzi, & S.Herment (Eds.), *Prosody and language in contact* (pp. 171–188). Springer.

Moradi, F., & Shahrokhi, M. (2014). The Effect of Listening to Music on Iranian Children's Segmental and Suprasegmental Pronunciation. *English Language Teaching*, *7*(6), 128–142.

Nault, D. R., & Munhall, K. G. (2020). Individual variability in auditory feedback processing: Responses to real-time formant perturbations capacity and their relation to perceptual acuity. *The Journal of the Acoustical Society of America*, *148*(6), 3709–3721.

Olson, D. J. (2014). Phonetics and technology in the classroom: A practical approach to using speech analysis software in second-language pronunciation instruction. *Hispania*, 47–68.

Peng, S. C., Chatterjee, M., & Lu, N. (2012). Acoustic Cue Integration in Speech Intonation Recognition With Cochlear Implants. *Trends in Amplification*, *16*(2), 67–82.

Peters, J. (2019). Fluency and speaking fundamental frequency in bilingual speakers of High and Low German. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 1–5). Melbourne, Australia.

Prieto, P. & Roseano, P. (2010). T*ranscription of the intonation of the Spanish language*. Lincom.

Repp, B. H. (1982). Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*(1), 81–110.

Shi, S., Zhang, J., & Xie, Y. (2014). Cross-language comparison of F0 range in speakers of native Chinese, native Japanese and Chinese L2 of Japanese: Preliminary results of a corpus-based analysis. In *Proceedings of the 9th International Symposium on Chinese Spoken Language Processing* (pp. 241–244). Singapore.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273–293.

So, C. K., & Best, C. T. (2011). Categorizing Mandarin tones into listeners' native prosodic categories: The role of phonetic properties. *Poznań Studies in Contemporary Linguistics*, *47*(1), 133.

So, C. K., & Best, C. T. (2014). Phonetic influences on english and french listeners' assimilation of mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*(2), 195–221.

Strange, W. (2007). Cross-language phonetic similarity of vowels. In O. S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (pp. 35–55). John Benjamins Publishing.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*(3), 434–464.

Trimble, J. C. (2013). Perceiving Intonational Cues in a Foreign Language : Perception of Sentence Type in Two Dialects of Spanish. In C. Howe, S. E. Blackwell, & M. L. Quesada (Eds.), *15th Hispanic Linguistics Symposioum* (pp. 78–92).

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, *122*(4), 2306–2319.

Yuan, C., González-Fuente, S., Baills, F., & Prieto, P. (2019). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, *41*(1), 5–32.

Yuan, J., Dong, Q., Wu, F., Luan, H., Yang, X., Lin, H., & Liu, Y. (2018). Pitch characteristics of L2 English speech by Chinese speakers: A large-scale study. In *Proceedings of the Annual Conference of the International Speech Communication Association* (pp. 2593–2597). Hyderabad, India.

Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B., & Trouvain, J. (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Proceedings of the 7th International Conference on Speech Prosody* (pp. 1037–1041). Dublin, Ireland.

Anexos

Anexo A

Perception of Spanish statements and questions by native Chinese speakers: Effect of tone language experience and individual features

Citado como:

Shang, P.; Roseano, P. & Elvira-García, W. (en revisión). Perception of Spanish Statements and Questions by Native Chinese Speakers: Effect of Tone Language Experience and Individual Features.

Perception of Spanish Statements and Questions by Native Chinese Speakers: Effect of Tone Language Experience and Individual Features

Peizhu Shang
Universitat de Barcelona, Spain

Paolo Roseano
Universidad Nacional de Educación a Distancia, Spain;
University of South Africa, South Africa

Wendy Elvira-García
Universitat de Barcelona, Spain

Abstract
This study examines several possible factors that may influence the perception of Spanish statements and questions by native and non-native listeners. To this end, 43 speakers of Peninsular Spanish and 75 speakers of Mandarin Chinese were presented with auditory stimuli in a gating paradigm for the identification of sentence type. The results showed that listeners' perception of the intonational cues for signaling the sentence type was highly dependent on their native language experience (namely, Chinese listeners differed significantly from native Spanish speakers in perceiving the initial F0 peak and the F0 downslope during the final stressed syllable). In addition to L1 background, other individual properties (e.g., proficiency level, pragmatic skill, and musical expertise) played a major role in perceiving the non-native intonation contrasts (Chinese listeners with higher proficiency, higher pragmatic skills, and musical training experience performed better). Besides, our data revealed that the position of the stress within the word (e.g., last syllable vs. penultimate syllable) was also a relevant factor for discriminating the sentence type of the one-word stimuli (sentences consisting of a paroxytone word had a significantly higher likelihood of being accurately perceived compared with sentences formed by an oxytone word). On the whole, thus, it seems that there is a complex integration mechanism for the cross-linguistic speech perception, in which both the language specificity and the interspeaker variability should be accounted for evaluating the non-native listeners' perceptual performance.

Keywords
Perception of intonation, Spanish, Chinese, tonal language effect, individual variability

Corresponding author:
Peizhu Shang, Department of Catalan Philology and General Linguistics, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, Barcelona 08007, Spain.
Email: pshangsh7@alumnes.ub.edu

1 Introduction

As it is well known, tone and intonational languages use fundamental frequency (F0) for different purposes: the first ones use F0 modulation to transmit lexical meanings whereas the seconds use it to transmit postlexical information. This causes severe difficulties when tonal native speakers try to learn an intonational language as a second (L2) or third language. Despite that research on the non-native perception of segments (e.g., Baigorri et al.; 2019; Escudero et al.; 2012), of stress (e.g., Lee et al.; 2019; Meng et al., 2020), of prominence (e.g., Bishop et al.; 2020; Luchkina & Cole, 2021), and of lexical tone (e.g., Ortega-Llebaria & Wu, 2021; Tsukada, 2018) has been extensively developed over the last decades, less is known about the cross-linguistic perception of intonation, particularly in environments of language contact between tone and non-tone languages. Moreover, an important limitation is that studies examining the L2 intonation perceived by tonal language speakers mostly have focused on Chinese learners of English and the effect of native language (L1) (e.g., Grabe et al., 2003; Liu & Rodriguez, 2012; Ortega-Llebaria et al., 2017; Zhang et al., 2010). To our knowledge, only three research papers have addressed the perception of Spanish intonation by Chinese listeners (Cortés Moreno, 1997, 2001; Li, 2020). Although the experimental methods used by the authors were different, the three papers, like most cross-linguistic studies of perception, regarded the L1 experience as the only decisive factor that influences the foreign listeners' perception performance. However, recent studies have shown that several inter-individual variables in relation to the biological properties (e.g., Baigorri et al., 2019; Richards et al., 2018; van Kesteren & Wiersinga-Post, 2007), the psychometric traits (e.g., Bishop et al., 2020; Jiang et al., 2015), and the musical experience (e.g., Bidelman et al., 2013; Wiener & Bradley, 2020) also play a major role in the native and non-native speech perception.

The present paper aims to shed light on the perception of Spanish intonation by native speakers of Mandarin Chinese who have acquired Peninsular Spanish as a foreign language. Of particular interest to us are how Chinese listeners' subjective judgments of Spanish sentence types differ from those of native Spanish speakers, and how their perception of the intonational contrasts is shaped by (a) their native tone language experience, and (b) several individual features. The rest of the paper is structured as follows. First, we review some preliminaries on the native and the non-native perception of Peninsular Spanish intonation (Section 1.1) and discuss the effect of some important factors that may affect the cross-linguistic perception (Section 1.2) and the specific research questions examined in our study (Section 1.3). Then, we describe the methodology used in this work (Section 2). Next, the statistical modeling of the data and the specific perceptual results are presented in Section 3. Finally, we discuss the implications of our findings in Section 4 and conclude the paper in Section 5.

1.1 Native and non-native perception of Peninsular Spanish intonation

In Spanish, statements and yes/no questions can be ambiguous from a morphosyntactic point of view, therefore intonation is often the primary cue to help listeners distinguish the two sentence types (Face, 2007). The intonation of statements and questions in Spanish are described to differ in four aspects: initial F0 peak, pitch movement in the medial position, pitch behavior during the final stressed syllable, and the final boundary tone (Face, 2004). Specifically, statements are characterized by an F0 contour that rises from the onset of the stressed syllable and has its peak in the poststressed syllable. And after the last prenuclear syllable, the contour undergoes a progressive downstepping till the utterance-final end (Estebas-Vilaplana & Prieto, 2010; Face, 2007, 2008; Hualde & Prieto, 2015). Yes/no questions in Peninsular Spanish are also known to have a rising pitch accent in the prenuclear position, but their initial F0 peak is higher and the rising movement starts at the offset of the accented syllable (Estebas-Vilaplana & Prieto, 2010; Face, 2007, 2008). After that, the intonation contour begins to fall gradually until the final nuclear syllable, and then

shows a sharp rise at the end of the utterance. In addition to these distinctions (i.e., initial peak and final pitch movement), Face (2004, 2005, 2007) proposed two other minor intonational differences between the two sentence types: (a) statements can have an F0 rises in medial position while yes/no questions do not. However, this potential difference can only be found when the utterance contains at least three-stressed words, and (b) statements frequently show a small rise in their final stressed syllable whereas yes/no questions have an F0 dip through that syllable.

Although the perception of intonation has received less attention than the perception of segmental features, some previous studies have made an attempt to investigate the perceived sentence types by native listeners of Peninsular Spanish (Face, 2005, 2007; Sensui, 1995, 2003). For instance, Face (2005) examined the effect of different intonational cues in perceiving the sentence type of Madrid Spanish by dividing each utterance into nine gates. The results showed that native Spanish listeners were able to correctly perceive 95% of the sentence type after hearing just the initial F0 peak, and this accuracy rate continued to increase after presenting the gate that contained the medial F0 pattern. Further, to compare the strength of each intonational cue, Face (2007) manipulated the height of some pitch points so that the intonational features presented in the utterance were in conflict. Results of this experiment demonstrated that the final pitch movement was the strongest cue for signaling the statement and the yes/no question, despite the heavy influence that the first peak and the median F0 pattern had on listeners' perceptual decisions.

While the perception of Spanish intonation by natives has been investigated, our knowledge of the non-native perception of these prosodic patterns is extremely limited. As far as we know, cross-linguistic perception of Spanish sentence types has been examined only for native English speakers (Trimble, 2013b; Zarate-Sandez et al., 2015) and the Chinese mainland (Li, 2020) and Taiwanese listeners (Cortés Moreno, 2001). For example, Zarate-Sandez et al. (2015) shifted the location of the F0 peak of Spanish statements and compared its perception between L2 English learners and native monolingual speakers of Spanish and English. Their results showed that low-proficiency English learners tended to align the prenuclear peak within the stressed syllable, while the two highly proficient English groups presented a more target-like perception, that is, preferred to use a later alignment of prenuclear peaks when perceiving the Spanish statement (Zarate-Sandez et al., 2015). These authors explained that the perceptual deviance in L2 Spanish, particularly at the initial stage of proficiency, may be attributed to the negative transfer of the L1 English prenuclear pitch accent which is realized within the stressed syllable in the statement.

Another finding of the cross-linguistic perception of Spanish intonation is found in Trimble (2013b). Using a gating experiment, the author revealed that most English speakers of Spanish relied heavily on the final pitch movement for perceiving the sentence type. Specifically, results showed that 58% of English learners failed to distinguish the Spanish statements from the yes/no questions without the final rise/fall. A similar test was conducted with Chinese learners of Spanish (Li, 2020). Interestingly, however, Chinese listeners achieved an overall higher perceptual accuracy (68.33%) solely upon hearing the first peak. The superior performance of Chinese speakers seems to support a recent claim that states that the long-term experience with a tone language can help listeners assimilate the non-native intonational contrasts (Meng et al., 2020; Ortega-Llebaria et al., 2017). But despite this L1 tonal enhancement, Chinese listeners were still significantly less accurate than native speakers in the use of the prenuclear peak to distinguish the sentence type (Li, 2020). A possible explanation for this is that the difference between statements and yes/no questions in Chinese are modeled through the global F0 variation, rather than the local change of tonal structures (i.e., pitch accent and boundary tone)-which is the most common case in intonation languages- (J. Yuan, 2011).

Furthermore, the studies of Li (2020), Trimble (2013b), and Zarate-Sandez et al. (2015) showed that both Chinese (tonal) and English (non-tonal) learners of Spanish can reach roughly the same accuracy level as native speakers after hearing the final pitch movement. These results are consistent with the findings of Gussenhoven and Chen (2000), in which speakers of three different

languages (i.e., Mandarin Chinese, Dutch, and Hungarian) were surprisingly able to use a higher final pitch, higher F0 peak, and later F0 peak as a cue to perceive questions in an unknown language. A generally accepted explanation for this is that such intonational forms, particularly the final rise/fall, maybe universal in the relationship with the paralinguistic meanings and derived from some biologically determined codes (Gussenhoven, 2004). For instance, the Frequency Code proposed that an increased F0 value expresses speaker's uncertainty and hence is an interrogative marker, whilst lower F0 signals the assertiveness and therefore is associated with the declarative encoding (Gussenhoven, 2002; Ohala, 1984).

1.2 Predicting cross-linguistic speech perception

1.2.1 Effect of native tone language experience. Despite their differences on deeper issues, speech perception theories show a strong agreement on the critical role of the L1 experience on the cross-linguistic perception. Take the Perceptual Assimilation Model (PAM) and its extension to L2 learning (PAM-L2), they posit that the perception of L2 segments is implicitly or explicitly constrained by the L1 phonological categories and phonetic properties (Best & Strange, 1992; Best & Tyler, 2007). Similarly, the PAM framework developed for the suprasegmentals (PAM-S) suggests that the perceptual performance of L2 listeners reflects the specific properties of their L1 phonological and phonetic systems (i.e., tone, pitch accent, and intonation) (Alexander & Wang, 2016; So & Best, 2010, 2014a). These claims have been supported by many studies from the prosodic and neurophysiological points of view. Particularly, in the field of suprasegmentals, it has been shown that Chinese listeners' perception of the native or non-native tone (Chang et al., 2017; Hallé et al., 2004; Krishnan et al., 2010), intonation (Ortega-Llebaria et al., 2017), and stress (Meng et al., 2020) is influenced by their native tone language background. For example, Ortega-Llebaria et al. (2017) examined the intonation perception of English words using the prime-target pairs and revealed some interesting differences in pitch processing between the tone and non-tone language groups. Specifically, their results showed that Chinese listeners detected faster the F0 mismatches than speakers of the non-tone group (English), suggesting that the long-term experience with a tone language can shape the way how listeners perceive the tonal contrasts inherent to a given language (Ortega-Llebaria et al., 2017).

This perceptual advantage enhanced by the tone language experience has been demonstrated in a vast majority of studies, even in the perception of local tonal events of a third foreign language. For instance, Wiener and Goss (2018) investigated the perception of Japanese pitch accents by native English speakers using an ABX discrimination task and found that the L1 English group improved apparently their accuracy after taking a 15-week L2 Mandarin tone course. For our purposes, and as we have discussed above (see Section 1.1), it is crucial that Chinese tonal listeners were more accurate than non-tone English speakers in perceiving the prenuclear difference between Spanish statements and questions, supposedly due to their broader experience in processing the pitch contrastive patterns. Similar conclusions have been reached in studies by Meng et al. (2020), Nguyễn et al. (2008), and Vickie and Andruski (2010), which showed that Cantonese and Mandarin listeners were more sensitive to the pitch shift of English lexical stress compared to the English and Vietnamese non-tone speakers. The differences, according to the authors, can be attributed to that tonal language speakers have more complicated and broader tonal categories and, therefore, they need to rely heavily on F0 in the speech (Meng et al., 2020). Moreover, Ortega-Llebaria and Wu (2021) examined the pitch perception in Mandarin and English words and nonwords, concluding that the perception of pitch in a non-tone foreign language was strongly affected by listeners' tonal language background. Specifically, their findings revealed that Chinese speakers were more sensitive to the acoustic salience of F0 contours in the pre-lexical processing because of the top-down feedback (Ortega-Llebaria & Wu, 2021).

Altogether, the research available seems to suggest that tone language experience can improve the perceptual and cognitive ability of pitch patterns, and it constitutes an extremely important factor in predicting the perception of intonational events in a non-tone foreign language.

1.2.2 Effect of individual features. In addition to L1 experience, cross-linguistic speech perception can also be modulated by some inter-individual factors. Possibly, the most studied variables in L2 perceptual studies are: (a) L1 background, (b) L2 proficiency, (c) age, (d) gender, (e) psychometric properties, and (f) musical experience. The critical role of the L1 background has been discussed in section 1.2.1, however, its effect can be adjusted by taking into account the variable of the level of proficiency, which is a more detailed characterization of learners' linguistic experience in the foreign language. The advantage of proficiency has been found in a great body of work on speech production (i.e., Graham & Post, 2018; Li & Post, 2014; Zarate-Sandez et al., 2015; Zhang & Lee, 2019), while it is somewhat unclear whether and how it will affect listeners' perceptual acquisition. A recent study by Li (2020) claimed that there was no significant relationship between the perception of Spanish yes/no questions and the language proficiency of Chinese students. In contrast, Zarate-Sandez et al. (2015) showed that English learners of Spanish were able to progressively fine-tune their perception of prenuclear pitch alignment and approach a native-like performance with increasing their proficiency in the L2. Further research from the neurobehavioral field also provides evidence of the correlates between perceptual ability and language proficiency. For example. Archila-Suerte et al. (2012, 2015) argued that high-proficiency adults were more accurate in between-categorizations of non-native novel sounds because of their ability in using the high-level cognitive process (i.e., attention) to perceive the different phonemic boundaries.

Regarding age, previous literature suggested that listeners who acquired a foreign language at an early stage of age have an enhanced perception and more target-like performance than those who acquired it after adulthood (Archila-Suerte et al., 2012, 2015; Baigorri et al., 2019; Richards et al., 2018). For instance, Baigorri et al. (2019) found that early Spanish-English bilinguals were more accurate than late Spanish-English bilinguals in a categorical discrimination task of English vowels. A similar benefit of early L2 acquisition was also observed in Archila-Suerte et al. (2015). By examining the brain activity of different regions, the authors concluded that the neural responses to L2 speech were primarily determined by the age of acquisition, in spite that their results showed no significant differences between early and late bilinguals on the behavioral perception of L2 sounds (Archila-Suerte et al., 2015). In comparison with the age of acquisition, however, the effect of listeners' age at the time of testing has received considerably less attention in perceptual research. Therefore, it is decided to involve this variable in the present study. Another biological factor that may explain the interspeaker variability in speech perception is gender. Bishop et al. (2020) found that female listeners statistically were more accurate than males in perceiving the prominence of stressed words in native English. However, as the authors noted, this difference was significant only in the semantic processing and probably not valid in other types of processing, such as the auditory temporal-order processing (van Kesteren & Wiersinga-Post, 2007).

Besides the linguistic and biological factors that have been discussed, a series of psychometric variables have been proposed to explain the perceptual differences as well. For instance, previous studies showed that some of the individual variability in prominence identification might be attributed to listeners' different ranges of "autistic traits" (Bishop, 2012, 2017; Bishop et al., 2020; Hurley & Bishop, 2016, Jun & Bishop, 2015). These traits were part of human cognitive processing styles measured by a non-diagnostic instrument, namely, Autism Spectrum Quotient (henceforth, AQ) (Baron-Cohen et al., 2001). One clear evidence of this is that English listeners who were higher in the AQ scores were less sensitive to the prosodic manipulation of prenuclear prominence in online lexical processing (Bishop, 2012, 2017). Moreover, in a recent perceptual study of English prominence, Bishop and colleagues (2020) used the communication subscale of AQ to evaluate the pragmatic skill and concluded that listeners who had higher AQ scores in this subscale (thus, poorer

in pragmatic skills) were less likely to perceive the prominence in words parsed into phonologically weak positions. The established relationship between the communication subscale of AQ and the so-called pragmatic skills has been used in other comparable studies to explain the cross-listener variations in speech. Similar to other listener-based variables such as age and gender, the psychometric properties were argued to have different effects among individuals depending on the prosodic features they were required to discriminate (Bishop et al., 2020, Turnbull, 2017).

The last factor worth highlighting is the role of musical experience. Cognitive experiments have shown that musical background can facilitate speakers' perceptual abilities in a wide range of auditory tasks. For example, listeners with music training experience were better in L2 pronunciation (Baills et al., 2021) and the discrimination of different speech features, such as phonemic vowel length contrasts (Cooper et al., 2017), synthetic syllables with different voice onset time (Zuk et al., 2013), and lexical tone variations (Alexander et al., 2005; Burnham & Brooker, 2002; Wiener & Bradley, 2020). Besides, it is argued that musicians were more sensitive to changes in the relative pitch structure and melodic contours both in their native (Fujioka et al., 2004; Schön et al., 2004) and non-native languages (Deguchi et al., 2012; Marques et al., 2007). However, the musician advantage is supposed to be significant only for auditory tasks that can be improved through years of musical practice and rely on the same skills that are needed for music (i.e., pitch and duration processing) (Başkent & Gaudrain, 2016; Besson et al., 2011; Fuller et al., 2014). In sum, the different types of predictors and their specific effects related to individuals and auditory tasks suggest a complex perceptual processing mechanism, which prompts us to examine the intonation perception from a broader point of view considering both the linguistic experience and the interspeaker variability.

1.3 The present study

Taking into account the above discussion, we can assume that the cross-linguistic perception is determined by many different types of factors. The overarching goal of the present study is to (a) explore the role of these different factors that play in the perception of intonation and (b) predict the perceptual difficulties of Chinese listeners of Spanish in comparison with the native Spanish subjects. To that end, we carried out a gating experiment to investigate the categorical perception of intonation patterns and contrastive structures of Spanish statements and questions. In particular, we examine two dimensions of factors: (a) being a native speaker of a tone language and (b) individual features, which include the level of proficiency, age, gender, AQ score, and musical experience. Therefore, we set three research questions about Chinese listeners' perception of Spanish intonation, neither of which has been fully answered for the Chinese-Spanish language pair.

1. Are Chinese listeners comparable to Spanish native speakers in perceiving the different intonational cues for Spanish statements and yes/no questions?
2. How is the categorical perception of Spanish sentence types shaped by listeners' native language background? Does long-term experience with a tone language help enhance speakers' perceptual and cognitive ability for the non-native intonation contours?
3. Up to which point is the Chinese listeners' perception of Spanish sentence types modulated by a complex set of inter-individual features, such as proficiency level, age, gender, AQ score, and musical experience?

2 Methodology

In order to answer the research questions posed above, we designed a perception test using the gating technique (Face, 2005, 2007, 2011). The following paragraphs state how the stimuli of this experiment were prepared and how the test was administered.

2.1 Materials and recordings

We used four pairs of broad focus statements and yes/no questions to generate the stimuli for the gating experiment (see Table 1). At the surface level, the statement and the yes/no question of each pair were morphosyntactically identical. The four sentence matches differed in their length and therefore in their number of stressed syllables. Specifically, we used sentences with two stressed words and sentences with one stressed word. For each length, we included two stress types for the utterance-final word: penultimate stressed syllable (paroxytone) and final stressed syllable (oxytone). The eight sentence items were comprised of words with high familiarity ratings or high frequency for the benefit of non-natives' comprehension during the auditory task (Tanaka & Terada, 2011).

Table 1. Categorization of the recorded stimuli.

Sentence type	N. of stress	Last word stress position	Stimulus
Statement	1	Oxytone	Alcalá.
Statement	1	Paroxytone	Sevilla.
Statement	2	Oxytone	Viene a Alcalá.
Statement	2	Paroxytone	Viene a Sevilla.
Yes/no question	1	Oxytone	¿Alcalá?
Yes/no question	1	Paroxytone	¿Sevilla?
Yes/no question	2	Oxytone	¿Viene a Alcalá?
Yes/no question	2	Paroxytone	¿Viene a Sevilla?

The DCT (Discourse Completion Task) method was used to elicit a natural production of the speech samples (Félix-Brasdefer, 2010). In specific, we created eight brief dialogues structured as situational contexts to elicit the neural statement and yes/no question intonation. A female native speaker of Barcelona (age at recording: 31) was asked to give her response at a normal speaking rate after presenting each context. The conversation was started by an interlocutor with whom the native speaker had a familiar relationship so that the politeness-related effects (i.e., power and social distance) could be neutralized (Borràs-Comes et al., 2015). The text of the dialogues was written on paper and distributed to the native speaker to enhance her memory of the responses and pragmatic understanding of the contexts. The Spanish native speaker performed this task twice and the first realization was preferentially used for our analysis unless there was a speech error in the first responses. The recordings took place in a quiet room using the microphone Rode Smartlav+ connected to the interface of Scarlett. Audio files were digitized at a sampling rate of 44.1 kHz and with a quantization precision of 16 bits. Each utterance was saved separately and annotated to a *TextGrid* object.

2.2 Stimuli creation

The natural recordings described above were used to create the stimuli of the test. In order to do so, a gating method was employed based on the rules proposed by Face (2005, 2007, 2011). Each of the four sentences with two stressed words was split into four gates. Contrary to what is usually done, we did not use segmental information, in order to create the steps, instead, we have used intonational events, which contain the three intonational differences between the statement and the yes/no question in Peninsular Spanish. As a result, gates coincided with known intonational events in Spanish phonology and therefore recordings with one and two stressed words had a different number of steps.

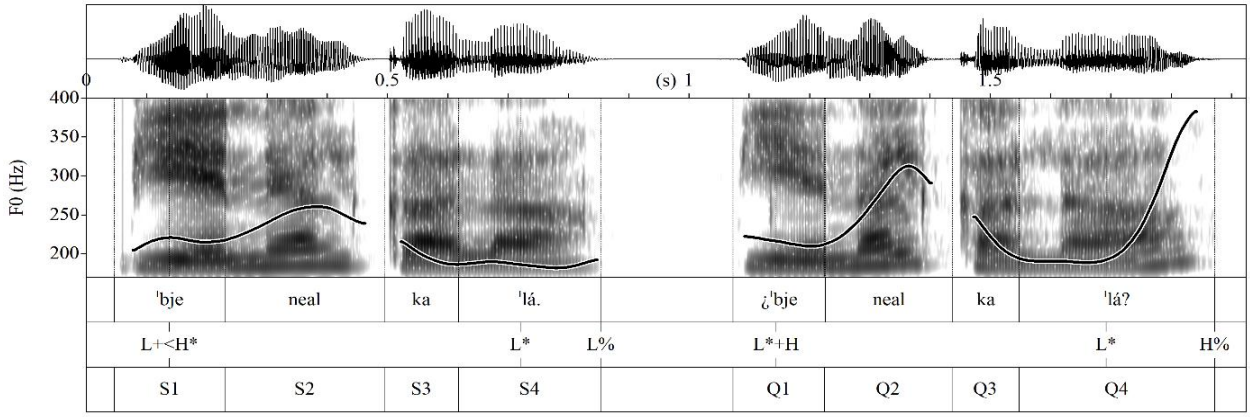


Figure 1. Intonation contours of the sentence pair with two stressed words aligned with the syllables and the gates.
S1, gate 1 of the statement; Q1, gate 1 of the yes/no question; and so on for the rest of the abbreviations.

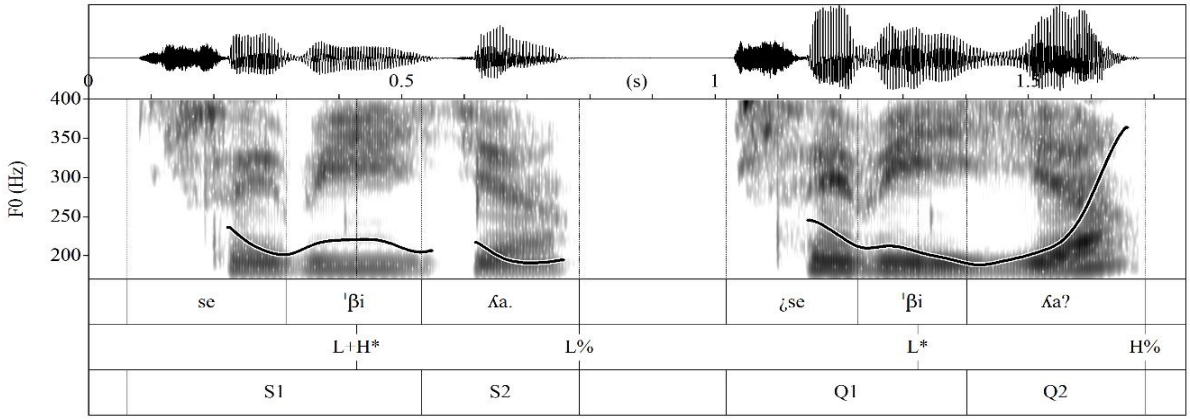


Figure 2. Intonation contours of the one-word statement and yes/no question aligned with the syllables and the gates.
S1, gate 1 of the statement; Q1, gate 1 of the yes/no question; and so on for the rest of the abbreviations

For sentences with two stressed words, Gate 1 consisted of the F0 contour during the first stressed syllable "Vie", where the statement showed a slight rise starting near the onset of the stressed syllable while the yes/no question no (see Figure 1). Gate 2 included the initial F0 peak, and it may sometimes contain an adjacent syllable of the following word because of the synalepha. This peak of the second gate is, according to the literature (Estebas-Vilaplana & Prieto, 2010; Face, 2007, 2011), the first strong differential cue in determining the type of the two-word sentences, given that questions have a different prenuclear pattern than statements in Peninsular Spanish. That is also the case in our study, for example, in Figure 1, the yes/no question has a 71 Hz higher F0 peak (335 Hz) than the statement (264 Hz). Gate 3 begins after the first peak and ends before the final syllable of the sentence. Hence, for sentences ending with a paroxytone word, the third gate included the F0 contour of the final stressed syllable, which unequivocally differentiates the yes/no question (rising contour) from the statement (low-falling contour). In the case of utterances ending with an oxytone word, gate 3 did not contain a stressed syllable. This means that gate 3 has more cues for the correct recognition in final paroxytone sentences than in oxytone ones. However, we decided to keep this division in order to have the same number of gates for the two stress types. This difference in gate 3 gives us the opportunity to explore the role of the final stressed syllable in perceiving the question-statement contrast. The last gate (gate 4) of the two-word sentences

consisted of the intonation patterns of the utterance-final syllable which traditionally have been referred to as the last but most typical cue for signaling the statement and the yes/no question.

Regarding the four sentences with one stressed word, we created two gates. The boundary of the gates was established just before the final syllable (see Figure 2). Similar to the sentences with two stressed words, the gating division of the two stress types in the one-word sentences was slightly different. In particular, the paroxytone word differed from the oxytone word in including the F0 contour of the stressed syllable in gate 1 (which was a relatively high-rising contour in statements whereas a low-falling contour in yes/no questions). The last gate (gate 2) was comprised of the F0 contours of the whole word, including the typical final pitch movement of statements and questions.

