Towards a Better Understanding of Novel Educational Technologies and the Impact of Feedback on Violin Learning

A Behavioral and Electrophysiological Account

Ángel David Blanco Casares

TESI DOCTORAL UPF / 2021

Thesis supervisor

Dr. Rafael Ramírez,

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Universitat Pompeu Fabra Barcelona

A mi padre,

Ángel Blanco Villaseñor (Quintanar de la Orden, 24 de Agosto de 1954 – Barcelona, 22 de Mayo de 2020)

Agradecimientos

Quiero agradecer esta tesis a todas las personas que han compartido conmigo estos años desde el Music and Machine Learning Lab. A Rafael Ramirez, por darme la oportunidad de colaborar y crecer como persona e investigador en el proyecto TELMI, y por toda su atenta ayuda y paciencia ayudándome a mejorar la calidad de cada uno de los manuscritos publicados y que han dado lugar a esta tesis. A Alfonso, por toda su pasión y entrega a la hora de crear un setup para poder medir los potenciales evocados de violinistas y así poder llevar adelante la colaboración con Brainlab. La cual fue, sin lugar a duda, el mayor reto que esta tesis tuvo que afrontar (y por ello también, la mayor fuente de orgullo). También por haber sido conejillo de indias repetidas veces en unos experimentos pilotos que sé que no fueron del todo agradables (muchas horas con un electroencefalograma y unos cascos que aprietan bastante la cabeza tocando aburridas “melodías” de cuatro notas sin parar...). A David, por haber estado siempre ahí en el día a día echando una mano para cualquier cosa, y por su habilidad y rapidez programando con Max/MSP el *online tone manipulation system* para el violín. Seguramente sin su ayuda no habría nunca haber podido comenzar los experimentos piloto tan rápido (ni habrían salido bien tan rápido). A Fabio, por su inseparable compañía de despacho desde el día uno de comenzar a trabajar en la tesis, su amistad y eternas conversaciones de sobremesa que sin duda han supuesto un profundo cambio en mi forma de entender la música, la ciencia y las tecnologías que se derivan de ellas. A Batuhan porqué aprender de estadística se hace más cómodo si es en su buena compañía. A Sergio y a Zacharias por sus importantes consejos como “hermanos mayores” que saben lo que es llevar una tesis doctoral. Fuera del Music and Machine Learning Lab también quiero agradecer a toda esa gente del Music Technology Group y de la UPF en general que ha estado a mi lado durante todos estos años inspirándome y ofreciéndome su ayuda desinteresada en multitud de ámbitos. Especialmente a Simone, por su paciencia y apoyo durante mis primeros encuentros (y desencuentros) con la estadística inferencial y el diseño experimental. A Rafa Caro, por su gran sabiduría musical, su profunda sensibilidad y amor divulgándola que ha afectado de sobremanera mi forma de entender el aprendizaje musical desde una perspectiva intercultural (y posiblemente influencie mis futuras decisiones laborales). A Albin, Xavier, Pritish por su gran amistad durante todos estos años. A Xavier Serra por darme la oportunidad desde un primer momento de conocer y participar en esta gran comunidad que es el MTG. También a Perfecto Herrera, no solo por ser de las primeras personas a quien acudí en busca de bibliografía para comenzar a plantearme que camino debía tomar en mi tesis, sino por haber sido mi puerta de entrada a la hora de conocer a Martí con los instrumentos Baschet, quien fue a su vez mi puerta de entrada al conjunto de Gamelan del Museu de la Música de Barcelona. Una gran cantidad de los momentos más estimulantes, apasionantes y felices de estos últimos años han sido a través del proceso de sumergirme en el mundo del Gamelan con gente tan increíble como Martí, Anna, Cora, Lluís, Gerard, Giuseppe, Rafa, Toko, Violeta, Judith, Carme, Ariana, Víctor, Jordi, Pamela, Patricia, Olga, Juan, Agustín y muchas personas más que, aunque no las haya citado saben que las tengo presente. Quiero agradecer esta tesis, no sin un gran dolor en el alma, a Agustín Martorell por toda su ayuda ofrecida durante la tesis, tanto con sus conocimientos sobre tecnología, ciencia y violín como por el acceso que me ofreció a datos de sus proyectos que de tanta utilidad han sido para esta tesis. Agustín, fuiste alguien que me devolvió las ganas de querer volver a la carrera únicamente para poder cursar tu asignatura y tenerte como profesor. Fue todo un orgullo haber podido tocar junto a ti estos años en gamelan. Nos dejaste demasiado pronto, descansa en paz Agustín.

Quiero también agradecer a todas las personas del equipo de Brainlab por toda la ayuda recibida y el cariño con el que me han acogido en su equipo. Principalmente a Jordi Costa y a Iria SanMiguel por toda la pasión, energía y tiempo que dedicaron a lograr que el proyecto saliese a delante y por toda su paciencia bajándome de las nubes continuamente cada vez que sufría de fuertes pareidolias y creía “ver cosas” en mis datos. A Natàlia, y especialmente a Nadia, por todo su apoyo enseñándome a colocar correctamente el electroencefalograma a los participantes y posteriormente ayudándome a entender cómo aplicar métodos de deconvolución a nuestras señales de EEG. Por último, no puedo olvidarme de Vittoria a quien, por ser violinista, le tocó participar en los primeros pilotos del experimento y no dudó en colaborar con una gran sonrisa en la cara.

No puedo irme sin agradecer esta tesis también a todos los participantes que voluntariamente se han ofrecido a participar en mis experimentos a cambio de un simple vaso de agua y efusivas señales de agradecimiento. Entre ellos muchos de mis grandes amigos de toda la vida que esta tesis no olvida. A Alfred, Morant, Javi, Martín, Judith, Dani, Laura, Piquín, Fabra, Medina, Astals, Vilaseca, Cabrera, Plech, Pau, Pol, Miqui, Mayor, Loza, Suri, Xavi, Meri, Alejandro, Piquín, Rita y todos aquellos de los que me pueda estar olvidando en este preciso momento.

Quiero también agradecer profundamente a quienes han sido con mayor probabilidad las mayores víctimas de esta tesis. Aquellas personas que han sufrido más de cerca las consecuencias de compartir su vida con alguien que ha dedicado tal innumerable cantidad de horas a su trabajo. A pesar de todo ello, le quiero agradecer a Rita por su compañía durante una importante parte de esta tesis, por todo su amor y paciencia conmigo y por tantos momentos de felicidad que nunca olvidaré. También a Olga, por la paciencia de haber tenido que pasar las fases más duras del confinamiento encerrada con un ser que no se despegaba del ordenador. Le agradezco por todo su amor, por todo el cariño ofrecido durante este último año de tesis que ha sido probablemente el más intenso de todos. También todos los increíbles momentos felices que hemos pasado juntos y haciendo música en general. El confinamiento sin ella habría sido una experiencia mucho menos agradable sin lugar a duda. He de agradecerle también que sin su increíble facilidad para el canto y nuestros largos debates acerca de "¿por qué la gente canta *mal*?" esta tesis no habría sido lo mismo. Quiero también agradecerle la tesis a Caracas (mi gatita), tantas noches de trabajo habrían sido diferentes sin su nocturna compañía en mi escritorio. También a Sisa (mi perrita), quien me recuerda que para llevar una vida más saludable hay que salir más a pasear y pasar menos tiempo sentado delante del ordenador.

Por último, quiero agradecerle esta tesis a mi familia. A mi madre, mi hermana y mi padre quienes han estado apoyándome desde el primer día (a pesar de no siempre entender muy bien que es exactamente lo que estoy haciendo). Pero sobre todo esta tesis va dedicada con todo mi amor a mi padre, quien dedicó una vida entera a la ciencia y a la academia y murió contestando a sus alumnos desde el hospital en medio de una pandemia global. ¿Sabes papá?

Quizás, uno de los mayores regalos que me ha dado esta tesis, ha sido la capacidad de poder comprender en profundidad tu trabajo. Sé que desde pequeño me lo has intentado explicar millones de veces, pero también es cierto la ciencia como práctica (al igual que la música) tiene un cierto componente tácito y encarnado que hace que hasta que no se practica no se *ve* y no se aprecie la utilidad de lo que uno hace. Quiero pedirte perdón por todas las veces que he podido ser injusto contigo. Nunca antes en mi vida me había sentido tan orgulloso de ti. Soy también consciente, ahora más que nunca, de la ilusión que te habría hecho estar presente el día de mi presentación de tesis doctoral. Un día antes que te fueses te prometí que ibas a estar allí conmigo ese día. Sigo sin albergar dudas de que así va a ser. Esta tesis va por ti.

Abstract

Mastering the violin and other bowed string instruments requires special considerations compared with other musical instruments. The process of good sound generation in the violin is a notoriously complex task that requires precise spatiotemporal control of bowing gestures. In addition, Unlike the piano and other keyed or fretted instruments like the guitar, pitch control in the violin is continuous and, thus, movements must be much more precise. This makes music production and intonation monitoring with those instruments much more dependent on auditory feedback. In this PhD thesis, we present the results of four experiments designed to study and evaluate the effects of feedback (visual or auditory) and educational technologies in both violin learners and professional violinists from a psychological and psychophysiological perspective. Results show promising perspectives for the development and advancement of this kind of new educational tools to help music students. At the same time, they highlight the need for a truly multidisciplinary enterprise not only, to design and correctly evaluate future educational technologies, but also to transcend our understanding of musical learning in all its deepest essences and facets.

Resumen

Dominar el violín y otros instrumentos de cuerda frotada requiere consideraciones especiales en comparación con otros instrumentos musicales. El proceso de generación de un buen sonido en el violín es una tarea notoriamente compleja que requiere un preciso control espaciotemporal de los movimientos del arco. Además, a diferencia del piano y otros instrumentos con teclas o trastes como la guitarra, el control del tono en el violín es continuo y, por lo tanto, los movimientos tienen que ser mucho más precisos. Esto hace que la producción musical y el monitoreo de la entonación con estos instrumentos sea mucho más dependiente del feedback auditivo. En esta tesis doctoral presentamos los resultados de cuatro experimentos diseñados para estudiar y evaluar los efectos del feedback (visual o auditivo) y de las tecnologías educativas tanto en estudiantes de violín como en violinistas profesionales, tanto desde una perspectiva psicológica como psicofisiológica. Los resultados muestran perspectivas prometedoras para el desarrollo y avance de este tipo de nuevas herramientas educativas para ayudar a los estudiantes de música. Al mismo tiempo, destacan la necesidad de una empresa verdaderamente multidisciplinaria no solo para diseñar y evaluar correctamente las tecnologías educativas futuras, sino también para trascender nuestra comprensión del aprendizaje musical en todas sus esencias y facetas más profundas.

Prefacio

He de confesar que realizar este doctorado se encontraba entre las últimas prioridades de mi vida allá por el año 2016 cuando finalicé el Máster in Sound and Music Computing (SMC) en la Universitat Pompeu Fabra. Ocho años antes, a pesar de las nefastas notas que había sacado en Física y Matemáticas en la selectividad, decidí estudiar Telecomunicaciones por encima de mi gran pasión (Filosofía) por dos simples motivos: uno era no morirse de hambre, y el segundo, si había suerte, consistía en encaminar los conocimientos de la carrera hacía mi segunda gran pasión: la música.

Lo que no esperé en ningún momento fue que Telecomunicaciones me ofreciese unos nexos tan fuertes con el mundo de la filosofía. Las matemáticas dejaron de parecerme ejercicios algorítmicos sin sentido que había que resolver para aprobar un examen y pasaron a convertirse en los cimientos de mi cosmovisión. Fascinado por la Teoría de la Información de Shannon, los descubrimientos de Hilbert, Gödel, Turing y su aplicación a las ciencias de la computación, a la física, a la electrónica, a la inteligencia artificial y a la relación de esta última con la neurociencia y el problema mente-cerebro. Tuve además la suerte de tener a alguien tan pasional como Joan Claudi como supervisor de mi proyecto de final de carrera destinado a la creación de un instrumento musical que funcionase por interacción natural utilizando la Kinect. También diseñamos un wav-to-midi converter que podía acoplarse al instrumento. Sé que aquel proyecto, a pesar de la falta de medios con la que fue realizado, fue la llave que me llevó a ser aceptado en el SMC Master y por eso le reservo aún un gran cariño.

En el SMC Master, profundicé tanto en la psicoacústica como en la cognición musical de la mano de profesores como Perfecto Herrera. Otros profesores, como Paul Verschure, me ayudaron a profundizar en los nexos entre tecnología, cognición y filosofía. Había descubierto que el mundo que mejor iba a adaptarse a mis incansables demandas de seguir aprendiendo desde una perspectiva amplia y lo más multidisciplinar posible era la academia. Sin embargo, cuando Rafael me propuso la idea de realizar un doctorado en el proyecto TELMI, sentí que no era ese el tipo de investigación que me hubiese gustado hacer. Sé que Rafael estaba, en parte, interesado en mí porqué había visto mi proyecto de final de carrera y era consciente de que sabía programar con la Kinect. Estaban planeando utilizar técnicas de captura de movimiento para poder traquear los movimientos del arco de los participantes. Sin embargo, mis intereses en aquel momento estaban alineados en lograr un doctorado en psicología o neurociencia de la percepción musical.

Comencé a aplicar en diferentes ofertas de doctorado. No tuve demasiada suerte en el Henkjan Honing's lab en Amsterdam, tampoco en un proyecto en Lyon/St Ettienne (France) acerca de la percepción auditiva de cocodrilos, y llegué tarde a enviar la application letter a una oferta de doctorado de un laboratorio en la Universitat de Barcelona (UB) dedicado a la percepción auditiva de movimientos autogenerados llamado Brainlab. Sorprendido por la existencia de un laboratorio trabajando en algo que me atraía tanto en Barcelona me puse en contacto con su coordinador, el Dr. Carles Escera, para explorar si el grupo podía ofrecerme algún tipo de posibilidad laboral. Brainlab se encontraba afincado en la facultad de Psicología en el Campus Mundet. Justo en la planta inferior del Departamento de Psicología Social y Psicología Cuantitativa donde residía el despacho de mi padre. No tengo palabras para explicar la nostalgia que se apoderó de mí el día que visité el campus para reunirme con Carles. Cada vez que de pequeño me enfermaba o me daba fiebre mi padre me llevaba con él a la facultad donde pasaba horas jugando al buscaminas en el ordenador de su despacho mientras él daba clase. Como si mi cuerpo no hubiese nunca olvidado esa asociación, de camino a mi cita con Carles, un fuerte debilitamiento, acompañado de una sensación febril, se apoderó de mí.

A pesar de la enriquecedora conversación que tuve con Carles no me gustó su recomendación de participar en su nuevo máster o la de colaborar con ellos hasta poder encontrar algún tipo de financiación (más adelante, la Dra María Teresa Anguera me contó que fue así como mi padre, en los años 70, logró entrar a formar parte de su grupo de investigación). Por aquella época andaba algo desesperado por volver a trabajar y a tener un sueldo. Fue en aquel momento cuando me llamó Rafael y me dijo: "¡Felicidades! ¡Has sido escogido para el doctorado en el proyecto TELMI!". Mi falta de ilusión ante la noticia debió ser muy patente en aquel instante. Sin embargo, Rafael añadió: "Hemos estado pensando acerca de aquello que comentaste de que tus intereses iban más centrados a la psicología y a la neurociencia y hemos pensado en adquirir un electroencefalograma para tu tesis. Lo puedes usar para recolectar y analizar datos electrofisiológicos de los violinistas". Aquello lo cambió todo. Rafael me ofreció la posibilidad de realizar mi doctorado en un contexto en el cual iba a tener máxima libertad y en el que me podía ver llevándolo hacia mis propios intereses. Es cierto, no iba a realizar mi doctorado en un contexto especializado en psicología y cognición como en Brainlab, pero siempre se me había dado bien aprender de forma autodidacta. Además, procesar y analizar datos de electroencefalografía parecía una tarea muy fácil comparada con el procesamiento de audio y música al que estaba acostumbrado. Filtramos en distintas bandas frecuenciales y observamos diferencias entre diferentes condiciones. ¿Qué podría salir mal? Casi todo. Pronto me vi lidiando con la interpretación de miles datos de electroencefalografía recolectados de mis participantes, en los que trataba de encontrar patrones que estuviesen reflejados en literatura previa. Todo ello en unas actitudes que parecían más propias a las de un agente de la serie CSI: Miami investigando un crimen, que a las de un científico que tratase de ser tomado en serio. Si no encontraba lo que quería en mis resultados era probablemente debido al ruido, me decía. Necesitaba aprender técnicas que me ayudasen a procesar mejor mis datos. Fue entonces cuando un conocido de mi novia por aquel entonces, Oscar Bedford, me recomendó asistir a unos talleres o workshops de EEG en Brainlab y ponerme en contacto con Jordi Costa. Cuando escribí a Jordi su contestación fue algo así como: "¡Anda! ¡Alguien del MTG! ¡Por fin! Llevábamos tiempo esperando iniciar una colaboración con vosotros. Te pongo en contacto con Iria SanMiguel a quien le puede interesar tu proyecto". Y así, sin habérmelo propuesto, acabé logrando el perfecto equilibrio entre las libertades ofrecidas por Rafael, y la experta supervisión de un grupo de investigación con el que anhelaba trabajar y que tenían una forma de operar radicalmente diferente a la que estaba acostumbrado. Al principio, no podía evitar ver en ellos una actitud un poco "tiquismiquis" a la hora de querer tenerlo todo controlado que contrastaba de sobremanera a la que estábamos acostumbrados en nuestro laboratorio. Después de cinco años de errores, pilotos fracasados, falsas expectativas, sudor y desesperación puedo decir que me he convertido también en uno de ellos (aunque todavía me ganan).