In both cases, the division of each gate for the four sentence pairs was realized using *Praat* (Boersma & Weenink, 2020) without modifying any acoustic signal across the whole utterance. The final stimuli were generated by successively accumulating the first gate, the second gate, and so on until involving the entire utterance. In sum, a total of 24 stimuli (four gate 1, gate 2, gate 3, and gate 4 for the two-word sentences; four gate 1 and gate 2 for the one-word sentences) were created for the gating experiment. Each stimulus was separated and saved as a new sound file, with 500 milliseconds of silence before and after the speech.

2.3 Questionnaire design and procedure

To ensure a minimum sample size and facilitate the data collection during the covid-19, a questionnaire was built up electronically using the online software *Alchemer* with customized solutions. The questionnaire consisted of three main sections. The first section provided the basic information on the socio-demographic and language background of the participants. The second part was designed exclusively for non-native Chinese listeners, and it comprised 10 items selected from the self-administered Autism Spectrum Quotient. The complete AQ involved a total of 50 tokens distributed in five subscales: social skills, communication, attention switching, attention to details, and imagination (Baron-Cohen et al., 2001). Out of those, we only used the 10 tokens of the communication subscale, since they were proposed to be a rough proxy for participants' pragmatic ability to engage the prosody in a specific context (Bishop et al., 2020). The responses of the 10 items were evaluated with a four-point Likert scale: strongly agree, slightly agree, slightly disagree, and strongly disagree, following the scoring method used in Lau et al. (2013). The total AQ score of each Chinese participant was the sum of the score for every item. Higher levels of scores indicated more "autistic-like" properties, and therefore, were considered to have inferior pragmatic skills (Lau et al., 2013; Yu, 2010). The third section was divided into three experimental sessions depending on the type of the auditory task. In this study, we particularly focused on the results of the gating test.

The complete questionnaire was prepared in Chinese and Spanish and administered in the language of the listener's preference. Therefore, the 10 AQ items (originally in English) were adapted to Chinese and Spanish taking the previous translations into consideration (see the Online Appendix A for more details) (Lau et al., 2013; Mengrong, 2008; The Autism Research Centre, 2010). The questionnaire was distributed separately in Chinese and Spanish via e-mail and social media platforms. Subjects had the option to choose to participate in one, two, or three tasks depending on their interests. Once the choice had been made, they were randomly assigned one of the three tasks. Participants were required to listen to the stimuli using the earphones in a quiet place. Moreover, they were instructed to listen to every audio sample once, however, they had the option of listening to it a second time if the recording did not play due to technical problems. Prior to the three formal tasks, participants performed a practice trial to get familiar with the procedure. The training items consisted of three stimuli that were not used in the tasks. Besides, given the difficulty that a gating task can be, the text of the entire utterance was displayed without punctuation marks. The task of the gating experiment was to identify whether the stimulus they

heard was “Statement” or “Yes/no question” by clicking a button on the screen. To avoid listeners’ response bias, the two perceptual choices and all speech stimuli were set in random order. A short break was inserted after each part of the questionnaire to avoid the fatigue of the listeners.

2.4 Participants

There were initially 76 Peninsular Spanish speakers and 117 Mandarin Chinese participants that completed and submitted the online questionnaire successfully. Among them, the data of two Spanish listeners were excluded because their age was over 60. As for Chinese participants, those who self-declared to have learned Spanish as a foreign language before reaching the age of 16 were removed from the final data to minimize the effect of age of acquisition. All Chinese listeners confirmed that they had been predominantly exposed to a Peninsular Spanish variety during the learning process. The proficiency of most Chinese participants (roughly 60%) was assessed using the level information of the last Spanish certificate DELE (Diploma of Spanish as a Foreign Language) that they held. As with the 31 Chinese speakers who did not have such a diploma, they were asked to self-evaluate their Spanish proficiency based on the language courses they had completed and according to the criteria of the six levels of European language proficiency. 8 Chinese listeners with a proficiency level of A1 and A2 in Spanish were removed because they were not enough to form a group. The selected Chinese and Spanish listeners were divided into four language groups according to their language proficiencies in Spanish: B1, B2, C1, and native.

Thus, participants of the gating test were 43 native Spanish listeners and 75 Chinese speakers. Among the Chinese participants, 22 listeners were at B1 level, 27 listeners were at B2 level, and 26 listeners were at C1 level. Ages of all participants ranged from 18 to 59 years with a mean age of 28.77 (SD=8.35). Among them, 28 were male, and 90 were female. None of the individuals in the four language groups reported any history of hearing or communication problems at the time of testing.

2.5 Data analysis

As it had been said, the data was gathered by means of *Alchemer*. The results were downloaded in a csv file and the subsequent analysis was carried out in R (R Core Team, 2020). To examine the effect of different factors on the perception of Spanish intonation, a binary logistic regression was carried out using the *glm* function in the base R package (R Core Team, 2020). Noticeably, in the gating task, the binary outcome variable “perceptual accuracy” was created based on listeners’ responses to each stimulus. The true and false answer was coded as 1 and 0, respectively. L1 (Chinese vs. Spanish), sentence type of the stimulus (statement vs. yes/no question), stress type of the final word (paroxytone vs. oxytone), music training (yes vs. no), and gender (male vs. female) were treated as binary predictors, while proficiency level (B1 < B2 < C1 < native) and gate (1 < 2 < 3 < 4 for two-word sentences; 1 < 2 for one-word sentences) were read as discrete ordinal variables. For the continuous predictors such as age, AQ score, and stimulus order, a z-transformation was performed based on the mean and standard deviation before entering the model.

The effect of native language experience and individual features were examined separately for the sentences with one and two stressed words. Thus, four multivariable general linear regression models (GLM) were built in the current study. The two models for analyzing the L1 effect were run with the factors of gate, L1, sentence type, stress type, stimulus order, and all their possible interactions, while the two for analyzing the induvial features were performed by substituting L1 for the variables of age, gender, proficiency level, AQ score, and music training experience.

Further, to find the best performing model, a stepwise backward elimination was carried out using the *stepAIC* function from *MASS* package to iteratively exclude the least contributive predictors (Kassambara, 2018; Ripley et al., 2013). That is to say, when the factors failed to achieve

the .05 level of significance, they were removed from the final model until all effects were statistically significant for our prediction. Additionally, for the interaction analysis, *emmeans* objects were created using the *contrast* function from the *emmeans* package to estimate the contrast of interests (Lenth et al., 2019). In particular, the *Tukey* method was applied to compare the effect of one predictor at each level of another by averaging over levels of the other variables in the model.

3 Results

3.1 Results of the sentences with two stressed words

3.1.1 Identification results depending on native language experience. Since the effects of stress type and stimulus order were so slight as to be statistically irrelevant to the prediction, they were not included in the final model. The likelihood ratio test of the selected model indicated that there was a highly significant effect between the perceptual accuracy and the variables of sentence type, gate, and L1, χ^2 (13) = 527.66, $p < .0001$. Further, the analysis of variance using the *anova.rms* function (Harrell Jr, & Hmisc, 2019) showed that the predictor of gate was strongly interacted with sentence type, χ^2 (3) = 35.09, $p < .0001$, and L1, χ^2 (3) = 18.20, $p < .0001$.

The output of the regression model is shown in Table 2. For ease of exposition, we discuss the interaction effects combined with the descriptive results in Table 3. In general, it is noticed that listeners were more accurate in perceiving the statements rather than the yes/no questions. However, the effect of sentence type was not distributed evenly across the four gates as revealed by the significant interactions between the two variables. Specifically, Table 3 shows that the perceptual accuracy of statements was strongly higher than the yes/no questions, especially in gate 1. To be more precise, we transformed the regression coefficient in Table 2 into odds ratios (OR) by exponentiating the logit of 1.3053, and the results indicated that statements were 3.69 times more likely to be accurately perceived than yes/no questions in the first gate, OR = exp (1.3053) = 3.69, $p < .0001$. As regards the other three gates, no statistically significant difference in accuracy was found between the two sentence types. These results, as a whole, are in good agreement with previous findings that suggested a psychological inclination for listeners to perceive an utterance as a statement (Face, 2005; Li, 2020; Trimble, 2013b), particularly when there was no apparent intonational cue for signaling the question, as for instance in gate 1. This perceptual bias existed consistently between the Chinese and Spanish listeners in our study, probably because the statement was the most neutral and unmarked modality in most human languages that carry little communicative burden in the speech.

Considering the interaction of L1 with gate, Table 3 indicates that native and non-native listeners had different ways of changes at each level of the gate, although both L1 groups improved the accuracy rate by accumulating the intonational cues in the stimuli. Specifically, in the first gate, the Chinese and the Spanish group showed no significant difference of accuracy in the question-statement identification (see Table 2). In the second gate, Spanish listeners were predicted to have a significantly (4.60 times) higher likelihood of making correct discriminations of the sentence type compared with Chinese speakers, OR = exp (1.5251) = 4.60, $p < .0001$, despite there also being significant progress for the Chinese group from gate 1 to gate 2, OR = exp (0.8321) = 2.30, $p < .0001$. The apparent increase of accuracy of both L1 groups in gate 2 was cued by the presence of the initial F0 peak, which was higher in the yes/no question and lower in the statement of Spanish. Similar results were also observed in Li (2020) and Trimble (2013b), which found an enhanced perception for the English and Chinese learners of Spanish after being presented with the first prenuclear peak of the utterance.

Table 2. Results of the logistic regression model testing the effect of L1 on the perception of sentences with two stressed words.

Fixed effect	*Coef*	*SE*	*Z*	*p*
(Intercept)	0.7770	0.209	3.71	<.001**
SentenceType *Gate (Q1 vs. S1)	-1.3053	0.195	-6.70	<.0001***
SentenceType *Gate (Q2 vs. S2)	0.3676	0.222	1.65	>.05
SentenceType *Gate (Q3 vs. S3)	-0.0825	0.287	-0.29	>.1
SentenceType *Gate (Q4 vs. S4)	-1.4066	1.124	-1.25	>.1
L1*Gate (Spanish1vs. Chinese1)	0.3142	0.202	1.55	>.1
L1*Gate (Spanish2 vs. Chinese2)	1.5251	0.284	5.37	<.0001***
L1*Gate (Spanish3 vs. Chinese3)	1.9698	0.478	4.12	<.0001***
L1*Gate (Spanish4 vs. Chinese4)	13.4462	294.463	0.05	>.1
Gate*L1(Chinese1 vs. Chinese2)	0.8321	0.173	4.80	<.0001***
Gate*L1(Chinese2 vs. Chinese3)	0.8412	0.195	4.30	<.0001***
Gate*L1(Chinese3 vs. Chinese4)	2.7613	0.582	4.74	<.0001***
Gate*L1(Spanish1 vs. Chinese2)	2.0430	0.303	6.74	<.0001***
Gate*L1(Spanish2 vs. Chinese3)	1.2859	0.521	2.47	>.05
Gate*L1(Spanish3 vs. Chinese4)	14.2381	294.463	0.05	>.1

Coef, estimated coefficients; *SE*, standard error; *Z*, z ratio; *p*, significant difference; Q1, gate 1 of the yes/no question; S1, gate 1 of the statement; Spanish1, the Spanish group in gate 1; Chinese1, the Chinese group in gate 1; and so on for the rest of the abbreviations.
* < 0.05, ** < 0.01, *** < 0.001.

Table 3. Descriptive statistics of the perceptual accuracy (%) after each of the four gates in the sentences with two stressed words.

Gate	Statement		Yes/no question		Combined total	
	CH	SP	CH	SP	CH	SP
1	62.67%	69.77%	31.33%	38.37%	47.00%	54.07%
2	63.33%	87.21%	70.00%	93.02%	66.67%	90.12%
3	84.00%	95.35%	80.67%	98.84%	82.33%	97.09%
4	99.33%	100.00%	97.33%	100.00%	98.33%	100.00%

CH, non-native Chinese group; SP, native Spanish group.

Next, from gate 2 to gate 3, the F0 down slope after the highest peak and before the final pitch change was referred to as the second intonational cue for the question-statement identification (with the yes/no question having a stepper F0 slope than the statement). Compared with the second gate, the likelihood of accurately perceiving the sentence type in gate 3 was 2.32 times greater for Chinese listeners, OR = exp (0.8412) = 2.32, p < .0001. Although Spanish listeners did not show significant progress after hearing gate 3, they statistically were still more accurate than Chinese listeners in the question-statement identification (see table 2). After gate 3, when the entire utterance was released, both native and non-native listeners were able to perceive the sentence type with a nearly 100% accuracy rate (see Table 3), suggesting that the final F0 rise may be a universal auditory cue for the activation of interrogative meanings. Finally, as opposed to gate 3, the odds of accurately perceiving the sentence type in gate 4 was significantly (15.82 times) greater for the Chinese group, OR = exp (2.7613) = 15.82, p < .0001. This finding confirms our previous predictions, showing that the final pitch movement was a strong intonational cue for the non-native listeners in the question-statement identification. While for the Spanish group, it may seem that the final F0 contour was not important as the initial peak for the perception of sentence type, because

their increase of perceptual accuracy was not significant after being presented with gate 4 (see Table 2). However, this impression has proved to be false by Face (2007, 2011), which demonstrated that the final pitch movement was the strongest cue for native Spanish listeners in determining the sentence type that can override any other previous cues of the sentence. In this study, to further confirm the strength of the final F0 contour, sentences with one stressed word were analyzed (see Section 3.2) due to their advantages in reducing the interference of other intonational cues in the perception.

3.1.2 Identification results depending on individual features. As discussed above, the cross-linguistic speech perception, rather than being solely determined by the L1 experience, was affected by several internal features specific to each individual. Thus, in this section, a new binary logistic regression was carried out by adding the listener-based variables (e.g., age, gender, proficiency level, music training, and AQ score) that allowed us to assess which factors explained the individual variations in the speech perception. The new GLM model was performed only for the data of the Chinese group. Apart from the significant interaction between sentence type and gate, χ^2 (3) = 24.24, $p < .0001$, Table 4 indicates that the perceptual accuracy of Chinese listeners was also highly predicted by their Spanish proficiency, χ^2 (2) = 14.09, $p < .001$, pragmatic skills χ^2 (1) = 31.18, $p < .0001$), and musical experience, χ^2 (1) = 6.04, $p< .05$. In contrast, the two predictors of age and gender statistically did not have an influence on the discrimination of questions and statements.

Table 4. Results of the logistic regression model testing the effect of individual features on the perception of sentences with two stressed words.

Fixed effect	*Coef*	*SE*	*Z*	*p*
(Intercept)	0.0960	0.233	0.41	>.1
SentenceType *Gate (Q1 vs. S1)	-1.3950	0.253	-5.51	<.0001***
SentenceType *Gate (Q2 vs. S2)	0.3210	0.254	1.26	>.1
SentenceType *Gate (Q3 vs. S3)	-0.2400	0.311	-0.77	>.1
SentenceType *Gate (Q4 vs. S4)	-1.4131	1.126	-1.26	>.1
Proficiency (B2 vs. B1)	0.5288	0.194	2.73	<.05*
Proficiency (C1 vs. B1)	0.7444	0.203	3.66	<.001**
Proficiency (C1 vs. B2)	0.2160	0.186	1.159	>.1
AQ_z	-0.4628	0.083	-5.58	<.0001***
MusicTrain (Yes vs. No)	0.4456	0.181	2.46	<.05*
Age_z	0.2084	0.172	1.21	>.1
Gender (Male vs. Female)	-0.1198	-0.200	-0.60	>.1

Coef, estimated coefficients; *SE*, standard error; *Z*, z ratio; *p*, significant difference; Q1, gate 1 of the yes/no question; S1, gate 1 of the statement; and so on for the rest of the abbreviations.
* < 0.05, ** < 0.01, *** < 0.001.

The interaction of sentence type with gate is consistent with the results of section 3.1.1, which showed a perceptual advantage of statements in the first gate where no apparent interrogative cue was provided in the utterance. Moreover, the main effect of proficiency level appears to suggest that the perceptual ability of the different intonational cues was positively related to listeners' development of proficiency in the non-native target language. For example, Figure 3 shows that highly proficient Chinese listeners (i.e., those of B2 or C1 level) estimated much more accurately

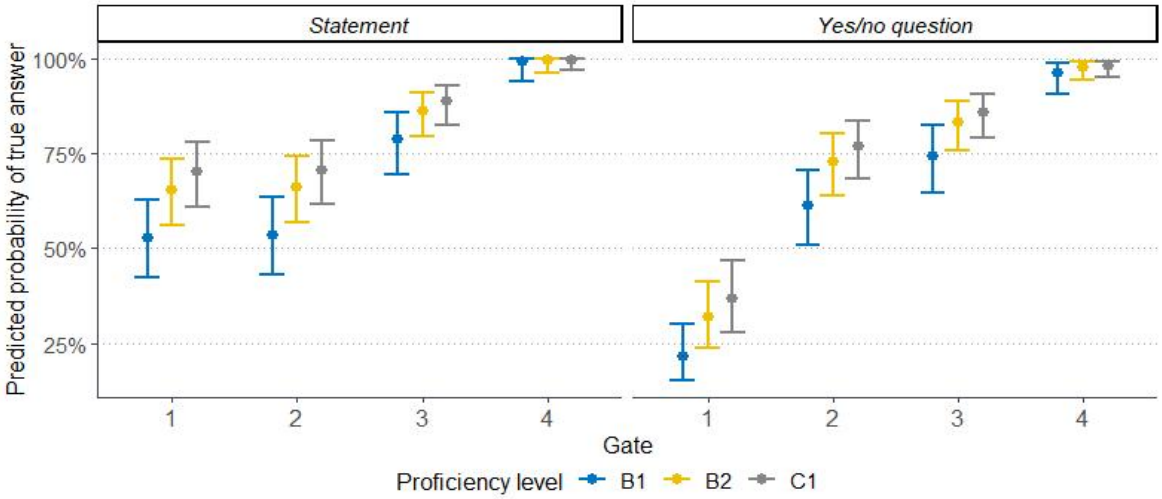


Figure 3. Effect of proficiency level across different categories of sentence type and gate in the logit model fitted for sentences with two stressed words.

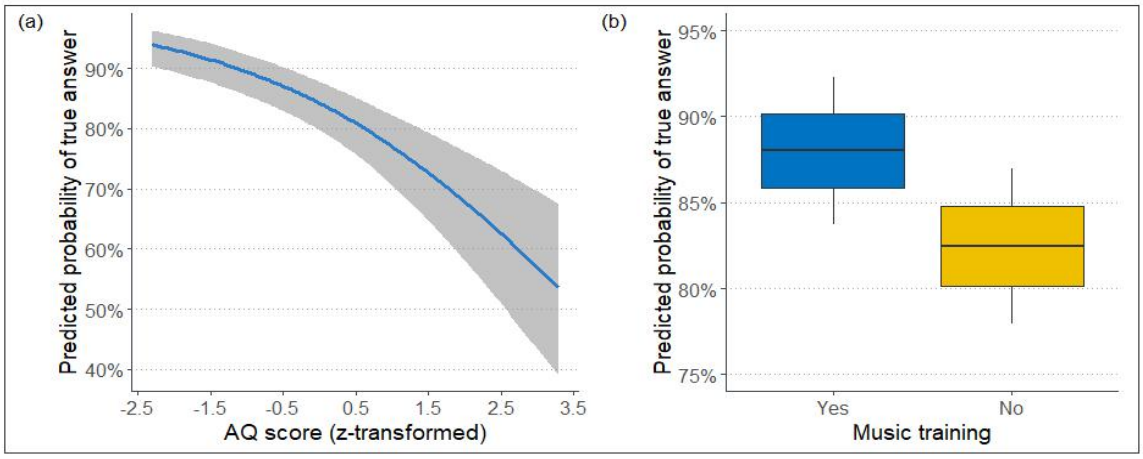


Figure 4. Effect of AQ score (a) and music training experience (b) in the logit model fitted for sentences with two stressed words.

the sentence type than those only with a B1 level in Spanish. In particular, the likelihood of accurately distinguishing the sentence type was receptively 1.70 times and 2.11 times higher if the Chinese listener belonged to the B2, OR = exp (0.5288) = 1.70, $p < .05$), and the C1 group, OR = exp (0.7444) = 2.11, $p < .001$, rather than the B1 group. Thus, overall, the results seem to support previous findings that proposed an enhancement of language proficiency for the cross-linguistic perception including the speech comprehensibility (Kraut & Wulff, 2013), the speech recognition in noise (Kilman et al., 2014; Smiljanić & Bradlow, 2011), and the between-categorization of the L2 sounds (Archila-Suerte, et al, 2010, 2012, 2015; Richards et al., 2018).

Further, consistent with previous studies (Bishop, 2017; Bishop et al., 2020; X. Yang et al., 2018), Figure 4a reveals a strong negative correlation between the AQ score and the probability of accurate perceptions of the sentence type. Specifically, the coefficient in Table 4 indicates that the likelihood of correctly perceiving the sentence type was 37% lower if there was a one standard deviation increase in Chinese listeners' AQ score, OR = exp (-0.4628) = 0.63, $p < .0001$. Moreover, as is clear from Figure 4b, Chinese listeners with musical training experience had a substantial benefit in perceiving the sentence type compared with those musically untrained subjects. All else being equal, the likelihood of accurately distinguishing the sentence type was 56% greater if the

listener had training experience with music, OR = exp (0.4456) = 1.56, $p < .05$. This result is as expected and highly consistent with previous findings that native and non-native listeners who had either a long-term or short-term musical practice outperformed non-musicians in several aspects of the speech, such as lexical tones (Chang et al., 2016; Wiener & Bradley, 2020; Zhao & Kuhl, 2015), pitch tracking accuracy (Bidelman et al., 2011; Chen et al., 2010; Schön et al., 2004), and melodic contours (Bradley, 2016; Fujioka et al., 2004; Nan et al., 2018). However, it should be mentioned that in this study subjects with music training experience accounted only for roughly one-third of the Chinese samples. A more balanced dataset therefore should be proposed for further reliable analysis.

3.2 Results of the sentences with one stressed word

3.2.1 Identification results depending on native language experience. To examine the strength of the final intonational cue in the native and the non-native perception, the data of the one-word sentences was analyzed by building a new GLM model with a similar fitting process. The variables of music training and stimulus order were excluded because their inclusion statistically did not contribute to the model's predictive accuracy. Overall, the significance test of the model showed that there was a highly significant effect between the perceptual accuracy of the one-word sentences and the four predictors of sentence type, gate, L1, and stress type, χ^2 (5) = 437.99, $p < .0001$.

Considering firstly the effect of sentence type, Table 5 indicates that listeners had significantly lower accuracy in perceiving the yes/no question rather than the statement. All else being equal, the yes/no question was predicted to have 85% lower odds of being correctly perceived compared to the statement over the utterance, OR = exp (-1.8959) = 0.15, $p < .0001$. The result is in parallel with our previous findings on the two-word sentences, suggesting that there might be a hierarchy of difficulties in perceiving the different sentence types in Spanish. Moreover, despite there being no significant interaction of sentence type with gate in the one-word sentences, the perceptual differences between the statement and the question were mainly shown on the first gate (see Table 6). From gate 1 to gate 2, all listeners experienced a highly significant increase in the likelihood that a given utterance is correctly perceived and achieved a nearly 100% accuracy rate. This result would, of course, provide evidence for the important role of the final F0 contour, overthrowing the view that the final pitch movement had little or less strength compared with other cues in perceiving the sentence type. In fact, for sentences with two stressed words, the seemingly lower weighting of the final gate was because there were apparent intonational makings (e.g., initial F0 peak) before the utterance-final end which could lead native listeners to make a confident perceptual decision without the presence of the final intonational cue.

Next, the positive coefficient of L1 (see Table 5) reveals that Spanish native listeners were significantly more accurate (2 times greater in odds ratios) than non-native Chinese learners in the overall identification of the one-word sentences, OR = exp (0.6919) = 2.00, $p < .01$. Apart from the L1 effect, a unique result was obtained for the variable of stress type. Specifically, Table 5 shows that the paroxytone word had a 2.32 times higher likelihood of being correctly perceived compared with the oxytone word, OR = exp (0.8422) = 2.32, $p < .01$. However, noticeably, the perceptual advantage of the paroxytone word was mainly restricted to the first gate (see Table 6). One possibility of this was related to the different phonological components involved in the first gate of the two stress types. Because the first gate of the paroxytone word "Sevilla" was divided including the F0 contour of the nuclear syllable "vi" (which was a low-falling contour in yes/no questions and a relatively high-rising contour in statements), it is speculated that listeners may probably make use of this intonational difference to improve their perceptual accuracy of the sentences with paroxytone. As with the oxytone word "Alcalá", however, the first gate did not contain the stressed syllable and there were no other apparent intonational cues during the F0 trajectories of "Alca". Therefore, it may be notoriously challenging for listeners to discriminate the intonation type in gate

1. Together, the perceptual distinction of the two stress types seems to imply that the F0 contour of the nuclear stressed syllable is also an effective cue for perceiving the one-word sentences in Peninsular Spanish.

Table 5. Results of the logistic regression model testing the effect of L1 on the perception of sentences with one stressed word.

Fixed effect	*Coef*	*SE*	*Z*	*p*
(Intercept)	0.7770	0.209	3.71	<.001**
SentenceType (Statement vs. Question)	-1.8959	0.221	-8.57	<.0001***
Gate (2 vs.1)	5.2728	0.597	8.83	<.0001***
L1 (Spanish vs. Chinese)	0.6919	0.220	3.15	<.01*
StressType (Paroxytone vs. Oxytone)	0.8422	0.217	3.89	<.01*

Coef, estimated coefficients; *SE*, standard error; *Z*, z ratio; *p*, significant difference.
* < 0.05, ** < 0.01, *** < 0.001.

Table 6. Descriptive statistics of the perceptual accuracy (%) after each of the two gates in the sentences with one stressed word.

Gate	Stress type	Statement		Yes/no question		Combined total	
		CH	SP	CH	SP	CH	SP
1	Paroxytone	78.67%	95.35%	34.67%	51.16%	57.33%	73.26%
1	Oxytone	65.33%	72.09%	25.33%	32.56%	45.33%	52.33%
2	Paroxytone	100.00%	100.00%	98.67%	100.00%	99.33%	100.00%
2	Oxytone	98.67%	100.00%	98.67%	100.00%	99.33%	100.00%

CH, non-native Chinese group; SP, native Spanish group.

3.2.2 Identification results depending on individual features. For the one-word sentences, an independent GLM model was run with all the listener-based factors included so that the effect of the individual features could be examined. The output in Table 7 shows that the effect of sentence type, gate, and stress type is consistent with the results of the two-word sentences (see Section 3.2.1), therefore, we did not discuss them in detail in this section. Regarding the individual features, it is generally the case that the variation of the non-native listeners' age and gender was not necessarily associated with the likelihood of the perceived sentence type. However, as with proficiency level, results in Figure 5 seem to suggest a general trend of perceptual enhancement with increasing learners' proficiency in the target language, although no statistical difference of accuracy was tested between the three Chinese groups. Here, the insignificance of the proficiency appears to be somewhat contradictory to the results of the two-word sentences, but it can be well explained by the gating structure of the one-word stimuli. Since there was little strong intonational cue involved in the first gate of the one-word sentences, a roughly comparable lower accuracy rate was achieved by the three Chinese groups. While in the second gate, the final pitch movement was so universally recognized that nearly all the Chinese listeners could make use of it for discriminating the sentence type.

Further, the pragmatic skill measured by the AQ score probably was the most distinctive feature in terms of the interspeaker variations in the cross-linguistic perception. Similar to our previous results, a strong negative correlation was found between the AQ score and the perceptual accuracy of the one-word stimuli. Specifically, Table 7 indicates that the odds of accurately discriminating the one-word sentences was 43% lower if there was a one standard deviation increase in the AQ

scale, OR = exp (-0.562) = 0.57, p < .0001. The last significant predictor for the perceived sentence type is musical experience, which had a similar effect size for the two lengths of sentences in our study. Compared with those who were naïve to music, Chinese listeners with musical expertise were 1.88 times more likely to make a correct identification of the one-word stimuli, OR = exp (0.6327) = 1.88, p < .05 (see Table 7). Similar findings can be observed in a vast of studies focused on the music-language interaction. A piece of recent evidence is that native speakers of English improved significantly their sensitivity to the pitch patterns of Mandarin tones after taking a short-time musical training combined with L2 classroom learning (Wiener & Bradley, 2020). As discussed above, the perceptual advantage shown by musicians can be explained by a transfer of auditory skills developed in music to speech perception, because music and speech share a similar set of acoustic and neutral resources in processing the signals (Bidelman et al., 2011; Kraus & Chandrasekaran, 2010; Patel, 2011).

Table 7. Results of the logistic regression model testing the effect of stress type and individual features on the perception of sentences with one stressed word.

Fixed effect	*Coef*	*SE*	*Z*	*P*
(Intercept)	0.3037	0.346	0.88	>.05
SentenceType (Question vs. Statement)	-1.9871	0.274	-7.25	<.0001***
Gate (2 vs.1)	5.1933	0.620	8.39	<.0001***
StressType (Paroxytone vs. Oxytone)	0.5734	0.263	3.89	<.05*
Proficiency (B2 vs. B1)	0.4547	0.336	2.18	>.1
Proficiency (C1 vs. B1)	0.5205	0.344	1.51	>.1
Proficiency (C1 vs. B2)	0.0658	0.313	0.21	>.1
MusicTrain (Yes vs. No)	0.6327	0.303	2.09	<.05*
AQ_z	-0.5620	0.1435	-3.92	<.0001***
Age_z	0.0181	0.298	0.06	>.1
Gender (Male vs. Female)	-0.2089	0.340	-0.61	>.1

Coef, estimated coefficients; *SE*, standard error; *Z*, z ratio; *p*, significant difference.
* < 0.05, ** < 0.01, *** < 0.001.

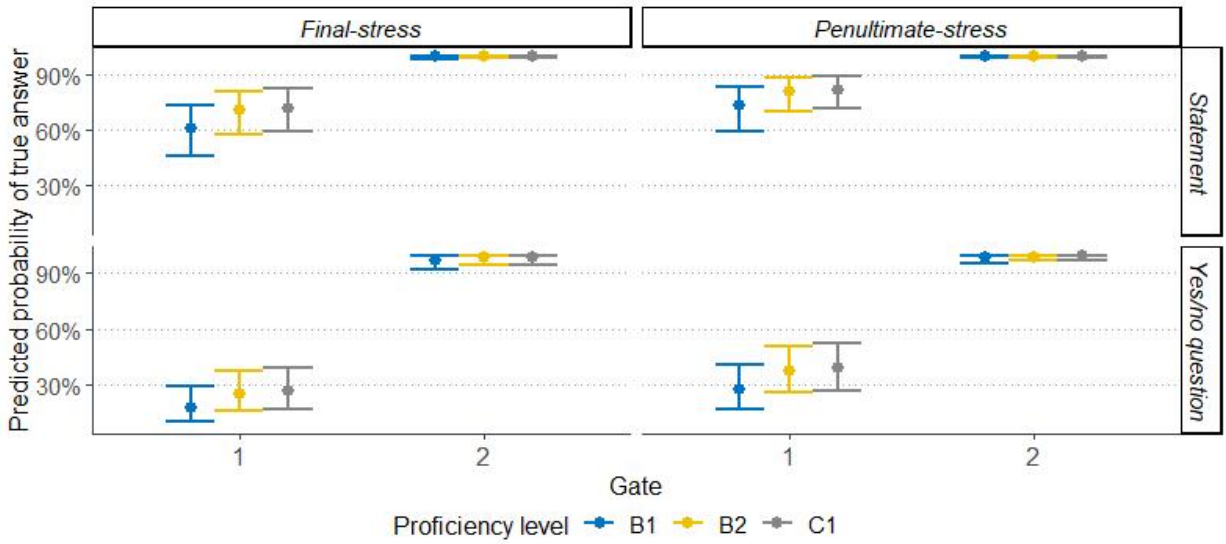


Figure 5. Effect of proficiency level across different categories of sentence type, gate, and stress type of the sentences with one stressed word.

3.3 Evaluation of model performance

For the estimation of the statistical validity, firstly, we used the method of K-fold cross validation to assess how well our generalized logistic regression models were performed in general when used to make predictions on the unseen data (R code is available in the Online Appendix B). The K value was fixed to 10, which empirically has been proved to be a good choice for the bias-variance trade-off (James et al., 2014). The 10-fold cross validation was performed using the *caret* package for R (Kuhn, 2008). For each of the four models, the dataset was randomly shuffled and partitioned into10 equally sized folds. Iteratively, nine of them were used as training data, and the remaining one was set for validation testing. Overall, for the two models fitted for the two-word sentences, the resulting average accuracy across 10 folds was respectively 83.51% and 82.50%, which are indicative of a good model fit. Similarly, the predictive accuracy computed for the two models of the one-word sentences were respectively 86.32% and 81.67%, suggesting that the predictions made in our study statistically are very reliable.

Further, we plotted the ROC curves (Receiver Operating Characteristic Curve) for the best training set to estimate how well the models discriminate the true and the false answers (R code is available in the Online Appendix B). Meanwhile, the area under the curve (AUC) was calculated and annotated in the plot using the *train* function from the *caret* package to give a clear assessment of the model's discriminative performance (Kuhn, 2008). According to previous literature, an AUC area of 0.5 indicates a poor random fit, while an area greater than 0.8 was considered to be an excellent model (Zhu et al., 2010). In our case, all the GLM models were calculated to have an AUC area greater than 0.8, which implies that our models were highly effective in classifying the binary outcome variable. Specifically, Figure 6 shows that the AUC area of the two models designed for predicting the perception of the one-word sentences were 0.8952 and 0.9056, respectively, while that of the two models fitted for the two-word sentences were respectively 0.8235 and 0.8160. Noticeably, we displayed the ROC plot only for the models fitted for the one-word sentences, because they shared roughly comparable tendency curves with the models fitted for the sentences with two stressed words.

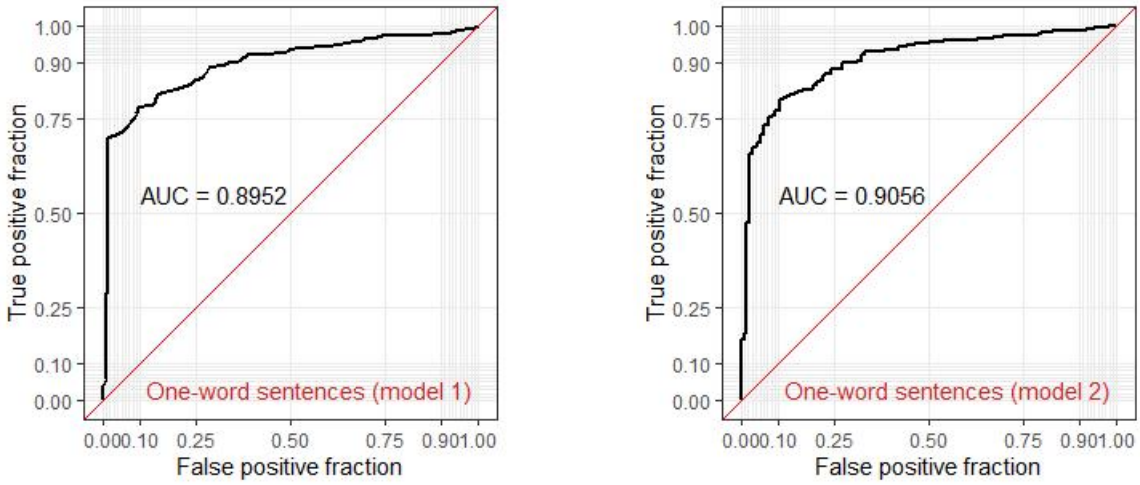


Figure 6. ROC curves for the regression models fitted for the one-word sentences depending on L1 (model 1) and individual features (model 2).

4 Discussion

The goal of this study is to examine the cross-linguistic perception of Spanish intonation by Chinese tonal speakers, and determine the factors that guide listeners' subjective judgments of the perceived

sentence type, as well as predict the relative perceptual difficulties of Chinese listeners compared to the native language users. To this end, we formulated three main research questions (as detailed in Section 1.3) to test whether and how the different intonational cues of the utterance, the native tone language experience, and various individual features mediate the perceptual processing of Spanish statements and questions. We began by analyzing the role of different intonational cues and the native language experience in perceiving the utterances with two-stressed words. Then, we examined the contribution of those factors that reflect the individual features to the likelihood that the type of the sentence is accurately identified. A similar analysis process was conducted for the utterances with one stressed word to further examine the cue strength of the final F0 contour. Finally, the statistical validity of the models was confirmed using two mathematical methods. The findings of our study in relation to the three research questions are discussed below.

We start to consider the first research question and argue primarily that the native and the non-native perception of Spanish sentence type was strongly influenced by the different intonational cues involved in the stimuli. Evidence for this is that both Chinese and Spanish listeners in our study progressively improved their perceptual accuracy by accumulating the number of gates in the utterance. This is highly consistent with the studies of Face (2005, 2007, 2011) that all intonational cues to sentence type contribute at least to some extent to the perception of Spanish statements and yes/no questions. However, as pointed out in previous cross-linguistic studies (e.g., Trimble, 2013; Li, 2020; Zarate-Sandez et al., 2015), native and non-native listeners differed in the ability to use the distinct intonational cues to help increase the likelihood that a given utterance is correctly perceived. A similar scenario was found in our data as well, since the prediction of the perceived sentence type was significantly determined by the interaction between the L1 group and gate, particularly in sentences with two stressed words. In general, the results of the experiment revealed an overall higher perceptual accuracy for the native listeners across every gate of the utterance compared with the non-native speakers. However, it is critical to note that the native and the non-native group differed significantly only in perceiving the initial peak and the F0 downslope during the nuclear syllable, while in the first and the final gate the two L1 groups had a statistically comparable performance in the two-word sentences. Similar difficulties of foreign listeners in perceiving the Spanish intonation were also observed in Trimble (2013) and Li (2020), and they probably were related to the different intonational forms embedded in learners' native prosodic system.

The perceptual deviation of Chinese listeners mostly was attributed to the fact that F0 contours in a tone language such as Mandarin are used primarily for encoding lexical meanings rather than for discriminating intonation types (J. Yuan, 2011). Therefore, Chinese speakers do not necessarily rely on the variation of pitch contours (e.g., initial peak height, final pitch movement) to make clear the sentence type (Shang et al., in press). But despite the typological difference in intonation mechanism, Chinese listeners in our study achieved roughly 67% accuracy rate as soon as they were presented with the first F0 peak. The result is perhaps surprising in comparison with the findings of Trimble (2013), which showed that more than half (58%) of the native English listeners failed to distinguish Spanish sentence types without the final pitch movement. However, and answering the second research question, the superior performance of Chinese learners is expected and could be reasonably explained by a positive transfer from L1 Mandarin. As shown in Section 1.3.1, listeners with a tonal language background were more sensitive than the non-tone speakers in perceiving the contrastive tonal features in native or non-native languages (Meng et al., 2020; Ortega-Llebaria et al., 2017; Ortega-Llebaria & Wu, 2021; Wiener & Bradley, 2020). Moreover, evidence from the psychophysiological area suggested that tone language speakers were similar to the musicians in the sense that they shared a neural architecture for processing signals, and therefore, were associated with higher auditory perceptual performance and enhanced general cognitive abilities than the non-tone speakers and nonmusicians in the perception of natural speech or complex music (Bidelman et al., 2013). This similarity would also explain why Chinese tonal listeners in our experiment

outperformed the English speakers of Spanish in previous parallel studies (Trimble, 2013).

Although the F0 initial peak and the F0 downslope are of great importance in perceiving the two-word sentences, it is only when the final pitch movement was presented that Chinese listeners could achieve an accuracy level similar to that of the native speakers. The significant improvement of Chinese listeners after hearing the final gate corroborates previous findings, showing that the final rise/fall was a strong intonational cue for the non-native speakers in disencoding the sentence type. As for Spanish listeners, it could seem that the final F0 cue plays a minor role in discriminating the utterances with more than one word since the native group reached a nearly complete accuracy without being presented with the final pitch movement. Nonetheless, by using the one-word stimuli, we refute this false impression caused by the significant influence of earlier cues, demonstrating that the utterance-final F0 movement was the greatest contributing factor to the likelihood that a given utterance is accurately recognized, either for native or non-native speakers of Spanish. This finding is in line with the results in Face (2007), which showed that the final F0 pattern was the strongest cue for signaling the sentence type that had the strength to change listeners' perceptual decisions made with earlier contradictory cues. More generally, the consistent cue weighting of the final F0 movement between listeners of different L1 seems to provide evidence for the cross-linguistic claim that there should be some language-universal effects on intonation forms and the pragmatic meanings they convey in human speech (Face, 2007; Gussenhoven & Chen, 2000; Gussenhoven, 2004).

In addition to the perceptual similarities and dissimilarities between the two L1 groups, the results of our experiment also revealed two interesting findings in relation to the linguistic properties of the stimuli. For one thing, we found that sentences ending with a paroxytone word were more likely to be perceived correctly than the ones ending with an oxytone word before presenting the final syllable of the utterance. As explained earlier, the perceptual difference probably is attributed to that the first gate of the paroxytone word was divided including the stressed syllable, while that of the oxytone word did not include it. In other words, the paroxytone word had a distinctive cue in the first gate to help increase its perceptual accuracy, since the statement in Peninsular Spanish is described to have an F0 rise during the final stressed syllable, as opposed to the yes/no question that is realized with a low F0 throughout this syllable (Face, 2007). However, the effect of stress type was apparent only for the one-word sentences (given that the nuclear pitch accent is L+H* for statements and L* for questions), while in the two-word utterances no significant difference of accuracy was observed depending on the stress type (because the nuclear pitch accent is L* for both questions and statements). A similar finding was also noticed in Face (2005), which showed that the F0 contour of the final stressed syllable was a weaker cue for discriminating the utterances with three stressed words, but it was of great importance for rating the naturalness of yes/no questions. For another, in line with earlier works on intonation perception (Face, 2005, 2007; Trimble, 2013b; Li, 2020), our study revealed that listeners generally were biased towards identifying stimuli as statements, particularly when there is no apparent interrogative cue involved in the gate. This parallel inclination between native and non-native listeners could be interpreted as a preference for unmarked forms in the communicative interaction (Face, 2007; Grosser, 2011). Since the statement is the most neutral and unmarked modality that carries the least communicative burden in most human languages, it would be, of course, more easily to be recognized than questions in cross-linguistic speech.

As regards the third research question, our study suggests that the perception performance of Chinese listeners was not solely determined by their native language experience, but also influenced by a variety of listener-specific factors related to language proficiency, pragmatic skill, and musical expertise. Considering firstly the proficiency, results of our experiment revealed a positive relationship between the perceptual ability and the proficiency in Spanish, showing that highly proficient Chinese listeners were more accurate in the categorical perception of Spanish intonational contrasts compared with those with lower proficiency. This finding is also supported by previous

neurobehavioral studies (Richards et al., 2018), which claimed that high proficiency could go hand in hand with greater activity in some later-developing brain regions related to higher-order cognition, such as attention and other abilities that may help to process the non-native speech. Moreover, the strong interaction between proficiency in Spanish and the number of gates in the two-word sentences, indicated that the effect of proficiency was highly dependent on the perceived F0 cues, and it was particularly apparent when Chinese speakers listen to the contrastive structures that were not present in their native category, for instance, the prenuclear peak and the F0 down slope during the final stressed syllable. Overall, from these findings, we can infer a developing intonation perception toward a more native-like system with increasing foreign listeners' proficiency which primarily is acquired through years of experience in natural linguistic environments.

Another individual property that considerably affects the cross-linguistic intonation perception is related to the AQ score, which has been used as a measurement of the subject's pragmatic skills in the literature (Bishop et al., 2020). Since the AQ is inversely associated with the pragmatic skill, Chinese listeners who were higher in the AQ score had poorer pragmatic skills and were less accurate in discriminating the target intonation patterns. The finding that non-native listeners' perception was strongly predicted by their pragmatic ability corroborates some recent perceptual studies focused on the individual differences (Bishop, 2012, 2017; Bishop et al., 2020; Jun & Bishop, 2015). For instance, Bishop et al. (2020) demonstrated that native English listeners with poorer pragmatic skills displayed a lower likelihood of perceiving the prominence patterns. Although the specific mechanism behind this individual variation is unclear to the researchers so far, it is at least confirmed that there should be a correlation between pragmatic skills and prosodic sensitivity (Bishop, 2017; Bishop et al., 2020).

Furthermore, our study reveals significant individual differences of musical experience in cross-linguistic speech perception. Chinese listeners who had music training experience were more accurate in the intonation processing of both one- and two-word utterances than those untrained. In line with previous research on music-language interaction, our results suggest that the explicit musical practice can enhance the perceptual ability of listeners by increasing their sensitivity to the acoustic cues that are also relevant to music (e.g., pitch and duration). The apparent advantage of musicians, as claimed in many interdisciplinary studies (Besson et al., 2011; Bidelman et al., 2011; Kraus & Chandrasekaran, 2010; A. D. Patel, 2011; Wiener & Bradley, 2020), can be attributed to a direct transfer of music training effects. Because music and speech share a common process for the signal processing, that is, rely on the same pool of acoustic parameters and neural resources, listeners may extend their auditory skills and sensitivity developed in one domain to another (Besson et al., 2011; Wiener & Bradley, 2020). Finally, in contrast, we did not find a significant effect of age and gender on the cross-linguistic intonation perception, despite there being a tendency that Chinese listeners who were female or older in age were more likely to give an accurate perceptual answer.

In sum, while previous cross-linguistic studies between Chinese and Spanish focused on the important role that L1 background play, the results of our study seem to support the claim of a complex integration mechanism in speech perception where intonational knowledge, native language experience, as well as a wide range of internal features unique to each individual are significant factors for predicting listeners' perception performance (Bishop et al., 2020; Cole, Mahrt, et al., 2017; Cole, Mo, et al., 2010; Röhr et al., 2020; Turnbull, 2017).

5 Conclusion

The present study was set out to explore a variety of possible factors that may affect the cross-linguistic perception of Spanish intonation and predict the relative perceptual difficulties of the non-native Chinese listeners compared to the native Spanish speakers.

The overall results demonstrated, firstly, that listeners' knowledge of Spanish intonational contours was highly significant for the subjective judgments of Spanish sentence type, and it was shaped by listeners' native language experience as well as several individual features including the level of proficiency, pragmatic skills, and musical expertise. As expected, Chinese listeners differed from native Spanish speakers mostly in perceiving the intonational events that were not involved in their native prosodic system. However, in comparison with the speakers of non-tonal languages (like English) in previous studies (Trimble, 2013b), Chinese listeners achieved a more target-like perception of the prenuclear stretch thanks to their L1 tonal experience.

Furthermore, analyses of the performance of Chinese and Spanish listeners seem to support the theory of linguistic meanings based on universal biological codes (Chen, 2009; Chen, Gussenhoven, & Rietveld, 2004), showing that there may be both universal and language-specific effects in the perception of intonation.

Additionally, the present study provides details for using the generalized logistic regression technique and two mathematical methods for evaluating the model performance, which may have important implications for the data modeling of binary outcome variables in perceptual research.

Funding

This work is supported by a Ph.D. research grant provided by China Scholarship Council (No.201808390059) to the first author.

ORCID iD

Peizhu Shang. https://orcid.org/0000-0001-5407-4476
Paolo Roseano. https://orcid.org/0000-0002-9743-9735
Wendy Elvira-García. https://orcid.org/0000-0001-7002-9851

Supplemental material

Supplemental material for this article is available online.

References

Alexander, J. A., Wong, P. C. M., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. In *9th European Conference on Speech Communication and Technology*, (pp. 397–400). 4–8 September 2005, Lisbon, Portugal.

Alexander, J., & Wang, D. Y. (2016). The perceptual assimilation model for suprasegmentals and cross-language lexical-tone identification. *The Journal of the Acoustical Society of America*, *139*(4), 2161.

Archila-Suerte, P., Zevin, J., Bunta, F., & Hernandez, A. E. (2012). Age of acquisition and proficiency in a second language independently influence the perception of non-native speech. *Bilingualism: Language and Cognition*, *15*(1), 190–201.

Archila-Suerte, P., Zevin, J., & Hernandez, A. E. (2015). The effect of age of acquisition, socioeducational status, and proficiency on the neural processing of second language speech sounds. *Brain and Language*, *141*, 35–49.

Archila, P., Ramos, A. I., Zevin, J., & Hernandez, A. E. (2010). How age of acquisition and proficiency predicts sound detection in bilinguals. In *Poster presented at Armadillo Conference.* College Station, Texas, USA.

Baigorri, M., Campanelli, L., & Levy, E. S. (2019). Perception of American – English Vowels by Early and Late Spanish – English Bilinguals. *Language and Speech*, *62*(4), 681–700.

Baills, F., Zhang, Y., Cheng, Y., Bu, Y., & Prieto, P. (2021). Listening to Songs and Singing Benefitted Initial Stages of Second Language Pronunciation but Not Recall of Word Meaning. *Language Learning*, *71*(2), 369–413.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251.

Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America*, *139*(3), EL51–EL56.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology*, *2*(94), 1–12.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*(3), 305–330.

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception. In M. Murry J. & B. Ocke-Schwen (Eds.), *Language Experience in Second Language Speech Learning: In Honour of James Emil Flege* (pp. 13–34). John Benjamins.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, *23*(2), 425–434.

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PloS One*, *8*(4), e60676.

Bishop, J. (2012). Focus, prosody, and individual differences in "autistic" traits: Evidence from cross-modal semantic priming. *UCLA Working Papers in Phonetics*, *111*, 1–26.

Bishop, J. (2017). Focus projection and prenuclear accents: Evidence from lexical processing. *Language, Cognition and Neuroscience*, *32*(2), 236–253.

Bishop, J., Kuo, G., & Kim, B. (2020). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics*, *82*, 100977.

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer programm]. Version 5.3.82. Retrieved from http://www.praat.org/

Borràs-Comes, J., Sichel-Bazin, R., & Prieto, P. (2015). Vocative intonation preferences are sensitive to politeness factors. *Language and Speech*, *58*(1), 68–83.

Bradley, E. D. (2016). Phonetic dimensions of tone language effects on musical melody perception. *Psychomusicology: Music, Mind, and Brain*, *26*(4), 337.

Burnham, D. K., & Brooker, R. (2002). Absolute pitch and lexical tones: Tone perception by non-musician, musician, and absolute pitch non-tonal language speakers. *Seventh International Conference on Spoken Language Processing*, 257–260.

Chang, D., Hedberg, N., & Wang, Y. (2016). Effects of musical and linguistic experience on categorization of lexical and melodic tones. *The Journal of the Acoustical Society of America*, *139*(5), 2432–2447.

Chang, Y. S., Yao, Y., & Huang, B. H. (2017). Effects of linguistic experience on the perception of high-variability non-native tones. *The Journal of the Acoustical Society of America*, *141*(2), EL120–EL126.

Chen, A. (2009). Perception of paralinguistic intonational meaning in a second language. *Language Learning*, *59*(2), 367–409.

Chen, A., Gussenhoven, C., & Rietveld, T. (2004). Language-specificity in the perception of paralinguistic intonational meaning. *Language and Speech*, *47*(4), 311–349.

Chen, J. K.-C., Chuang, A. Y. C., McMahon, C., Hsieh, J.-C., Tung, T.-H., & Li, L. P.-H. (2010). Music training improves pitch perception in prelingually deafened children with cochlear implants. *Pediatrics*, *125*(4), e793–e800.

Cole, J., Mahrt, T., & Roy, J. (2017). Crowd-sourcing prosodic annotation. *Computer Speech & Language*, *45*, 300–325.

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, *1*(2), 425–452.

Cooper, A., Wang, Y., & Ashley, R. (2017). Thai rate-varied vowel length perception and the impact of musical experience. *Language and Speech*, *60*(1), 65–84.

Cortés Moreno, M. (1997). Sobre la percepción y adquisición de la entonación española por parte de hablantes nativos de chino. *Estudios de Fonética Experimental*, *9*, 67–134.

Cortés Moreno, M. (2001). Percepción y adquisición de la entonación española en enunciados de habla espontánea: el caso de los estudiantes taiwaneses. *Estudios de Fonética Experimental*, *11*, 90–119.

Deguchi, C., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., & Colombo, L. (2012). Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: Behavioral, electrophysiological and psychoacoustic study. *Brain Research*, *1455*, 75–89.

Escudero, P., Simon, E., & Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics*, *40*(2), 280–288.

Estebas-Vilaplana, E., & Prieto, P. (2010). Castilian Spanish intonation. In P. Prieto & P. Roseano (Eds.), *Transcription of intonation of the Spanish language* (pp. 17–48).

Face, T. L. (2004). The intonation of absolute interrogatives in Castilian Spanish. *Southwest Journal of Linguistics*, *23*(2), 65–80.

Face, T. L. (2005). F0 Peak Height and the Perception of Sentence Type in Castilian Spanish. *Revista Internacional de Lingüística Iberoamericana*, *3*(2), 49–65.

Face, T. L. (2007). The role of intonational cues in the perception of declaratives and absolute interrogatives in Castilian Spanish. *Estudios de Fonética Experimental*, *16*, 186–225.

Face, T. L. (2008). *The intonation of Castilian Spanish declaratives and absolute interrogatives*. Lincom Europa.

Face, T. L. (2011). *Perception of Castilian Spanish Intonation: Implications for Intonational Phonology*. Lincom Europa.

Face, T. L., & Prieto, P. (2007). Rising accents in Castilian Spanish: A revision of Sp_ToBI. *Journal of Portuguese Linguistics*, *6*(1), 117–146.

Félix-Brasdefer, J. C. (2010). Data collection methods in speech act performance. *Speech Act Performance: Theoretical, Empirical and Methodological Issues*, *26*(41), 69–82.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *The Journal of the Acoustical Society of America*, *95*(6), 3623–3641.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience*, *16*(6), 1010–1021.

Fuller, C. D., Galvin III, J. J., Maat, B., Free, R. H., & Başkent, D. (2014). The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience*, *8*, 179.

Gareth James, Trevor Hastie, Robert Tibshirani, D. W. (2014). *An Introduction to Statistical Learning: With Applications in R By*. Springer.

Grabe, E., Rosner, B. S., García-Albea, J. E., & Zhou, X. (2003). Perception of English Intonation by English , Spanish , and Chinese Listeners. *Language and Speech*, *46*(4), 375-401.

Graham, C., & Post, B. (2018). Second language acquisition of intonation: Peak alignment in American English. *Journal of Phonetics*, *66*, 1–14.

Grosser, W. (2011). On the acquisition of tonal and accentual features of English by Austrian learners. In A. Leather & J. & J (Eds.), *Second-language speech* (pp. 211–228). De Gruyter Mouton.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

Gussenhoven, C. (2002). Intonation and Interpretation : Phonetics and Phonology. *International Conference on Speech Prosody 2002*, 47–57.

Gussenhoven, C., & Chen, A. (2000). Universal and language-specific effects in the perception of question intonation. In *6th International Conference on Spoken Language Processing* (pp. 91–94). 16–20 October 2000, Beijing, China

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421.

Harrell Jr, M. F. E., & Hmisc, D. (2021). Package 'rms.' *R package version 6.2-0.* Retrived from https://cran.r-project.org/web/packages/rms/rms.pdf

Hualde, J. I., & Prieto, P. (2015). Intonational variation in Spanish: European and American varieties. In F. Sonia & P. Pilar (Eds.), *Intonation in Romance* (pp. 350–391). Hardcover.

Jiang, J., Liu, F., & Wan, X. (2015). Perception of Melodic Contour and Intonation in Autism Spectrum Disorder : Evidence From Mandarin Speakers. *Journal of Autism and Developmental Disorders*, *45*(7), 2067–2075.

Jun, S.-A., & Bishop, J. (2015). Priming implicit prosody: Prosodic boundaries and individual differences. *Language and Speech*, *58*(4), 459–473.

Kilman, L., Zekveld, A., Hällgren, M., & Rönnberg, J. (2014). The influence of non-native language proficiency on speech perception performance. *Frontiers in Psychology*, *5*(651), 1–9.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605.

Kraut, R., & Wulff, S. (2013). Foreign-accented speech perception ratings: A multifactorial case study. *Journal of Multilingual and Multicultural Development*, *34*(3), 249–263.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26.

Lau, W. Y. P., Gau, S. S. F., Chiu, Y. N., Wu, Y. Y., Chou, W. J., Liu, S. K., & Chou, M. C. (2013). Psychometric properties of the Chinese version of the Autism Spectrum Quotient (AQ). *Research in Developmental Disabilities*, *34*(1), 294–305.

Lee, G., Shin, D.-J., & Garcia, M. T. M. (2019). Perception of lexical stress and sentence focus by Korean-speaking and Spanish-speaking L2 learners of English. *Language Sciences*, *72*, 36–49.

Lee, H. Y., & Song, J. (2019). Evaluating Korean learners' English rhythm proficiency with measures of sentence stress. *Applied Psycholinguistics*, *40*(6), 1363–1376.

Lenth, R., Singmann, H., & Love, J. (2019). Emmeans: Estimated marginal means, aka least-squares means. *R package version 1.3.4*. Retrieved from https://cran.r-project.org/web/ packages/ emmeans/ emmeans.pdf

Li, A., & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German learners of English. *Studies in Second Language Acquisition*, *36*(2), 223–255.

Li, W. (2020). *Producción y percepción de la entonación del español por los sinohablantes: patrones entonativos prototípicos de la declarativa neutra, la interrogativa absoluta neutra y la declarativa con foco estrecho* (Unpublished doctoral dissertation). Department of Spanish Philosophy, Universitat Autónoma de Barcelona, Spain.

Liu, C., & Rodriguez, A. (2012). Categorical perception of intonation contrasts: Effects of listeners' language background. *The Journal of the Acoustical Society of America*, *131*(6), EL427–EL433.

Luchkina, T., & Cole, J. S. (2021). Perception of Word-level Prominence in Free Word Order Language Discourse. *Language and Speech*, *64*(2), 381–412.

Marques, C., Moreno, S., Luís Castro, S., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *19*(9), 1453–1463.

Meng, Y., Zhang, J., Liu, S., & Wu, C. (2020). Influence of different acoustic cues in L1 lexical tone on the perception of L2 lexical stress using principal component analysis: an ERP study. *Experimental Brain Research*, *238*(6), 1489–1498.

Mengrong, L. (2008). Chinese Version of the Autism Spectrum Quotient for Adults: Predictive Validity and Correlation Analysis. *Journal of Special Education Research*, *33*(1), 73–92.

Nan, Y., Liu, L., Geiser, E., Shu, H., Gong, C. C., Dong, Q., Gabrieli, J. D. E., & Desimone, R. (2018). Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proceedings of the National Academy of Sciences*, *115*(28), E6630–E6639.

Nguyễn, T. A.-T., Ingram, C. L. J., & Pensalfini, J. R. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics*, *36*(1), 158–190.

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, *63*(3), 324–346.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*(1), 1–16.

Ortega-Llebaria, M., Nemogá, M., & Presson, N. (2017). Long-term experience with a tonal language shapes the perception of intonation in English words: How Chinese-English bilinguals perceive "Rose?" vs. "Rose." *Bilingualism*, *20*(2), 367–383.

Ortega-Llebaria, M., & Wu, Z. (2020). Chinese-English Speakers' Perception of Pitch in Their Non-Tonal Language: Reinterpreting English as a Tonal-Like Language. *Language and Speech*, *January*. https://doi.org/10.1177/0023830919894606

Ortega-Llebaria, M., & Wu, Z. (2021). Chinese-English Speakers' Perception of Pitch in Their Non-Tonal Language: Reinterpreting English as a Tonal-Like Language. *Language and Speech*, *64*(2), 467–487.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, 142.

Shang, P., Elvira-García, W., & Roseao, P. (in press). La modalidad interrogativa en español y en chino: un enfoque funcionalista. *Círculo de Lingüística Aplicada a La Comunicació*.

Richards et al. (2018). Age of acquisition and proficiency in a second language independently influence the perception of non-native speech. *Physiology & Behavior*, *176*(5), 139–148.

Röhr, C. T., Brilmayer, I., Baumann, S., Grice, M., & Schumacher, P. B. (2020). Signal-driven and expectation-driven processing of accent types. *Language, Cognition and Neuroscience*, 1–27.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music facilitates pitch processing in language. *Psychophysiology*, *41*, 341–349.

Sensui, H. (1995). Percepción de la entonación interrogativa del español: un estudio experimental. *Sophia Linguistica*, *38*, 1–23.

Sensui, H. (2003). A pilot case study on sentence pattern perception of Spanish. *PoLAR*, 1731–1734.