El aprendizaje natural que conllevan cinco años de tesis se ve reflejado en la sucesión de manuscritos publicados durante este tiempo. Esto me llevó a reescribir para esta tesis una buena parte de mi primera publicación *Evaluation of a Sound Quality Visual Feedback System for Bow Learning Technique in Violin Beginners: An EEG Study.* Acceder de nuevo a procesar y analizar esos datos desde una óptica más actual lo encontré una tarea de suma importancia y relativamente fácil a pesar de todas las complicaciones y limitaciones de ese trabajo. Así, a pesar de muchas cosas acerca de las que pueda no estar orgulloso de ese trabajo, fue una parte indispensable de mi proceso de aprendizaje. Y sí de aprender es de lo que trata esto de hacer un doctorado, ¿por qué no incluirlo, aunque sea con unas ligeras modificaciones más simples?

Por último, y dado que me imaginé en el futuro a muchos más estudiantes de doctorado en mi situación y sin un background en psicología y neurociencia, en el capítulo dos de esta tesis decidí ofrecer un resumen de todos aquellos distintos paradigmas de aprendizaje motor que pueden serles de utilidad a la hora de interpretar mejor sus resultados y ser conscientes de la importancia que el diseño experimental va a tener en sus resultados. En aras de ser breve, sé que me dejé cosas en el tintero. Sin embargo, visto ahora con perspectiva, pude haber dicho quizás incluso demasiadas. Supongo que después de todos estos años de duro trabajo me merecía escribir una tesis a mi estilo. Y, muy a pesar de todas las cosas que todavía cambiaría de esta tesis, puedo decir que me siento orgulloso del resultado, aunque mucho más del proceso.

Ángel David Blanco Casares
Barcelona, 5 de Julio de 2021

Table of contents

“One must learn to love.— This is what happens to us in music: first one has to learn to hear a figure and melody at all, to detect and distinguish it, to isolate it and delimit it as a separate life; then it requires some exertion and good will to tolerate it in spite of its strangeness, to be patient with its appearance and expression, and kindhearted about its oddity:—finally there comes a moment when we are used to it, when we wait for it, when we sense that we should miss it if it were missing: and now it continues to compel and enchant us relentlessly until we have become its humble and enraptured lovers who desire nothing better from the world than it and only it.— But that is what happens to us not only in music: that is how we have learned to love all things that we now love. In the end we are always rewarded for our good will, our patience, fairmindedness, and gentleness with what is strange; gradually, it sheds its veil and turns out to be a new and indescribable beauty:— that is its thanks for our hospitality. Even those who love themselves will have learned it in this way: for there is no other way. Love, too, has to be learned.”

Friedrick Nietszche, Die fröhliche Wissenschaft, p. 334

1. INTRODUCTION

1.1 Music learning and music technologies

What distinguishes a good musician from a bad musician? A beginner from an expert? What is it that makes music better for some than for others? Is it simply a matter of many hours of practice? Certainly, the number of hours of practice devoted to one instrument can directly predict achievement (Ericsson, Krampe, & Tesch-Römer, 1993; Sloboda et al., 1996). Ericsson et al. reported that those violin students who had the potential for careers as international soloists, nominated by their professors, accumulated 10,000 hours of practice over their lifespan at the age of 21 years. Exactly twice as many hours as average violin students of the same age (5,000 hours). Considering that those students started, on average, to play the violin at the age of 6 or 7 years old, the accumulation of 10,000 hours of practice up to the age of 21 requires 2 hours of practice with the instrument per day every day of the year. Practice hours that are not always inherently rewarding to young children and require parental support and direct supervision (Sloboda and Howe, 1991).

Practice, by itself, however, doesn't seem to be a magic wand to musical perfection. Recent studies have found that practice can only account for 30% of the variation in performance quality (Hambrick et al., 2014). As the authors affirm (page 43): “The evidence is quite clear that some people do reach an elite level of performance without copious practice, while other people fail to do so despite copious practice”. What are the rest of the factors that could be influencing performance quality? Between them are those related to starting at a young age (Simonton, 1991) or those that could be even genetically inherited (Coon and Carey, 1989), and socio-motivational factors (Sloboda and Howe, 1991). However, as motor learning researchers know well, it is not only about the number of hours someone dedicates to practice but about the quality of the practice itself and how those hours have been spent. Hallam (1997) found that an important part of music students left errors uncorrected when practicing a new piece. Those errors tended to become permanent and left uncorrected if they were not explicitly insisted on considering them. External feedback plays an important role in creating awareness of the performance of music students and enhancing their metacognitive skills. That is, those skills related to planning, monitoring, and evaluation of performance (Hallam, 2001).

How music instructors deliver external feedback about their performance to students can strongly influence their views and attitudes toward practice. However, external feedback does not have to be exclusively offered by the music instructor. Self-recording of performance (video or audio) could easily provide feedback. Previous research (Kepner, 1986; Bundy, 1987) found that high school instrumentalists were more able to identify musical errors when hearing tape recordings of their own performances than when actually performing the pieces. However, in absence of a teacher, it requires the student to be his/her own judge which may be problematic.

Current music technologies provide us with objective measures of student improvement in specific music tasks. Thus, such technologies can allow us to monitor the learning process of music students in order to provide better and personalized learning strategies. In addition, objective measures about music students' performance may serve as additional information which could complement the verbal feedback given by the teacher. It is important to note that this type of technology does not displace the importance of the music teacher but can even enhance the effectiveness of his/her classes with students. For example, the stopwatch is a very simple technology that allows time to be measured with a high degree of sensitivity. Both professional swimmers and their coaches find it very useful to keep track of the competitor's marks, and based on them, plan the training. Additionally, this feedback can be used as positive reinforcement for the swimmer, thus increasing his/her motivation levels, which will lead him to continue surpassing his previous records in a kind of virtuous cycle.

This thesis has been written in the context of the European TELMI Project (Technology Enhanced Learning of Musical Instrument Performance). The TELMI Project had the general objectives to design and implement new technologies for music learning and training (based on multi-modal-feedback technologies, such as audio, image, video, and motion), together with the evaluation of their pedagogical effectiveness. Using the violin as a case of study, the aim of this thesis has been devoted to the later one. In this thesis, however, we want not only to evaluate some of the outcomes arising from the TELMI project in recent years, but we also want to show how these technologies can be an important tool for research related to music education. Not only from a pedagogical or behavioral perspective but also from a cognitive perspective, allowing the investigation of neural correlates of learning, feedback monitoring, and error correction processes through the analysis of electrophysiological signals. An important part of this thesis was developed in collaboration with the Brainlab from the UB neuroscience institute thus reflecting its multidisciplinary nature.

1.2 Motivation

External feedback can be delivered to students in multiple ways. It can be delivered as Knowledge of Results (KR), which is defined as any source of feedback delivered by external sources that offer information about the consequences of an action (Salmoni, Schmidt and Walter 1984; Schmidt and Lee 1999; Swinnen 1996; Magill 2001). That is, a number representing the time in seconds you spent swimming 100 meters or a binary "correct" or "incorrect" signal after answering a test question. It can also be delivered as Knowledge of Performance (KR) which is defined as any source of feedback that offers information about the development of the action (Gentile 1972, Salmoni, Schmidt and Walter 1984). For example, when violin instructors try to improve violin posture in their students by correcting them during the exercises. It can be delivered online, offline, only after positive outcomes, through different modalities (auditory, visual, tactile). The design of a feedback technology always starts from a previous hypothesis (or prejudices) about what it is considered that the student needs the most. However, we can be wrong and feedback could distract participants from their main goal, could create dependency or could be irrelevant. Evaluation, thus, should always be an important part of the process in the development of educational tools.

One way to evaluate music technologies could be through controlled, randomized studies as has been made in the motor learning literature for decades. Is thus, important to learn about the lessons learned by that field while considering, at the same time, the particularities of music performance. Certainly, this type of experimental research has always attracted the suspicion of pedagogues. And they are right. The students in real life always behave differently from how they behave in a laboratory and laboratory experiments are not always extrapolated to the day-to-day life of a student at a music conservatory. It would be also unfair, however, to deny that pedagogy is not constantly being influenced by experimental research in motor learning and psychology paradigms (from behaviorism to constructivism and recent psychological theories of learning).

Thus, without being integrated into a pedagogy that includes how to use them, feedback technologies are an empty tool. The important thing is that for the first time they could help to obtain scientific evaluations about motor-auditory activities that were previously evaluated only by subjective means. Goldin and Rouse (2000) found that the adoption of "blind" auditions in orchestras increased the probability of women being hired. Not only subjective evaluations are subject to strong subjective biases (Wadell et al. 2017), traditional teaching methods of musical performance movement may not be based on the understanding of its biomechanics components but on the subjective and vague perception of human movement (Brandfonbrener, 2004). Therefore, the design and evaluation of music technologies are not only favored by pedagogy and scientific literature about motor learning, but the evaluation of music technologies itself already implies new knowledge that feeds on and allows create collaborative networks between pedagogues, neuroscientists, psychologists, and engineers. In addition, the collection of electrophysiological signals from both expert and beginner musicians and those of their performance can help us better understand how different learning processes evolve and measure possible conditions of stress and excessive workload of the participants that could be detrimental to their learning.

1.3 Contributions

This thesis has carried out multiple experiments with professional violinists, amateur musicians, and complete beginners. We have evaluated them while they were both playing or learning to play the violin and we have captured data and information about their performance throughout different learning blocks. We have listed below some of the main contributions of this thesis:

1. We have offered several datasets with data collected about: sound quality, kinematic variables of movement with the bow, intonation errors, and their evolution through different stages and after having received different types of treatment. Also, electrophysiological data collected from the participants during the experiments. All the data collected is freely available at Zenodo to be downloaded by anyone with an interest in it.
2. We have provided a state-of-the-art methodology and state-of-the-art for anyone who needs to evaluate newly designed music technologies to be aware of the methodological limitations, obstacles, and considerations they must consider before embarking on the task.
3. Throughout our experiments, we have demonstrated the importance of this type of musical technology both for musical learning and for the evaluation of possible different treatments. At the same time, we have found descriptors both concerning the quality of the sound and kinematic movements with the bow that can be important to evaluate the practice with the violin.
4. In collaboration with the Brainlab from the UB, we have designed a setup that allows not only to collect electrophysiological information from participants while playing the violin but also to perform tone manipulations, contributing to the advancement of research in performance science, expanding the field of study typically dedicated to studying the piano, to instruments like the violin. The violin is interesting not only because of the greater dependence on auditory feedback that playing them entails but also for the study of online error correction processes that cannot be studied on the piano.

5. Finally, we have validated the setup with violinists and cellists and we have studied some of the electrophysiological processes related to error monitoring and paved the way for future research about the learning of sensorimotor associations in music learning with a fretless instrument.

1.4 Organization of the Thesis

This thesis can be seen as consisting of 4 main parts.

The first part of the thesis comprises chapter 2. In chapter *2. State of the Art*, we offer a summary of all the possible roles of augmented feedback technologies in music and motor learning. We also offer a summary of different motor learning and cognitive paradigms that hopefully will help those with no background in the aforementioned topics to fully understand the rest of the thesis.

In the second part of the thesis, we will show the results of two experiments designed with the purpose of understanding the effects of augmented feedback in bow learning technique with the violin. The second part contains chapters 3 and 4.

1. Chapter *3. Evaluation of a Sound Quality Visual System for Bow Learning Technique in Violin Beginners: An EEG Study (New Reviewed Version from 2021)* consists of a new reviewed version (not peer-reviewed) of an already published paper in *Frontiers in Psychology* (Blanco & Ramirez, 2019). This work is a first pilot study where we evaluate the effects of sound quality feedback in violin beginners during the process of learning to make full bow exercises. We also measured their cortical activity with EEG and compared their results with those of a group of experts.

2. Chapter *4. Real-Time Sound and Motion Feedback for Violin Bow Technique Learning: A Controlled Randomized Trial* is a paper already published in *Frontiers in Psychology* (Blanco, Tassani & Ramirez, 2021a). This work is an extension of chapter 3 with a larger sample of beginner and experts' participants (although without collecting EEG data)

where we evaluate some of the main outcomes of the Telmi project: SkyNote. A system capable of offering real-time sound quality feedback together with kinematic feedback related to participant's bow movement.

In the third part of the thesis, we will evaluate the effects of augmented feedback in learning a more "musical" activity such as intonation. We will also explore the electrophysiological processes of error-monitoring and detection in experts violinists and cellists.

1. Chapter *5. Effects of Visual and Auditory Feedback in Violin and Singing Voice Pitch Matching Tasks* is a paper already published in *Frontiers in Psychology* (Blanco, Tassani & Ramirez, 2021b). This work explores the effects of real-time visual feedback for learning intonation skills (offered by SkyNote too) and compares it with the effects of aural feedback with a similar timbre to that produced by beginners, both with the violin and the voice.

2. Chapter *6. Online Tone Manipulation in Violin Performance* is a study that has not been peer-reviewed yet. This work is the result of a collaboration between the MTG of the UPF (Angel David Blanco, David Dalmazzo, Alfonso Perez & Rafael Ramirez) and the Brainlab of the UB (Jordi Costa & Iria SanMiguel). In this work, we study the electrophysiological correlates of error-monitoring, detection, and correction in string-instrument performance. For that purpose, we have designed a setup that allows us to online manipulate notes of a melody during the performance of experts violinists and cellists.

Finally, the last part of the thesis comprises chapter 7 with the final conclusions. In chapter *7. Conclusions* we will offer a brief summary of the results of all the experiments and we will also discuss possible future directions.

1.5 References

Bundy, O. R. (1987). Instrumentalists' perception of their performance as measured by detection of pitch and rhythm errors under live and recorded conditions. (Doctoral dissertation, Pennsylvania State University). Dissertation Abstracts International, 48-10A, 2567.

Burkholder, K. R., & Brandfonbrener, A. G. (2004). Performance-related injuries among student musicians at a specialty clinic. *Medical Problems of Performing Artists*, *19*(3), 116-122.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review*, *100*(3), 363.

Gentile, A. M. (1972). A working model of skill acquisition with application to teaching. *Quest*, *17*(1), 3-23.

Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of" blind" auditions on female musicians. *American economic review*, *90*(4), 715-741.

Hallam, S. (1997). What do we know about practicing? Towards a model synthesising the research literature. *Does practice make perfect? Current theory and research on instrumental music practice*, *1*, 179-231.

Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Deliberate practice: Is that all it takes to become an expert?. *Intelligence*, *45*, 34-45.

Howe, M. J., & Sloboda, J. A. (1991). Young musicians' accounts of significant influences in their early lives. 2. Teachers, practising and performing. *British Journal of Music Education*, *8*(1), 53-63.