Smiljanić, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *The Journal of the Acoustical Society of America*, *130*(6), 4020–4031.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273–293.
So, C. K., & Best, C. T. (2014). Phonetic influences on english and french listeners' assimilation of mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*(2), 195–221.
Tanaka-Ishii, K., & Terada, H. (2011). Word familiarity and frequency. *Studia Linguistica*, *65*(1), 96–116.
R Core Team. (2020). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria: R Foundation for Statistical Computing. http://softlibre.unizar.es/manuales/aplicaciones/r/fullrefman.pdf/
The Autism Research Centre. (2010). *Coeiciente del Espectro Autista para Adultos (AQ) Más de 16 años* (pp. 1–5). Cambrige, UK, University of Cambrige. Avalible from http://espectroautista.info/AQ-es.html
Trimble, J. C. (2013a). *Acquiring Variable L2 Spanish Intonation in a Study Abroad Context* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis, USA.
Trimble, J. C. (2013b). Perceiving Intonational Cues in a Foreign Language : Perception of Sentence Type in Two Dialects of Spanish. In C. Howe, S. E. Blackwell, & M. L. Quesada (Eds.), *Proceedings of the 15th Hispanic Linguistics Symposium* (pp. 78–92). 6–9 October 2013, University of Georgia, USA.
Tsukada, K. (2018). The Perception of Mandarin Lexical Tones by Native Speakers of Burmese. *Language and Speech*, *62*(4), 1–16.
Turnbull, R. (2017). The role of predictability in intonational variability. *Language and Speech*, *60*(1), 123–153.
van Kesteren, M. T. R., & Wiersinga-Post, J. E. C. (2007). Auditory temporal-order thresholds show no gender differences. *Restorative Neurology and Neuroscience*, *25*(2), 119–122.
Vickie, Y. Y., & Andruski, J. E. (2010). A cross-language study of perception of lexical stress in English. *Journal of Psycholinguistic Research*, *39*(4), 323–344.
Wiener, S., & Bradley, E. D. (2020). Harnessing the musician advantage: Short-term musical training affects non-native cue weighting of linguistic pitch. *Language Teaching Research*, November. https://doi.org/10.1177%2F1362168820971791
Wiener, S., & Goss, S. (2018). Perceptual assimilation of non-native prosodic cues: Cross-linguistic effects of lexical F0 learning. In *9th International Conference on Speech Prosody* (pp. 947–951). 13–16 June 2018, Poznań, Poland.
Yang, X., Minai, U., & Fiorentino, R. (2018). Context-sensitivity and individual differences in the derivation of scalar implicature. *Frontiers in Psychology*, *9*, 1720.
Yu, A. C. L. (2010). Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PLoS ONE*, *5*(8).
Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America*, *130*(6), 4063–4069.
Zarate-Sandez, G. A., Morales-Front, A., & Sanz, C. (2015). *Perception and production of intonation among English-Spanish bilingual speakers at different proficiency levels* (Doctoral dissertation). Department of Spanish and Portuguese, Georgetown University, USA.
Zhang, J., & Lee, S. (2019). Acquisition of English speech rhythm by Chinese learners of English at different English proficiency levels. *Phonetics and Speech Sciences*, *11*(4), 71–79.
Zhang, S., Li, K., Lo, W.-K., & Meng, H. (2010). Perception of English suprasegmental features by non-native Chinese learners. In *5th International Conference on Speech Prosody* (pp. 1–4). 11–14 May 2010, Doubletree Magnificent Mile, Chicago, USA.
Zhao, T. C., & Kuhl, P. K. (2015). Effect of musical experience on learning lexical tone categories. *The Journal of the Acoustical Society of America*, *137*(3), 1452–1463.

Zhu, W., Zeng, N., & Wang, N. (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. In *NESUG Proceedings: Health Care and Life Sciences* (pp. 1–9). 14-17 November 2010, Baltimore, Maryland.

Zuk, J., Ozernov-Palchik, O., Kim, H., Lakshminarayanan, K., Gabrieli, J. D. E., Tallal, P., & Gaab, N. (2013). Enhanced syllable discrimination thresholds in musicians. *PloS One*, *8*(12), e80546.

Anexo B

Dynamic cue weighting in the perception of Spanish intonation contrasts: Differences between tonal and non-tonal language listeners

Citado como:

Shang, P.; Roseano, P. & Elvira-García, W. (en revisión). Dynamic cue weighting in the perception of Spanish intonation contrasts: Differences between tonal and non-tonal language listeners.

Article Cover Page

Title: Dynamic cue weighting in the perception of Spanish intonation contrasts: Differences between tonal and non-tonal language listeners

Peizhu Shang[a]*, Paolo Roseano[b,c], Wendy Elvira García[a]

*Corresponding author: pshangsh7@alumnes.ub.edu, Tel: +34 644 326 548
[a] Universitat de Barcelona, Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain
[b] Universidad de Educación a Distancia, C. de Bravo Murillo 38, 28015 Madrid, Spain
[c] University of South Africa, Preller Street, Muckleneuk Ridge, 0002 Pretoria, South Africa

Abstract
Research shows that listeners from different language backgrounds use multiple acoustic cues to identify phonemic contrasts and sound changes. However, cue weighting differences between tonal and non-tonal languages in intonation perception remain relatively unexplored. This study aimed to fill this gap by examining how Chinese- and Spanish- native listeners use three acoustic cues (f0, duration, and intensity) to perceive question-statement contrasts in Spanish. To this end, two identification tests were conducted using synthesized stimuli, in which the f0 contour covaried with duration or intensity in sentence-final syllables. Whereas changes in the three acoustic properties significantly affected Spanish listeners' performance, Chinese listeners were only sensitive to f0 and duration modulations. Crucially, this sensitivity to f0 and duration cues was higher for Spanish listeners; however, they required significantly higher f0 contours to identify utterances as questions, compared with Chinese speakers. Moreover, the trade-off between primary and secondary cues revealed that intonation perception is a dynamic process wherein listeners can take advantage of the covariations of multiple cues to recognize the intended phonological category. The findings also showed that intonation perception was mainly influenced by native language experience; however, it was modulated by individual differences and stress patterns.

Keywords
Intonation perception, acoustic cue weighting, tonal and non-tonal languages, phonetic trading relation, individual differences

Dynamic cue weighting in the perception of Spanish intonation contrasts: Differences between tonal and non-tonal language listeners

1.0 Introduction

Intonation, a suprasegmental feature of speech, can be used for several purposes, such as signaling sentence types and focusing differences in a linguistically structured manner (Ladd, 1996; Levis, 1999) as well as conveying emotional and attitudinal meanings in the socioculturally oriented interface between prosody and pragmatics (Rodero, 2011; Wichmann & Blakemore, 2006). In any case, intonation is realized using multidimensional acoustic properties. The continuous values along each acoustic dimension can be used by listeners as cues to recognize a particular linguistic category (cf. Kong & Edwards, 2016; Kuang & Cui, 2018, 2019; Peng et al., 2012; Toscano & McMurray, 2010). While listeners can take advantage of multiple acoustic cues to identify the linguistic category, the contribution or weight of these cues is different to listeners' assessment because their attention is selective (Clayards, 2018; Holt & Lotto, 2006; Kuang & Cui, 2018; Tillman et al., 2017; Toscano & McMurray, 2010).

Furthermore, cross-linguistic research shows that listeners from different language backgrounds may exhibit differences in their cue weighting paradigms for a phonological contrast (cf. Dmitrieva, 2019, for stop-voicing contrast by native and Russian speakers of English; Hattori & Iverson, 2005, for /r/-/l/ contrasts by native and Japanese speakers of English; and Lee et al., 2013, for voiceless stops by speakers of two Korean dialects) depending on the influence of their prior linguistic experience and native phonetic and phonological conventions. Nonetheless, most cue-weighting studies have focused on the perception of segmental features between first (L1) and second (L2) language speakers of English; thus, other language pairs remain relatively unexplored in the suprasegmental domain. Therefore, this study attempts to bridge this gap by examining how L1 and Chinese L2 speakers of Spanish—a new pairing between tonal and non-tonal languages—rely on multiple acoustic cues to perceive Spanish intonation contrasts, and how these cues interact and combine during listeners' perceptual process.

1.1 Multiple cue weighting in intonation perception

Question-statement contrast can be encoded by several resources, other than intonation contour. For instance, in Peninsular Spanish, yes-no questions differ from statements in terms of the word order: questions follow a verb-subject-object order, whereas statements follow a subject-verb-object order (Haverkate, 2006). In Mandarin Chinese, the best-known characteristic of yes-no questions is the presence of the final particle *ma* (Yuan, 2011). However, although syntactic or lexical mechanisms can encode sentence types, intonation alone can achieve the same effect. Questions that differ from statements only in their intonation contour are known as "intonation questions," which are of particular interest in this study.

Similar to other perceptual domains, multiple sources of information must be combined in the perception of intonation. For instance, normal-hearing English listeners can use variations in f0, intensity, and duration to identify question-statement contrasts, depending on listening conditions, stimulus type, and the chronological age of the participants (Peng et al., 2012). Similar integration of acoustic cues for perceiving intonation categories has been observed in native listeners of Mandarin Chinese (Yuan, 2006), Cantonese (Ma et al., 2008, 2011), German (Niebuhr, 2007), and South Korean (Chang, 2013). However, Lieberman (1967) noted that most English listeners could effectively recognize the difference between questions and statements based solely on the f0 contour during the final 150 to 200 ms of an utterance. Face (2005, 2007) reported similar findings for Madrid Spanish, in which the final pitch movement was the most reliable cue for signaling the question-statement contrasts that could override any previous contradictory cue. The importance of f0 pitch as the main cue for questionhood encoding was also confirmed by Bolinger (1978), who analyzed roughly 250 languages and found that 70% of

human languages use a final pitch rise to mark questions; the rest, by contrast, may rely on the increase of the overall pitch level for question encoding. Spanish is a good example of the first category, in contrast to Mandarin Chinese, a typical tonal language, that corresponds to the second category and differs from intonational languages in exhibiting a tone-dependent intonation mechanism. That is, provided there is no particle at the end of sentences, the lexical tone type of the final word can influence intonation in Chinese. For instance, Yuan (2006, 2011) showed that the question intonation in Chinese was perceived more accurately by L1 listeners if the utterance-final tone was falling (Tone4); conversely, this was harder to recognize if there was a rising tone at the end (Tone2).

Although intonation is perceived primarily as a function of the f0 modulation in tonal and non-tonal languages, other secondary cues, such as duration and intensity, can also influence listeners' judgments by conveying important, even if less reliable, acoustic information (Ortega-Llebaria et al., 2017). Feng et al. (2019) reported that an increase in the duration and intensity of the sentence-final word significantly contributed to the likelihood that a given sentence was perceived as a question in American English. A similar duration and intensity effect is expected in the two language groups examined in this study, as intonation questions in Spanish and Chinese have a longer duration and a higher intensity in the vocalic part of the sentence-final word (cf. Romera Barrios et al., 2007, for Barcelona Spanish, and Yuan, 2006, for Mandarin Chinese). However, the perceptual weighting of these acoustic cues (i.e., f0, duration, and intensity) may differ between L1 and L2 listeners from different language backgrounds. Studies supporting this notion found that Chinese learners of English could use the intensity of the final word as a secondary cue to signal question-statement contrasts in English, whereas English L1 speakers ignored the intensity changes when the f0 contour was strongly represented (Morrow & Liu, 2013). This finding for English L1 listeners is in line with the results of Peng et al. (2012). By contrast, Feng et al. (2019) reported that English L1 listeners were significantly affected by duration and intensity variations in question-statement recognition, whereas Chinese L2 learners were not. Although it remains unclear why such studies reached different results with the same experimental approach, the small samples used in their research (5 and 14 participants, respectively) may have affected the robustness of their findings.

1.2 Debate on the tonal language benefit in f0 perception

Given that Mandarin Chinese is a tonal language, its native speakers are constantly exposed to f0 modulations at the lexical-tone and sentence-intonation level (Liu & Pell, 2012). The strong informational emphasis of f0 pitch in Mandarin has led some researchers to explore whether Chinese listeners have enhanced behavioral sensitivity to the processing of pitch events, compared with speakers of non-tonal languages. Chang et al. (2017) found that Mandarin listeners exhibited higher identification rates for high-variability Cantonese-level tones, compared with English listeners, because they had been more frequently exposed to lexical tones. Deroche et al. (2019), Ortega-Llebaria et al. (2017), and Xu et al. (2006) showed that Chinese listeners had a higher sensitivity to f0 direction changes in English words and detected f0 mismatches more rapidly, compared with non-tonal language listeners. Similar findings were reported by Bidelman et al. (2013), who indicated that tonal language (Cantonese) listeners and musicians shared enhanced perceptual and cognitive abilities in lower-level pitch perception, thereby outperforming English speakers in processing the f0 information necessary for basic auditory and music perception. Even though these studies have reported parallel findings, the hypothesis that speaking a tonal language increases listeners' sensitivity to pitch perception is yet to be verified. For instance, Chien et al. (2020), Liang and Heuven (2007), So and Best (2010), and Tsukada (2018) provided counterexamples to this hypothesis, demonstrating that speakers with tone language experience were not necessarily better at perceiving non-native lexical tones or native intonation contrasts, compared with non-tonal language listeners.

A possible explanation for these inconsistent results may be that the pitch enhancement exhibited by tonal language listeners depends on the internal f0 representation of the stimuli, and may not extend to all pitch dimensions. The f0 advantage of tonal language listeners is often present in perception tasks, where pitch information is processed in a manner consistent with the auditory demands of their native language (Bidelman et al., 2013; Deroche et al., 2019; Krishnan et al., 2009). This notion is supported by Bidelman et al.'s (2013) study, which showed that the higher discrimination ability shown by Cantonese listeners for perceiving large pitch incongruences, disappeared when perceiving subtle pitch deviations smaller than the f0 differences between Cantonese-level tones. Relative to the lexical tone perception, the higher complex pitch processing ability required by fine-grained f0 perception may explain the absence of tonal language's hypothesized advantages in this situation. Similarly, the lack of f0 sensitivity among Chinese listeners in some prior perception studies may be attributed to the lack of compatibility between the auditory skills necessary for curvilinear Mandarin tones and the expertise required for perceiving discrete music-level tones and linear intonation patterns. Thus, the f0 advantage of tonal language speakers is limited to linguistically relevant features of their native pitch events (with which they have broader experience), such as dynamic f0 contrastive patterns with particular curvatures and direction changes (Deroche et al., 2019).

Further, findings from neurobehavioral research support the notion of a tonal language benefit under certain pitch conditions by demonstrating that the neural substrates for processing higher- and lower-level pitches differ during perception (Doherty et al., 2004; Gandour, 2009; Zatorre & Gandour, 2008). More crucially, Chien et al. (2020) found that listeners from tonal and non-tonal languages had cross-linguistic commonalities and dissociations in f0 processing, depending on the type of pitch events perceived. Consistent with Friederici (2011) and Kreitewolf et al. (2014), Chien et al. (2020) revealed that pitch perceived as lexical tone overlapped with intonation in Mandarin and German listeners' left-brain regions, whereas it activated an additional semantic area only in tonal language speakers. By contrast, pitch processed as intonation activated bilateral brain areas and was generalized to tonal and non-tonal language listeners, regardless of their language-specific realization of intonation. Thus, the results seem to suggest that the neural enhancement of pitch encoding exhibited by tonal language speakers is present only in the perception of contrastive tone patterns, but not in the prosodic evaluation of intonation categories at the sentence level.

The potential f0 conflicts in the tone-dependent intonation system of Chinese may also explain the cross-linguistic differences in pitch perception reported in prior studies. Unlike intonational languages that use pitch primarily to signal intonation, the f0 contours in Chinese have the dual linguistic function of conveying both lexical tone and sentential intonation meanings (Lin, 2018). Thus, Chinese listeners may experience additional challenges in intonation processing when f0 is used simultaneously to encode information in the two dimensions (Yuan, 2011). Liang and Heuven (2007) provided supportive evidence for this view by showing that Mandarin L1 listeners were less sensitive to intonation but more sensitive to lexical tones, compared with L2 speakers of a non-tonal language. This result has been attributed to the fact that f0 is primarily perceived at the lexical level for tonal language listeners, ascribing their longer processing times and reduced sensitivity to pitch perceived at the sentence-intonation level (Liang & Heuven, 2007). Similar findings were observed by Chen (2005) and Gussenhoven and Chen (2000), who posited that listeners of non-tonal languages (Dutch and Hungarian) were more effective in using prosodic differences for question-statement identification of nonsensical stimuli because they were less bothered by the lexical tone interference, compared with Chinese listeners. However, findings of recent research on whether tonal language experience increases listeners' behavioral sensitivity to dynamic pitch perceived as intonation, remain controversial. Unlike the aforementioned studies, Feng et al. (2019) posited that Mandarin L2 speakers of English were more sensitive to f0 linear changes perceived in English question-statement intonation categories, compared with L1 speakers. However, the generalizability of their findings remains unverified due to the small sample size in their experiment.

1.3 The dynamic nature of speech perception

Speech perception is a dynamic process in which multiple cue sources are integrated to help recognize listeners' intended linguistic targets. The manner in which these cues interact and combine during this process is often described using phonetic trading relation—a phenomenon in which one acoustic cue's shift of values can be compensated by another cue's opposing change, such that the original percept is recovered (cf. Repp, 1982, for more details). One instance of this trade-off for intonation perception was reported by Peng et al. (2012), who found that English L1 listeners increased their reliance on duration and intensity cues to compensate for the f0 loss under spectrally degraded conditions. Similar compensatory behavior for weakened acoustic properties has been observed in the perception of whispered speech (Jiao & Xu, 2019) and many coarticulatory processes, such as vowel nasalization (Beddor & Krakow, 1999), voicing on initial stops (Repp, 1982), and ongoing sound changes (cf. tense vs. lax registers in Kuang & Cui, 2018). However, perceptual compensation only occurs among listeners who are sensitive to both acoustic properties under the trading relations (Hodgson & Miller, 1996). For instance, speakers with a smaller, barely noticeable difference in F1 discrimination tasks (i.e., with higher perceptual acuity to F1) were found to have made greater acoustic compensations in response to F1 formant perturbations in English vowels (Villacorta et al., 2007). Similarly, Nault and Munhall (2020) detected a significant positive correlation between individuals' compensatory changes to real-time formant perturbations and their perceptual acuity, which was evaluated by their discrimination performance on a vowel continuum.

In the suprasegmental domain, little research has been conducted on the relationship between perception and compensations for prosodic variations. As an indirect example of such claim, Feng et al. (2019) found that English L1 listeners, who were more sensitive to duration and intensity modulations compared with Chinese L2 speakers, used higher final f0 contours to recognize questions as an acoustic compensation for the weakened duration and intensity cues under certain experimental conditions. From the standpoint of prior compensation studies (Nault & Munhall, 2020; Villacorta et al., 2007), English L1 listeners exhibited more significant responses to the trading relations between f0 and duration (intensity), which suggests they should correlate with higher sensitivity to these acoustic properties involved for the identification of intonation contrasts. Surprisingly, the findings of Feng et al. (2019) contradicted this prediction by showing that English L1 listeners were less sensitive to sentence-final f0 linear changes, compared with Chinese L2 learners. Given that the small sample size of Feng et al.'s (2019) study may have affected its reliability, parallel investigations are conducted in the present study for the Chinese-Spanish language pair to test whether listeners' performance confirms the relationship between perceptual sensitivity and compensatory ability proposed in prior research.

1.4 Objectives and hypotheses

The main goal of this study is to investigate how L1 and Chinese L2 speakers of Spanish use f0, duration, and intensity cues to perceive the question-statement contrast in Peninsular Spanish. We focus on: (a) whether listeners' perception of Spanish intonation contrasts is influenced by changes in the three acoustic properties; (b) which language group is more sensitive to f0, duration, and intensity modulations in sentence-final syllables; and (c) how these multiple cues are combined and interact during listeners' perceptual process. Moreover, prior studies show that individual features, such as age and gender, may affect speech perception by fine-tuning listeners' perceptual ability and pragmatic skills (Bishop et al., 2020; Patel & Grigos, 2006; Ryalls et al., 1994). Hence, this study also examines (d) to which extent listeners' perception performance is predicted by their L1 experience, age, and gender. Accordingly, we designed two perceptual experiments in which stimuli were synthesized by manipulating the values of the three acoustic properties.

Given prior neurobehavioral findings (Chien et al., 2020), we hypothesized that tonal (Chinese) and non-tonal (Spanish) language listeners have comparable sensitivity to the f0 perceived as intonation because they share a general bilateral network for intonation processing. However, relative to listeners of non-tonal languages (specifically, Spanish L1 listeners), we hypothesized that Chinese L2 listeners use lower final f0 contours to recognize questions, given the influence of language-specific intonation categories (Feng et al., 2019; Liu & Rodriguez, 2012). Following cue weighting studies in L1 and L2 English (e.g., Feng et al., 2019; Morrow & Liu, 2013; Peng et al., 2012), we hypothesized that the perception of Spanish sentence types could be influenced by changes in secondary cues (duration and intensity), on which L1 Spanish listeners would rely more heavily than Chinese L2 listeners. Additionally, given the dynamic nature of speech processing, we also hypothesized that listeners could take advantage of the trading relations between acoustic cues to help recognize the intonation category. Finally, in addition to L1 background, we hypothesized that individual differences in age and gender could affect listeners' performance during question-statement identification.

2.0 Method

The cue weighting paradigm for the two language groups was identified by analyzing listeners' responses in the two perceptual experiments. In Experiment 1, the stimuli consisted of a set of one-word sentences that varied in the final f0 contour and duration simultaneously, whereas in Experiment 2, the auditory stimuli differed in the two acoustic dimensions of f0 and intensity.

2.1 Participants

Forty-eight native listeners of Peninsular Spanish (hereafter, SN) and ninety-five native listeners of Mandarin Chinese (hereafter, CN) participated in this study. The participants were recruited by word of mouth or through advertisements posted on social media accounts of the Phonetics Laboratory of the Universitat de Barcelona or elsewhere. All participants gave informed consent electronically prior to inclusion in the test, and the study design was approved by the local ethics committee. Generally, but not always, participants were compensated monetarily for their time.

To control for the effect of age of acquisition, data on two CN listeners who started studying Spanish before the age of 16 were removed from the analysis. Data on two SN listeners aged over 60 were also excluded. Moreover, data on CN participants (N = 5) with an A1 or A2 proficiency level in Spanish were also excluded for their lack of knowledge regarding Spanish intonation. The rest of the CN listeners had an intermediate (B1), advanced (B2), or superior (C1) level of proficiency and confirmed that Peninsular Spanish was the language variety to which they had been predominantly exposed during the learning process. Therefore, finally, there were 39 SN listeners and 78 CN listeners in Experiment 1, whose ages ranged from 18 to 59 years (N = 117, 90 women, 28 men, M_{age} = 28.27, SD = 8.43). Experiment 2 had 33 SN listeners and 77 CN listeners, whose ages ranged from 19 to 58 years (N = 110, 84 women, 26 men, M_{age} = 28.02, SD = 8.46). No participants reported any history of hearing or communication disorders at the time of testing.

2.2 Stimuli synthesis

The intonation of Spanish yes-no questions and broad focus statements differs in the intonational prenucleus and nucleus (Face, 2011; Face & Prieto, 2007). Since the focus of this study is the nucleus, we only used sentences with a single stressed word as stimuli such that the prenucleus effect could be excluded. Each sentence comprised one trisyllabic word: *Sevilla* (penultimate syllable stress—paroxytone) and *Alcalá* (final syllable stress—oxytone). The original recordings of the two items were obtained using a discourse completion task (Félix-

Brasdefer, 2010). A female native speaker of Peninsular Spanish (age at the time of the recording: 31) was required to produce the two one-word sentences in a broad focus statement context. The speaker was required to speak at a normal rate. The recordings were done in a quiet room using a Rode Smartlav+ microphone connected to a Scarlett Solo interface. Speech files were digitized at a sampling rate of 44.1 kHz with a quantization precision of 16 bits. The final syllable of both one-word sentences was chosen for acoustic manipulation. The target stimuli were created by manipulating f0, duration, and intensity. Prior to pitch manipulation, the non-final segments of the two one-word sentences were normalized at a 70 dB sound pressure level and used as a reference for the parametric variation of intensity in Experiment 2.

2.2.1 F0 manipulation

The f0 contour of the word-final syllable was replaced by a multi-step f0 continuum using the "to manipulation" function in Praat software (Boersma & Weenink, 2020). The pitch contour of the last syllable was stylized into two pitch points (defined as A1 and A2), and the values between them were defined by interpolation (Zarate-Sandez et al., 2015). As shown in Fig. 1, the start point, A1, was fixed at the beginning of the final syllable's vowel, keeping the pitch height similar to the original sentence. Thus, the f0 values of A1 for *Alcalá* and *Sevilla* were 196 Hz and 200 Hz, respectively. The endpoint, A2, was anchored at the last regular glottal pulse observed in the spectrogram and had an f0 value nearly identical to A1 (3 Hz difference). The f0 continuum of A2 was manipulated upward nine times and downward once, with a 20 Hz step size (greater than the slightest pitch variation normal-hearing adults can perceive). Hence, 11 f0 steps spanning over 200 Hz were generated for the final syllable of each word. The 22 f0 contours (2 stress positions*11 f0 steps) with different terminal pitches were further used as the basis for manipulating the duration and intensity.

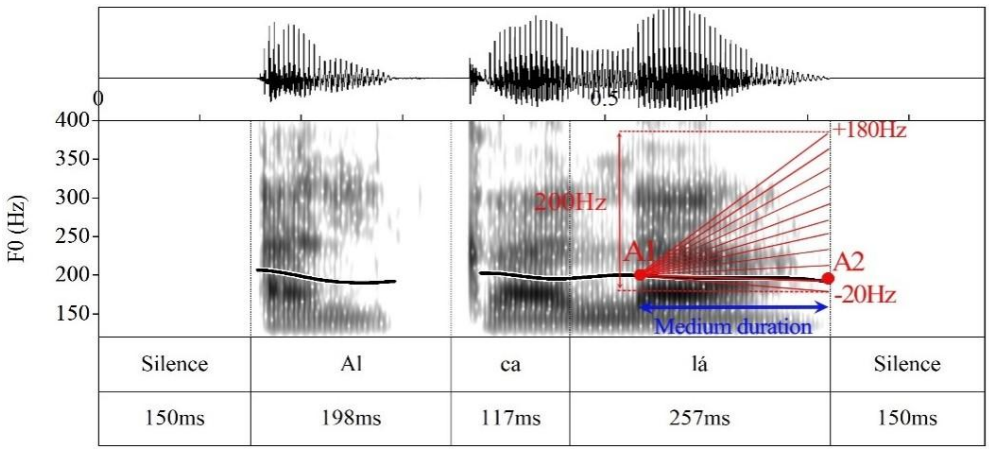


Fig. 1. Schematic representation of pitch manipulation in the oxytone word Alcalá. The start point A1 was 196 Hz. The endpoint A2 was manipulated from 176 Hz to 376 Hz, with a 20 Hz step size. The original duration of the final vowel (185 ms) was set as the medium duration for the stimulus Alcalá.

2.2.2 Duration manipulation

In Experiment 1, the 22 pitch stimuli described in the previous section were used to add a duration manipulation following a procedure similar to that used by Feng et al. (2019). Three duration levels were established in the study: short, medium (i.e., original), and long. As shown in Fig. 1, the original vowel duration of the statement's last syllable was selected as the medium value for each word. The specific duration values of the stimuli at the three duration levels are shown in Appendix A. The long duration was created by incorporating Praat 50 ms from the center of the nucleus of the original vowel in the last syllable. The +50 ms value was used because previous acoustic findings of Spanish intonation show that the final vowel of yes-no

questions was roughly 40–70 milliseconds longer, depending on the stress type of the final word, compared with that of statements (cf. Romera Barrios et al., 2007, for the intonation analysis of Barcelona Spanish). The short duration was generated by removing 40 ms from the same vowel nucleus. The reduced value was defined by referring to the shortest production of a statement by the same Spanish speaker. Since each duration level was combined with 11 different pitch contours, Experiment 1 included 66 stimuli (2 stress positions*3 durational levels*11 f0 steps).

2.2.3 Intensity manipulation

In Experiment 2, the same set of 22 pitch contours in Section 2.2.1 was used to manipulate intensity. The intensity of the word-final syllable was synthesized using the "constant amplification" function in Cool Edit Pro 2.1 software (Syntrillium Software Corporation, 2003). Three levels of intensity were created by applying -7 dB, 0 dB, and +7 dB changes to the final syllable vowel (cf. schematic representations of the intensity manipulation in Appendix B). The manipulated values employed the normalized intensity of the non-final part of the sentence (70 dB, as in Section 2.1.1) as a reference level. Hence, the low, medium (original), and high intensities were set at 63 dB, 70 dB, and 77 dB, respectively. Thus, Experiment 2 had 66 auditory stimuli (2 stress positions*3 intensity levels*11 f0 steps).

2.3 Procedure

Data of the perception tests were collected via an online survey, developed electronically using the web-based software Alchemer. The survey comprised three sections. The first section collected participants' socio-demographic and linguistic background information. The second (third) section contained the 66 auditory stimuli of Experiment 1 (2). The survey was administered in each listener's preferred language (Chinese or Spanish). The stimulus order of each experiment was randomized, and the text of each stimulus was displayed on the screen without punctuation marks. Listeners could participate in one or both experiments, depending on their interests. Those who chose to participate in only one experiment were randomly assigned to one of the two auditory experiments. They were required to listen to the stimuli using earphones in a quiet place. Listeners were also instructed to listen to every stimulus once; however, they could listen to it a second time if the recording failed to play for technical reasons.

Prior to the experiments, participants performed practice trials to familiarize themselves with the procedure. The perceptual answers of the two experiments were arranged on a five-point Likert scale to capture listeners' perceptual changes more accurately. Listeners were tasked to identify which of the five descriptions of the sentence type was closer to the stimulus they had just heard, by selecting one of the following options: "statement," "more statement than question," "either statement or question," "more question than statement," and "question." A short break was inserted after each section to compensate for fatigue.

2.4 Statistical analysis

We conducted all statistical analyses using R software (R Core Team, 2020). Since the two experiments' outcome variable was arranged on a five-point ordinal scale, proportional-odds logistic regression (POLR) was used to model the data. However, from the Brant test results, some explanatory variables in the POLR models rejected the parallel regression assumption. Hence, for the final estimates, we employed a more flexible ordered generalized linear model (OGLM) to relax the parallel lines assumption for variables where it was violated and allowed every effect to have separate coefficients for the five levels of the dependent variable (Abrudan et al., 2020; Williams, 2016): "statement" (level 1), "more statement than question" (level 2), "either statement or question" (level 3), "more question than statement" (level 4), "question" (level 5). Two OGLM models were built separately for the data of Experiments 1 and 2 using the oglmx package for R (Carroll, 2018). Stress Type (penultimate vs. final), L1 Group (CN vs.

SN), and Gender (male vs. female) were included as binary predictors, whereas Duration (short < medium < long) and Intensity (63 dB < 70 dB < 77 dB) were treated as discrete ordinal variables. Moreover, f0 Change, Age, and Stimulus Order were included as continuous predictors; the latter two were z-transformed and centered on their mean before being integrated into the model.

A stepwise backward elimination was performed to get the best model fit. The stepAIC function from the MASS package was used to iteratively exclude the least contributive predictors of the model (Kassambara, 2018; Ripley et al., 2013). Predictors that failed to achieve the .05 level of significance were removed, until all factors were statistically significant for the prediction. Finally, all samples were randomly split into a training and testing set to examine the statistical validity. The training set was constructed to estimate all the effects included in the model. The testing set was used to evaluate the testing error by computing the logarithmic (log) loss on the folds from the training set (Brown, 2020).

3.0 Results

3.1 Results of Experiment 1

The results of the OGLM model fitted for Experiment 1 revealed that listeners' perception of the five possible sentence types was significantly predicted by Stimulus Order [$\chi^2(1) = 7.73$, $p <. 01$], Age [$\chi^2(1) = 60.13$, $p <. 001$], Gender [$\chi^2(1) = 4.29$, $p <. 05$], Stress Type [$\chi^2(1) = 78.60$, $p <. 001$], and the three two-way interactions: L1 Group × f0 Change [$\chi^2(1) = 28.05$, $p <. 001$], L1 Group × Duration [$\chi^2(2) = 28.15$, $p <. 001$], and f0 Change × Duration [$\chi^2(2) = 16.75$, $p <. 001$]. Since changes in the dependent variables impact the response scale differently, we discuss these relationships by focusing on their margins in our outcome of interest. The full marginal effects of Experiment 1 are given in Appendix C.

Table 1. Summary of the marginal effects in the OGLM model, fitted for Experiment 1 (Marginal effects and *t*-stat. ***$p < 0.001$; ** $p < 0.01$; * $p < 0.05$; '.' $p < 0.1$).

Predictor	Level 1	Level 2	Level 3	Level 4	Level 5
Order	0.00381**	0.00928**	0.00391**	-0.01100**	-0.00600**
Age	-0.01241***	-0.03021***	-0.01272***	0.03581***	0.01954***
GenderFemale	-0.00640*	-0.01530*	-0.00607*	0.01818*	0.00959*
StressPenultimate	-0.01307***	-0.03171***	-0.01332***	0.03754***	0.02056***
f0	-0.00210***	-0.00512***	-0.00216***	0.00607***	0.00331***
L1SN	0.14632***	0.24241***	0.034729***	-0.28640***	-0.13706***
L1SN×f0	-0.00026***	-0.00063***	-0.00027***	0.00075***	0.00041***
L1SN×DurMedium	-0.01137*	-0.02892	-0.01420.	0.03386.	0.02064.
L1SN×DurLong	-0.02868***	-0.07777***	-0.04775***	0.08662***	0.06757***
f0×DurMedium	-0.00006	-0.00014	-0.00006	0.00016	0.00009
f0×DurLong	-0.00023***	-0.00056***	-0.00024***	0.00066***	0.00036***

Note: Levels 1–5, in order, signify the five possible responses in the perception test: "statement," "more statement than question," "either statement or question," "more question than statement," and "question."

Because Age and Gender are individual predictors that do not interact with any other variable, the marginal effects presented in Table 1 can be directly interpreted as coefficients, which represent the predicted change in the five response levels by a unit change in the variable. To better understand their effects, we graphed the predicted probability plot for each variable. Specifically, Fig. 2 shows that when the listener was female and older, the predicted probability of making question-like answers increased, whereas that of statement-like responses decreased.

This finding implies that, holding constant all other variables, older and female listeners were more likely to recognize one-word sentences as questions, compared with males and younger ones. We posited that the age effect was likely because older listeners, particularly those of the CN group, had higher proficiency in Spanish, which made them more capable of using acoustic cues for signaling question intonation. To validate this proposal, a Kruskal-Wallis test by ranks was performed (Pohlert, 2014); the results confirmed that there was a significant correlation between CN listeners' Age and Spanish proficiency in Experiment 1 [$\chi^2(2) = 554.60$, $p < .0001$]. Thus, the effect of Age on question-statement identification can be considered a perceptual benefit of higher language proficiency in a non-native listening environment. Another significant individual predictor of Experiment 1 was Stress Type, whose coefficients in Table 1 were negative in levels 1 and 2 and positive in levels 4 and 5. This result suggests that, all else being equal, words with penultimate stress were more likely to be perceived by listeners as questions, compared with those with final stress. The reasons listeners were more prone to perceive Spanish paroxytone words as questions are discussed in Section 4.0.

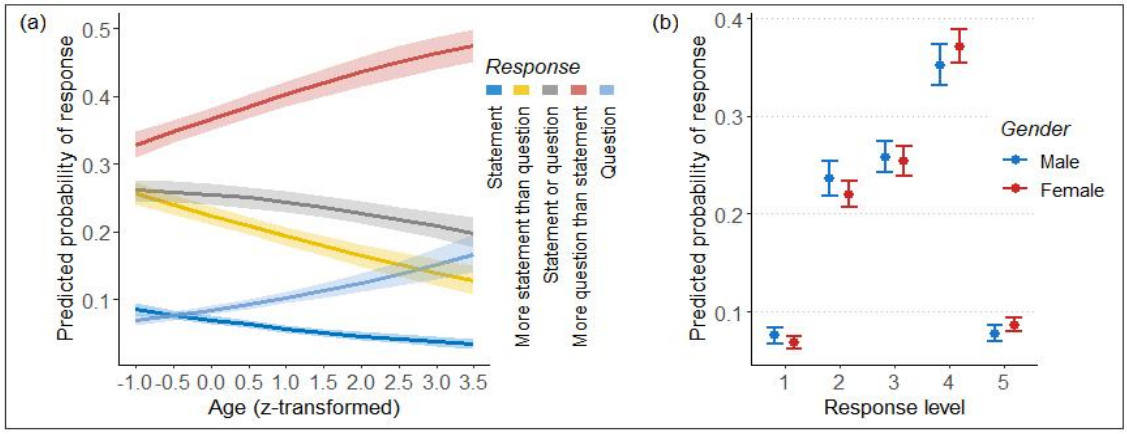


Fig. 2. Average marginal effects of Age and Gender on the outcome variable, holding constant all other variables in Experiment 1.

Regarding the marginal effects of the f0 cue, Table 1 indicates that, on average across all the variables in the model, a unit increase in the final f0 contour significantly decreased the likelihood of having statement-like responses (i.e., levels 1 and 2) while increasing the possibility that a stimulus was perceived as a question-like sentence (i.e., levels 4 and 5). Moreover, after confirming the general predictions, critical cross-linguistic differences emerged between the CN and SN groups during question-statement identification. The marginal effects of L1 reflected that, holding all other variables at their means, CN listeners were significantly more likely to have a question-like response, and needed a lower f0 increase for question recognitions than SN listeners. For instance, Fig. 3 shows that when the final f0 contour was increased by 120 Hz, the sentence type most frequently predicted by CN listeners was "question"; meanwhile, SN listeners needed an f0 change greater than 140 Hz for "question" to become their most frequently predicted answer in the short and medium durations. One possible reason for these differences in f0 use relates to the three-way contrast of intonational phrase (IP)-final boundary tones in Peninsular Spanish. This language variety, apart from the low (L%) and high (H%) boundary tones, involves mid-level tones to signal uncertainty statements (!H%) and statements of the obviousness (L!H%) (Estebas-Vilaplana & Prieto, 2010). From native (SN) listeners' perspectives, both low- and mid-level tones correspond to the statement category (whether broad focus statements, as in L%, or epistemically biased statements, as in !H% and L!H%). Thus, the lower likelihood of question-like responses exhibited by SN listeners is probably because they categorized the final mid-level pitch as a biased statement. By contrast, CN listeners may have interpreted it as a question because any tone not strictly low was perceived as high, thereby associated to a question category.

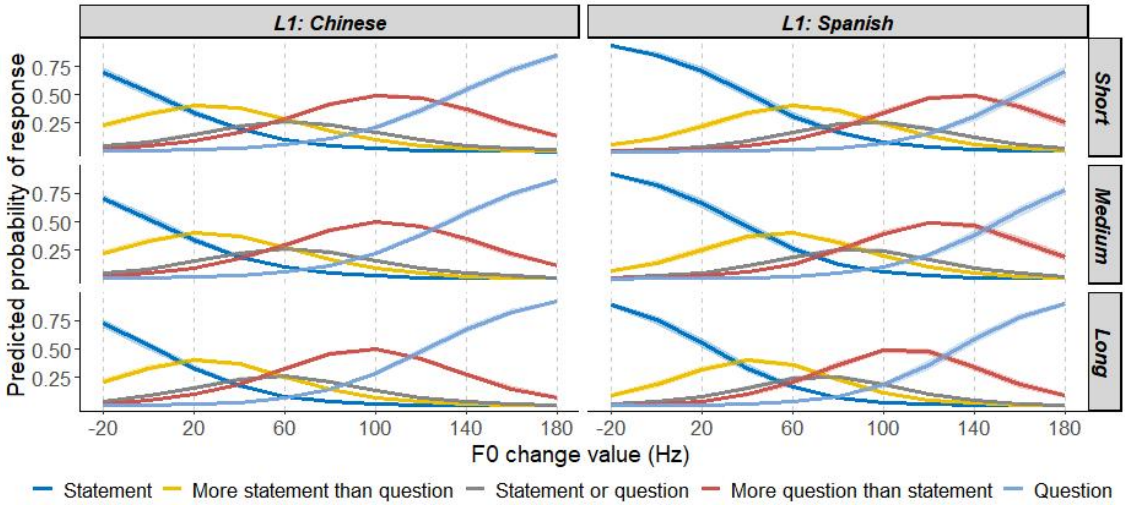


Fig. 3. Effect displays between L1 Group, f0 Change, and Duration for each response level, holding constant all other variables in Experiment 1.

While f0 Change and L1 Group alone can contribute significantly to the model predictions, the higher-order interaction between L1 Group × f0 Change revealed that the marginal effects of both variables clearly depended on the level of the other variable. Since the models were dummy-coded in the study and the CN group was set as the reference level for the L1 Group factor, the significant marginal effects of L1SN × f0 in Table 1 can be interpreted as: the slope for the question-statement identification on f0 pitch was significantly steeper for the SN group, compared with the CN group. In general, this result corroborates the previous hypothesis, showing that the impact of the f0 cue on the perception of intonation contrasts was significantly modulated by listeners' L1 background. However, contrary to prior assumptions, our findings seem to suggest that SN listeners were more sensitive than CN listeners to f0 modulations perceived as intonation because they changed the perceived phonological categories more rapidly overall when the f0 contour of sentence-final syllables varied.

In addition to the interaction with the f0 cue, the significance of the higher-order term L1 Group × Duration indicated that CN and SN listeners' likelihood of choosing each of the five possible responses changed depending on variations in the duration of the sentence-final syllables. The post-hoc results showed that SN listeners had significantly greater probabilities of making a higher level of response in the long-duration patterns relative to stimuli with a short duration (β = .931, SE = .099, $p < .0001$). A significant identification improvement was also observed for the CN group from the short to long-duration condition (β = .303, SE = .069, $p < .0001$). Nonetheless, as shown in Fig. 4, the overall magnitude of perceptual changes caused by duration modulations was less apparent in CN listeners, compared with SN listeners. This finding was also consolidated by their prior coefficients of the contrast between the short and long duration levels. Thus, our results suggested that SN listeners were more sensitive than CN speakers to duration modulations, although listeners of the two L1 groups could use the duration cue to a different extent, to perceive intonation contrasts in Spanish. The full multiple comparison results of Duration by L1 Group are available in Appendix D.

Further, the average marginal contribution of the interaction term f0 × DurLong in Table 1 reveals that the slope for question-statement identification as a function of the f0 pitch was significantly steeper in the long duration condition, compared with the short duration. Particularly, Fig. 4 shows that CN and SN listeners could use a lower f0 contour to hear the question when the duration of the final syllable was increased. Conversely, in the short duration condition, a higher final f0 contour was required for question recognitions to compensate for the weakened duration cue. Both can be considered an acoustical consequence of phonetic trading relations, which assumes that changes in one acoustic cue can be counteracted by opposing

changes in another cue such that phonetic quality is preserved (Repp, 1982). Consistent with previous speech perception studies (Feng et al., 2019; Kuang & Cui, 2018; M. Yang & Sundara, 2019), our findings showed that listeners could use the covariations between acoustic cues to better recognize the intended linguistic targets. However, based on the overall magnitude of slope changes in identification curves over the three duration levels by the L1 group (Fig. 4), SN listeners, who were more sensitive to the duration cue (cf. previous paragraph for discussions), exhibited greater compensations for duration changes, compared with CN listeners. This finding corroborates previous research (Nault & Munhall, 2020), which demonstrated that listeners with higher perceptual sensitivity were more likely to make larger compensations for acoustic variations in auditory processing.

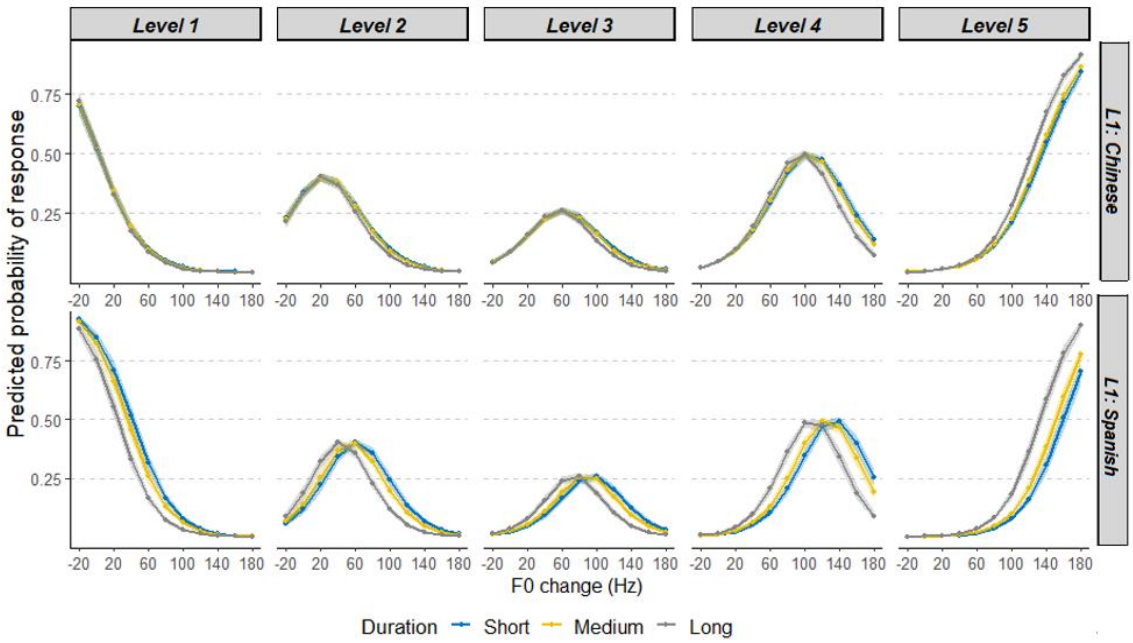


Fig. 4. Average marginal effects of duration on each L1 Group while varying the final f0 contour of the one-word sentences in Experiment 1.

3.2 Results of Experiment 2

In Experiment 2, the predictors of Gender and Stimulus Order did not reach statistical significance and were thereby excluded from the model. The analysis of the OGLM model indicated that the perception of the five possible sentence types was significantly predicted by Age [$\chi^2(1) = 9.00$, $p <. 01$] and by the L1 Group × f0 Change [$\chi^2(1) = 34.83$, $p <. 0001$] and f0 Change × Intensity [$\chi^2(2) = 13.33$, $p <. 01$] two-way interactions. Although the three-way interaction between L1 Group × Intensity × Stress Type was not statistically significant, we nevertheless included it in the model because (a) the effects of Intensity [$\chi^2(2) = 16.80$, $p <. 001$] and L1 Group [$\chi^2(1) = 26.18$, $p <. 0001$] were significantly modulated by Stress Type, and (b) the three-way interaction term can help explain possible differences in intensity weighting between CN and SN listeners. The full model results of Experiment 2 are given in Appendix E.

First, the positive coefficients of Age at levels 4 and 5 (Table 2) corroborate the data of Experiment 1, confirming that older listeners were more likely to recognize one-word sentences as yes-no questions than younger listeners. Similarly, this effect can be explained by the strong positive correlation between CN listeners' Age and Spanish proficiency in Experiment 2 [$\chi^2(2) = 158.19$, $p < .0001$]. The effect of f0 on the question-statement identification is also in line with that in Experiment 1. For instance, Fig. 5 shows that the probability of question-like answers increased as the final f0 contour was raised. Moreover, the negative margins of the SN Group on levels 4 and 5 (Table 2) indicated that CN listeners were significantly more likely to perceive

one-word sentences as yes-no questions, compared with SN listeners, probably because they categorized both mid- and high-level pitches in Spanish as question intonation.

Table 2. Summary of the marginal effects in the OGLM model, fitted for Experiment 2 (Marginal effects and *t*-stat. ***$p < 0.001$; ** $p < 0.01$; * $p < 0.05$; '.' $p < 0.1$).

Predictor	Level 1	Level 2	Level 3	Level 4	Level 5
Age	-0.00488**	-0.01125**	-0.00667**	0.01456**	0.00825**
L1SN	0.13833***	0.22080***	0.04999***	-0.28048***	-0.12864***
f0×L1SN	-0.00030***	-0.00069***	-0.00041***	0.00090***	0.00051***
f0×70dB	-0.00004	-0.00010	-0.00006	0.00012	0.00007
f0×77dB	-0.00020***	-0.00045***	-0.00027***	0.00058***	0.00033***
L1SN×70dB	-0.01510.	-0.03680.	-0.02615	0.04654.	0.03151
L1SN×77dB	0.00327	0.00746	0.00425	-0.00967	-0.00531
L1SN×StressP	-0.02046**	-0.05025**	-0.03682*	0.06296**	0.04457*
StressP×70dB	-0.00197	-0.00456	-0.00276	0.00590	0.00339
StressP×77dB	-0.01604**	-0.03879*	-0.02695*	0.04921**	0.03258*
StressP×70dB:L1SN	-0.00037	-0.00086	-0.00051	0.00111	0.00063
StressP×77dB×L1SN	-0.01426	-0.03488	-0.02510	0.04404	0.03019

Note: Levels 1–5, in order, signify the five possible responses in the perception test: "statement," "more statement than question," "either statement or question," "more question than statement," and "question." StressP signifies one-word sentences with penultimate syllable stress.

Next, the average marginal effects of the higher-order term f0 × L1SN in Table 2 showed that the SN group had significantly steeper identification functions for each f0 pitch setting, compared with the CN group. This result strengthens the findings of Experiment 1, demonstrating that SN listeners were more sensitive than CN speakers to f0 linear changes perceived as Spanish intonation. Additionally, consistent with Experiment 1, the results of Experiment 2 revealed that, all else being equal, SN listeners needed higher terminal pitches to perceive questions than CN listeners, due to the influence of the three-way contrast of IP-final boundary tones in their L1 intonation system. As shown in Fig. 5, CN listeners needed a final f0 rise of 120 Hz for making "question" the most possible sentence type, whereas SN listeners required a final f0 change greater than 140 Hz for "question" to become their most frequently predicted answer in the 63 dB and 70 dB conditions.

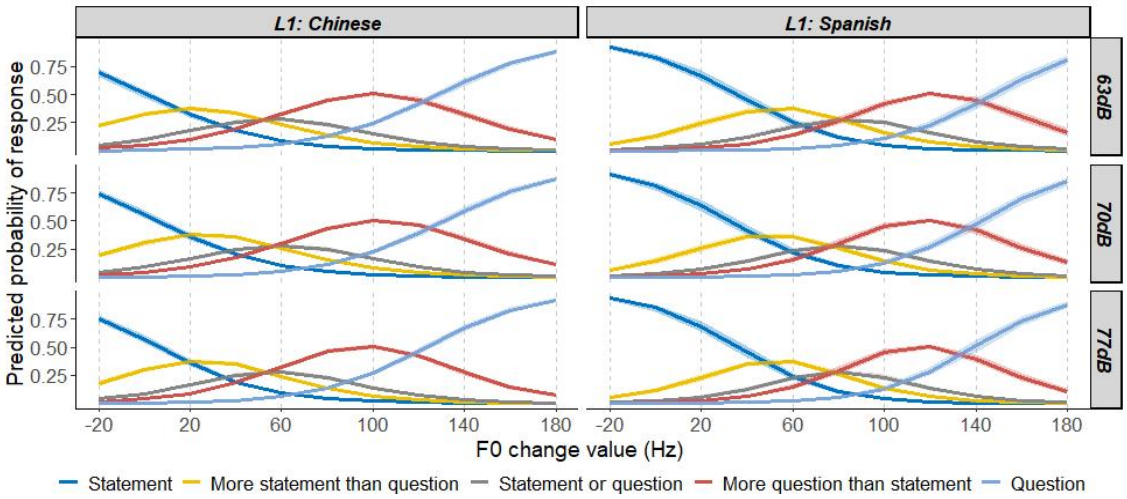


Fig. 5. Effect displays between L1 Group, f0 Change, and Intensity for each response level, holding constant all other variables in Experiment 2.

Similar to the f0 Change × Duration interaction in Experiment 1, a significant trade-off was found between the two acoustic properties of f0 Change × Intensity. Specifically, the positive margins of f0 × 77 dB at levels 4 and 5 (Table 2) indicated that the slope of the identification curves as a function of the f0 Change was significantly steeper at 77 dB than at 63 dB, implying that listeners could rely on fewer f0 cues for question recognition when the intensity of word-final syllables was increased to the highest level (77 dB). This finding serves as additional evidence for phonetic trading relations and can be interpreted as a counter-directional compensatory behavior of listeners in response to the enhanced or weakened acoustic cues, which are relevant for the recognition of a specific phonological category.

Before interpreting the three-way interaction, we compared the marginal effects of the SN group in 70 dB and 77 dB conditions (Table 2), which suggested that SN listeners were more likely to perceive words as questions at 70 dB instead of 77 dB. This result may seem surprising, but the modulation of Stress Type explains it further. For the sake of simplicity, we discuss the relationship of the three variables with reference to Fig. 6. For instance, CN and SN listeners exhibited a higher likelihood of question-like responses in different ways when the word's stress was on the penultimate syllable—paroxytone—and the intensity was manipulated from lower to higher levels. Nevertheless, this Intensity effect was not found when participants listened to words with final stresses—oxytone. Moreover, as shown in Fig. 6, the SN group had a steady and significant stepwise increase in question identifications when the intensity of word-final syllables increased from 63 dB to 77 dB (β = .462, SE = .155, p < .01). By contrast, the effect of Intensity was weaker and irregular on the performance of CN listeners, as the coefficient of the contrast between 63 dB and 77 dB (β = .222, SE = .098, p = .06) reveals. The full multiple comparison results are available in Appendix F. Together, the three-way interaction analysis revealed that listeners used the intensity cue inconsistently depending on the word's stress. In addition, the SN group relied more heavily on increased intensity to perceive paroxytone words as yes-no questions, suggesting that SN listeners were more sensitive to the intensity cue than CN listeners.

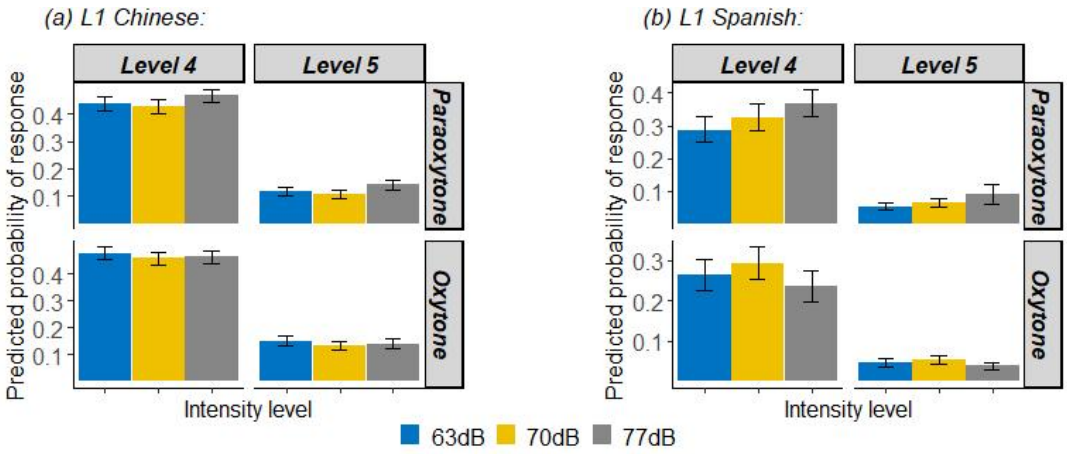


Fig. 6. Effect displays for the three-way interaction between L1 Group, Intensity, and Stress Type in levels 4 (more statement than question) and 5 (question) of Experiment 2.

3.3 Estimation of model performance

The log loss metric is among the most selective functions to estimate the quality of probabilistic predictions; it works particularly well when there are multi-class classification problems (Brown, 2020). Therefore, to assess the predictive accuracy of the models, log loss values were computed on the testing set using the mlogLoss function from the ModelMetrics package for R (Hunt, 2020). Log loss values of the OGLM model were 1.0483 and 1.0286 for Experiments 1

and 2, respectively.

Due to log loss measures testing error and model uncertainty, the lower the value, the better the prediction. A model considered good should have a log loss that is at least less than the "dumb" or non-informative log loss. The "dumb" log loss was calculated assuming that the number of classes (M) was uniformly distributed in the data (Brown, 2020). In this study, we have M = 5 response levels, each with a 20% chance of occurrence. Hence, the "dumb" log loss of the model, fitted for Experiments 1 and 2, was calculated using Equation (1).

$$\log loss = -\ln\left(\frac{1}{M}\right) = -\ln\left(\frac{1}{5}\right) = \ln(5) = 1.6094 \tag{1}$$

However, in fact, the five classes of the outcome variable did not occur equally in the data; thus, the non-informative log loss of the model was further computed, assuming a specified non-uniform distribution of listeners' responses. The non-informative log loss for both models was calculated by specifying the probability of each class. For instance, in Experiment 1, 21.77%, 16.49%, 10.88%, 24.13%, and 26.74% of responses indicated, respectively, "statement," "more statement than question," "either statement or question," "more question than statement," and "question." Thus, the non-informative log loss for the model of Experiment 1 was calculated using Equation (2).

$$\begin{aligned} \log loss = -\begin{pmatrix} 0.2117\ln(0.2117) + 0.1649\ln(0.1649) + 0.1088\ln(0.1088) \\ +0.2413(0.2413) + 0.2674\ln(0.2674) \end{pmatrix} \\ = 1.5662 \end{aligned} \tag{2}$$

Similarly, the non-informative log loss for the model of Experiment 2 was 1.6113. Overall, the average log loss for the two models was effectively reduced to 1.0483 and 1.0286 relative to the "dumb" and non-informative log loss. Therefore, the models for these two experiments can be considered accurate and predictive, outperforming "dumb" and non-informative models.

4.0 Discussion

The present study investigated the perceptual weighting of the three acoustic cues—f0, duration, and intensity—in perceiving Spanish intonation contrasts by CN and SN listeners. The stimuli were created by manipulating the f0 contours progressively from falling to rising directions while changing the duration or intensity of word-final syllables. Unlike prior studies, that typically provided two opposite perceptual options (i.e., statement vs. question), we used a five-point Likert response scale to capture listeners' perceptions more precisely.

In general, the results of the two auditory experiments demonstrated that f0 contour was the primary cue for both CN and SN listeners' question-statement identification of Peninsular Spanish, followed by duration changes. The parametric variation of intensity seemed to be the acoustic factor that contributed the least to question-statement identification, being used most variably by listeners in the prosodic evaluation of intonation categories. Beyond the cross-linguistic commonality of cue ranking, listeners of both language groups exhibited great differences in their use of the f0, duration, and intensity modulations for perceiving intonation categories, as well as in their compensatory behavior in response to the trading relations between multiple acoustic cues.

4.1 The role of f0 contour in intonation perception

SN listeners perceived yes-no questions using significantly higher final f0 contours, compared

with CN listeners, especially when the value of secondary cues (i.e., duration or intensity) was reduced. This finding confirms our predictions and is consistent with the study of Feng et al. (2019), who found that English L1 listeners had a significantly higher offset f0 frequency for question identifications, compared with Chinese L2 speakers, except in the long-duration condition. The situation may be attributed to SN listeners being more capable—compared with CN speakers—of compensating for the weakened secondary cues and f0 loss caused by the pitch synthesis because of their broader linguistic experience with Spanish. Moreover, the exact mapping between f0 levels and Spanish intonation categories may have required SN listeners to use higher terminal pitches to signal questions to distinguish them from epistemically biased statements encoded with mid-level tones (e.g., uncertainty statements expressed with !H% or statements of the obviousness expressed with L!H%). However, intonation types in Chinese are not determined by the surface f0 contours at the end of the sentence. Thus, CN listeners may interpret any IP-final boundary tone not strictly low (e.g., !H% or L!H%) as high, thereby categorizing it as a question. CN listeners' tendency to categorize both mid- and high-level pitches as questions also explained their greater probability of identifying yes-no questions, compared with SN listeners.

Apart from auditory categorization differences, the significantly steeper identification curves exhibited by the SN group in the two experiments suggested that SN listeners were more sensitive than CN listeners to f0 linear transitions perceived as intonation. Thus, the previously proposed tonal language benefit may be significant only in specific dimensions of pitch perception, such as naturally-produced f0 contrastive tone patterns that are closely linked to their native pitch events (e.g., Chandrasekaran et al., 2007; Deroche et al., 2019; Hallé et al., 2004; Ortega-Llebaria & Wu, 2021), but not in f0 linear transitions perceived at the sentence-intonation level. Recent studies supporting this idea include that of Chien et al. (2020), who demonstrated that pitch processing as intonation involved bilateral cortical areas among listeners of both tonal (Chinese) and non-tonal (German) languages, regardless of the different realizations of intonation in their L1. By contrast, the processing of f0 pitch as lexical tone engaged semantic areas specific to Chinese listeners, in addition to the left-hemispheric phonological system activated by tonal and non-tonal language listeners (Chien et al., 2020).