Kepner, C. B. (1986). The effect of performance familiarity, listening condition, and type of performance error on correctness of performance error detection by 50 high school instrumentalists as explained through a sensory blocking theory. (Doctoral dissertation, Kent State University, Kent, OH). Dissertation Abstracts International, 47-05A, 1643.

Kruse-Weber, S., & Parncutt, R. (2014). Error management for musicians: an interdisciplinary conceptual framework. *Frontiers in psychology*, *5*, 777.

Magill, R. A. (2001). *Motor learning: Concepts and applications* (6th ed.). Madison, WI: Brown & Benchmark.

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: a review and critical reappraisal. *Psychological bulletin*, *95*(3), 355.

Schmidt, R. A., & Lee, T. D. (1999). *Motor control and learning: A behavioral emphasis* (3rd ed.). Champaign, IL: Human Kinetics.

Simonton, D. K. (1991). Emergence and realization of genius: The lives and works of 120 classical composers. *Journal of Personality and Social Psychology*, *61*(5), 829.

Sloboda, J. A., Davidson, J. W., Howe, M. J., & Moore, D. G. (1996). The role of practice in the development of performing musicians. *British journal of psychology*, *87*(2), 287-309.

Swinnen, S. P. (1996). Information feedback for motor skill learning: A review. In H. N. Zelaznik (Ed.), *Advances in motor learning and control* (pp. 37-66). Champaign, IL: Human Kinetics.

Waddell, G., & Williamon, A. (2017). Eye of the beholder: Stage entrance behavior and facial expression affect continuous quality ratings in music performance. *Frontiers in Psychology*, *8*, 513.

2. STATE OF THE ART

2.1 Introduction

When evaluating the augmented feedback offered by some new educational music technology using, for example, controlled randomized designs, we must be aware of all the many factors that can influence its success or failure. For example, the same experiment can be designed in a certain way that makes it impossible or very difficult to succeed without any kind of feedback in learning the task to be carried out. In non-ecological situations (as has been the case on many occasions in the motor learning literature), the advantages of using a specific feedback technology to learn a certain skill may be too overrated. On the other hand, the feedback offered may be useless. If the task is too easy for the participants of the experiment, it is very easy that in a short time the participants achieve skill results that can no longer be improved, and therefore the effect of the evaluated technology is null. This is what is called the "ceiling effect". This can happen in the other direction as well: a task that is too complicated can impose limitations on the improvement of skills in the short periods of time in which these types of controlled experiments are normally carried out. Because a task may seem complicated or simple to students of varying skill or experience levels, controlling the musical abilities of the participants under study can be critical. Last, and worst, increased feedback could actually impede learning. For example, a musician with little experience could have serious difficulties evaluating her tuning skills. An electronic tuner can make a musician with little experience learn much faster to tune her instrument and with greater precision, than one who does not use it. However, learning to tune the instrument with a tuner is not the type of learning desired and can make the musician very dependent on the tuner. There are, however, very simple strategies that could help the student to improve their tuning skills using a tuner without creating dependency. Like for example, using it *after* having tuned the instrument as an evaluative tool and not *during* the process.

With this, we want to make clear how the effects that we find in our results when evaluating a technology that offers augmented feedback can be affected both by the strategies we use when evaluating it, as well as by the level of expertise of the participants. Also, by the complexity and the type of the task to learn and/or the ecological context itself in which the experiment is being carried out and the learning is evaluated. All these variables can make augmented feedback play different roles. According to Magill (1994, 2001) and Batalla (2005) (see Figure 1):

1. Augmented Feedback can be *essential* for motor learning: this can be the case in those situations in which the participant cannot make use of their own intrinsic feedback (as in the case of proprioceptive dysfunction), or when they do not even have the skills necessary metacognitive or a strengthened schema to be able to correctly evaluate the results of their actions (as in the previously mentioned case of the musician who does not know how to tune his instrument).
2. Augmented Feedback may not be essential for motor learning: contrary to the previous situation, when intrinsic feedback is clear, available and the participant is able to recognize and learn from their successes and mistakes by themselves, the increased feedback is no longer essential for motor learning.
3. Augmented Feedback could aid motor learning: even in the former case that Augmented Feedback is not essential for learning, it could be the situation that its use helps students to maintain high levels of motivation during the long hours of practice required. Also, in some cases, it could even shorten their practice time by helping them to detect their weaknesses more quickly.
4. Augmented Feedback could impede motor learning: this can occur in situations where Augmented Feedback could lead to dependency. When this occurs, one tends to observe poorer results in retention tests when removing the Augmented Feedback.

In order to offer a concise summary of the different pedagogical strategies among which we can use technologies that offer feedback, we have decided to classify them into two types of strategies: “errorless” learning strategies and “errorful” learning strategies. Contradictory as it may seem, these types of paradigms have evolved in parallel, adapting to the different theories of motor learning of the time (behaviorism, cognitive and ecological approaches) and continuously offering mixed results about the effectiveness of each. We will also talk about the possible role of expectancies and explicit and implicit motor learning.

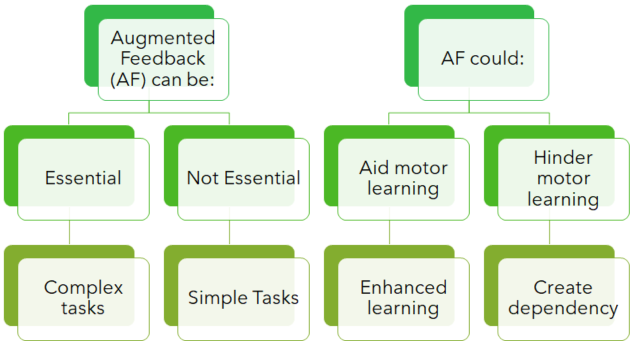

Figure 1: Augmented feedback can act in many different ways when being evaluated. The complexity of the task and the participants' level of expertise are some of the variables that can most influence its effects.

2.2 Errorless and Errorful Learning Strategies

In 1890, William James (1890) proposed the response-chaining or reflex-chaining hypothesis to describe movement control. Later, in 1906, Sherrington proposed the reflex theory of motor control (Sherrington, 1906). From their perspective, reflexes are the building blocks of movement even those of greater complexity. Simple reflexes were combined and chained into greater actions through stimulus-response. Posterior behaviorists' approaches of learning like classical and operant conditioning offered a deep understanding of the processes of learning, reinforcement, and feedback (especially in the animal kingdom). According to Skinner (1953), from a behaviorist perspective reinforcement can strengthen the desired behavior (contributing to learning) while punishment inhibits it. Thus, errors in performance can be equated with punishment that inhibits behavior but does not contribute to learning. From a pedagogical point of view using this perspective, learning should be done step by step because learning is seen as a chain of reflexes that need to be reinforced one by one. An error is seen as a consequence of moving too fast from one step to the next without having learned the previous prerequisite behaviors needed to advance. Giving participants detailed instructions for the correct performance, trying to avoid errors in performance, and offering as much feedback as we can are the typical strategies that would be offered by errorless learning advocates.

Errorless learning and behaviorism have largely influenced pedagogical views of music in the XX century (Kruse-Weber & Parncutt, 2014). Starting with a slow tempo and gradually increasing it to faster tempos, learning multi-step tasks, not at once but step by step (backward or forward chaining), stop playing at the same moment an error has been detected to start again from the beginning, and placing feedback and reinforcement at the center of the learning process have been common. However, new cognitive paradigms of motor learning such as the Schmidt's schema theory (Schmidt, 1975), together with new views on the functioning of human memory (Miller, George A., 1956, Schneider and Shiffrin (1977); Battig, 1972; Batting, 1966) predicted that variability of practice (which tends to produce more errors) plays a major role in motor learning than just errorless practice.

These new paradigms also called into question the effectiveness and the excessive importance that was tended to be given to feedback and reward in the motor learning literature (Swinnen, 1996, Batalla, 2005). According to Swinnen, many of the studies conducted previously had been conducted under highly controlled conditions where the learning process of people with sensory deficits was mimicked. These types of conditions increased the need for feedback and reinforcement in the participants. On the other hand, according to Batalla, previous motor learning literature tended not to distinguish between *motor learning* and *transient performance effects*, two concepts that are central today in motor learning literature. While continuously offering feedback to the participants could result in a clear improvement in the participants' performance, the introduction into the experiments of transfer and retention phases showed that, in many situations, less feedback helped in retention phases (Lee, White and Carnahan 1990; Winstein, Pohl and Lewthwaite 1994; Park Shea and Wright 2000). Importantly, participants receiving less feedback performed *worse* than participants receiving more feedback during the acquisition phases.

Despite this, it has continued to be evidenced for a multitude of other types of tasks that sometimes more feedback can be better than less feedback (Wulf, Shea, and Matschiner, 1998). According to Magill (2001), there is no optimum frequency of augmented feedback for learning any kind of skill, but it is dependent on the complexity of the task to learn. Magill understands complexity as the number of components of the task that could be given augmented feedback. Thus, isomorphic tasks need less augmented feedback than tasks that could be informed about different aspects. Anderson, Magill, and Sekiya (2001) conclude that according to the review of the literature we can conclude that in complex tasks (or in those where participants have less experience) learning is improved with more frequent augmented feedback, while in simple tasks or more familiar to participants, are more effective those conditions where feedback is less frequent. Other similar views were offered by Guadagnoli and Lee (2004). This view considers the amount of information available from a performance, understanding information as both errors (overestimated effects of the action), or surprises (underestimating actual outcomes). According to this view, too little or too much information can be suboptimal for learning. Support from this view in empirical investigations can be seen in Goh et al (2012) and Onla-or & Winstein (2008).

2.3 Cognitive Load Theory

Thinking about our capacity to process information offers us a different perspective to interpret the results seen until now. Chandler and Sweller (1991) distinguished between two different types of cognitive load. On the one hand, intrinsic cognitive load refers to the level of difficulty associated with a specific task, and it may not be altered by an instructor. On the other hand, extraneous cognitive load relates to all the unnecessary cognitive load which is generated by the way information is presented. For example, one can teach trigonometry using a verbal medium and without the use of drawings, however, extraneous cognitive load can be reduced by using a visual medium instead of a verbal one. The implications of CLT for learning are important because it is known that a high working memory load can disrupt the way attention is focused (Lavie, 2005). In an experiment by De Fockert et al (2001), it was possible to show through fMRI scans that participants under a high load condition, tended to pay more attention to distractor cues while performing a classification task.

Regarding the effects of extraneous cognitive load in motor learning, Maxwell et al (2001) hypothesized that errorless learning strategies would prevent explicit or declarative knowledge via hypothesis testing strategies, thereby reducing dependence on working memory resources. Poolton et al (2005) tested the hypothesis in a golf putting task constraining the environment during the first phases of learning to minimize performance error. They also tested the effect of an additional cognitive task on putting performance to assess the level of reliance on working memory. They found that participants from the errorless group were less affected by the addition of the cognitive task suggesting that they were using fewer working memory resources.

A new type of cognitive load was posteriorly proposed by Sweller, Van Merriënboer, and Paas(1998) and is called germane cognitive load. The germane cognitive load could be understood as “productive load” and it is supposed to be related to the processing and construction of schemas. This perspective could explain how it is possible, that in some circumstances, low variability and complete guidance feedback hinder the transfer of learning. Although these strategies may reduce the difficulty of the task, as a side-effect they could also reduce germane cognitive load and hinder learning and retention (Van Merrienboer et al, 2016).

Another different effect derived from limited working memory was found by Mayer and Moreno (1998). It was found that explaining a concept using animation and narration together was more effective than animation and text. They explained these findings based on the idea that visual and auditory channels are processed separately and thus facilitating a “split-attention effect”. Theories like the Wickens multiple resource theory (Wickens, 1980, 2008) or threaded cognition (Salvucci & Taatgen, 2008) employ the concept of multiple resources and add context to the split-attention effect. These theories suggest that visual (ambient and focal) and auditory stimuli will not interfere with one another when being perceived. However, we can also find other kinds of stimuli treated independently like tactile perception (Boles, et al 2007), but also some exceptions (for example, see Colavita, 1974) This is an important dimension that could affect the way we deliver feedback for motor learning. The modality by which real-time feedback is delivered (visual, auditory, or tactile) can avoid the use of the same resource at the same time and thus, decreasing extraneous cognitive load. Using the visual modality may be more useful when spatial resolution is needed (Ernst & Banks, 2002), and auditory modality when temporal accuracy is required (Repp & Penel, 2002). Hanson et al (2009) showed that in tasks where modalities are combined, reaction time changed significantly when participants were required to split their attention between more than one modality. Horrey and Wickens (2004) found that two tasks that needed to be performed one with focal vision and the other one with ambient vision were more efficient than two tasks involving ambient vision but less effective than two tasks. There is not much research done about the relationship between real-time feedback modality, cognitive load, and music learning but interestingly, some of the first attempts are related to the use of real-time feedback technologies for instrumental learning in the violin (see Rose Mary Grace Johnson, 2014), and for expressive percussion performance (Brandmeyer et al, 2011).

2.4 Motivational Factors and Expectancies

Recent research has been changing its focus to new variables that also influence motor learning such as attention, confidence, and motivation. Janelle, et al (1995 and 1997) showed that leaving the participants to choose for themselves when they received the feedback, turned out to be more effective than a condition in which 100% of the trials had feedback and another with 50% of the trials receiving feedback. Several other studies found that providing feedback only after “good” trials is more effective compared with groups receiving feedback only after “poor” trials (Badami et al, 2012; Chiviacowsky et al 2007; Chiviacowsky et al 2009). In the meantime, Wulf and colleagues were the first ones to report experimentally the beneficial effects of learning with an external focus of attention. Using the same ski-simulator described previously, they found that when participants were instructed to focus on the pressure exerted on the wheels of the ski-simulator, or the markers attached to a balance platform as opposed to their feet balance learning was enhanced. (Wulf, 1998). According to the constrained action hypothesis (Wulf, McNevin, & Shea, 2001; Wulf, Shea & Park, 2001), an internal focus of attention (like asking a soccer player to focus on the movement of his/her feet when shooting) induces a conscious type of control which cause individuals to interfere with the automatic control processes of their motor system. The beneficial effects of learning with an external focus of attention have been well documented for a diverse amount of different tasks and through participants of all ages (for a review see Lohse, Wulf & Lewthwaite, 2012; Marchant, 2011; Wulf 2007a, 2013). It also leads to inefficient activation of the muscular system and promotes automaticity of performance (Wulf and Lewthwaite, 2016).

Dopaminergic systems in the brain are relevant to motor, cognitive and motivational functioning (Nieoullon & Coquerel, 2003; Wise, 2004) and it is known that dopamine plays an important role in outcome expectations (which are central in the placebo response) (Lidstone, Schulzer, Dinelle, Mak, Sossi, Ruth, et al., 2010; Wager & Atlas, 2015). It is known that expectations can influence working memory, long-term memory, and attentional capture (Bollinger, Rubens, Zanto, & Gazzaley, 2010; Shomstein & Johnson, 2013). The OPTIMAL theory of motor learning (Wulf and Lewthwaite, 2016) proposes that motor learning is mediated by both attentional and motivational variables such as autonomy, confidence, and external locus of focus. According to the authors, considering these three variables could lead to positive incomes that enhance expectancies and leads to a virtuous cycle of motor learning. However, it may also occur in the opposite way. Negative outcomes in learning tend to decrease expectancies leading to a vicious cycle where learning deteriorates. From this perspective, it is important when designing educational plans to consider if they promote an internal or external locus of focus, give enough autonomy to the learners, and promote confidence in their own abilities.

2.5 Cognitive Neuroscience of Motor Learning

2.5.1 Adaptation Paradigms

Most of the studies of motor learning from a neuroscientific point of view have studied the so-called *Motor adaptation paradigms.* As we will show, these paradigms can offer us new perspectives when understanding the implications of explicit and implicit learning. Motor adaptation usually relies on tasks where participants need to reach a target by moving their arms. A common example is *force-field adaptation (*FFA) tasks where the participant needs to reach targets by using a robotic arm. Once the participant has been adapted to use the system to reach targets, then the system imposes a force proportional to the current speed of the hand and directed orthogonally to the direction of the arm movement. After introducing the force participants start to produce movement errors in the direction of the force field. After training, and once participants had learned to adapt to the force field, it is removed leading to reaching errors in the opposite direction until, again, the participant re-adapts to the absence of force. Another common example we can find is *visuomotor rotation adaptation* (VRA) tasks where participants do the reaching movements without the direct vision of their hand. Instead of that, they receive online feedback of a cursor on a screen representing the location of their hand. This cursor can be rotated a relative number of degrees resulting, again, in reaching errors related to the direction of the previous adaptation. A recent review of motor adaptation paradigms can be found in Krakauer et al (2019).