According to prior neurobehavioral findings, the CN and SN listeners in this study should have comparable f0 perception performance as tonal and non-tonal language speakers share the same neural network for intonation processing. However, contrary to the predictions, CN listeners showed lower sensitivity to f0 pitch perceived as intonation than SN listeners; this could be because they were exposed to a non-native environment (unlike the Chinese auditory stimuli used in Chien et al., 2020) where they had limited experience in perceiving f0 contours along the lines of language-specific and well-defined intonation categories. Moreover, according to the functional view, if some phonetic cues are exploited in one dimension of grammar, they will not be used to the same degree in another area of phonology (Gandour et al., 1995; Liang & Heuven, 2007; Seddoh, 2002). Thus, the dual linguistic function (e.g., tone and intonation) of f0 pitch in Mandarin Chinese may have influenced CN listeners' auditory processing of different pitch events. Accordingly, Chen (2005), Gussenhoven and Chen (2000), and Liang and Heuven (2007) argued that the preference of tonal language listeners to primarily perceive the f0 information bearing on word meanings (i.e., lexical tone) was a critical factor for their decreased sensitivity to f0 variations processed as intonation. By extension, we posit that CN listeners' increased effort when processing the pitch patterns of non-native lexical stress may have resulted in their weaker ability to decode the f0 cues at the sentence-intonation level.

4.2 The role of secondary cues in intonation perception

In addition to the primary cue of f0 contour, duration serves as a secondary cue for both CN and SN listeners during question-statement identification. Long-duration patterns could significantly increase the possibility of question answers for both L1 groups vis-à-vis stimuli with a short

duration. However, compared with the SN group, the overall magnitude of improvements over the three duration levels was smaller in the CN group, confirming our predictions that CN listeners were less sensitive to the duration modulations in Spanish sentence types, compared with SN listeners. A similar scenario was observed regarding the perceptual weighting of the intensity cue. Only SN listeners were sensitive to intensity variations in paroxytone words, whereas the performance of CN listeners was not influenced by changes in intensity when perceiving paroxytone or oxytone words. These cross-linguistic differences in intonation cue weighting between native and non-native listeners were generally consistent with Feng et al.'s (2019) findings; however, Feng et al. (2019) did not find a significant effect of the duration or intensity on question-statement recognition by Chinese learners of English. Further, comparing the statistical validity of two secondary cues (i.e., duration and intensity), this study posited that intensity was a less prominent cue than duration in question-statement identification for the SN group, which was used variably by listeners depending on the word's stress position.

Similar perceptual compensation explanations are adopted here for SN listeners' greater reliance on those secondary cues (i.e., duration or intensity). Since pitch contours synthesized in this study differed from intonation patterns of Spanish statements and questions produced in natural speaking environments (not wholly linear at the end of the utterance), listeners may increase the weight of other secondary cues to compensate for the loss of f0 information and enhance their perceptual decisions (Feng et al., 2019). Nonetheless, SN listeners may know how to compensate for acoustic changes in speech more accurately than CN listeners because they have more extensive experience with the phonetic patterns of Spanish intonation categories. Beyond the differences in compensatory ability, the cue weighting paradigm in CN listeners' native prosodic system may induce their reduced sensitivity to duration and intensity cues. Since f0 pitch plays an overwhelming role in Mandarin, CN listeners may be less likely to attend to those less prominent acoustic features when identifying an intonation category, especially in non-native environments (e.g., Chang & Yao, 2007; Feng et al., 2019; Jiao & Xu, 2019). Overall, the findings of perceptual cue weighting are consistent with the results of previous research (e.g., Feng et al., 2019; Holt & Lotto, 2006; Kuang & Cui, 2018; Meng et al., 2020; Peng et al., 2012), where listeners from different language backgrounds differed in the way they used multiple sources of acoustic cues to specify the best exemplar of speech categories.

4.3 The implications of the trade-off in speech perception

From the above discussion, it is clear that the degree to which different acoustic cues can contribute to listeners' assessment of a phonological contrast may vary, since their attention is selective (Kuang & Cui, 2018). However, how do listeners integrate and coordinate the weighting of these cues to arrive at an overall estimate of an utterance's intonation category? A clue to answering this question may lie in the significant interaction effects between primary and secondary cues on question-statement recognition. Specifically, our study revealed that the effect of f0 contour strongly covaried with the duration or intensity cues during the perceptual process. Since f0, duration, and intensity are relevant cues for identifying the question-statement contrast, the attenuation in one acoustic dimension of these cues could be compensated by providing greater contributions from the other such that the original percept can be recovered. This trade-off relationship between two or more acoustic properties is consistent with the core idea of phonetic trading relations proposed by Repp (1982). Similar phenomena have been widely documented in identifications of phonetic segments, such as ongoing sound changes in Southern Yi (cf. tense vs. lax register in Kuang & Cui, 2018), consonant-vocal syllable distinctions in English (cf. *ba* vs. *wa* in Boardman et al., 1994), and stop-consonant voicing contrasts in American English (cf. Holt et al., 2001; Jacewicz et al., 2009).

Previous research has shown that listeners are sensitive to the multiplicity and covariations of acoustic-phonetic information for a given speech category, and can actively use a compensatory strategy to ensure the success of perceiving an intended linguistic category (e.g., Feng et al.,

2019; Hodgson & Miller, 1996; Holt et al., 2001; Holt & Lotto, 2006; Kuang & Cui, 2018; Peng et al., 2012). However, Hodgson and Miller (1996) argued that auditory compensation was contingent upon listeners' sensitivity to the co-varying cues that are consistently correlated with the recognition of a given phonetic segment. This notion has been supported by recent empirical studies (Nault and Munhall, 2020; Villacorta et al., 2007), which further demonstrated that listeners with higher perceptual sensitivity were more likely to make greater compensations for acoustic changes in speech. Thus, in this study, SN listeners were more susceptible to phonetic trading relations and better at compensating for acoustic changes in Spanish sentence types, probably because of their greater auditory sensitivity than CN listeners to acoustic-phonetic properties encoded in their L1 intonation system. These findings may serve as a good example of phonetic trading relations used as a tool to assess listeners' sensitivity to multiple acoustic cues of speech.

4.4 The role of individual and prosodic features in intonation perception

In line with our prior hypothesis, question-statement identification in Spanish is not solely determined by listeners' native language experience, but is also modulated by individual differences in age and gender. Older and female listeners were more likely to recognize one-word sentences as yes-no questions than younger and male speakers in the same acoustical condition. Given that there was a strong correlation between the age of CN listeners and their proficiency in Spanish, the age effect could be alternatively explained by older CN listeners' higher proficiency in Spanish, as they had considerable exposure to distributions of acoustic features in the target phonetic environment, which made them more capable of using multiple cues for recognizing question-statement contrasts. Nevertheless, this relationship between the two variables is likely limited to the CN group. Thus, further research is needed to verify this correlation and examine the respective age effect on intonation perception for native and non-native listeners. Moreover, women's greater likelihood of question answers can be attributed to their higher pragmatic skills than men in speech communication (evaluated by Bishop et al., 2020, using the communication subscale of the Autism-Spectrum Quotient Baron-Cohen et al., 2001), which may make them more sensitive to the mapping between prosodic patterns and pragmatic meanings. Notably, this gender effect was significant only in Experiment 1, whereas it did not reach statistical significance in Experiment 2. Therefore, further studies should be conducted with an equal number of male and female listeners to better understand this gender difference in the prosodic evaluation of intonation categories.

The present study also showed that listeners' perception performance is subject to the words' prosodic structure, such as stress position. In Experiment 1, paroxytone words were more likely to be perceived as yes-no questions, compared with oxytone words. This could be because the paroxytone is the most frequent unmarked stress pattern in Spanish, which is why, perhaps, adult listeners could process it more easily (Defior & Serrano, 2017; Roca, 2019). Moreover, the potential f0 conflicts between stress and intonation bearing on the last syllable of oxytone words may have made it more difficult for CN listeners to perceive the f0 pitch at the sentence-intonation level. In Experiment 2, the effect of intensity amplification was significant only for the perception of paroxytone words. Although the reasons for this perceptual difference in stress type are not fully understood, we speculate that the acoustic interference of the stress on the last syllable may lead to listeners' irregular use of intensity to perceive oxytone words. Further studies are needed to validate the effect of stress on intensity perception.

5.0 Conclusion

This study examined CN and SN listeners' perceptual weighting of three acoustic cues—f0, duration, and intensity—in Spanish question-statement identification. The overall findings revealed cross-linguistic commonalities and dissociations of intonation processing between

tonal and non-tonal language listeners. While CN and SN listeners could use f0 and duration cues to varying degrees in order to recognize intonation contrasts, only SN listeners were sensitive to intensity modulations of paroxytone words. More importantly, SN listeners showed a higher sensitivity to the acoustic changes of f0, duration, and intensity than CN listeners, probably because of their superior compensatory capacity and broader linguistic experience with Spanish prosody. The greater f0 sensitivity exhibited by SN listeners during intonation perception tentatively rejects the general hypothesis of tonal language benefit in pitch perception, as the f0 advantage of CN listeners was not generalized to all dimensions of pitch patterns. Moreover, beyond listeners' L1 background, the cue weighting paradigms of CN and SN listeners could be influenced by certain individual and prosodic features. Finally, the trade-off between primary (f0) and secondary (duration and intensity) cues confirms that rather than a single f0 modulation, the perception of intonation contrasts involves the dynamic interplay of multidimensional acoustic cues (Levis, 1999; Ma et al., 2008; Niebuhr, 2007).

This study had several limitations. First, even though f0 contour, duration, and intensity are the most salient cues for perceiving intonation, recognizing question-statement contrasts is not limited to these acoustic properties. Other contextual variables, such as speaking rate and phonetic environment, may also play a role in the specification of the best exemplars of intonation categories. Second, this study used one-word sentences to examine the cue weighting differences between CN and SN listeners in question-statement identification. Thus, whether these results can be generalized to sentences with more than one stressed word remains to be seen. Third, although the findings of this study support the existence of a phonetic trading relationship between primary (i.e., f0) and secondary (i.e., duration or intensity) cues, it remains unclear whether this relationship exists between the duration and intensity cues. Future studies using different experimental designs (e.g., including stimuli with covariations in duration and intensity) could bridge these limitations. Additional limitations include the imbalance between male and female listeners, which may limit the powers of the present sample and explain the nonsignificant gender effect in Experiment 2. Finally, despite positing that the acoustic interference of final stress may induce perceptual differences in the two stress types, how stress and intonation are encoded in parallel via changes in acoustic cues and how stress processing affects the perception of acoustic patterns as intonation, remains to be elucidated.

CRediT authorship contribution statement

Peizhu Shang: Conceptualization, Methodology, Software, Formal analysis, Writing -Original Draft, Visualization. **Paolo Roseano**: Methodology, Writing - Review & Editing, Supervision, Resources. **Wendy Elvira-García**: Writing - Review & Editing, Supervision, Resources, Visualization.

Funding

This work was supported by the China Scholarship Council [grant number 201808390059].

Acknowledgments

We would like to express our gratitude to Yiqiu Shi (Universitat de Barcelona) and Songlin Wang (Universitat de Barcelona) for helping us share the experimental information with the study's participants. They would also like to thank Xinyi Li (Technische Universität München) for his assistance with specific mathematical issues. We would also like to thank all the Chinese and Spanish participants of the study.

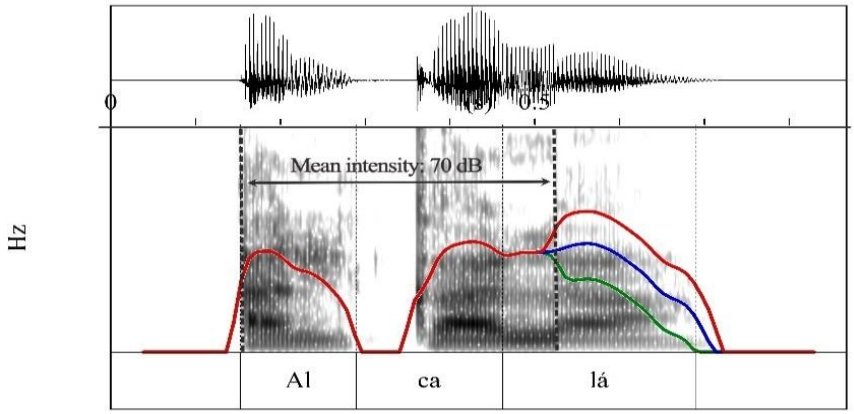
Declaration of interest

None.

Appendix A. Detailed manipulation values of the duration in the final vowel of the two one-word sentences.

Duration	Alcalá	Sevilla
Short	145 ms	130 ms
Medium (original)	185 ms	170 ms
Long	235 ms	220 ms

Appendix B. Schematic representation of intensity manipulation for the one-word sentence "Alcalá," keeping the original pitch contour of the statement (the green, blue, and red intensity curves in the final vowel have a value of 63 dB, 70 dB, and 77dB, respectively).



Appendix C. Summary of the marginal effects in the OGLM model fitted for Experiment 1 (***$p < 0.001$; ** $p < 0.01$; * $p < 0.05$; '.' $p < 0.1$).

Marginal effects on Pr (Outcome = Statement)				
Predictors	Marg. Eff.	SE	*t*	*p*
Order	0.00381	0.00138	2.77	**
Age	-0.01241	0.00166	-7.50	***
GenderFemale	-0.00640	0.00318	-2.01	*
StressPenultimate	-0.01307	0.00317	-4.12	***
L1SN	0.14632	0.01494	9.80	***
f0	-0.00210	0.00008	-25.99	***
L1SN×f0	-0.00026	0.00005	-5.12	***
L1SN×DurMedium	-0.01137	0.00576	-1.97	*
L1SN×DurLong	-0.02868	0.00454	-6.32	***
f0×DurMedium	-0.00006	0.00006	-1.02	0.308
f0×DurLong	-0.00023	0.00006	-3.87	***
Marginal effects on Pr (Outcome = More statement than question)				
Predictors	Marg. Eff.	SE	*t*	*p*
Order	0.00928	0.00337	2.75	**
Age	-0.03021	0.00397	-7.61	***
GenderFemale	-0.01530	0.00746	-2.05	*
StressPenultimate	-0.03171	0.00763	-4.16	***
L1SN	0.24241	0.01463	16.57	***
f0	-0.00512	0.00017	-30.45	***
L1SN×f0	-0.00063	0.00012	-5.17	***
L1SN×DurMedium	-0.02892	0.01526	-1.89	0.058.
L1SN×DurLong	-0.07777	0.01333	-5.83	***
f0×DurMedium	-0.00014	0.00014	-1.02	0.308
f0×DurLong	-0.00056	0.00014	-3.87	***

Marginal effects on Pr (Outcome = Either statement or question)				
Predictors	Marg. Eff.	SE	*t*	*p*
Order	0.00391	0.00141	2.77	**
Age	-0.01272	0.00184	-6.93	***
GenderFemale	-0.00607	0.00281	-2.05	*
StressPenultimate	-0.01332	0.00326	-4.08	***
L1SN	0.034729	0.00596	5.82	***
f0	-0.00216	0.00015	-14.60	***
L1SN×f0	-0.00027	0.00005	-5.08	***
L1SN×DurMedium	-0.01420	0.00860	-1.65	0.099.
L1SN×DurLong	-0.04775	0.01092	-4.37	***
f0×DurMedium	-0.00006	0.00006	-1.02	0.308
f0×DurLong	-0.00024	0.00006	-3.77	***
Marginal effects on Pr (Outcome = More question than statement)				
Predictors	Marg. Eff.	SE	*t*	*p*
Order	-0.01100	0.00396	-2.78	**
Age	0.03581	0.00472	7.59	***
GenderFemale	0.01818	0.00888	2.05	*
StressPenultimate	0.03754	0.00900	4.17	***
L1SN	-0.28640	0.01753	-16.34	***
f0	0.00607	0.00021	29.30	***
L1SN×f0	0.00075	0.00014	5.19	***
L1SN×DurMedium	0.03386	0.01757	1.93	0.054.
L1SN×DurLong	0.08662	0.01313	6.60	***
f0×DurMedium	0.00016	0.00016	1.02	0.308
f0×DurLong	0.00066	0.00017	3.88	***
Marginal effects on Pr (Outcome = Question)				
Predictors	Marg. Eff.	SE	*t*	*p*
Order	-0.00600	0.00217	-2.76	**
Age	0.01954	0.00257	7.60	***
GenderFemale	0.00959	0.00453	2.12	*
StressPenultimate	0.02056	0.00496	4.14	***
L1SN	-0.13706	0.00937	-14.63	***
f0	0.00331	0.00011	30.04	***
L1SN×f0	0.00041	0.00008	5.21	***
L1SN×DurMedium	0.02064	0.01203	1.72	0.086.
L1SN×DurLong	0.06757	0.01530	4.35	***
f0×DurMedium	0.00009	0.00009	1.02	0.308
f0×DurLong	0.00036	0.00009	3.85	***

Appendix D. Results of multiple comparisons of the three duration levels by the L1 group.

	CN group				SN group			
Contrast	*β*	SE	z. ratio	*P*	*β*	SE	z. ratio	*p*
Medium-Short	0.066	0.068	0.98	>.1	0.285	0.098	2.92	< .01
Long-Short	0.303	0.069	4.37	<.0001	0.931	0.099	9.42	< .0001
Long-Medium	0.236	0.069	3.42	<.01	0.647	0.099	6.56	< .0001

Appendix E. Summary of the marginal effects in the OGLM model fitted for Experiment 2 (***$p < 0.001$; ** $p < 0.01$; * $p < 0.05$; '.' $p < 0.1$).

Marginal effects on Pr (Outcome = Statement)

Predictors	Marg. Eff.	SE	*t*	*p*
Age	-0.00488	0.00164	-2.98	**
L1SN	0.13833	0.01782	7.76	***
f0×L1SN	-0.00030	0.00005	-5.61	***
f0×70dB	-0.00004	0.00005	-0.76	0.448
f0×77dB	-0.00020	0.00006	-3.45	***
L1SN×70dB	-0.01510.	0.00780	-1.94	0.052.
L1SN×77dB	0.00327	0.01031	0.32	0.751
L1SN× StressPenultimate	-0.02046	0.00747	-2.74	**
StressPenultimate×70dB	-0.00197	0.00710	-0.28	0.782
StressPenultimate ×77dB	-0.01604	0.00611	-2.63	**
StressPenultimate ×70dB:L1SN	-0.00037	0.01359	-0.03	0.978
StressPenultimate ×77dB×L1SN	-0.01426	0.01090	-1.31	0.190

Marginal effects on Pr (Outcome = More statement than question)

Predictors	Marg. Eff.	SE	*t*	*p*
Age	-0.01125	0.00377	-2.99	**
L1SN	0.22080	0.01735	12.73	***
f0×L1SN	-0.00069	0.00012	-5.69	***
f0×70dB	-0.00010	0.00013	-0.76	0.448
f0×77dB	-0.00045	0.00013	-3.46	***
L1SN×70dB	-0.03680	0.02000	-1.84	0.066.
L1SN×77dB	0.00746	0.02320	0.32	0.748
L1SN× StressPenultimate	-0.05025	0.01937	-2.59	**
StressPenultimate×70dB	-0.00456	0.01656	-0.28	0.783
StressPenultimate ×77dB	-0.03879	0.01540	-2.52	*
StressPenultimate ×70dB:L1SN	-0.00086	0.03143	-0.03	0.978
StressPenultimate ×77dB×L1SN	-0.03488	0.02824	-1.24	0.217

Marginal effects on Pr (Outcome = Either statement or question)

Predictors	Marg. Eff.	SE	*t*	*p*
Age	-0.00667	0.00226	-2.96	**
L1SN	0.04999	0.00660	7.57	***
f0×L1SN	-0.00041	0.00007	-5.67	***
f0×70dB	-0.00006	0.00007	-0.76	0.448
f0×77dB	-0.00027	0.00008	-3.41	***
L1SN×70dB	-0.02615	0.01666	-1.57	0.116
L1SN×77dB	0.00425	0.01269	0.33	0.738
L1SN× StressPenultimate	-0.03682	0.01698	-2.17	*
StressPenultimate×70dB	-0.00276	0.01021	-0.27	0.787
StressPenultimate ×77dB	-0.02695	0.01233	-2.19	*
StressPenultimate ×70dB:L1SN	-0.00051	0.01883	-0.03	0.978
StressPenultimate ×77dB×L1SN	-0.02510	0.02402	-1.04	0.296

Marginal effects on Pr (Outcome = More question than statement)

Predictors	Marg. Eff.	SE	*t*	*p*
Age	0.01456	0.00487	2.99	**
L1SN	-0.28048	0.02174	-12.90	***
f0×L1SN	0.00090	0.00016	5.72	***
f0×70dB	0.00012	0.00016	0.76	0.448
f0×77dB	0.00058	0.00017	3.46	***
L1SN×70dB	0.04654	0.02439	1.91	0.056.
L1SN×77dB	-0.00967	0.03012	-0.32	0.748
L1SN× StressPenultimate	0.06296	0.02293	2.75	**
StressPenultimate×70dB	0.00590	0.02136	0.28	0.783

StressPenultimate ×77dB	0.04921	0.01894	2.60	**
StressPenultimate ×70dB:L1SN	0.00111	0.04064	0.03	0.978
StressPenultimate ×77dB×L1SN	0.04404	0.03425	1.29	0.198
Marginal effects on Pr (Outcome = Question)				
Predictors	Marg. Eff.	SE	*t*	*p*
Age	0.00825	0.00276	2.99	**
L1SN	-0.12864	0.01024	-12.56	***
f0×L1SN	0.00051	0.00009	5.77	***
f0×70dB	0.00007	0.00009	0.76	0.448
f0×77dB	0.00033	0.00010	3.45	***
L1SN×70dB	0.03151	0.02004	1.57	0.115
L1SN×77dB	-0.00531	0.01606	-0.33	0.741
L1SN× StressPenultimate	0.04457	0.02083	2.14	*
StressPenultimate×70dB	0.00339	0.01250	0.27	0.786
StressPenultimate ×77dB	0.03258	0.01485	2.19	*
StressPenultimate ×70dB:L1SN	0.00063	0.02321	0.03	0.978
StressPenultimate ×77dB×L1SN	0.03019	0.02889	1.05	0.296

Appendix F. Results of multiple comparisons of the three intensity levels by L1 group and stress type of the one-word sentences.

		CN group				SN group			
	Contrast	*β*	SE	z. ratio	*p*	*β*	SE	z. ratio	*p*
Paroxytone word	63dB-70dB	0.115	0.095	1.21	> .1	-0.207	0.153	-1.35	> .1
	63dB-77dB	-0.222	0.098	-2.28	> .05	-0.462	0.155	-2.99	< .01
	70dB-77dB	-0.337	0.097	-3.47	<. 01	-0.255	0.155	-1.64	> .1
Oxytone word	63dB-70dB	0.152	0.098	1.56	> .1	-0.163	0.153	-1.07	> .1
	63dB-77dB	0.108	0.099	1.09	> .1	0.168	0.156	1.08	> .1
	70dB-77dB	-0.044	0.098	-0.45	> .1	0.331	0.155	2.14	> .05

Reference

Abrudan, I.-N., Pop, C.-M., & Lazăr, P.-S. (2020). Using a General Ordered Logit Model to Explain the Influence of Hotel Facilities, General and Sustainability-Related, on Customer Ratings. *Sustainability*, *12*(21), 9302. https://doi.org/10.3390/su12219302.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251. https://pubmed.ncbi.nlm.nih.gov/11280420/.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, *23*(2), 425–434. https://doi.org/10.1162/jocn.2009.21362.

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PloS One*, *8*(4), e60676. https://doi.org/10.1371/journal.pone.0060676.

Bidelman, G. M., Krishnan, A., & Gandour, J. T. (2011). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *European Journal of Neuroscience*, *33*(3), 530–538. https://doi.org/10.1111/j.1460-9568.2010.07527.x.

Bishop, J., Kuo, G., & Kim, B. (2020). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics*, *82*. https://doi.org/10.1016/j.wocn.2020.100977.

Boardman, I., Grossberg, S., & Cohen, M. (1994). *Neural Dynamics of Phonetic Trading Relations for Variable-Rate CV Syllables*. Boston University Center for Adaptive Systems and Department of Cognitive and Neutral Systems. https://hdl.handle.net/2144/2175.

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program] (Version 5.3.82). Retrieved from http://www.praat.org.

Bolinger, D. (1978). Intonation across languages. In Joseph H. Greenberg (Ed.), *Universals of Human Language Volume 2: Phonology* (pp. 471–524). Stanford University Press.

Brown, M. K. W. (2020). *Evaluating an Ordinal Output using Data Modeling, Algorithmic Modeling, and Numerical Analysis*. Master thesis. Department of Mathematics and Statistics, Murray State University. https://digitalcommons.murraystate.edu/etd/168.

Carroll, N. (2018). Oglmx: A Package for Estimation of Ordered Generalized Linear Models. R package version 3.0.0.0. https://cran.r-project.org/web/packages/oglmx/index.html.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport*, *18*(18), 1963–1967. https://doi.org/10.1097/wnr.0b013e3282f213c5.

Chang, C. B., & Yao, Y. (2007). Tone production in whispered Mandarin. In *16th International Congress of Phonetic Science*s (pp. 1085–1088). https://escholarship.org/content/qt1581q7qr/qt1581q7qr.pdf.

Chang, S. E. (2013). Effects of Fundamental Frequency and Duration Variation on the Perception of South Kyungsang Korean Tones. *Language and Speech*, *56*(2), 211–228. https://doi.org/10.1177/0023830912443951.

Chen, S. H. (2005). The effects of tones on speaking frequency and intensity ranges in Mandarin and Min dialects. *Journal of the Acoustical Society of America*, *117*(5), 3225–3230. https://doi.org/10.1121/1.1872312.

Chien, P. J., Friederici, A. D., Hartwigsen, G., & Sammler, D. (2020). Neural correlates of intonation and lexical tone in tonal and non-tonal language speakers. *Human Brain Mapping*, *41*(7), 1842–1858. https://doi.org/10.1002/hbm.24916.

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *Journal of the Acoustical Society of America*, *144*(3), EL172–EL177. https://doi.org/10.1121/1.5052025.

Defior, S., & Serrano, F. (2017). Learning to Read Spanish. In Verhoeven, L., & Perfetti, C. (Eds.), *Learning to Read across Languages and Writing Systems* (pp. 243–269). Cambridge University Press. https://assets.cambridge.org/97811070/95885/frontmatter/9781107095885_frontmatter.pdf.

Deroche, M. L. D., Lu, H. P., Kulkarni, A. M., Caldwell, M., Barrett, K. C., Peng, S. C., Limb, C. J., Lin, Y. S., & Chatterjee, M. (2019). A tonal-language benefit for pitch in normally-hearing and cochlear-implanted children. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-018-36393-1.

Divenyi, P., Greenberg, S., & Meyer, G. (Eds.) (2006). Dynamics of speech production and perception (Vol. 374). Amsterdam: Ios Press. https://ebooks.iospress.nl/ISBN/978-1-58603-666-9.

Doherty, C. P., West, W. C., Dilley, L. C., Shattuck-Hufnagel, S., & Caplan, D. (2004). Question/statement judgments: an fMRI study of intonation processing. *Human Brain Mapping*, *23*(2), 85–98. https://dx.doi.org/10.1002%2Fhbm.20042.

Estebas-Vilaplana, E., & Prieto, P. (2010). Castilian Spanish intonation. In P. Prieto & P. Roseano (Eds.), *Transcription of intonation of the Spanish language* (pp. 17–48). München: LINCOM Europa. http://prosodia.upf.edu/home/arxiu/activitats/4th_workshop/protegit/transcription_intonation_spanish_prieto_roseano_lincom.pdf.

Face, T. L. (2005). F0 Peak Height and the Perception of Sentence Type in Castilian Spanish. *Revista Internacional de Lingüística Iberoamericana*, *3*(2), 49–65. https://dialnet.unirioja.es/servlet/articulo?codigo=1342296.

Face, T. L. (2007). The role of intonational cues in the perception of declaratives and absolute interrogatives in Castilian Spanish. *Estudios de Fonética Experimental*, *16*, 186–225.

Face, T. L. (2011). *Perception of Castilian Spanish Intonation: Implications for Intonational Phonology*. München: LINCOM Europa.

Face, T. L., & Prieto, P. (2007). Rising accents in Castilian Spanish: A revision of Sp_ToBI. *Journal of Portuguese Linguistics*, *6*(1), 117–146.

Félix-Brasdefer, J. C. (2010). Data collection methods in speech act performance. In Flor, A. M., & Juan, E. U. (Eds.), *Speech Act Performance: Theoretical, Empirical and Methodological Issues* (Vol. 26, pp. 69–82). Amsterdam: John Benjamins. https://benjamins.com/catalog/lllt.26.

Feng, J., Tao, S., Wu, X., Alsbury, K., & Liu, C. (2019). The effects of amplitude and duration on the perception of English statements vs questions for native English and Chinese listeners. *Journal of the Acoustical Society of America*, *145*(5), EL449–EL455. https://doi.org/10.1121/1.5109046.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological Reviews*, *91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011.

Gandour, J., Larsen, J., Dechongkit, S., Ponglorpisit, S., & Khunadorn, F. (1995). Speech prosody in affective contexts in Thai patients with right hemisphere lesions. *Brain and Language*, *51*(3), 422–443. https://doi.org/10.1006/brln.1995.1069.

Gandour, J. T. (2009). Neural substrates underlying the perception of linguistic prosody. In Gussenhoven, C., & Riad, T. (Eds.), *Experimental Studies in Word and Sentence Prosody* (pp. 3–26). Berlin: De Gruyter Mouton. https://www.degruyter.com/document/doi/10.1515/9783110207576/html.

Gussenhoven, C., & Chen, A. (2000). Universal and language-specific effects in the perception of question intonation. In *6th International Conference on Spoken Language Processing* (pp. 91–94). https://www.isca-speech.org/archive/pdfs/icslp_2000/gussenhoven00_icslp.pdf.

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421. https://doi.org/10.1016/S0095-4470(03)00016-0.

Hattori, K., & Iverson, P. (2010). Examination of the relationship between L2 perception and production: an investigation of English/r/-/l/perception and production by adult Japanese speakers. In *Second Language Studies: Acquisition, Learning, Education and Technology*, 1–4.

Haverkate, H. (2006). Aspectos pragmalingüísticos de la interrogación en español con atención especial a las secuencias de preguntas. *Cultura, Lenguaje y Representación*, 3(3), 27–40. https://www.e-revistes.uji.es/index.php/clr/article/view/1317.

Hodgson, P., & Miller, J. L. (1996). Internal structure of phonetic categories: Evidence for within‐category trading relations. *Journal of the Acoustical Society of America*, *100*(1), 565–576. https://psycnet.apa.org/doi/10.1121/1.415867.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America*, *109*(2), 764–774. https://doi.org/10.1121/1.1339825.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, *119*(5), 3059–3071. https://doi.org/10.1121/1.2188377.