Some common characteristic we find in both FFA and VRA tasks is that, considering a constant perturbation for each trial, the error-reduction process results to be exponential while, if the perturbation varies randomly, the motor system tends to adapt to the average value of the perturbation (Hadjiosif AM, et al; 2015). These results tend to be explained from a Bayesian perspective of learning (Korenberg AT, Grahramani Z; 2002). From this perspective, the reason why errors are not corrected by the motor system in a single step is due to two factors: noise in issuing the desired motor common command, and uncertainty related to the perception of the error. Participants thus have to estimate the properties of the newly imposed perturbation taking into consideration both the noise in the motor system and changes in the environment. For example, in a target shooting exercise with darts being held outdoors, both occasional wind and “internal” noise due to distraction can make you miss your shot. However, that is not enough reason to change the already learned motor plan in the next trial. The prediction that in conditions of high uncertainty the motor system adapts less to a given error has been tested and validated empirically on multiple occasions (Burge J, et al; 2008; K. Wei, K. Körding; 2009).

Bayesian perspectives of learning and state-space models of adaptation derived from them characterize adaptation by a simple learning rule. However, there is evidence that adaptation may be supported by distinct processes operating in parallel (Krakauer et al; 2019): a “fast” process which learns as fast as is forgotten (retention decays rapidly), and a “slow” process which learns slowly but shows higher retention rates. McDougle et al (2015) suggested that these processes could be related to implicit and explicit learning mechanisms making their dissociation much easier.

In a prominent study, Mazzoni and Krakauer (2006) addressed the study of the implicit and explicit components of learning in a VRA task. According to the authors, VRA tasks are particularly amenable to dissociate those components in learning as an explicit strategy just involves aiming at a different location than the target. In that study, participants were told an explicit strategy about how to counter a 45º perturbation by aiming to a neighboring target 45º in the opposite direction. Interestingly, although participants were able to move the cursor towards the intended target, as the number of trials progressed the movements started to drift away from the target and the performance tended to worsen. At a certain moment, participants were instructed to stop using the strategy, and then, the adaptation process became as usual. This drift phenomenon, which is involuntary, is the most elegant proof of the existence of implicit processes driving adaptation. A common way of interpreting this phenomenon requires the differentiation of two different kinds of error systems in the brain: *reward errors* and *sensory-prediction errors*.

2.5.2 Reward vs Sensory-Prediction Errors

Reward or task errors represent the failure to achieve the desired movement goal while sensory-prediction errors represent a discrepancy between the actual motor command and the appropriate motor command. As demonstrated in Mazzoni and Krakauer (2006), although participants were not experiencing task errors after start using the explicit strategy, the mismatch of the intended direction of the hand and the observed direction of the cursor (that is, sensory-prediction errors) drove the implicit recalibration. Several studies have shown that implicit adaptation appears to be indifferent to reward (van der Kooij, et al, 2018; Hirashima & Nozaki, 2012; Cashaback et al, 2017) confirming more the dissociation of these two types of learning systems, and that reward and punishment seem to operate on the explicit process.

Task and reward errors can be easily removed in adaptation paradigms by removing online feedback of the cursor and providing binary feedback (hit/miss) at the end of the movement (Izawa et al, 2011; Nikooyan & Ahmed; 2015; Shmuelof et al, 2012). For example, Shmuelof et al (2012) found that learning from reward errors enhanced retention when removing visual feedback of the cursor contradicting previous interpretations that the explicit component of learning is poorly retained. On the other hand, Izawa et al (2011) reported how reward learning in adaptation leads to a narrow generalization and insignificant aftereffects when compared to implicit adaptation.

The identification of these two kinds of learning components and their interaction is crucial for the future development and evaluation of music technologies for singing and intonation accuracy. Intonation is also composed of reward and sensory prediction errors which can be isolated by offering binary feedback to learners or by real-time visualization of pitch in a screen as in the classic Seashore Tonoscope (1902).

However, and keeping discrepancies aside the retention effects of each learning component, one important lesson we can learn from these adaptation paradigms is to review the false misconception that motor learning is an implicit process. As mentioned in Krakauer et al (2019), although motor learning can be implicit, this does not mean that they were learned implicitly at the beginning. Although learning of novel tools in amnesic patients and serial reaction time tasks were thought to be driven by purely implicit processes (Milner 2005), recent work reinforces the idea that both combinations of instruction and implicit learning are essential for amnesic patients (Roy & Park, 2010) and that serial reaction time task also include explicit learning components (Frensch, Lin & Buchner, 1998; for a review see Krakauer et al, 2019). These two components can be both of particular importance for what has been called error-based learning.

2.5.3 Neurophysiological Bases of Error-Based Learning

Support for the notion that the explicit component plays a more important role during the early phases of adaptation comes from the fact that it seems that the Basal Ganglia is more active during the early phases but less in later phases (Krakauer et al, 2004; Seidler et al, 2006; Shadmehr et al, 1994). Basal Ganglia is known for its key role in reinforcement learning by dopaminergic neuron activity which is thought to encode reward prediction errors to select or inhibit wanted movements (Schultz 2016; Albin et al 1989). Patients with Huntington's disease, (a disease caused by the gradual degeneration of parts of the basal ganglia) tend to exhibit normal adaptation but reduced long-term learning (Smith & Shadmehr 2005). Similar results have been found for Parkinson's patients who exhibit normal adaptation but impaired savings (Leow et al, 2012, Marinelli et al, 2009) reinforcing the idea that reward prediction errors play a central role in retention. The basal ganglia-thalamocortical circuits have the M1 as a primary output target which receives dopaminergic inputs capable of modulating its synaptic plasticity (Dumas et al, 2012; Hosp and Luft, 2013; Hosp et al 2011; Molina-Luna et al 2009). As we will see in posterior chapters, the Anterior Cingulate Cortex (ACC) also plays an important role in generating the feedback-related negativity (f-ERN) seen in electroencephalographic activity after losses or punishments. All of these components can be characterized to rely in part on the so-called Dorsolateral Frontal Cortex (DFC) and being involved in strategic processes and selecting environmental goals in motor control (Willingham, 1999).

On the other hand, the Posterior Parietal Cortex (PPC) which receives inputs from the cerebellum is thought to be related to processes of perceptual motor-integration and sensory-prediction errors (Willingham, 1999). Online correction of movements seems to be highly dependent on the cerebellum (Blackemore et al. 2001; Miall et al. 1993) and several studies have repeatedly reported deficits of adaptation in patients with hereditary cerebellar ataxia (Tseng et al. 2007; Taylor et al 2010; Criscimanga-Hemminger et al. 2010). Support for the hypothesis that sensory-based prediction errors and cerebellum rely on implicit processes was found by Taylor et al (2010). Taylor et al. replicated the study of Mazzoni and Krakauer (2006) on patients with cerebellar ataxia. They found that after receiving the explicit strategy both the control and the cerebellar group improved their results. However, due to implicit adaptation, the control group started to show worsening performance while the cerebellar group (who cannot implicitly learn) did not.

Although we have been using adaptation paradigms as an example of how these two systems interact, error-based learning has also been studied in laboratory tasks with sequence learning paradigms, learning new controllers from zero or motor acuity in animals (see Krakauer et al. 2019 for a review). Each of these paradigms drives changes at different stages of motor planning offering significant insights to our understanding of motor learning when studied jointly.

Until now, we have been talking about explicit learning, reward, and task errors as part of one learning component of error-based learning. However, some reviews treat them separately and classify motor learning as being underlined by four different processes which can also be differentiated neurophysiologically: error-based learning (via sensory-prediction errors), use-dependent learning (via mere repetition), reinforcement learning (via reinforcement and task and reward errors) and strategy learning (via explicit use of strategies) (Spampinato and Celnik 2021). Spampinato and Celnik defend that each one of these components contributes differently to the process of learning not only at the different stages of learning but also depending on the type of motor skill (sequence, adaptation discrete *de Novo,* continuous *de Novo*…). Only taking all of these considerations in mind would allow us to be effective when studying and learning a new motor skill. And, by doing that, by extending the study of new activities of motor learning we are contributing to an essential to the global understanding of motor learning. Musical instruments represent from our point of view a great opportunity to study the interaction and coupling of motor learning with auditory perception.

2.6 Summary

In this chapter, we have briefly reviewed the different ways in which augmented feedback can influence the learning process through the perspective of various motor learning paradigms. We have also seen how the type of task, its complexity and the skill level of the participants can influence the results we find. There is no universal rule or method that will assure us of the best way to improve the quality of learning for a particular student in a particular skill. However, cognitive sciences can offer a strategic map on which pedagogues and designers of educational technologies can use in order not to start from a vacuum and consider in advance all the ways in which their intervention can affect the development of students.

2.7 References

Albin, R. L., Young, A. B., & Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. *Trends in neurosciences*, *12*(10), 366-375.

Anderson, D. I., Magill, R. A., & Sekiya, H. (2001). Motor learning as a function of KR schedule and characteristics of task-intrinsic feedback. *Journal of motor behavior*, *33*(1), 59-66.

Badami, R., VaezMousavi, M., Wulf, G., & Namazizadeh, M. (2012). Feedback about more accurate versus less accurate trials: Differential effects on self-confidence and activation. *Research quarterly for exercise and sport*, *83*(2), 196-203.

Batalla Flores, A. (2005). *Retroalimentación y aprendizaje motor: influencia de las acciones realizadas de forma previa a la recepción del conocimiento de los resultados en el aprendizaje y la retención de habilidades motrices*. Universitat de Barcelona.

Battig, W. F. (1966). Evidence for coding processes in “rote” paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, *5*(2), 177-181.

Battig, W. F. (1972). Interference During Learning as a Sources of Facilitation in Subsequent Retention and Transfer.

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport*, *12*(9), 1879-1884.

Boles, D.B. , Bursk, J.H. , Phillips, J.B. , & Perdelwitz, J.R. (2007). Predicting dual-task performance with the multiple resource questionnaire. Human Factors, 49, 32-45.

Brandmeyer, A., Timmers, R., Sadakata, M., & Desain, P. (2011). Learning expressive percussion performance under different visual feedback conditions. *Psychological Research*, *75*(2), 107-121.

Burge, J., Ernst, M. O., & Banks, M. S. (2008). The statistical determinants of adaptation rate in human reaching. *Journal of vision*, *8*(4), 20-20.

Cashaback, J. G., McGregor, H. R., Mohatarem, A., & Gribble, P. L. (2017). Dissociating error-based and reinforcement-based loss functions during sensorimotor learning. *PLoS computational biology*, *13*(7), e1005623.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and instruction*, *8*(4), 293-332.

Chiviacowsky, S., & Wulf, G. (2007). Feedback after good trials enhances learning. *Research quarterly for exercise and sport*, *78*(2), 40-47.

Chiviacowsky, S., Wulf, G., Wally, R., & Borges, T. (2009). Knowledge of results after good trials enhances learning in older adults. *Research quarterly for exercise and sport*, *80*(3), 663-668.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*(2), 409-412.

Criscimanga-Hemminger SE, Bastian AJ, Shadmehr R. (2010) Size of error affects cerebellar contributions to motor learning. Journal of Neurophysiology, 103(4), 2275–2284.

Dumas, E. M., van den Bogaard, S. J., Ruber, M. E., Reilmann, R., Stout, J. C., Craufurd, D., ... & Roos, R. A. (2012). Early changes in white matter pathways of the sensorimotor cortex in premanifest Huntington's disease. *Human brain mapping*, *33*(1), 203-212.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429-433.

Frensch, P. A., Lin, J., & Buchner, A. (1998). Learning versus behavioral expression of the learned: The effects of a secondary tone-counting task on implicit learning in the serial reaction task. *Psychological research*, *61*(2), 83-98.

Goh, H. T., Sullivan, K. J., Gordon, J., Wulf, G., & Winstein, C. J. (2012). Dual-task practice enhances motor learning: a preliminary investigation. *Experimental brain research*, *222*(3), 201-210.

Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of motor behavior*, *36*(2), 212-224.

Hadjiosif, A. M., & Smith, M. A. (2015). Flexible control of safety margins for action based on environmental variability. *Journal of Neuroscience*, *35*(24), 9106-9121.

Hanson, J. V., Whitaker, D., & Heron, J. (2009). Preferential processing of tactile events under conditions of divided attention: effects of divided attention on reaction time. *Neuroreport*, *20*(15), 1392.

Hirashima, M., & Nozaki, D. (2012). Distinct motor plans form and retrieve distinct motor memories for physically identical movements. *Current Biology*, *22*(5), 432-436.

Horrey, W. J., & Wickens, C. D. (2004). Driving and side task performance: The effects of display clutter, separation, and modality. *Human factors*, *46*(4), 611-624.

Hosp, J. A., & Luft, A. R. (2013). Dopaminergic meso-cortical projections to m1: role in motor learning and motor cortex plasticity. *Frontiers in neurology*, *4*, 145.

Hosp, J. A., Pekanovic, A., Rioult-Pedotti, M. S., & Luft, A. R. (2011). Dopaminergic projections from midbrain to primary motor cortex mediate motor skill learning. *Journal of Neuroscience*, *31*(7), 2481-2487.

Izawa, J., & Shadmehr, R. (2011). Learning from sensory and reward prediction errors during motor adaptation. *PLoS computational biology*, *7*(3), e1002012.

James, W. (1890). *Habit*. H. Holt.

Janelle, C. M., Barba, D. A., Frehlich, S. G., Tennant, L. K., & Cauraugh, J. H. (1997). Maximizing performance feedback effectiveness through videotape replay and a self-controlled learning environment. *Research quarterly for exercise and sport*, *68*(4), 269-279.

Janelle, C. M., Kim, J., & Singer, R. N. (1995). Subject-controlled performance feedback and learning of a closed motor skill. *Perceptual and motor skills*, *81*(2), 627-634.

Johnson, R. M. G. (2014). *In touch with the wild: Exploring real-time feedback for learning to play the violin* (Doctoral dissertation, UCL (University College London)).

Korenberg, A. T., & Ghahramani, Z. (2002). A Bayesian view of motor adaptation. *Current Psychology of Cognition*, *21*(4/5), 537-564.

Krakauer, J. W., Ghilardi, M. F., Mentis, M., Barnes, A., Veytsman, M., Eidelberg, D., & Ghez, C. (2004). Differential cortical and subcortical activations in learning rotations and gains for reaching: a PET study. *Journal of neurophysiology*, *91*(2), 924-933.

Krakauer, J. W., Hadjiosif, A. M., Xu, J., Wong, A. L., & Haith, A. M. (2019). Motor learning. *Compr Physiol*, *9*(2), 613-663.

Kruse-Weber, S., & Parncutt, R. (2014). Error management for musicians: an interdisciplinary conceptual framework. *Frontiers in psychology*, *5*, 777.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in cognitive sciences*, *9*(2), 75-82.

Lee, T. D., White, M. A., & Carnahan, H. (1990). On the role of knowledge of results in motor learning: Exploring the guidance hypothesis. *Journal of Motor Behavior*, *22*(2), 191-208.

Leow, L. P., Beckert, L., Anderson, T., & Huckabee, M. L. (2012). Changes in chemosensitivity and mechanosensitivity in aging and Parkinson's disease. *Dysphagia*, *27*(1), 106-114.

Lidstone, S. C., Schulzer, M., Dinelle, K., Mak, E., Sossi, V., Ruth, T. J., ... & Stoessl, A. J. (2010). Effects of expectation on placebo-induced dopamine release in Parkinson disease. *Archives of general psychiatry*, *67*(8), 857-865.

Lohse, K. R., Wulf, G., & ITE, R. L. W. (2012). Attentional focus affects movement efficiency. In *Skill acquisition in sport* (pp. 66-84). Routledge.

Magill, R. A. (1994). The influence of augmented feedback on skill learning depends on characteristics of the skill and the learner. *Quest*, *46*(3), 314-327.

Magill, R. A. (2001). *Motor learning: Concepts and applications* (6th ed.). Madison, WI: Brown & Benchmark.

Marchant, D. C. (2011). Attentional focusing instructions and force production. *Frontiers in psychology*, *1*, 210.

Marinelli, L., Crupi, D., Di Rocco, A., Bove, M., Eidelberg, D., Abbruzzese, G., & Ghilardi, M. F. (2009). Learning and consolidation of visuo-motor adaptation in Parkinson's disease. *Parkinsonism & related disorders*, *15*(1), 6-11.

Maxwell, J. P., Masters, R. S. W., Kerr, E., & Weedon, E. (2001). The implicit benefit of learning without errors. *The Quarterly Journal of Experimental Psychology Section A*, *54*(4), 1049-1068.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of educational psychology*, *90*(2), 312.

Mazzoni, P., & Krakauer, J. W. (2006). An implicit plan overrides an explicit strategy during visuomotor adaptation. *Journal of neuroscience*, *26*(14), 3642-3645.

McDougle, S. D., Bond, K. M., & Taylor, J. A. (2015). Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning. *Journal of Neuroscience*, *35*(26), 9568-9579.

Miall, R. C., Weir, D. J., Wolpert, D. M., & Stein, J. F. (1993). Is the cerebellum a smith predictor?. *Journal of motor behavior*, *25*(3), 203-216.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Milner, B. (2005). The medial temporal-lobe amnesic syndrome. *Psychiatric Clinics*, *28*(3), 599-611.

Molina-Luna, K., Pekanovic, A., Röhrich, S., Hertler, B., Schubring-Giese, M., Rioult-Pedotti, M. S., & Luft, A. R. (2009). Dopamine in motor cortex is necessary for skill learning and synaptic plasticity. *PloS one*, *4*(9), e7082.

Nieoullon, A., & Coquerel, A. (2003). Dopamine: a key regulator to adapt action, emotion, motivation and cognition. *Current opinion in neurology*, *16*, S3-S9.

Nikooyan, A. A., & Ahmed, A. A. (2015). Reward feedback accelerates motor learning. *Journal of neurophysiology*, *113*(2), 633-646.

Onla-Or, S., & Winstein, C. J. (2008). Determining the optimal challenge point for motor skill learning in adults with moderately severe Parkinson's disease. *Neurorehabilitation and neural repair*, *22*(4), 385-395.

Park, J. H., Shea, C. H., & Wright, D. L. (2000). Reduced-frequency concurrent and terminal feedback: a test of the guidance hypothesis. *Journal of motor behavior*, *32*(3), 287-296.

Poolton, J. M., Masters, R. S. W., & Maxwell, J. P. (2005). The relationship between initial errorless learning conditions and subsequent performance. *Human movement science*, *24*(3), 362-378.

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1085.

Roy, S., & Park, N. W. (2016). Effects of dividing attention on memory for declarative and procedural aspects of tool use. *Memory & cognition*, *44*(5), 727-739.

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological review*, *115*(1), 101.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological review*, *82*(4), 225.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological review*, *84*(1), 1.

Schultz, W. (2016). Reward functions of the basal ganglia. *Journal of neural transmission*, *123*(7), 679-693.

Seashore, C. E. (1902). A voice tonoscope. *University of Iowa Studies in Psychology*, *3*, 18-28.
Seidler, R. D., Noll, D. C., & Chintalapati, P. (2006). Bilateral basal ganglia activation associated with sensorimotor adaptation. *Experimental brain research*, *175*(3), 544-555.

Shadmehr, R., & Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *Journal of neuroscience*, *14*(5), 3208-3224.

Sherrington, C. S. (1906). Observations on the scratch-reflex in the spinal dog. *The Journal of physiology*, *34*(1-2), 1-50.

Shmuelof, L., Huang, V. S., Haith, A. M., Delnicki, R. J., Mazzoni, P., & Krakauer, J. W. (2012). Overcoming motor "forgetting" through reinforcement of learned actions. *Journal of Neuroscience*, *32*(42), 14617-14621a.

Shomstein, S., & Johnson, J. (2013). Shaping attention with reward: Effects of reward on space-and object-based selection. *Psychological science*, *24*(12), 2369-2378.

Skinner, B. F. (1953). Some contributions of an experimental analysis of behavior to psychology as a whole. *American Psychologist*, *8*(2), 69.

Smith, M. A., & Shadmehr, R. (2005). Intact ability to learn internal models of arm dynamics in Huntington's disease but not cerebellar degeneration. *Journal of neurophysiology*, *93*(5), 2809-2821.

Spampinato, D., & Celnik, P. (2021). Multiple motor learning processes in humans: Defining their neurophysiological bases. *The Neuroscientist*, *27*(3), 246-267.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*(3), 251-296.

Swinnen, S. P. (1996). Information feedback for motor skill learning: A review. In H. N. Zelaznik (Ed.), *Advances in motor learning and control* (pp. 37-66). Champaign, IL: Human Kinetics.

Taylor, J. A., Klemfuss, N. M., & Ivry, R. B. (2010). An explicit strategy prevails when the cerebellum fails to compute movement errors. *The Cerebellum*, *9*(4), 580-586.

Tseng, Y. W., Diedrichsen, J., Krakauer, J. W., Shadmehr, R., & Bastian, A. J. (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *Journal of neurophysiology*, *98*(1), 54-62.

Van der Kooij, K., Oostwoud Wijdenes, L., Rigterink, T., Overvliet, K. E., & Smeets, J. B. (2018). Reward abundance interferes with error-based learning in a visuomotor adaptation task. *PLoS One*, *13*(3), e0193002.
van Merrienboer, J. J. (2015). Worked examples in the classroom. *Perspectives on medical education*, *4*(6), 282-283.

Wager, T. D., & Atlas, L. Y. (2015). The neuroscience of placebo effects: connecting context, learning and health. *Nature Reviews Neuroscience*, *16*(7), 403-418.

Wei, K., & Kording, K. (2009). Relevance of error: what drives motor adaptation?. *Journal of neurophysiology*, *101*(2), 655-664.
Wickens, C. D. (1980). The structure of attentional resources. *Attention and performance VIII*, *8*, 239-257.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors*, *50*(3), 449-455.

Willingham, D. B., & Goedert-Eschmann, K. (1999). The relation between implicit and explicit learning: Evidence for parallel development. *Psychological Science*, *10*(6), 531-534.

Winstein, C. J., Pohl, P. S., & Lewthwaite, R. (1994). Effects of physical guidance and knowledge of results on motor learning: support for the guidance hypothesis. *Research quarterly for exercise and sport*, *65*(4), 316-323.

Wise, R. A. (2004). Dopamine, learning and motivation. *Nature reviews neuroscience*, *5*(6), 483-494.

Wulf, G. (2013). Attentional focus and motor learning: a review of 15 years. *International Review of sport and Exercise psychology*, *6*(1), 77-104.

Wulf, G., & Lewthwaite, R. (2016). Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychonomic bulletin & review*, *23*(5), 1382-1414.

Wulf, G., & Su, J. (2007). An external focus of attention enhances golf shot accuracy in beginners and experts. *Research quarterly for exercise and sport*, *78*(4), 384-389.

Wulf, G., Höß, M., & Prinz, W. (1998). Instructions for motor learning: Differential effects of internal versus external focus of attention. *Journal of motor behavior*, *30*(2), 169-179.

Wulf, G., McNevin, N., & Shea, C. H. (2001). The automaticity of complex motor skill learning as a function of attentional focus. *The Quarterly Journal of Experimental Psychology Section A*, *54*(4), 1143-1154.

Wulf, G., Shea, C. H., & Matschiner, S. (1998). Frequent feedback enhances complex motor skill learning. *Journal of motor behavior*, *30*(2), 180-192.

Wulf, G., Shea, C., & Park, J. H. (2001). Attention and motor performance: preferences for and advantages of an external focus. *Research quarterly for exercise and sport*, *72*(4), 335-344.

Zanto, T. P., Rubens, M. T., Bollinger, J., & Gazzaley, A. (2010). Top-down modulation of visual feature processing: the role of the inferior frontal junction. *Neuroimage*, *53*(2), 736-745.

Blanco, A. D., & Ramirez, R. (2019). Evaluation of a sound quality visual feedback system for bow learning technique in violin beginners: An EEG study. *Frontiers in psychology, 10,* 165.

3. EVALUATION OF A SOUND QUALITY VISUAL FEEDBACK SYSTEM FOR BOW LEARNING TECHNIQUE IN VIOLIN BEGINNERS: AN EEG STUDY (NEW REVIEWED VERSION FROM 2021)

3.1 Abstract

Current music technologies can assist in the process of learning to play a musical instrument and provide objective measures for evaluating the improvement of music students in concrete music tasks. In this paper, we investigated the effects of a sound quality visual feedback system (SQVFS) in violin learning. In particular, we studied the EEG activity of a group of participants with no previous violin playing experience while they learned to produce a stable sound (regarding pitch, dynamics, and timbre) in order to find motor learning biomarkers in a music task. Eighteen subjects with no prior experience in violin playing were divided into two groups: participants in the first group (experimental group, N = 9) practiced with instructional videos and offline feedback from the SQVFS provided in alternation with their performance, while participants in a second group (control group, N = 9) practiced with the instructional videos only. A third group of violin experts (players with more than 6 years of experience) performed the same task for comparative purposes (N = 7). All participants were asked to perform 20 trials (4 blocks of 5 trials) consisting of a violin bowing exercise while their EEG activity and their produced sound was recorded. Significant sound quality improvements along the session were found in all participants with the exception of participants in the expert group. In addition, participants in the experimental group showed increased interest in the learning process and significant improvements at the last block not present in the control group. A significant correlation between the levels of frontal gamma band power and the sound improvement along the task was found in both the experimental and control group. This result is consistent with the temporal binding model which associates gamma band power with the role of integrating (binding) information processed in distributed cortical areas. Task complexity demands more cognitive resources, more binding and thus, gamma band power enhancement, which may be reduced as the demanded task begins

to be automated as it is likely to be the case in both beginners groups.

3.2 Introduction

3.2.1 Feedback in Motor Learning

There is ample literature reporting on the effects of feedback in motor learning tasks. From the first attempts to experimentally test Thorndike's theory of learning (Trowbridge and Cason, 1932), to more recent work (Newell, 1974; Salmoni et al., 1984; Schmidt et al., 1989; Winstein, 1991), studies consistently show how feedback during motor learning increases the rate of improvement over trials. However, the effects on retention and long-term learning are less clear. Approaches providing infrequent feedback have shown improvements in retention phases usually performed 24 h after the experiment (for a review of feedback studies see Winstein, 1991). Still, it is important to consider that this kind of research has focused on studying the effect of feedback in controlled environments where the effect of intrinsic feedback (e.g., visual, auditory, proprioceptive) pertaining to the outcome movement is minimized. This kind of experimental design may imitate the process of learning in a person with sensory deficits who is unable to use intrinsic feedback and depends on the extrinsic feedback (i.e., related to the result of the action) given by the experimenter. On the other hand, motivation is very important in learning (Elwell and Grindley, 1938). Some researchers have attempted to control the motivational effects of feedback in experimental setups where subjects were asked to improve their performance or were given explicit goals (Locke, 1966; Locke and Bryan, 1966) in order to find a significant goal effect. Although those results may relate with the learning of motor skills (e.g., in sports, music), extrinsic feedback could play a different role depending on the task performed, so more specific research is needed in order to understand better the impact of feedback technologies in music students.

3.2.2 Technology-Enhanced Music Learning

Mastering the violin and other bowed-string instruments require special considerations compared with other musical instruments. As opposed to the piano, for instance, pitch control in the violin is continuous and correct intonation is an important issue. In addition, the process of good sound generation in the violin is a notorious complex task which requires precise spatiotemporal control of bowing gestures (Schoonderwaldt and Demoucron, 2009). Acquiring correct bowing motor skills require many hours of practice in which aural feedback is crucial for students to adjust their motor gestures to generate good sound. According to Konczak and Jaeger (2009) novice players need approximately 700 practice hours to achieve bowing skills comparable to those of experts. Moreover, string players have the highest risk of playing-related musculoskeletal injuries/disorders (PRMDs) with the neck and shoulder being the main body parts affected (Middlestadt and Fishbein, 1989).

A recent survey on Australian higher education music students showed how the use of Youtube and self-recording has become common practice among them (Zhukov, 2015). Youtube offers videos of professional musicians performing music repertoire pieces as a model for students while self-recording has become an important tool for self-evaluation. Previous research (Kepner, 1986; Bundy, 1987) found that high school instrumentalists were more able to identify musical errors when hearing tape recordings of their own performances than when actually performing the pieces. In particular, Bundy (1987) explains the obtained results by a sensory blocking theory which hypothesizes that when musicians are concerned with monitoring a big number of sensory aspects involved in performance (like sight-reading or finger movements) the sense of hearing, which is perceived to be of lesser importance, is blocked. However, recent research (Hewitt, 2001) studied the effects of listening to a model (i.e., an expert reference performance), listening to oneself on audiotape, and self-evaluation on junior high school instrumentalists, concluding that there is a significant interaction effect for modeling and self-evaluation. However, self-evaluation (which in the case of the study consisted on the Woodwind Brass Solo Evaluation Form Saunders and Holahan, 1997) or self-recording on their own were not found to be effective strategies for improving music performances. Although self-recording may be important, in absence of a teacher it requires the student to be his/her own judge which may be problematic. The superiority of highly trained musicians encoding spectral and temporal features of music-sound compared with non-musicians has been found in a large number of neuroscientific studies (Besson et al., 1994; Koelsch et al., 1999; Pantev et al., 2001; Tervaniemi et al., 2005; Hutchins and Peretz, 2012). For example, Koelsch et al. (1999) demonstrated, using electrophysiological and behavioral data, that highly trained violin players are able to detect automatically undetectable pitch differences for nonmusicians.

Current music technologies provide us with objective measures of student improvement in specific music tasks. Thus, such technologies can allow us to monitor the learning process of music students in order to provide better and personalized learning strategies. In addition, objective measures about music students' performance may serve as additional information which could complement the verbal feedback given by the teacher. In the past, the role of feedback in music learning has been addressed mainly to study the effects of real-time visual feedback (RTVF) in singing. Welch et al. (1989) studied the effect of a feedback system called SINGAD (Singing Assessment and Development) in 32 primary school children aged 7 years. The system provided a real-time F0 trace plotted against time together with the target notes displayed in order to guide time and pitch accuracy. The study reported improved pitch accuracy by using the system. Previous research has studied the effect of using different kinds of interfaces and different kinds of feedback in singing voice (Thorpe, 2002; Welch et al., 2004; Wilson et al., 2008; Leong and Cheng, 2014), trombone (Schlegel and Gregory Springer, 2018), piano (Hamond, 2017), and violin (Wang et al., 2012). Although there are differences in the way RTVF may improve performance, most of the previous studies reported beneficial effects of RTVF in learning. An extensive review on feedback and technology applied to music learning can be found in Hamond (2017). The same author also investigated the nature and application of combined visual-auditory feedback generated by technology systems in higher education piano learning and teaching contexts. As suggested by self-reports collected from music students, the feedback provided could increase conscious-awareness of their own performance. As related by one of those students: “Sometimes you know in your mind what you want to do, […] but sometimes you do not realize exactly what you're doing in practice[…]. So, when you hear, you can clearly see what you are doing and what you're not” (Hamond, 2017, p. 278).

Regarding violin learning, special efforts have been done to offer different kinds of feedback, not only on the produced pitch but also on timbre, good posture, and bowing technique. The i-Maestro project (Ng and Nesi, 2008) was one of the first steps in that direction offering tools based on gesture analysis and audio processing. More recently the TELMI project has developed tools for providing feedback on timbre quality, pitch and timing accuracy, posture and bowing techniques, and musical expression (Ortega et al., 2017; Dalmazzo et al., 2018; Giraldo et al., 2018; Zacharias et al., 2018). Optical motion capture combined with sensors has also been used to extract bowing parameters from violin performance (Schoonderwaldt and Demoucron, 2009; Deutsch, 2011) allowing to study and compare the motor patterns of professional and student violinists. Tracking violin performance using low-cost methods has also been investigated by Perez Carrillo and Wanderley (2012) through the sole use of audio signal and a system trained on empirical data previously collected with a highly accurate sensing system. Pardue et al. (2015) also explored low-cost methods using a resistive fingerboard and four optical reflectance sensors placed on the bow stick. Some attempts have been done in order to evaluate motion capture techniques to teach violin skills. For example, Van Der Linden et al. (2011) used a wearable system to teach good posture and bowing technique to novice violin students and found a larger improvement when compared with a control group of subjects who received the same number of training sessions using conventional teaching techniques. One possible limitation of the previously mentioned study is that the quality of generated sound is not taken into account, while in violin learning the production of a good sound is one of the main reasons for learning a correct bowing technique.