Hunt, T. (2020). ModelMetrics: Rapid calculation of model metrics. R package version 1.2.2.2. https://cran.r-project.org/web/packages/ModelMetrics/ModelMetrics.pdf.

Jacewicz, E., Fox, R. A., & Lyle, S. (2009). Variation in stop consonant voicing in two regional varieties of American English. *Journal of the International Phonetic Association*, *39*(3), 313–334. https://doi.org/10.1017/S0025100309990156.

Jiao, L., & Xu, Y. (2019). Whispered Mandarin has no production-enhanced cues for tone and intonation. *Lingua*, *218*, 24–37. https://doi.org/10.1016/j.lingua.2018.01.004.

Kassambara, A. (2018). *Machine learning essentials: Practical guide in R*. Sthda. https://www.datanovia.com/en/product/machine-learning-essentials-practical-guide-in-r/?url=/5-bookadvisor/54-machine-learning-essentials/.

Kreitewolf, J., Friederici, A. D., & von Kriegstein, K. (2014). Hemispheric lateralization of linguistic prosody recognition in comparison to speech and speaker recognition. *NeuroImage*, *102*, 332–344. https://doi.org/10.1016/j.neuroimage.2014.07.038.

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics*, *23*(1), 81–95. https://doi.org/10.1016/j.jneuroling.2009.09.001.

Kuang, J., & Cui, A. (2018). Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics*, *71*, 194–214. https://doi.org/10.1016/j.wocn.2018.09.002.

Ladd, D. R. (1996). Intonational phonology. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511808814.

Lee, H., Politzer-Ahles, S., & Jongman, A. (2013). Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of Phonetics*, *41*(2), 117–132. https://doi.org/10.1016/j.wocn.2012.12.002.

Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly*, *33*(1), 37–63. https://doi.org/10.2307/3588190.

Liang, J., & Heuven, V. J. (2007). Chinese tone and intonation perceived by L1 and L2 listeners. In C. Gussenhoven & T. Riad (Eds.), *Experimental Studies in Word and Sentence Prosody Volume 2* (pp. 27–62). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110207576.1.27.

Lieberman, P. (1967). Intonation, perception, and language. M.I.T. Research Monograph, 38, xiii, 210.

Lin, M. (2018). *Tone and intonation processing: From ambiguous acoustic signal to linguistic representation* (Doctoral dissertation). Leiden University. https://hdl.handle.net/1887/66615.

Liu, P., & Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavior Research Methods*, *44*(4), 1042–1051. https://doi.org/10.3758/s13428-012-0203-3.

Ma, J. K. Y., Ciocca, V., & Whitehill, T. L. (2008). Acoustic cues for the perception of intonation in Cantonese. In *9th Annual Conference of the International Speech Communication Association* (pp. 520–523). https://www.isca-speech.org/archive_v0/archive_papers/interspeech_2008/i08_0520.pdf.

Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2011). The perception of intonation questions and statements in Cantonese. *Journal of the Acoustical Society of America*, *129*(2), 1012–1023. https://doi.org/10.1121/1.3531840

Matthews, W. J., Stewart, N., & Wearden, J. H. (2011). Stimulus Intensity and the Perception of Duration. *Journal of Experimental Psychology: Human Perception and Performance*, 37(1), 303–313. https://doi.org/10.1037/a0019961.

Meng, Y., Zhang, J., Liu, S., & Wu, C. (2020). Influence of different acoustic cues in L1 lexical tone on the perception of L2 lexical stress using principal component analysis: an ERP study. *Experimental Brain Research*, *238*(6), 1489–1498. https://doi.org/10.1007/s00221-020-05823-w.

Morrow, K., & Liu, C. (2013, June). Intonation perception in English: Effects of stimulus amplitude and listeners′ language background. In *Proceedings of Meetings on Acoustics* (Vol. 19, 1–5). Acoustical Society of America. https://doi.org/10.1121/1.4800156.

Nault, D. R., & Munhall, K. G. (2020). Individual variability in auditory feedback processing: Responses to real-time formant perturbations capacity and their relation to perceptual acuity. *The Journal of the Acoustical Society of America*, *148*(6), 3709–3721. https://doi.org/10.1121/10.0002923.

Niebuhr, O. (2007). Categorical perception in intonation: A matter of signal dynamics? In *INTERSPEECH 8th Annual Conference of the International Speech Communication*

Association (pp. 642–645). https://www.isca-speech.org/archive/pdfs/interspeech_2007/niebuhr07_interspeech.pdf.

Ortega-Llebaria, M., Nemogá, M., & Presson, N. (2017). Long-term experience with a tonal language shapes the perception of intonation in English words: How Chinese-English bilinguals perceive "Rose?" vs. "Rose." *Bilingualism*, *20*(2), 367–383. https://doi.org/10.1017/S1366728915000723.

Ortega-Llebaria, M., & Wu, Z. (2021). Chinese-English Speakers' Perception of Pitch in Their Non-Tonal Language: Reinterpreting English as a Tonal-Like Language. *Language and Speech*, *64*(2), 467–487. https://doi.org/10.1177/0023830919894606

Patel, R., & Grigos, M. I. (2006). Acoustic characterization of the question–statement contrast in 4-, 7- and 11-year-old children. *Speech Communication*, *48*(10), 1308–1318. https://doi.org/10.1016/j.specom.2006.06.007.

Peng, S. C., Chatterjee, M., & Lu, N. (2012). Acoustic Cue Integration in Speech Intonation Recognition With Cochlear Implants. *Trends in Amplification*, *16*(2), 67–82. https://doi.org/10.1177/1084713812451159.

Pohlert, T. (2014). *The pairwise multiple comparison of mean ranks package (PMCMR*). R package. https://cran.microsoft.com/snapshot/2015-03-19/web/packages/PMCMR/vignettes/PMCMR.pdf

Repp, B. H. (1982). Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*(1), 81–110. https://psycnet.apa.org/doi/10.1037/0033-2909.92.1.81.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'mass.' *Cran R*, *538*, 113–120. ftp://192.218.129.11/pub/CRAN/web/packages/MASS/MASS.pdf.

Roca, I. (2019). Spanish Word Stress: an updated multidimensional account. In Goedemans, R., Heinz, J., & van der Hulst, H. (Eds.), *The Study of Word Stress and Accent: Theories, Methods and Data* (pp. 256–292). Cambridge University Press.

Rodero, E. (2011). Intonation and emotion: influence of pitch levels and contour type on creating emotions. *Journal of Voice*, *25*(1), e25–e34. https://doi.org/10.1016/j.jvoice.2010.02.002.

Romera Barrios, L., Fernández-Planas, A. M., Salcioli Guidi, V., Carrera Sabaté, J., & Montes de Oca, D. R. (2007). Una muestra del español de Barcelona en el marco AMPER. *Estudios de Fonética Experimental*, *16*(16), 147–184. https://raco.cat/index.php/EFE/article/view/140051.

Ryalls, J., Dorze, G. Le, Lever, N., Ouellet, L., & Larfeuil, C. (1994). The effects of age and sex on speech intonation and duration for matched statements and questions in French. *Journal of the Acoustical Society of America*, *95*(4), 2274–2276. https://doi.org/10.1121/1.408639

Seddoh, S. A. (2002). How discrete or independent are "affective prosody" and "linguistic prosody"? *Aphasiology*, *16*(7), 683–692. https://doi.org/10.1080/02687030143000861.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273–293. https://doi.org/10.1177/0023830909357156.

Steppling, M. L., & Montgomery, A. A. (2002). Perception and production of rise-fall intonation in American English. *Perception and Psychophysics*, *64*(3), 451–461. https://doi.org/10.3758/BF03194717.

Syntrillium Software Corporation. (2003). Cool Edit Pro. Version 2.1. Phoenix.

R Core Team (2021). R: A language and environment for statistical computing. Version 1.4.1103 Foundation for Statistical Computing, Vienna, Austria. Available at: www.R-project. org.

Tillman, G., Benders, T., Brown, S. D., & Ravenzwaaij, D. van. (2017). An evidence accumulation model of acoustic cue weighting in vowel perception. *Journal of Phonetics*, *61*, 1–12. https://doi.org/10.1016/j.wocn.2016.12.001.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. In *Cognitive Science* (Vol. 34, Issue 3, pp. 434–464). https://doi.org/10.1111/j.1551-6709.2009.01077.x.

Tsukada, K. (2018). The Perception of Mandarin Lexical Tones by Native Speakers of Burmese. *Language and Speech*, *62*(4), 1–16. https://doi.org/10.1177/0023830918806550.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *Journal of the Acoustical Society of America*, *122*(4), 2306–2319. https://doi.org/10.1121/1.2773966.

Wichmann, A., & Blakemore, D. (2006). The prosody-pragmatics interface. *Journal of Pragmatics*, *38*(10), 1537–1541. https://doi.org/10.1016/j.pragma.2006.02.009.

Williams, R. (2016). Understanding and interpreting generalized ordered logit models. *Journal of Mathematical Sociology*, *40*(1), 7–20. https://doi.org/10.1080/0022250X.2015.1112384.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America*, *120*(2), 1063-1074. https://doi.org/10.1121/1.2213572.

Yuan, J. (2004). *Intonation in Mandarin Chinese: Acoustics, perception, and computational modeling* (Doctoral dissertation). New York: Cornell University.

Yuan, J. (2006). Mechanisms of question intonation in Mandarin. In M. B. Huo Q., C. ES., & L. H. (Eds.), *International Symposium on Chinese Spoken Language Processing* (pp. 19-30). Lecture Notes in Computer Science, vol 4274. Berlin, Heidelberg: Springer. https://doi.org/10.1007/11939993_7.

Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America*, *130*(6), 4063–4069. https://doi.org/10.1121/1.3651818.

Zarate-Sandez, G. A., Morales-Front, A., & Sanz, C. (2015). *Perception and production of intonation among English-Spanish bilingual speakers at different proficiency levels* (Doctoral dissertation). Washington: Georgetown University. https://repository.library.georgetown.edu/handle/10822/761510.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1087–1104.

Anexo C

Corpus, contextos, cuestionarios e informantes del experimento de producción y percepción

Anexo C1: Formulario de autorización

UNIVERSITAT DE BARCELONA

Facultat de Filologia
Gran Via de les Corts Catalanes, 585
08007 Barcelona
Tel. +34 934 035 650
labphon@ub.edu
http://stel.ub.edu/labfon

LF Laboratori de Fonètica

D/Dª _____, mayor de edad, provisto/a de DNI/NIE núm. _____ con domicilio en _____, nacido/a en _____ el día ___ / ___ / _____

MANIFIESTO

I.- Que, a petición de la Sra. Peizhu Shang (DNI: E84839638), investigadora predoctoral del programa de doctorado "Ciencia Cognitiva y Lenguaje" de la Universitat de Barcelona, **acepto participar como informante en los estudios que darán lugar a su tesis doctoral**.

II.- Que con esta finalidad autorizo a la Universitat de Barcelona para que, a través del Laboratori de Fonètica de esta universidad, pueda grabar mi voz, y reproducirla, editarla y distribuirla, en cualquier formato y por cualquier medio, con fines exclusivamente docentes y de investigación o divulgación y sin ánimo de lucro.

Los datos personales del estudio, así como las voces grabadas, se podrán hacer públicas, sin ánimo de lucro, en actividades de ámbito académico y de investigación no lucrativas, así como en la tesis doctoral de la interesada.

Y para que así conste a todos los efectos oportunos, firmo la presente autorización en

Barcelona, el día ___ / ___ / _____.

_____
(Nombre y firma)

Anexo C2: Corpus y contextos

Este anexo incluye las situaciones contextuales usadas para elicitar el corpus del experimento de producción y percepción. Los diálogos de interacción se presentan a continuación en español y en chino.

a. Corpus y contextos para el experimento de producción

1) **Contextos para las preguntas totales informativas**

- ENTREVISTADOR: Estás en una frutería y cuando vas a pagar, te das cuenta de que sólo llevas un billete de 100 euros. Te preocupa que el dependiente no tenga cambio de un billete tan grande y le preguntas si tiene cambio. （Traducción en chino: 你在水果店正准备付钱的时候，发觉自己身上只带了一张一百欧的纸币。你担心店员找不开这么大面额的钱，于是主动向他询问是否有零钱找零。）

  RESPUESTA: *¿Tiene cambio?* （Traducción en chino: 您有零钱吗？）

- ENTREVISTADOR: Entras en una frutería donde no has estado nunca y le preguntas al dependiente si tiene mandarinas. （Traducción en chino: 你去了一个之前从没有去过的水果店，然后你向售货员询问他们店里是否有苹果。）

  RESPUESTA: *¿Tiene mandarinas?* （Traducción en chino: 您这儿有柑橘吗？）

- ENTREVISTADOR: Entras en una frutería donde no has estado nunca y le preguntas al dependiente si tiene melón. （Traducción en chino: 你去了一个之前从没有去过的水果店，然后你向售货员询问他们店里是否有西瓜。）

  RESPUESTA: *¿Tiene melón?* （Traducción en chino: 您这儿有柑橘吗？）

2) **Contextos para las preguntas parciales informativas**

- ENTREVISTADOR: Has quedado con dos buenos amigos para ir al cine a las seis de la tarde. Pero al llegar la hora, sólo se presenta Andrés y no ves al otro amigo. Pregunta a Andrés cuándo viene el otro amigo. （Traducción en chino: 你和两个

好朋友约好了下午六点去看电影。但到了约定的时间，只有安德列斯来了，另一个却没有到。于是你向安德列斯询问另一个朋友他什么时候来。）

RESPUESTA: *¿Cuándo viene?* （Traducción en chino: 他什么时候来呢？）

- ENTREVISTADOR: Has quedado con dos buenas amigas para hacer compras esta tarde. Pero al llegar la hora, sólo se presenta una chica y no ves a la otra amiga, Marina. Pregúntale dónde está Marina. （Traducción en chino: 你和两个好朋友约好了今天下午一起购物。但到了约定的时间，只有一个女生来了。于是你像她询问另一个朋友玛丽娜在哪里。）

  RESPUESTA: *¿Dónde está Marina?* （Traducción en chino: 玛丽娜在哪里？）

- ENTREVISTADOR: Has quedado con dos buenos amigos para hacer compras esta tarde. Pero al llegar la hora, sólo se presenta un chico y no ves al otro, Manuel. Pregúntale dónde está Manuel. （Traducción en chino: 你和两个好朋友约好了今天下午一起购物。但到了约定的时间，只有一个男生来了。于是你像他询问另一个朋友马努埃尔在哪里。）

  RESPUESTA: *¿Dónde está Manuel?* （Traducción en chino: 马努埃尔在哪里呢？）

3) **Contextos para las preguntas disyuntivas informativas**

- ENTREVISTADOR: Has invitado a buen amigo a tu piso para una cena. Después de acabar los platos principales, le preguntas si quiere tarta o helado de postre. （Traducción en chino: 你邀请了一个好朋友去你家吃晚餐。吃完主菜后，你向他询问甜点是想要吃蛋糕还是冰激凌。）

  RESPUESTA: *¿Quieres tarta o helado?* （Traducción en chino: 你想吃蛋糕还是冰激凌？）

- ENTREVISTADOR: El propietario de tu piso te dice que la bañera está rota y que

va a venir más tarde un trabajador para repararla. Como el martes no podrás estar en casa, pregunta al propietario si el trabajador viene el lunes o el martes. （Traducción en chino: 房东告诉你浴室的水管设备坏了，稍后会有一位工人来进行修缮。但因为周二你不在家，你向房东询问工人是周一来还是周二来。）

RESPUESTA: *¿Viene el lunes o el martes?* （Traducción en chino:他周一来还是周二来？）

- ENTREVISTADOR: Has invitado a buen amigo a tu piso para una cena. Después de acabar los platos principales, le preguntas si quiere melón o melocotón de postre. （Traducción en chino: 你邀请了一个好朋友去你家吃晚餐。吃完主菜后，你问她甜点是想要吃西瓜还是桃子。）

  RESPUESTA: *¿Quieres melón o melocotón?* （Traducción en chino: 你想吃西瓜还是桃子？）

4) **Contextos para las preguntas totales confirmatorias**

- ENTREVISTADOR: Estás hablado con tu compañero de piso y ves que está cubierto con dos mantas. Tú infieres que tu compañero de piso de debe de tener frío y le preguntas si es así. （Traducción en chino: 你正在和你的室友讲话，你瞧见他身上披了三条毯子。你猜想他应该是有些冷，于是你向他发问确认。）

  RESPUESTA: *¿Tienes frío?* （Traducción en chino: 你是不是冷？）

- ENTREVISTADOR: Estás haciendo los deberes de Matemáticas con un buen amigo en la biblioteca. Al llegar el mediodía, oyes que su estómago hace ruido de vez en cuando, tú infieres que debe de tener hambre y le preguntas si es así. （Traducción en chino: 你和一个好朋友在图书馆做英语作业。中午到了，你听见他的肚子不停地发出响声。你推测他应该是饿了，于是你向他询问确认。）

  RESPUESTA: *¿Tienes hambre?* （Traducción en chino: 你是不是饿了？）

- ENTREVISTADOR: Estás hablado con tu compañero de piso y te das cuenta de

que tiene mucho sudor en la frente. Tú infieres que debe de tener calor y le preguntas si es así. （Traducción en chino: 你正在和室友讲话，你发现他额头上有很多汗。你猜想他应该是有点热，于是你向他询问确认。）

RESPUESTA: *¿Tienes calor?* （Traducción en chino: 你是不是热？）

5) **Contextos para las preguntas confirmatorias de apéndice**

- ENTREVISTADOR: Te encuentras con un tu compañero de piso en la puerta de tu edificio y ves que sale con un carrito de la compra. Tú infieres que debe de salir al mercado y le preguntas si es así. （Traducción en chino: 你在楼下的大门口碰见了你的室友，你看见他带着一个购物小推车准备出去。你推测他应该是要去超市，于是你向他发问以证实你的猜想。）

  RESPUESTA: *Sales al mercado, ¿no?* （Traducción en chino: 你这是要去超市，是嘛？）

- ENTREVISTADOR: Vas a una discoteca con un amigo y ves que baila muy bien, así que supones que practica baile a menudo (mucho) y le preguntas si es así. （Traducción en chino: 你和一个好朋友一起去舞厅，你看他跳舞跳得非常好。你推测他应该是经常跳舞，于是你向他发问以证实你的猜想。）

  RESPUESTA: *Bailas a menudo, ¿verdad?* （Traducción en chino: 你经常跳舞，对吧？）

- ENTREVISTADOR: Antes de ir a la universidad, encuentras que tu compañero de piso está tendido en el sofá y tiene un cara roja y llena de sudor. Tú infieres que debe de estar enfermo y le preguntas si es así. （Traducción en chino: 在去学校前，你发现你的室友躺在沙发上满头大汗，脸色红的异常。你推测他应该是不舒服，于是你向他发问确认。）

  RESPUESTA: *No te encuentras bien, ¿verdad?* （Traducción en chino: 你不舒服，是吗？）

b. Corpus y contextos para el experimento de percepción

1) **Contextos para las declarativas neutras**

- ENTREVISTADOR: ¿Cuál es la capital de Andalucía? (Tiene que ser un extranjero que emita la pregunta, porque si no, existe el riesgo de que salga con entonación de obviedad)
  INFORMANTE: *Sevilla*.

- ENTREVISTADOR: ¿Qué hace tu prima la semana que viene?
  INFORMANTE: *Viene a Sevilla*.

- ENTREVISTADOR: ¿Dónde nació el novelista Cervantes?
  INFORMANTE: *Alcalá*.

- ENTREVISTADOR: ¿Qué hace tu prima la semana que viene?
  INFORMANTE: *Viene a Alcalá*.

2) **Contextos para las interrogativas totales neutras** en que uno pide informaciones sobre todo el evento, mientras que el otro, da una respuesta confirmatoria

- INFORMANTE: *¿Sevilla?*
  ENTREVISTADOR: Sí, Sevilla.

- INFORMANTE: *¿Viene a Sevilla?*
  ENTREVISTADOR: Sí, viene a Sevilla.
- INFORMANTE: *¿Alcalá?*
  ENTREVISTADOR: Sí, Alcalá.

- INFORMANTE: *¿Viene a Alcalá?*
  ENTREVISTADOR: Sí, viene a Alcalá.

Anexo C3: Cuestionario sociolingüístico para el experimento de producción

Este anexo contiene los cuestionarios elaborados en español y en chino para recopilar los datos sociolingüísticos de los informantes españoles y chinos, respectivamente, del experimento de producción.

a. Cuestionario para los hablantes nativos de castellano

Queridos amigos/gas,

¡Hola!

Soy investigadora predoctoral del Laboratorio de Fonética de la Universitat de Barcelona. Ahora estamos proponiendo un estudio experimental del español peninsular y por favor, necesitamos que nos ayudes a completar el grupo de control como hablantes nativos. Si aceptas colaborar, tendrás que rellenar el siguiente formulario ante de emprender el experimento de producción. ¡ Ojo! Este cuestionario no es un examen ni una prueba, por ello, te agradecemos que seas sincero y tomes la libertad en contestar.

Los datos del estudio se utilizarán con fines estrictamente docentes y de investigación y sin ánimo de lucro.

El cuestionario y el experimento te ocuparán aproximadamente treinta minutos.

¡MUCHAS GRACIAS por tu participación!

Saludos cordiales,

Peizhu Shang

Octubre de 2019

1. Nombre y apellidos: .....
2. Edad: .....
3. Sexo:
   - □ Femenino
   - □ Masculino
4. Lugar de procedencia (ciudad, provincia y país): .....

5. Domicilio actual: .....
6. Lengua(s) materna(s): .....
7. Lengua que se usa más en su vida diaria (si es algún dialecto, indícalo): .....
8. Lugar de nacimiento de la madre y lengua: .....
9. Lugar de nacimiento del padre y lengua: .....
10. Profesión actual: .....
11. Nivel de estudios:
    - □ Grado
    - □ Máster
    - □ Doctorado
    - □ Otros
12. ¿Qué más lenguas extranjeras dominas o hablas en su vida o trabajo (según el orden temporal del aprendizaje)? .....
13. ¿Tienes algún problema de audición o de habla?
    - □ Sí
    - □ No

..... (lugar), el día ..... / ..... / .....

….....

(Nombre y firma)

b. Cuestionario para los aprendices chinos del español peninsular

亲爱的朋友们：

您们好！

我是巴塞罗那大学语音实验室的博士研究人员。我们现在进行一项关于伊比亚半岛西班牙语的研究，十分需要以汉语为母语的西班牙语学习者帮助我们完成实验。如果您同意合作，麻烦您请在正式实验开始之前认真填写以下问卷。请注意，该问卷既不是考试也不是测验，所以请您诚实自由地回答下列空白问题。

调查数据将严格用于教学和研究目的，并在非营利的基础上使用。我们保证您的个人信息和数据不会被泄漏。

问卷填写和实验完成共计大约需要 30 分钟。

十分感谢您的参与！

祝好，

尚培珠

2019 年 10 月

1. 姓名：.....
2. 年龄：.....
3. 性别：
   - □ 女
   - □ 男
4. 出生地（国家、省份及城市）：.....
5. 现居住地址：.....
6. 母语：.....
7. 在家中使用频率最高的语言（如为方言，请详细指出）：.....
8. 职业：.....
9. 教育程度：
   - □ 本科

□ 硕士

□ 博士

□ 其他

10. 您从几岁开始学习西班牙语？.....

11. 截止目前，您在西班牙一共生活了多少个月？.....

12. 您是否具有 DELE 语言证书，如果有，请指明其具体等级：.....

13. 如果不具有 DELE 证书，请根据欧洲共同语言参考标准自我衡量您目前的西班牙语语言水平（A1-C2）：.....

14. 您在学习或工作生活中还会说其他语言吗？如果有，请具体指明：.....

15. 您是有任何听力或者嗓音问题吗？

□ 有

□ 没有

16. 您是否会演奏某种乐器？如果有，请具体指明：.....

17. 您在生活中经常唱歌吗？

□ 经常

□ 很少

□ 几乎不唱

..... （地点），..... 年 ..... 月 ..... 日

.....

(签字)

Anexo C4: Cuestionario del Cociente de Espectro Autista para los adultos

Este anexo incluye los ítems de la subescala de comunicación del Cociente de Espectro Autista (Autism Spectrum Quotient, abreviado AQ) para los adultos (Baron-Cohen, Wheelwright, Hill, et al., 2001). Los diez ítems, originalmente en inglés, se han traducido al español y al chino para facilitar la comprensión de los participantes de este estudio. En general, el grado del fenotipo del autismo relacionado a la comunicación es mayor (y, por tanto, menor en la habilidad pragmática) cuando un participante responde "acuerdo parcial" y "acuerdo total" a las preguntas 1-6, o "desacuerdo parcial" y "desacuerdo total" a las preguntas 7-10.

a. Los diez ítems en español:

		Acuerdo total	**Acuerdo parcial**	**Desacuerdo parcial**	**Desacuerdo total**
1	A menudo otras personas me dicen que lo que he dicho es maleducado, a pesar de que yo en realidad no creo que sea así.	○	○	○	○
2	Me gusta la cháchala social.	○	○	○	○
3	Cuando yo hablo no siempre es fácil para los demás meter la pata.	○	○	○	○
4	A menudo me ocurre que no sé cómo mantener una conversación. con otra persona.	○	○	○	○
5	Me resulta fácil "leer entre líneas" o captar el doble sentido, cuando alguien me está hablando.	○	○	○	○
6	Soy capaz de darme cuenta si una persona que me está escuchando se aburre.	○	○	○	○
7	Cuando hablo por teléfono no estoy seguro de cuando es mi turno para hablar.	○	○	○	○
8	A menudo soy el último en entender la gracia de un	○	○	○	○

	chiste.				
9	Soy bueno en las charlas sociales.	○	○	○	○
10	La gente a menudo me dice que vuelvo una y otra vez sobre el mismo tema.	○	○	○	○

b. Los diez ítems en chino:

		完全同意	稍微同意	稍微不同意	完全不同意
1	虽然我自认为很有礼貌，但还是经常被其他人告知我说了不礼貌的话。	○	○	○	○
2	我喜欢社交闲谈。	○	○	○	○
3	当我说话时，别人常常不是很容易能插得上话。	○	○	○	○
4	我时常发现我不知如何使对话持续下去。	○	○	○	○
5	与人谈话时，我能很轻易地察觉对方的言外之意。	○	○	○	○
6	我知道如何辨别对方是否已经厌倦听我说话。	○	○	○	○
7	当我讲电话时，我不太确定什么时候该我接话	○	○	○	○
8	我常常是最后一个捕捉到笑话笑点的人。	○	○	○	○
9	我擅长社交闲谈。	○	○	○	○
10	我常被告知总是重复地说同样的事。	○	○	○	○

Anexo C5: Estructura y diseño de la encuesta de percepción

Este anexo incluye las capturas de pantalla que ilustran el contenido y la estructura de la encuesta perceptiva. La encuesta cuenta con varias secciones que se dedican a la recopilación de los datos sociolingüísticos de los oyentes, los ítems del AQ y las tres tareas de percepción. La encuesta se administra en el idioma preferido de cada participante (chino o español). A continuación, se presenta la versión en español.

(a) Enlace para la encuesta perceptiva en chino mandarín: https://survey.alchemer.com/s3/6035555/ExperimentPercepCH-3task

(b) Enlace y capturas de pantalla para la encuesta perceptiva en español: https://survey.alchemer.com/s3/6034068/ESpercep-3task

Datos personales

1. Nombre y apellido *

2. Edad *

3. Sexo *

- ◯ Mujer
- ◯ Hombre
- ◯ Ninguno/Prefiero no decirlo

4. Ciudad de procedencia en España *

5. Además del español, ¿hablas alguna otra lengua? Indícalas si tienes un nivel inicial, intermedio o avanzado. Escribe las otras lenguas que conoces con su nivel en la ventana "Comments".
*

Inglés	-- Please Select --
Francés	-- Please Select --
Alemán	-- Please Select --
Italiano	-- Please Select --
Portugués	-- Please Select --

Comments

6. ¿Tienes algún problema de audición o de habla? *

- ◯ Sí
- ◯ No

7. ¿Has estudiado música o canto en una escuela de música? *

- ◯ Sí
- ◯ No

8. ¿Tocas algún instrumento musical? *

- ◯ Sí
- ◯ No

9. ¿Con qué frecuencia escuchas música? *

- ◯ Diariamente
- ◯ 5-6 días por semana
- ◯ 3-4 días per semana
- ◯ 1-2 días per semana
- ◯ Ocasionalmente
- ◯ Nunca

Next

17%

Save and continue later ▾

Experimento perceptivo

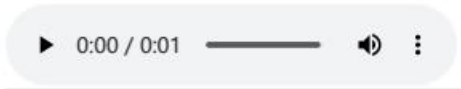
Gating task

A continuación oirás un total de 28 frases. Algunas las oirás completas, hasta el final, mientras que de otras escucharás solo una parte. Mientras los estés escuchando, dispondrás del texto completo (sin puntuación) en la pantalla. Después de escuchar el audio, tendrás que decidir si has oído una declarativa o interrogativa.

¡Atención! La respuesta de declarativa o interrogativa no siempre aparecen en el mismo orden. Asegúrate de leer la respuesta antes de pinchar.

¿LISTO? ¡¡VAMOS!! (La página y los audios pueden tardar un poco en cargar.)

***Ejemplo**: Pulsa "Play" para escuchar la frase "Viene a Sevilla".



¿Qué tipo oracional crees que se corresponde el fragmento escuchado? *

◯ Interrogativa ◯ Declarativa

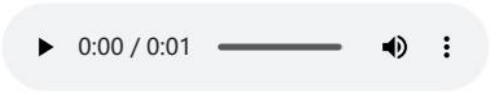
Ahora empezamos con las frases cortadas.

1. Pulsa "Play" para escuchar la frase "Viene a Alcalá".

¿Qué tipo oracional crees que se corresponde el fragmento escuchado? *

◯ Interrogativa ◯ Declarativa

2. Pulsa "Play" para escuchar la frase "Sevilla".



¿Qué tipo oracional crees que se corresponde el fragmento escuchado? *

◯ Declarativa ◯ Interrogativa

(……)

28. Pulsa "Play" para escuchar la frase "Viene a Sevilla".

¿Qué tipo oracional crees que se corresponde el fragmento escuchado? *

◯ Declarativa ◯ Interrogativa

¡Muy bien! ¡Ya has terminado esta tarea!

Ahora, si te apetece, puedes descansar un rato.

Pulsa "**Next**" para continuar. ¡ÁNIMO!