The work of Romaní et al. (2015) aimed to identify audio descriptors, extracted from the recordings of professional musicians while playing single notes, maximally correlated with their own subjective opinions about the quality of the produced sound. Some of the features that showed higher correlations were those characterizing pitch stability and dynamic stability. This research led to the implementation of Cortosia (Korg, 2018) an app owned by the Korg company, which aims to provide students with visual feedback about the quality of their produced sound. More recently, Giraldo et al. (2018) investigated the application of machine learning techniques to obtain sound quality model and implemented a real-time feedback system for enhancing violin learning. However, no studies have been done until now to evaluate the pedagogical effectiveness of such systems.

One could be tempted to offer simultaneous real-time feedback in violin learning environments (e.g., violin-bow orientation, bowing trajectory, and timbre quality). However, a common concern found in user studies offering several simultaneous feedback is that participants usually have difficulties dealing with them (Van Der Linden et al., 2011; Johnson et al., 2012; Johnson, 2014). Delivering the different feedback separately at different times and as requested by the user could be one possibility to resolve that problem, as has been the approach in the TELMI project. Another common concern is the potential dependency that feedback systems could create on students.

Recent research (Brandmeyer et al., 2011) has evaluated the effects of RTVF on expressive percussion performance interpreting their results using the Cognitive Load Theory (CLT) (Paas et al., 2003). In their work, they differentiate between three different kinds of cognitive load: intrinsic, extraneous and germane. Intrinsic cognitive load is associated with the difficulty of the particular task whereas extraneous cognitive load relates to the manner in which information is received. On the other hand, germane cognitive load relates to the mental resources involved in learning in general, independently of the task. Brandmeyer et al. (2011) found empirically that too many visual elements can create a high extraneous cognitive load in participants, dividing their attention and leading to poorer learning outcomes. However, apart from behavioral measures, no other measures were used to evaluate the amount of cognitive load participants were experiencing. Physiological measures can provide objective measures of the mental work a person is experiencing while learning. Recently, the neural activity associated with learning tasks has been investigated by the neuroscientific community using both functional neuroimaging and electroencephalography (EEG) techniques. In particular, EEG is the most common technique used to study cognitive load from brain activity and one of the most feasible among other electro-physiological measures (Miller, 2001).

3.2.3 E-Learning Systems Inspired in Brain Activity (EEG)

Event-related (de)synchronization (ERS/ERD) is a well-established measure for the quantification of changes in different frequency bands of the EEG signal. It reflects the decrease (desynchronization) or increase (synchronization) in a band power during a test (time period where the subject is performing a specific task which demands cognitive load) compared with a reference baseline (time period without any task demands). This is usually done for each electrode. A positive ERD/ERS value means a decrease in a band power (desynchronization, ERD) while a negative value indicates an increase in band power (ERS). It has been reported repeatedly for several researchers that alpha and theta band activity (8–13 Hz and 4–7 Hz, respectively) is very sensitive to task difficulty or cognitive load in a wide variety of task demands (Klimesch, 1999; Gevins and Smith, 2003; Neubauer et al., 2006). Generally, as cognitive load increases, frontal midline theta band increases, and posterior alpha band decreases. Larger alpha band ERD has been associated with highly intelligent subjects and good performance (Jaušovec and Jaušovec, 2004). Explanations of this phenomenon are usually delegated to the neural efficiency hypothesis which assumes that high alpha band power reflects cortical inhibition. On the other hand, theta has been investigated for its implications in memory performance (Raghavachari et al., 2006) showing strong increases in the frontal area during the encoding and retention period (Maurer et al., 2015). Thus, an alpha band power decrease at posterior sites (larger alpha band ERD) and a frontal theta increase represent a general index for cognitive demands. Some research also highlights the importance of gamma band waves (30–100 Hz) which its enhancement is observed within a task-specific spatial distribution (Fitzgibbon et al., 2004) and seems to be correlated with cognitive load in humans (Howard et al., 2003). The temporal binding model gives gamma band the responsible role of integrating (binding) information processed in distributed cortical areas. Task complexity demands more cognitive resources, more binding and thus, gamma band power enhancement. Interestingly, some research has found that subjects with musical training show enhanced induced gamma band activity (Shahin et al., 2008; Trainor et al., 2009) suggesting it reflects a superior binding of acoustical features (e.g., pitch, timbre, harmony) and processes also thought to be enhanced by music training, e.g., anticipation, expectation and attention (Bhattacharya et al., 2001; Sokolov et al., 2004; Gurtubay et al., 2006).

The viability of the use of EEG to test the effectiveness of learning materials designs has been provided by some studies (Antonenko and Niederhauser, 2010; Antonenko et al., 2010). Thanks to the measure of participants' cognitive load it is possible to assess which learning strategy seems to work better in concrete situations. On the other hand, some studies have also started to investigate the potential of real-time monitoring of mental workload to improve human performance. For instance, Kohlmorgen et al. (2007) describes a system to reduce distractions while driving by monitoring mental workload.

EEG has also been used to improve music performance through the use of an increasingly popular technique called neurofeedback. It consists of learning, through visual or auditory feedback, how to modify voluntarily your own mental activity. Several studies have reported improvements in the music performance of those musicians who received a neurofeedback session on the theta /alpha protocol (i.e., learning how to maximize the theta to alpha ratio) before a performance, compared with other groups who received different kinds of relaxing techniques like the Alexander technique or different neurofeedback protocols (Bazanova et al., 2009; Gruzelier, 2009). Similar results have also been found for dancers (Raymond et al., 2005). According to the authors, the production of theta waves with eyes closed is related to the hypnogogic process which at the same time is associated with an improvement of the creative process and well-being of users.

Other studies have tried the use of theta-EEG and EMG biofeedback with violinists while they perform, with positive results (Silvana et al., 2008). The pre-recorded sound of applauses as feedback gave the musician the opportunity to recognize which is the adequate mental and muscular state needed for optimum performance. The reason to train theta during the performance was that some investigations have found enhanced theta activity in highly-skilled professional musicians (Klimesch et al., 1997; Bazanova and Aftanas, 2006). According to the neural efficiency hypothesis experts should show lower brain activation (which means higher theta power and more efficient networks), and thus, training students to learn how to use their brain more efficiently could lead to an enhancement of their performance.

The relationship between EEG power changes and proficiency have also been reported in sports activities such as rifle marksmanship (Haufler et al., 2000; Kerick et al., 2004), archery (Salazar et al., 1990; Landers et al., 1994) and golf (Crews and Landers, 1993; Babiloni et al., 2008). This research shows how the most predictive data of expertise is recorded before the skilled movements occur, in what is called the "pre-shot routine." For instance, it has been shown that the magnitude of the increase in theta power before the shot is correlated with the accuracy of the shot. Berka et al. (2010) tracked the learning process of beginners in rifle marksmanship while firing a total of 40 shots and correlated the accuracy of the results with the EEG power activity, finding increases in theta and high theta Bands (6–7 Hz) just as experts showed during all their trials. They also compared the results of the learning group with another one which, additionally, received a neurofeedback training based on the same frequency bands showing how the neurofeedback group obtained significantly better results. Similar results were also found by Gentili et al. (2008) where subjects had to learn and interact with new tools. They found increases in alpha and theta band power in the frontal and temporal lobes during movement planning (i.e., just before the movement, like in the pre-shot routine).

However, in a recent study (Gutierrez and Ramírez-Moreno, 2016) changes in brain activity associated with the progression of the learning experience were estimated with different results. They monitored the process of learning to typewrite using the Colemark keyboard layout, which is an alternative to the QWERTY layout, finding a decreasing trend of the beta and gamma bands. They interpreted beta band decrease as a result of long-duration repetitive hand movements, similar to results found by as Niemann et al. (1991) and Erbil and Ungan (2007), and explained the gamma band decrease as a consequence of the temporal binding model previously mentioned, which associates gamma band activity with coupling perception and learning, as reported by Gruber and Müller (2005).

Recently, the study of movement-related cortical potentials reported lower amplitudes of MRCP for expert performers compared with novice performers (Di Russo, Pitzalis, et al., 2005, Fattapposta et al., 1996, Hatta et al., 2009, Kita et al., 2001, Wright et al., 2012). MRCP seems to be related with amplitudes at the beta band (Tan et al., 2014, Torrecillos., 2015) which would allow explaining in part, the reported results of Gutierrez and Ramirez-Moreno (2016).

3.2.4 Aims of the Present Work

The aim of this work is to contribute to the understanding of the effects of feedback in music learning from an electrophysiological point of view. For this purpose, we have evaluated the effectiveness of using a sound quality visual feedback system (SQVFS) to improve the quality of sound produced by of novice violin players while their EEG activity and the violin sound they produced was recorded. These recorded data provides non-invasive biomarkers of motor learning in a musical task. Participants (with no previous experience with violin or any other bowed string instrument) were asked to produce a stable and sustained violin sound on an open string (i.e., the second string in the violin). The choice of using an open string was to allow participants to exclusively concentrate their attention to control the bow movement. This task requires to control and change the pressure of the bow along the whole movement due to the fact that bow pressure requires to be heavier at the frog and lighter at the tip. If the pressure of the bow is not constant along the movement both pitch and energy of the produced tone could change. For that reason, we hypothesized that the use of dynamic stability and pitch stability audio descriptors, as Romaní et al. (2015) did, to measure sound quality among trials would allow us to track improvement through the session. We also offered the numerical result of the descriptors as feedback to the participants (i.e., the SQVFS).

Participants were divided into two groups. Both of them had access to learning materials and reference videos during the experiment, but in addition one of the groups received offline feedback about the quality of their performance given by the SQVFS. The quality of the produced sound, as well as the EEG activity of each participant, was recorded during 4 blocks of 5 trials each (20 trials in total). An additional group of violin experts was considered in the experiment for comparative purposes. Data recollected in this study is publicly available in Zenodo (Casares and Ramírez, 2018) and the code to analyze it in Github (Blanco, 2018).

3.3 Materials and Methods

3.3.1 Participants

The study was carried out in the recording studio located in the Information and Communication Technologies Engineering (ETIC) department of the Universitat Pompeu Fabra, Barcelona and included the participation of twenty-five right-handed subjects. Participants conceded their written consent and procedures were approved by the Conservatoires UK Research Ethics committee on 04/04/2017, following the guidelines of the British Psychological Society. Participants provided information about their musical skills, main instrument and years of music training. Those with extensive experience in violin playing were included in the expert group [EG;6 male, 1 female; mean age: 35.2 (9.01); mean years studying violin: 7.6 (2.19)]. Participants with no violin (or viola, double-bass or cello) experience were included in the beginner's group. This last group, was randomly divided in two groups: the Feedback Group [FG; 6 male, 3 female; mean age: 27.57 (4.46); all of them were musicians with several years of experience, mean: 9 (5.07)] practiced with instructional videos and offline feedback from the SQVFS reflecting the quality of their produced sound, while the control group [CG; 8 male, 1 female; mean age: 27.2 (2.28)] practiced with the instructional videos only. All participants were musicians with several years of experience, mean: 10.8 (4.65).

3.3.2 Materials

EEG data were acquired using the Emotiv EPOC EEG device. The Emotiv EPOC consists of 16 wet saline electrodes, located at the positions AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 according to the international 10-20 system (see Figure 1). The two remaining electrodes located at P3 and P4 are used as reference. The data acquired were digitized using the embedded 16-bit ADC with 128 Hz sampling frequency per channel and sent to the computer via Bluetooth. The Emotiv Control Panel software was used to monitor visually the impedance of the electrodes contact to the scalp. The data were recorded using the OpenViBE platform

(Renard et al., 2010) and later processed in EEGLAB (Delorme and Makeig, 2004) under the Matlab environment (MATLAB, 2010).

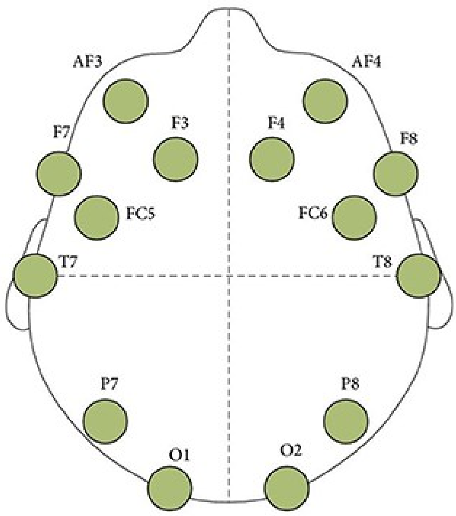

Figure 1. Emotiv EPOC electrodes aligned with positions in the 10–20 system.

A Zoom H4N handy recorder was used to record the audio of each trial which was processed in Matlab using the "Yin pitch estimation toolbox" (Llimona, 2015) in order to extract audio features for assessing sound quality and provide feedback to participants. Yin is a widely used algorithm to estimate fundamental frequency both in speech and music (De Cheveigne and Kawahara, 2002).

Visual feedback provided to the BF group consisted of graphs generated in Matlab showing the sound quality score in the y-axis and the trial number in the x-axis. Feedback was intended to allow participants to monitor their progress and compare their performance to that of an expert participant who previously did the experiment (also plotted in the feedback screen) (see Figure 2).

Instructional videos about basic violin playing techniques, e.g., stance, violin position, bow position, and grip, were used to provide participants with basic information. The videos were collected from the web (Sassmannshaus, 2018) (see Figure 3). In addition, we recorded a reference video of the requested task performed by a professional violin player. The produced video was shown to all participants to explain the task to be performed. The video can be found in Zenodo (Casares and Ramírez, 2018).

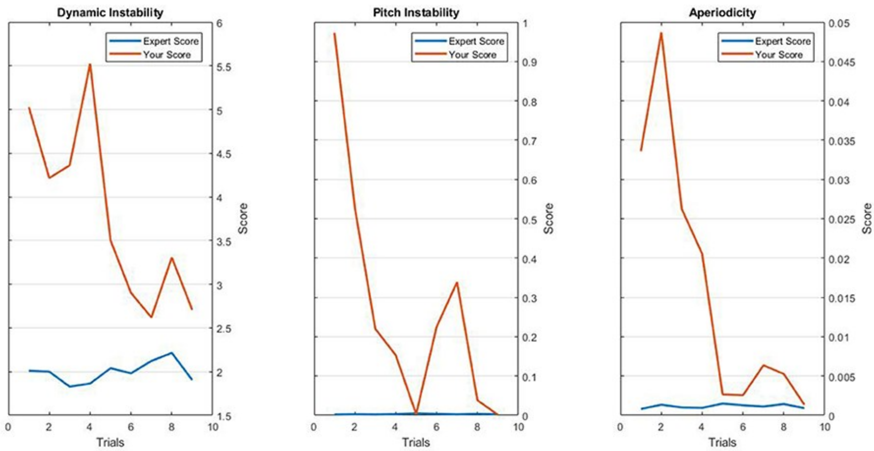

Figure 2. Example of the results in dynamic instability, pitch instability and aperiodicity shown to a subject just after performing trial number nine.

EEG acquisition and audio processing were performed on different laptops (PC1 and PC2, respectively). To synchronize audio and EEG data PC2 sent markers to OpenVibe in PC1 through OSC everytime a new trial began and ended. The experimenter controlled the display of instructional videos and the reference expert video for both BNF and BF groups and sound quality visual feedback for the BF group (see Figure 4).