Next

33%

Save and continue later ▾

Experimento perceptivo

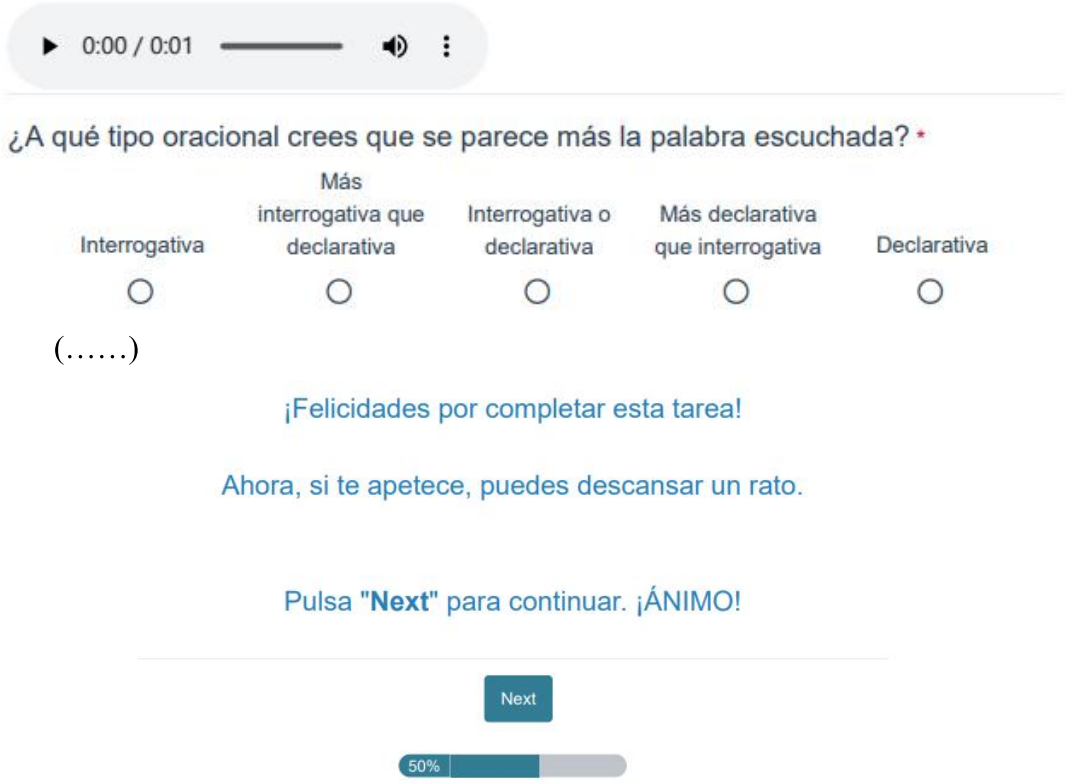
F0_dur task

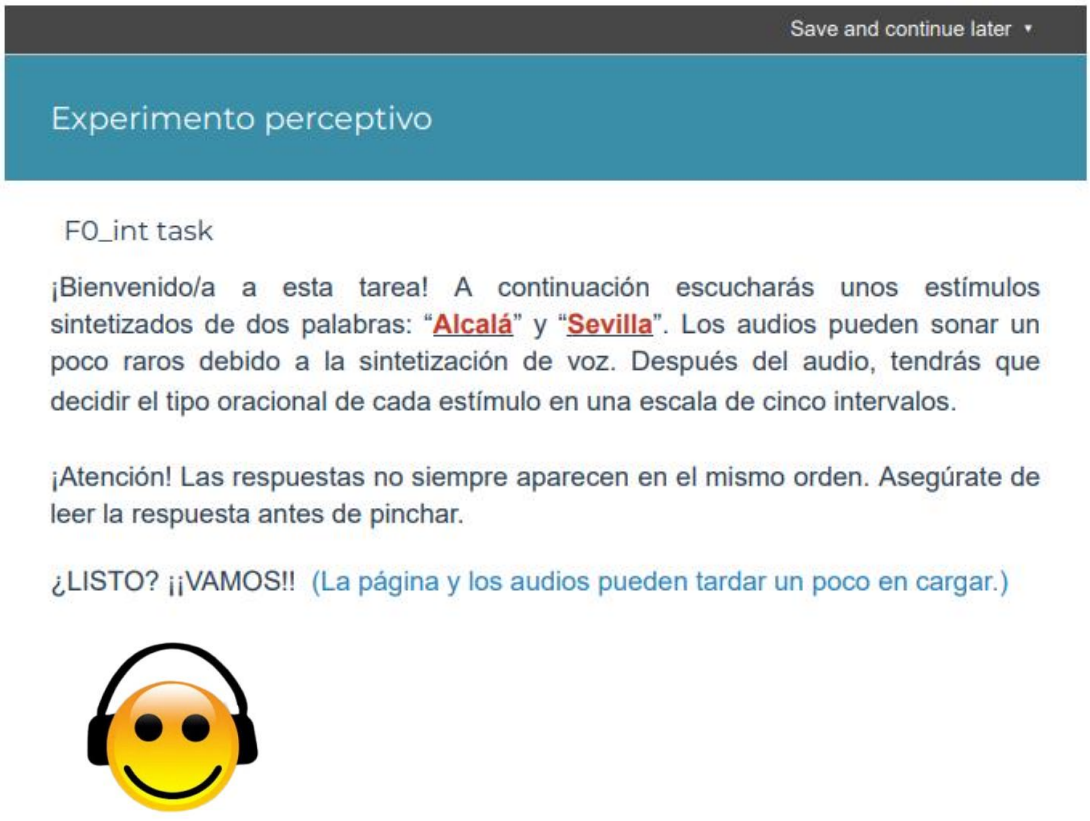
¡Bienvenido/a a esta tarea! A continuación escucharás unos estímulos sintetizados de dos palabras: "**Alcalá**" y "**Sevilla**". Los audios pueden sonar un poco raros debido a la sintetización de voz. Después del audio, tendrás que decidir el tipo oracional de cada estímulo en una escala de cinco intervalos.

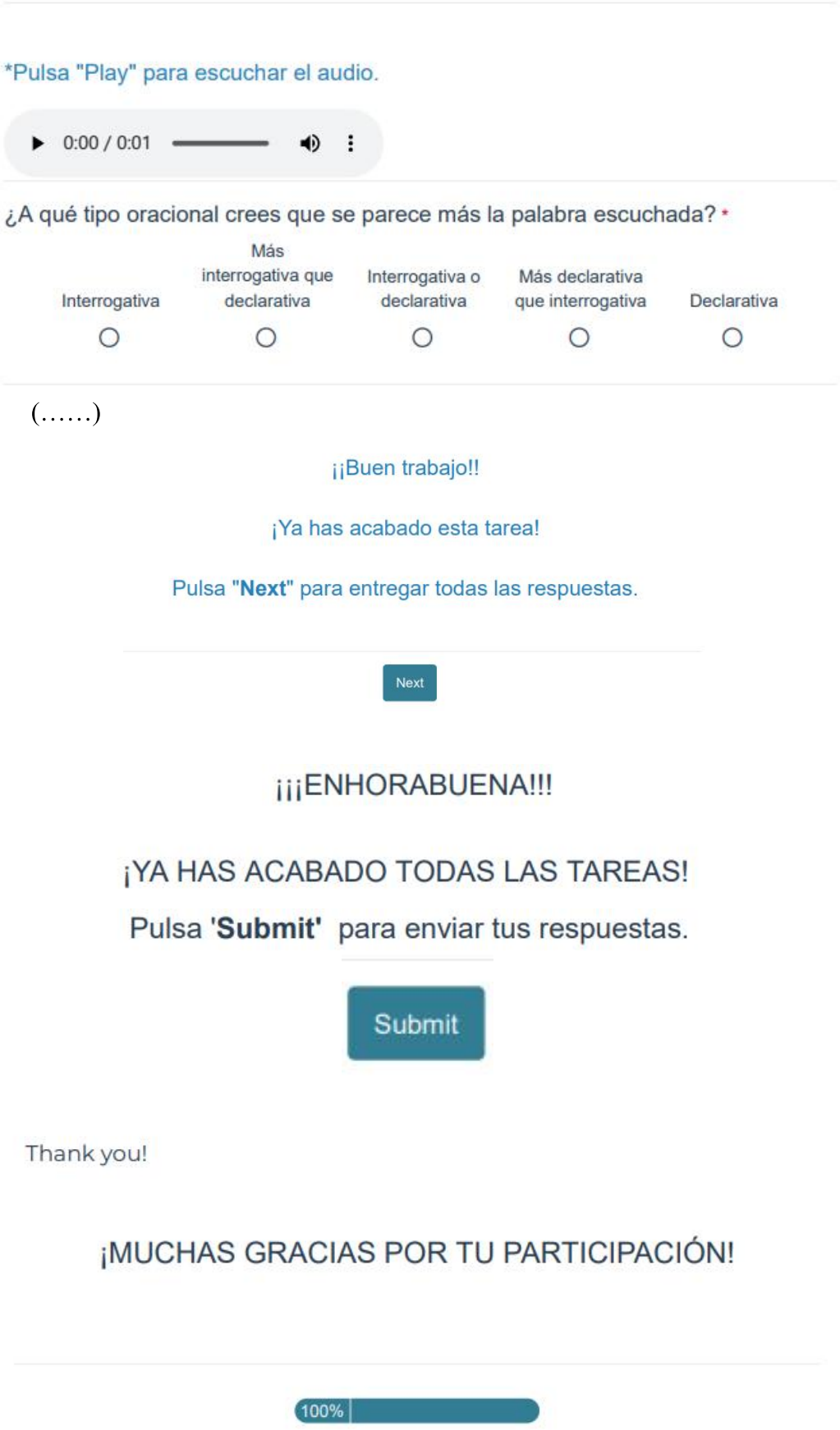
¡Atención! Las respuestas no siempre aparecen en el mismo orden. Asegúrate de leer la respuesta antes de pinchar.

¿LISTO? ¡¡VAMOS!! (La página y los audios pueden tardar un poco en cargar.)

*Pulsa "Play" para escuchar el audio.


0:00 / 0:01
¿A qué tipo oracional crees que se parece más la palabra escuchada? *
Interrogativa
Más interrogativa que declarativa
Interrogativa o declarativa
Más declarativa que interrogativa
Declarativa
(……)
¡Felicidades por completar esta tarea!
Ahora, si te apetece, puedes descansar un rato.
Pulsa "Next" para continuar. ¡ÁNIMO!
Next
50%


Save and continue later
Experimento perceptivo
F0_int task
¡Bienvenido/a a esta tarea! A continuación escucharás unos estímulos sintetizados de dos palabras: "Alcalá" y "Sevilla". Los audios pueden sonar un poco raros debido a la sintetización de voz. Después del audio, tendrás que decidir el tipo oracional de cada estímulo en una escala de cinco intervalos.
¡Atención! Las respuestas no siempre aparecen en el mismo orden. Asegúrate de leer la respuesta antes de pinchar.
¿LISTO? ¡¡VAMOS!! (La página y los audios pueden tardar un poco en cargar.)


*Pulsa "Play" para escuchar el audio.
0:00 / 0:01
¿A qué tipo oracional crees que se parece más la palabra escuchada? *
Interrogativa
Más interrogativa que declarativa
Interrogativa o declarativa
Más declarativa que interrogativa
Declarativa
(……)
¡¡Buen trabajo!!
¡Ya has acabado esta tarea!
Pulsa "Next" para entregar todas las respuestas.
Next
¡¡¡ENHORABUENA!!!
¡YA HAS ACABADO TODAS LAS TAREAS!
Pulsa 'Submit' para enviar tus respuestas.
Submit
Thank you!
¡MUCHAS GRACIAS POR TU PARTICIPACIÓN!
100%

Anexo C6: Lista de los estímulos del test de percepción fonológica

Este anexo comprende la transcripción ortográfica de los estímulos creados para el test de percepción fonológica junto a su código de identificación correspondiente. El primer dígito (D o I) corresponde a la modalidad oracional (declarativa o interrogativa total). El segundo (1 o 2) corresponde al número de palabras acentuadas del enunciado. El penúltimo (A o L) se codifica el tipo acentual de la última palabra (aguda o llana), y el último (1, 2, 3, 4) indica el número de *gates* del estímulo.

Código	**Estímulo**
D1_A1	Alca…
D1_A2	Alcalá.
D1_L1	Sevi…
D1_L2	Sevilla.
D2_A1	Vie…
D2_A2	Viene a al…
D2_A3	Viene a alca…
D2_A4	Viene a Alcalá.
D2_L1	Vie…
D2_L2	Viene a…
D2_L3	Viene a Sevi…
D2_L4	Viene a Sevilla.
I1_A1	¿Alca…?
I1_A2	¿Alcalá?
I1_L1	¿Sevi…?
I1_L2	¿Sevilla?
I2_A1	¿Vie…?
I2_A2	¿Viene a al…?
I2_A3	¿Viene a alca…?
I2_A4	¿Viene a Alcalá?
I2_L1	¿Vie…?
I2_L2	¿Viene a…?
I2_L3	¿Viene a Sevi…?
I2_L4	¿Viene a Sevilla?

Anexo C7: Lista de los estímulos de los test de percepción fonético-acústica

En este anexo se facilita la lista de códigos para identificar la condición acústica de los estímulos creados para los test perceptivos de carácter fonético-acústico. El primer dígito (A o S) representa el contenido ortográfico del enunciado (Alcalá o Sevilla). Desde el segundo hasta el antepenúltimo se codifica el valor de manipulación de la F0 aplicado al enunciado (desde -20 Hz hasta +180 Hz). El último en el anexo a (c, m o l) corresponde al nivel de duración de la sílaba final (corta, media o larga), mientras en el anexo b corresponde al nivel de intensidad de la sílaba final (63 dB, 70 dB o 77 dB).

a. Códigos para los estímulos que varían en la F0 y la duración

A-20_c	A-20_m	A-20_l
A+0_c	A+0_m	A+0_l
A+20_c	A+20_m	A+20_l
A+40_c	A+40_m	A+40_l
A+60_c	A+60_ m	A+60_l
A+80_c	A+80_m	Al+80_l
A+100_c	A+100_m	A+100_l
A+120_c	A+120_m	A+120_l
A+140_c	A+140_m	A+140_l
A+160_c	A+160_m	Al+160_l
A+180_c	A+180_m	A+180_l
S-20_c	S-20_m	S-20_l
S+0_c	S+0_m	S+0_l
S+20_c	S+20_m	S+20_l
S+40_c	S+40_m	S+40_l
S+60_c	S+60_m	S+60_l
S+80_c	S+80_m	S+80_l
S+100_c	S+100_m	S+100_l
S+120_c	S+120_m	S+120_l
S+140_c	S+140_m	S+140_l
S+160_c	S+160_m	S+160_l
S+180_c	S+180_m	S+180_l

b. Códigos para los estímulos que varían en la F0 y la intensidad

A-20_63	A-20_70	A-20_77
A+0_63	A+0_70	A+0_77
A+20_63	A+20_70	A+20_77
A+40_63	A+40_70	A+40_77
A+60_63	A+60_70	A+60_77
A+80_63	A+80_70	A+80_77
A+100_63	A+100_70	A+100_77
A+120_63	A+120_70	A+120_77
A+140_63	A+140_70	A+140_77
A+160_63	A+160_70	A+160_77
A+180_63	A+180_70	A+180_77
S-20_63	S-20_70	S-20_77
S+0_63	S+0_70	S+0_77
S+20_63	S+20_70	S+20_77
S+40_63	S+40_70	S+40_77
S+60_63	S+60_70	S+60_77
S+80_63	S+80_70	S+80_77
S+100_63	S+100_70	S+100_77
S+120_63	S+120_70	S+120_77
S+140_63	S+140_70	S+140_77
S+160_63	S+160_70	S+160_77
S+180_63	S+180_70	S+180_77

Anexo C8: Informantes del test de percepción fonológica

Este anexo incluye las informaciones básicas de los informantes que han participado en el test perceptivo de los patrones entonativos fonológicos del español.

a. Perfiles de los informantes nativos de español

ID	Sexo	Edad	Experiencia del aprendizaje musical	Tocar instrumento musical	Frecuencia de escuchar música
1	F	31	FALSE	FALSE	1
2	M	27	TRUE	TRUE	5
3	F	26	FALSE	FALSE	5
4	F	54	FALSE	FALSE	2
5	F	37	FALSE	FALSE	2
6	M	22	FALSE	FALSE	5
7	M	26	FALSE	FALSE	5
8	F	27	FALSE	FALSE	3
9	M	46	TRUE	FALSE	4
10	F	45	FALSE	FALSE	5
11	F	46	TRUE	TRUE	4
12	M	55	FALSE	FALSE	1
13	F	41	TRUE	FALSE	3
14	F	35	FALSE	FALSE	4
15	M	40	FALSE	FALSE	4
16	F	24	FALSE	FALSE	3
17	F	26	FALSE	FALSE	5
18	F	30	FALSE	TRUE	5
19	F	24	FALSE	FALSE	5
20	F	23	FALSE	FALSE	5
21	F	34	FALSE	FALSE	4
22	F	38	FALSE	FALSE	1
23	F	30	TRUE	TRUE	5
24	F	41	FALSE	FALSE	2
25	F	26	FALSE	FALSE	3
26	F	39	FALSE	FALSE	4
27	M	20	FALSE	TRUE	5
28	M	43	FALSE	FALSE	5
29	F	37	FALSE	FALSE	5
30	F	20	FALSE	FALSE	5
31	F	29	FALSE	FALSE	5
32	M	34	FALSE	FALSE	5
33	M	49	FALSE	FALSE	5
34	F	47	FALSE	FALSE	5
35	M	59	FALSE	FALSE	2
36	F	25	FALSE	FALSE	5
37	M	34	TRUE	TRUE	5

38	F	33	FALSE	FALSE	2
39	M	28	FALSE	FALSE	4
40	F	33	FALSE	TRUE	5
41	F	47	FALSE	FALSE	5
42	M	30	FALSE	FALSE	5
43	F	52	FALSE	FALSE	4

Notas: Las abreviaturas F y M representan el sexo femenino y masculino, respectivamente. Los números 1, 2, 3, 4 y 5 de la columna "Frecuencia de escuchar música" se corresponden a las respuestas de "Ocasionalmente", "1-2 días por semana", "3-4 días por semana", "5-6 días por semana" y "diariamente", respectivamente.

b. Perfiles de los aprendices sinohablantes de español

ID	Sexo	Edad	Nivel de competencia del español	Experiencia del aprendizaje musical	Tocar instrumento musical	Frecuencia de escuchar música	Puntuaciones totales en el test de AQ
44	F	33	B1	FALSE	FALSE	1	22
45	F	24	B1	FALSE	TRUE	5	16
46	F	25	B1	FALSE	FALSE	1	14
47	F	24	B1	FALSE	TRUE	4	11
48	F	29	B1	FALSE	FALSE	1	17
49	F	20	B1	FALSE	TRUE	3	18
50	F	22	B1	FALSE	FALSE	5	10
51	F	27	B1	FALSE	TRUE	5	19
52	F	35	B1	FALSE	FALSE	1	20
53	F	31	B1	FALSE	TRUE	4	21
54	F	18	B1	FALSE	TRUE	3	21
55	F	22	B1	FALSE	FALSE	1	24
56	M	20	B1	TRUE	TRUE	5	15
57	F	29	B1	FALSE	FALSE	2	20
58	F	30	B1	FALSE	FALSE	1	22
59	F	28	B1	FALSE	FALSE	1	18
60	F	20	B1	FALSE	FALSE	5	19
61	F	19	B1	FALSE	TRUE	4	13
62	F	20	B1	FALSE	FALSE	3	25
63	F	19	B1	FALSE	FALSE	4	19
64	F	19	B1	FALSE	FALSE	4	11
65	M	20	B1	FALSE	FALSE	5	19
66	F	24	B2	FALSE	FALSE	4	19
67	F	23	B2	FALSE	TRUE	5	15
68	M	26	B2	FALSE	FALSE	3	25
69	M	25	B2	FALSE	TRUE	5	14
70	F	22	B2	FALSE	FALSE	5	21
71	M	24	B2	FALSE	FALSE	3	25
72	F	22	B2	FALSE	FALSE	3	20
73	F	24	B2	FALSE	TRUE	3	20

74	F	28	B2	FALSE	TRUE	1	21
75	F	33	B2	FALSE	FALSE	2	22
76	F	23	B2	FALSE	FALSE	3	24
77	F	22	B2	FALSE	FALSE	5	24
78	M	25	B2	FALSE	FALSE	5	22
79	M	22	B2	FALSE	FALSE	5	17
80	F	27	B2	FALSE	TRUE	5	19
81	F	25	B2	FALSE	FALSE	5	21
82	M	28	B2	FALSE	FALSE	1	18
83	F	29	B2	FALSE	FALSE	1	20
84	F	25	B2	FALSE	FALSE	5	16
85	F	20	B2	FALSE	FALSE	3	23
86	F	27	B2	FALSE	FALSE	1	24
87	F	23	B2	FALSE	FALSE	5	17
88	F	27	B2	FALSE	FALSE	1	21
89	F	29	B2	FALSE	TRUE	1	18
90	F	22	B2	FALSE	FALSE	5	21
91	F	20	B2	FALSE	TRUE	5	19
92	F	27	B2	FALSE	FALSE	5	20
93	F	26	C1	FALSE	FALSE	4	19
94	F	25	C1	FALSE	FALSE	5	22
95	F	25	C1	FALSE	FALSE	4	33
96	M	25	C1	FALSE	FALSE	5	21
97	F	30	C1	FALSE	FALSE	2	18
98	M	25	C1	FALSE	FALSE	3	21
99	F	23	C1	TRUE	TRUE	5	22
100	M	26	C1	TRUE	TRUE	4	21
101	M	27	C1	FALSE	FALSE	5	14
102	F	26	C1	FALSE	FALSE	3	23
103	F	32	C1	FALSE	FALSE	1	24
104	F	24	C1	FALSE	FALSE	5	25
105	F	22	C1	FALSE	FALSE	4	13
106	F	28	C1	FALSE	TRUE	1	28
107	F	29	C1	TRUE	TRUE	3	16
108	F	35	C1	FALSE	FALSE	1	19
109	F	27	C1	FALSE	FALSE	5	13
110	F	22	C1	FALSE	FALSE	5	12
111	F	23	C1	FALSE	FALSE	3	23
112	F	27	C1	FALSE	FALSE	1	26
113	F	26	C1	FALSE	FALSE	5	13
114	M	24	C1	FALSE	TRUE	2	17
115	F	26	C1	FALSE	FALSE	4	18
116	F	24	C1	FALSE	TRUE	5	16
117	F	25	C1	FALSE	FALSE	2	15
118	M	24	C1	FALSE	FALSE	5	24

Notas como lo anterior.

Anexo C9: Informantes de los test de percepción fonético-acústica

Este anexo se subdivide en dos partes dado que los informantes para los dos test de percepción acústica son diferentes. La primera es la identificación de los estímulos que varían en la F0 y la duración, y la segunda, es la percepción de los estímulos que varían en la F0 y la intensidad.

a. Perfiles de los informantes en el primer test perceptivo

ID	Sexo	Edad	L1	Nivel de competencia del español
1	M	31	Español	Nativo
2	F	27	Español	Nativo
3	F	26	Español	Nativo
4	F	54	Español	Nativo
5	M	37	Español	Nativo
6	M	22	Español	Nativo
7	F	26	Español	Nativo
8	M	27	Español	Nativo
9	F	46	Español	Nativo
10	F	45	Español	Nativo
11	M	46	Español	Nativo
12	F	55	Español	Nativo
13	M	41	Español	Nativo
14	F	40	Español	Nativo
15	F	24	Español	Nativo
16	F	26	Español	Nativo
17	F	30	Español	Nativo
18	F	24	Español	Nativo
19	F	23	Español	Nativo
20	F	34	Español	Nativo
21	F	38	Español	Nativo
22	F	30	Español	Nativo
23	F	41	Español	Nativo
24	M	26	Español	Nativo
25	F	46	Español	Nativo
26	M	39	Español	Nativo
27	F	58	Español	Nativo
28	F	37	Español	Nativo
29	F	25	Español	Nativo
30	M	45	Español	Nativo
31	F	59	Español	Nativo
32	M	25	Español	Nativo
33	M	34	Español	Nativo
34	M	28	Español	Nativo

35	F	22	Español	Nativo
36	F	42	Español	Nativo
37	M	23	Español	Nativo
38	F	30	Español	Nativo
39	M	45	Español	Nativo
40	F	33	Chino	B1
41	F	24	Chino	B1
42	F	25	Chino	B1
43	F	24	Chino	B1
44	F	29	Chino	B1
45	F	20	Chino	B1
46	F	22	Chino	B1
47	F	27	Chino	B1
48	F	25	Chino	B1
49	F	22	Chino	B1
50	F	20	Chino	B1
51	F	31	Chino	B1
52	F	20	Chino	B1
53	F	18	Chino	B1
54	F	20	Chino	B1
55	F	22	Chino	B1
56	M	20	Chino	B1
57	F	29	Chino	B1
58	F	30	Chino	B1
59	F	28	Chino	B1
60	F	20	Chino	B1
61	F	19	Chino	B1
62	F	20	Chino	B1
63	F	19	Chino	B1
64	F	19	Chino	B1
65	F	23	Chino	B1
66	M	20	Chino	B1
67	F	24	Chino	B2
68	F	23	Chino	B2
69	M	26	Chino	B2
70	M	25	Chino	B2
71	F	22	Chino	B2
72	M	24	Chino	B2
73	F	22	Chino	B2
74	F	24	Chino	B2
75	F	28	Chino	B2
76	F	33	Chino	B2
77	F	23	Chino	B2
78	F	22	Chino	B2
79	M	25	Chino	B2
80	M	22	Chino	B2

81	F	27	Chino	B2
82	F	25	Chino	B2
83	M	28	Chino	B2
84	F	29	Chino	B2
85	F	23	Chino	B2
86	F	27	Chino	B2
87	F	29	Chino	B2
88	M	19	Chino	B2
89	F	30	Chino	B2
90	F	22	Chino	B2
91	F	26	Chino	C1
92	F	25	Chino	C1
93	F	25	Chino	C1
94	M	25	Chino	C1
95	F	30	Chino	C1
96	M	25	Chino	C1
97	F	23	Chino	C1
98	M	26	Chino	C1
99	M	27	Chino	C1
100	F	26	Chino	C1
101	F	32	Chino	C1
102	F	28	Chino	C1
103	F	22	Chino	C1
104	F	28	Chino	C1
105	F	29	Chino	C1
106	F	35	Chino	C1
107	F	25	Chino	C1
108	F	23	Chino	C1
109	F	22	Chino	C1
110	F	23	Chino	C1
111	F	27	Chino	C1
112	F	26	Chino	C1
113	M	24	Chino	C1
114	F	26	Chino	C1
115	M	25	Chino	C1
116	F	22	Chino	C1
117	F	25	Chino	C1

Notas: Las abreviaturas F y M representan el sexo femenino y masculino, respectivamente.

b. Perfiles de los informantes en el segundo test perceptivo

ID	Sexo	Edad	L1	Nivel de competencia del español
1	F	31	Español	Nativo
2	M	27	Español	Nativo
3	F	26	Español	Nativo
4	F	54	Español	Nativo
5	F	37	Español	Nativo
6	M	22	Español	Nativo
7	M	26	Español	Nativo
8	F	27	Español	Nativo
9	M	46	Español	Nativo
10	F	45	Español	Nativo
11	F	46	Español	Nativo
12	M	55	Español	Nativo
13	F	41	Español	Nativo
14	M	40	Español	Nativo
15	F	24	Español	Nativo
16	F	26	Español	Nativo
17	F	30	Español	Nativo
18	F	24	Español	Nativo
19	F	23	Español	Nativo
20	F	34	Español	Nativo
21	F	38	Español	Nativo
22	M	46	Español	Nativo
23	M	20	Español	Nativo
24	M	58	Español	Nativo
25	M	43	Español	Nativo
26	F	25	Español	Nativo
27	F	20	Español	Nativo
28	F	45	Español	Nativo
29	F	29	Español	Nativo
30	M	34	Español	Nativo
31	M	49	Español	Nativo
32	F	47	Español	Nativo
33	F	45	Español	Nativo
34	F	33	Chino	B1
35	F	24	Chino	B1
36	F	25	Chino	B1
37	F	24	Chino	B1
38	F	29	Chino	B1
39	F	20	Chino	B1
40	F	22	Chino	B1
41	F	25	Chino	B1
42	F	35	Chino	B1
43	F	22	Chino	B1

44	F	20	Chino	B1
45	F	20	Chino	B1
46	F	20	Chino	B1
47	F	22	Chino	B1
48	M	20	Chino	B1
49	F	29	Chino	B1
50	F	30	Chino	B1
51	F	28	Chino	B1
52	F	19	Chino	B1
53	F	20	Chino	B1
54	F	19	Chino	B1
55	M	19	Chino	B1
56	M	20	Chino	B1
57	F	26	Chino	B1
58	F	31	Chino	B1
59	F	24	Chino	B2
60	F	23	Chino	B2
61	M	26	Chino	B2
62	M	25	Chino	B2
63	F	22	Chino	B2
64	M	24	Chino	B2
65	F	22	Chino	B2
66	F	24	Chino	B2
67	F	28	Chino	B2
68	F	33	Chino	B2
69	F	23	Chino	B2
70	F	22	Chino	B2
71	M	25	Chino	B2
72	M	22	Chino	B2
73	F	27	Chino	B2
74	F	25	Chino	B2
75	M	28	Chino	B2
76	F	29	Chino	B2
77	F	25	Chino	B2
78	F	20	Chino	B2
79	F	27	Chino	B2
80	F	23	Chino	B2
81	F	27	Chino	B2
82	F	29	Chino	B2
83	F	30	Chino	B2
84	F	24	Chino	B2
85	F	20	Chino	B2
86	F	26	Chino	C1
87	F	25	Chino	C1
88	F	25	Chino	C1
89	M	25	Chino	C1

90	F	30	Chino	C1
91	M	25	Chino	C1
92	F	23	Chino	C1
93	M	26	Chino	C1
94	M	27	Chino	C1
95	F	26	Chino	C1
96	F	32	Chino	C1
97	F	24	Chino	C1
98	F	28	Chino	C1
99	F	25	Chino	C1
100	F	23	Chino	C1
101	F	27	Chino	C1
102	F	22	Chino	C1
103	F	23	Chino	C1
104	F	26	Chino	C1
105	F	20	Chino	C1
106	M	25	Chino	C1
107	F	22	Chino	C1
108	F	23	Chino	C1
109	F	22	Chino	C1
110	F	20	Chino	C1

Notas: Las abreviaturas F y M representan el sexo femenino y masculino, respectivamente.