3.3.3 Methods

Due to the nature of the experiment, it was not possible to conduct a double-blind study. In order to avoid unconscious bias during the instructions given to participants, both beginner groups (i.e., FG and CG) watched the same set of instructional videos on violin and bow position and stance with a total duration of 10 min (Sassmannshaus, 2018). Participants watch the videos while the EEG device was positioned on their heads. Once the setup of the EEG device and the videos were finished, participants proceeded to perform the violin bowing exercise which consisted of the alternation of eight up and down bowing movements using the full length of the bow with the goal of producing a sound in the A open string. Participants were asked to produce a stable and sustained sound at the same tempo as the reference video. Participants were also asked to minimize blinking and facial movements during the exercise to avoid artifacts in the EEG signal.

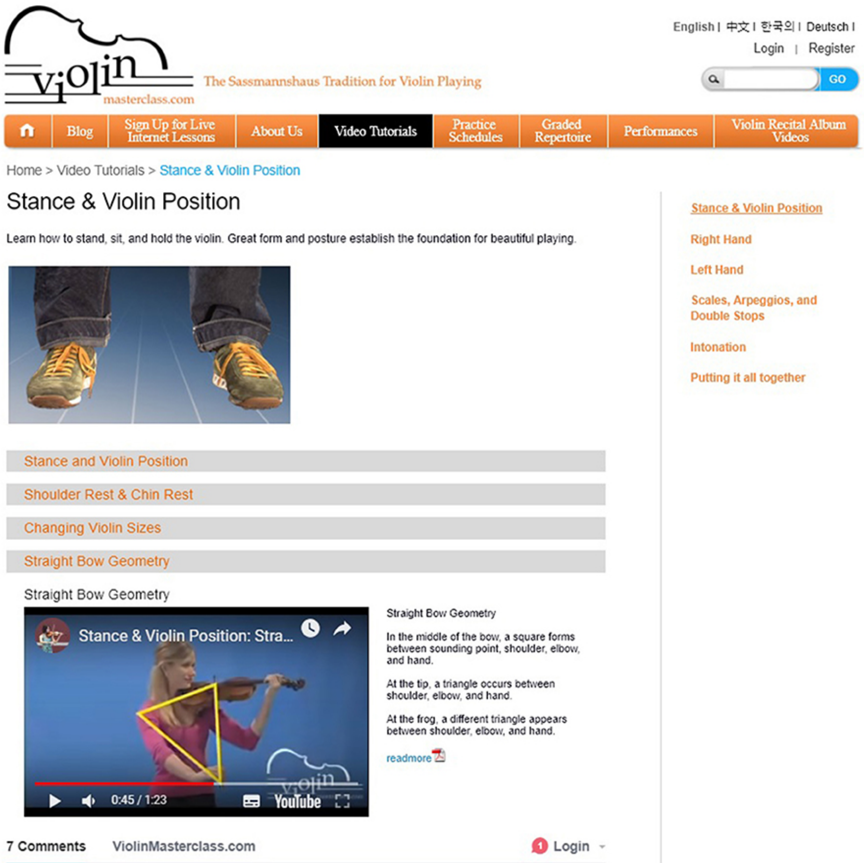

Figure 3. Instructional videos on stance, violin position, straight bow geometry and bow grip were collected from Violinmasterclass.com.

The blocks of trials were named as follows: The Early block (trials from 6 to 10), the Middle block (trials from 11 to 15), and the Late block (trials from 15 to 20). The total duration of the experiment was approximately 45 min. The first block of trials, where both groups of beginners did not have the option to rewatch instructional videos or offline feedback from the SQVFS, was used as a baseline to compute the amount of change in both sound quality and EEG waves around the rest of the blocks. From the early block on, FG and CG had the option to rewatch both instructional and/or reference expert videos as many times as they wanted for the rest of the trials. In addition, the FG group had the opportunity to receive offline feedback from the SQVFS visualizing the dynamic stability, pitch stability, and aperiodicity scores of their performance for each trial. The number of times a participant requested the learning materials were also recorded.

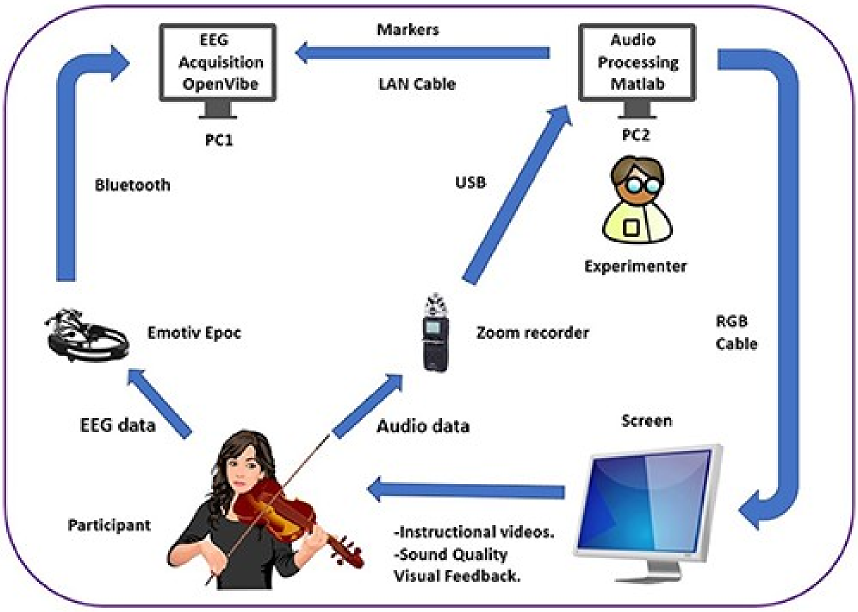

Figure 4. Setup of the experiment. EEG data and audio data from the participant are processed separately in different computers that are communicated through OSC. The experimenter controlled the display of instructional videos (including the reference expert video) for both BNF and BF groups and sound quality visual feedback for the BF group.

3.3.4 Extraction of Audio Features

Violin sounds generated by participants were recorded for each trial with a sampling rate (SR) of 44,100 samples. The Yin algorithm was used to extract sound descriptors from the audio signal of each trial using a windows size of 33 ms and a hop size of 0.7 ms. Three different parameters were computed for each window: instantaneous power, fundamental frequency (f0) in cents (reference: 440) and aperiodicity. The quality of the sound recorded in one trial may be assessed through sound descriptors such as dynamic stability (see 1) or pitch stability (see 2) by computing the standard deviation of both f0 and power throughout the trial (Romaní et al., 2015). Aperiodicity was also included as a descriptor (details about how aperiodic power is computed can be found in De Cheveigne and Kawahara, 2002). See Equations (1–3) for a formal definition of these descriptors.

$$dynamicStability = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Pi-\mu)^2} \quad (1)$$

$$pitchStability = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f0i-\mu)^2} \quad (2)$$

$$aperiodicity = \frac{aperiodicPower}{totalPower} \quad (3)$$

Where N is the number of samples in a trial. Pi is instantaneous power in Db. f0i is the instantaneous fundamental frequency in Hz and μ is the mean value of, respectively, the power (Equation 1) or the fundamental frequency (Equation 2) calculated over the trial. Note that in this definition of the descriptors lower values indicate more stability while higher values indicate less stability.

3.3.5 Sound Quality Analysis

Because we wanted to evaluate the importance of the sound quality descriptors extracted to differentiate between beginners and experts, we performed a 2 × 4 multivariate mixed-design with Group (both groups of beginners and experts) as between-subject factors and Condition (Baseline, Early, Middle, Late) as the within-subject factor, and the results of dynamic stability, pitch stability, and aperiodicity for each condition as dependent variables. Post-hoc tests using the Tukey method for multiple comparisons with Bonferroni correction were performed between the groups of participants. The descriptors that showed significant differences between the experts and both groups of beginners were considered good performance evaluators.

To study the impact of feedback in our beginner's groups we needed to look for possible interactions between both beginner's groups and conditions for those variables previously considered. For that purpose, we performed a 2 × 5 mixed-design with Group (this time only FG and CG) as the between-subject factor and Condition as the within-subject factor. A posterior simple main effect analysis was performed on each group. Pairwise comparison tests were performed between the conditions using the Bonferroni correction.

We removed those values from a condition that were labeled as outliers (values bigger than three interquartile ranges) inside each participant’s group. All the results presented in the following sections were Greenhouse-Geisser corrected.

3.3.6 EEG power computation

The EEG data were visually inspected to reject those periods contaminated from noise and non-stereotyped muscle artifacts. After the general observation that electrodes F3 and F4 were the least affected by noise, we decided to focus our analysis on them.

Data preprocessing was performed offline using EEGlab v2021.0 software (Delorme and Makeig, 2004) running on Matlab R2017a. For each subject and every single trial, the power spectral density (PSD) was computed from activity in each electrode using Welch's overlapped segment averaging estimator using a window size of the length of the trial (EEGlab’s function spectopo.m). Four frequency bands were averaged and extracted corresponding to theta (4–8 Hz), alpha (8–13 Hz), beta (13–24 Hz), and gamma (30–50 Hz).

3.3.7 EEG Analysis

First, to find possible electrophysiological differences between beginners and experts, we performed a 2 x (2 x 4) multivariate mixed-design analysis with Group (both groups of beginners and experts) as between-subject factor and Electrode (F3 and F4) and Block (Baseline, Early, Middle, and Late) as within-subject factors, and theta, alpha, beta and gamma as dependent variables. Post-hoc tests using the Tukey method for multiple comparisons with

Bonferroni correction were performed between the groups of participants. In case of finding significant differences between electrodes, a simple main effect analysis would be performed for each one and those frequency bands that showed differences between conditions and/or groups.

To study the impact of feedback in our beginner's groups we needed to look for possible interactions between both beginner's groups and conditions for those electrodes and bands previously considered. For that purpose, we performed a 2 × 5 mixed-design with Group (this time only FG and CG) as the between-subject factor and Condition as the within-subject factor. A posterior simple main effect analysis was performed on each group. Pairwise comparison tests were performed between the conditions using the Bonferroni correction.

We removed those values from a condition that were labeled as outliers (values bigger than three interquartile ranges) inside each participant's group. All the results presented in the following sections were Greenhouse-Geisser corrected.

3.4 Results

3.4.1 Differences in Sound Quality Between Beginners and Experts

Experts showed lower results of dynamic stability, pitch stability, and aperiodicity than beginners. Post-hoc tests showed significant results between beginners and experts in dynamic stability ($p < .0001$), pitch stability ($p < .0001$), and aperiodicity ($p = .002$). However, beginners showed a decreasing trend of results across blocks that was not seen in the experts (see Figure 5). Univariate tests showed significant effects of Block and a significant Block * Group interaction at the three descriptors (see Table 1). Simple main effect analysis yielded significant effects of Block for the beginners at dynamic stability, $F(1,93) = 15$, $p < .0001$, eta squared = .51, at pitch stability, $F(1.53) = 31$, $p < .0001$, eta squared = .68, and at aperiodicity, $F(1) = 16.48$, $p = .001$, eta squared = .54. Pairwise comparisons showed significant differences for dynamic stability between the Baseline the rest of blocks (Baseline > Early, $p = .01$; Baseline > Middle, $p = .008$; Baseline > Late, $p = .001$), for pitch stability between the Baseline and the rest of blocks (Baseline > Early, $p < .0001$; Baseline > Middle, $p < .0001$; Baseline > Late, $p < .0001$), and for aperiodicity between the Baseline and the rest of blocks (Baseline > Early, $p = .008$; Baseline > Middle, $p = .005$; Baseline > Late, $p = .006$).

3.4.2 Effects of Augmented Feedback in Sound Quality

Both beginners' groups showed a decreasing trend of their results across blocks. However, the FG seemed to maintain the improvement until the Late block while results from the CG stabilized at the Middle block (See Figure 6). Univariate tests yielded significant effects of Block for dynamic stability, $F(1.9) = 14.26$, $p < .0001$, eta squared = .52, for pitch stability, $F(1.47) = 29.33$, $p < .0001$, eta squared = .523, and for aperiodicity $F(1) = 15$, $p < .001$. No significant Block * Group interaction was found. Simple main effect analysis yielded significant effects of Block for the CG at dynamic stability, $F(1.8) = 5.5$, $p = .02$, eta squared = .48, at pitch stability, $F(1.45) = 12.09$, $p = .005$, eta squared = .668, and at aperiodicity $F(1.14) = 6.57$, $p = .036$, eta squared = .52. Pairwise comparisons showed only significant differences for pitch stability between the Baseline and the Early block (Baseline > Early, $p = .045$) and between the Baseline and the Middle block (Baseline > Middle, $p = .036$). Simple main effects of Block were also found for the FG at dynamic stability, $F(1.56) = 9.97$, $p < .005$, eta squared = .58, at pitch stability, $F(1.44) = 18.14$, $p = .001$, eta squared = .56, and at aperiodicity, $F(1) = 9$, $p = .018$, eta squared = .56. Pairwise comparisons showed significant differences for dynamic stability between the Baseline and the Late block (Baseline > Late, $p = .015$) and for pitch stability between the Baseline and the rest of blocks (Baseline > Early, $p = .031$; Baseline > Middle, $p = .01$; Baseline > Late, $p = .017$).

No significant differences in three independent t-tests were found in the relative difference of the Late and the Baseline block between the FG and CG for each sound quality descriptor.

3.4.3 Electrophysiological Differences Between Beginners and Experts

We found an asymmetry of power between the F3 and F4 electrodes in all the frequency bands. The average power measured at each

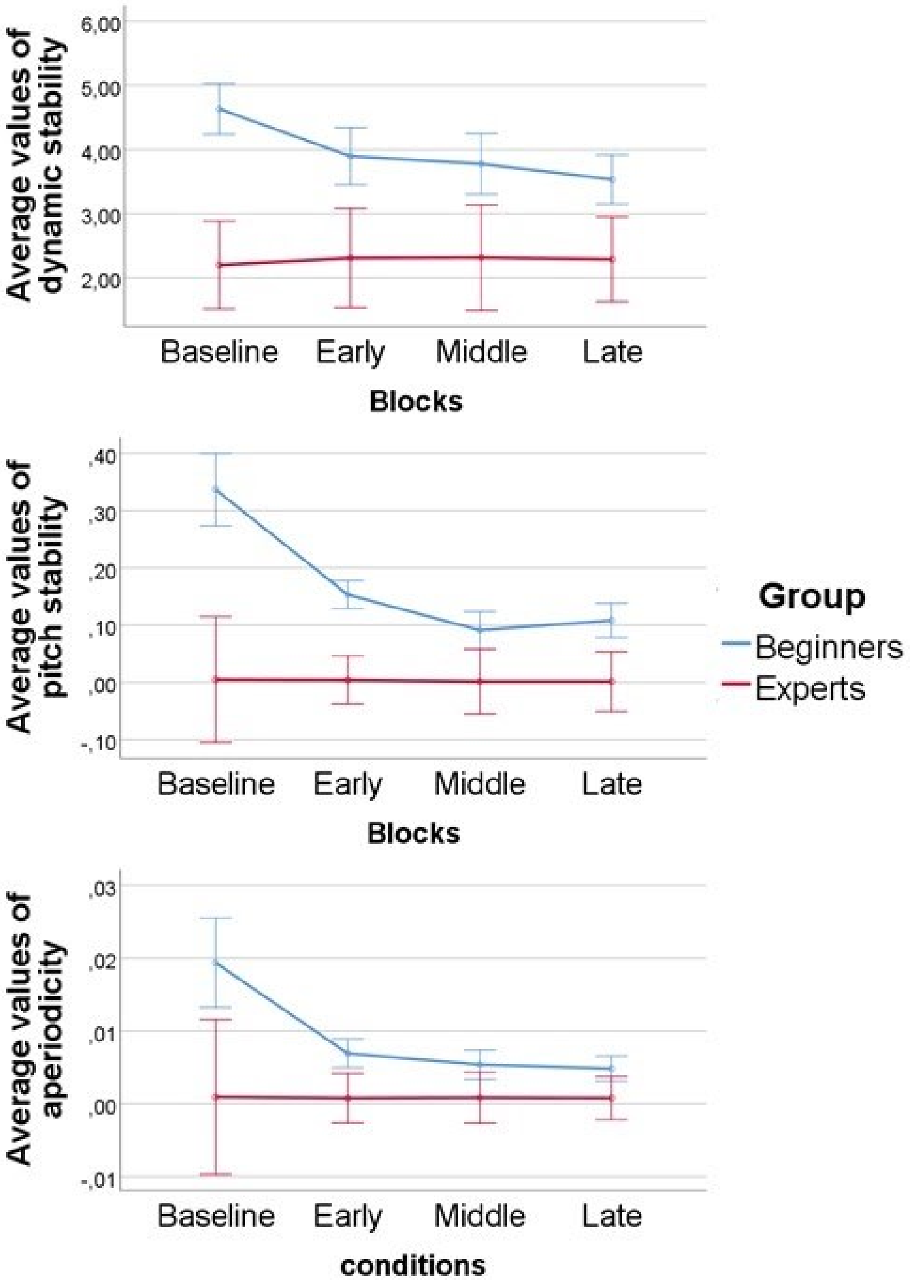

Figure 5: Average values of dynamic stability, pitch stability and aperiodicity across blocks in beginners and experts.

band was much lower at F3 compared with F4 (-1.14 dBs for theta, -1.9 dBs for alpha, -1.84 dBs for beta, and -1.6 dBs for gamma). Multivariate tests of within-subject effects showed a significant effect of Electrode, Pillai's Trace = .754, F(4) = 9.19, p = .001, eta squared = .75. Univariate tests also showed a significant effect of Electrode at all the frequency bands (p < .0001 in all tests). This power asymmetry tended to be higher for the expert group than for the beginner group. These differences were accentuated in alpha and beta (see Figure X). Univariate tests of within-subject effects only showed a significant Electrode * Group interaction at gamma, F(1) = 4.72, *p* = .046, eta squared = .24. Differences did not reach the significant threshold for beta, F(1) = 4.48, p = 0.051, eta squared = .23. Post-hoc tests between groups for the F3 electrode yielded significant differences at alpha (*p* = 0.037) and beta (*p* = 0.018). No significant differences were found at gamma (*p* = 0.056). Post-hoc tests between groups for the F4 electrodes did not yield any significant differences.

Beginners also showed a general trend of power desynchronization through the different blocks of the experiment that was not seen in the experts. This desynchronization tended to be also accentuated in the higher ranges of frequencies such as beta and gamma and in the F3 electrode (see Figure 7). Multivariate analysis of within-subject effects yielded a significant Block * Group interaction, Pillai's Trace = .482, F(12) = 2.1, p = 0.02, eta squared = .161. However, univariate tests did not yield any significant Block * Group interaction for any range of frequency. Simple main effect analysis for each electrode and group yielded only significant effects of Block for beginners at the F3 electrode in the gamma range, F(1.89) = 3.63, p = 0.043, eta squared = .219. Pairwise comparisons showed significant differences between the Baseline and the Middle block (p = 0.031). No significant effects of Block were found in the F4 electrode for beginners. Experts showed significant effects of Block neither in the F3 nor F4 electrode.

Beginners also showed a general trend of power desynchronization through the different blocks of the experiment that was not seen in the experts. This desynchronization tended to be also accentuated in the higher ranges of frequencies such as beta and gamma and in the F3 electrode (see Figure 8). Multivariate analysis of within-subject effects yielded a significant Block * Group interaction, Pillai's Trace = .482, F(12) = 2.1, p = 0.02, eta squared = .161. However, univariate tests did not yield any significant Block * Group interaction for any range of frequency. Simple main effect analysis for each electrode and group yielded only significant effects of Block for beginners at the F3 electrode in the gamma range, F(1.89) = 3.63, p = 0.043, eta squared = .219. Pairwise comparisons showed significant differences between the Baseline and the Middle block (p = 0.031). No significant effects of Block were found in the F4 electrode for beginners. Experts showed significant effects of Block neither in the F3 nor F4 electrode.

Univariate Tests					
Source	Measure	df	F	Sig.	Partial Eta Squared
condition	dynInstab	1,949	3,842	,032	,176
	pitchInstab	1,535	10,218	,001	,362
	aperiodicity	1,098	5,398	,028	,231
condition * Groups	dynInstab	1,949	5,686	,008	,240
	pitchInstab	1,535	9,696	,001	,350
	aperiodicity	1,098	5,201	,031	,224
Error(condition)	dynInstab	35,074			
	pitchInstab	27,631			
	aperiodicity	19,758			

Table 1: Univariate tests of within-subject effects for the 2 × 4 multivariate mixed-design with Group (both groups of beginners and experts) as between-subject factors and Condition (Baseline, Early, Middle, Late) as the within-subject factor, and the results of dynamic stability, pitch stability, and aperiodicity for each condition as dependent variables.

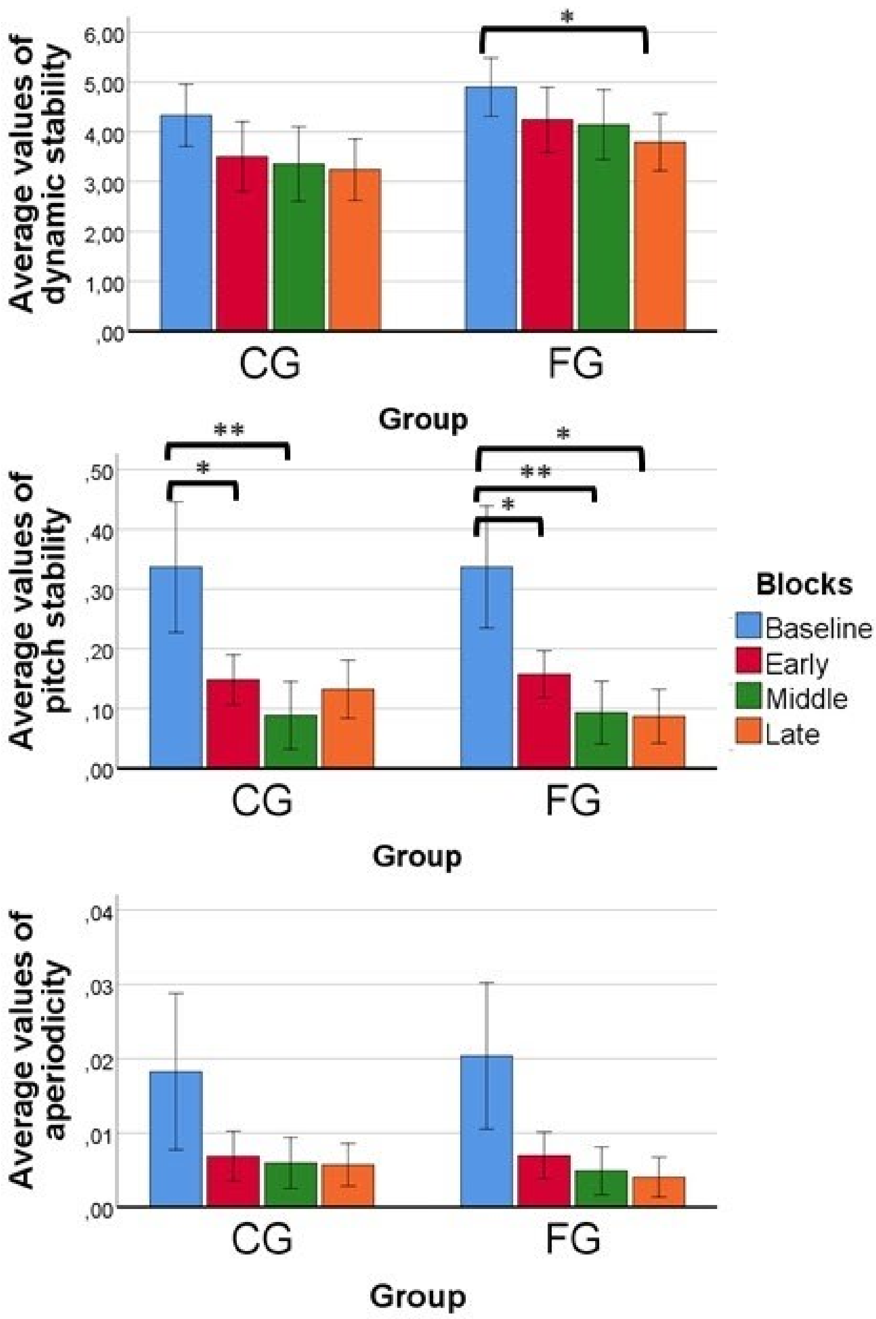

Figure 6: Average values of dynamic stability, pitch stability and aperiodicity across blocks in the FG and CG. $^{*}p \leq 0.05$, $^{**}p \leq 0.01$

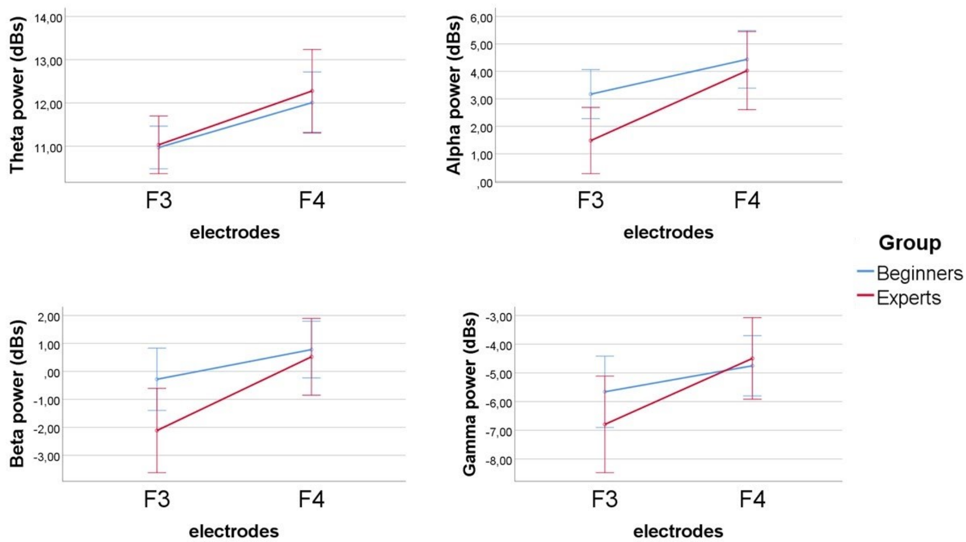

Figure 7: Differences of power between the F3 and F4 electrodes for each frequency range (theta, alpha, beta, and gamma). Both beginners and experts showed a power asymmetry between the F3 and F4. However, experts showed higher desynchronizations at alpha, beta, and gamma.

3.4.4 Effects of Augmented Feedback

Both the FG and CG showed the power asymmetry between F3 and F4 (F3 < F4). Multivariate tests of within-subject effects showed an effect of Electrode, Pillai's Trace = .796, F(4) = 5.84, *p* = .029, eta squared = .8. Univariate tests yielded significant effects for Electrode at all the frequency bands (theta, *p* = .004; alpha, *p* = .006, beta, *p* = .026; gamma, *p* = .022). Both groups of beginners also showed the general trend of power desynchronization through the blocks of the experiment, especially at the gamma range. This effect seemed more accentuated for the FG (see Figure 9). Multivariate tests of within-subject effects yielded an effect of Block, Pillai's Trace = .792, F(12) = 2.3, *p* = .013, eta squared = .3. Univariate tests yielded a significant effect of Block at gamma, F(1.7) = 4, *p* = .04, eta squared = .3. No significant interactions were found. However, simple main effect analysis for the F4 electrode yielded a significant effect of Block at gamma for the FG, F(2.39) = 5.81, *p* = .009, eta squared = .45. Pairwise comparisons showed significant differences between the Baseline and the Early block at gamma (*p* = .021). Simple main effect analysis for the F3 electrode yielded a significant multivariate effect of Block for the FG, Pillai's Trace = .97, F(12) = 2.4, p = .013, eta squared = .32. Pairwise comparisons showed significant differences between the Baseline and the Early block at gamma (p = .021).

3.4.5 Correlations

We performed three Pearson correlations for each electrode and group (beginners, experts) between the three audio descriptors and the levels of gamma power. We found a significant correlation between pitch stability and gamma power at F3 in beginners, r = .28, p = .028, N = 61, and a significant correlation between dynamic stability and gamma power at F3 in experts, r = 632, p = .001, N = 23 (see Figure 10). No significant correlations were found at F4.

3.4.6 Learning Materials

The number of times each participant requested each learning material (instructional videos, reference video, or their score evaluated with audio descriptors) was compared between the two different beginner groups (FG and CG). We found that the FG showed a tendency to request more the reference video than the CG. On average, the BF group requested the reference video 25.8% more times than the BNF group (see Figure 11). Those differences were significant $t(16) = -2.44$, p = .02. No significant differences were found in the number of times they requested the learning materials.

The FG group also had the possibility to request the audio-based automatic evaluation of their performance produced by the system. A paired sampled t-test was performed between the number of times the FG group requested the reference video with the number of times they requested the audio evaluation. No significant differences were found.

3.5 Discussion

In this work, we have used audio features like pitch stability and dynamic stability to measure sound quality as has been done in previous related work (Romaní et al., 2015; Giraldo et al., 2018). We have found that the aperiodicity measure is also a reliable indicator and offers extra information not found in the rest of the descriptors. However, in this work we have not only shown how these descriptors could be useful to discriminate between those sounds performed in the violin by experts and beginners (i.e., a good or bad sound) but, in addition, we have used them to track the amount of learning of 18 participants, with no prior experience neither with the violin nor any bowed-string instrument, during 20 trials while learning to produce a stable and sustained sound in an open string. Allowing us to study objectively the impact of feedback technologies in the process of learning to produce a good sound with the violin.

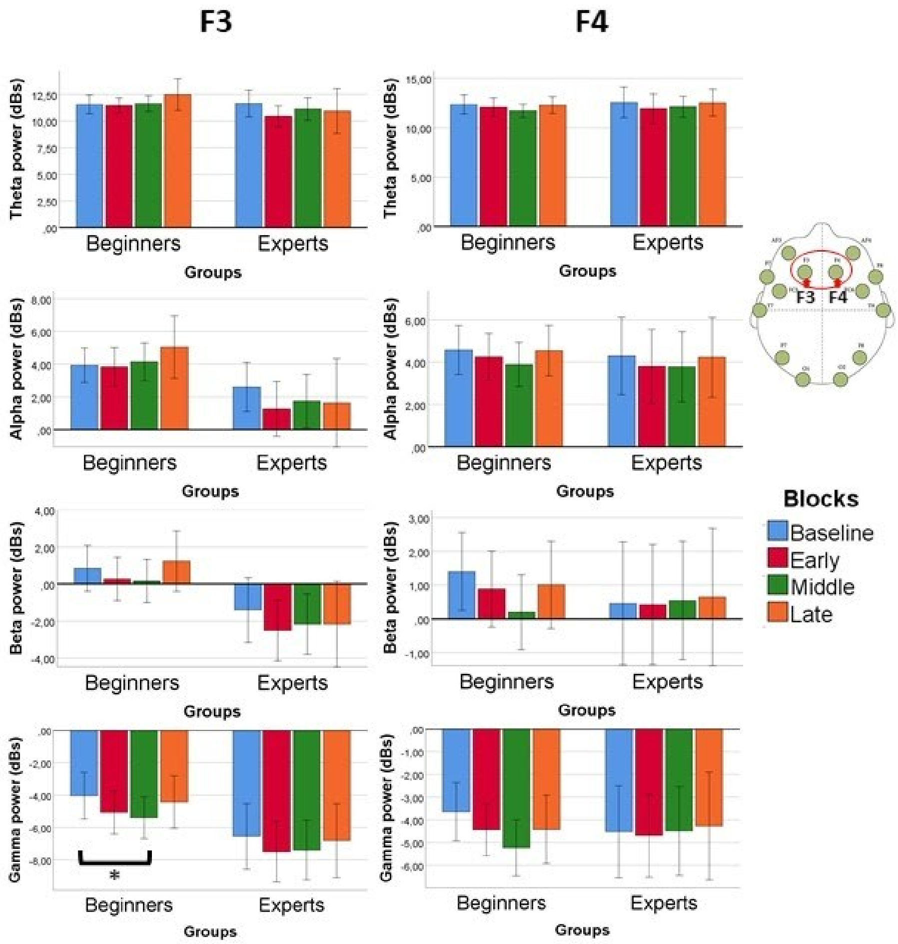

Figure 8. Differences of power between conditions for each electrode and groups. Beginners tended to show significant desynchronizations between blocks not seen in the experts especially at F3. Significant differences were found at F3 between the Baseline and the Middle block at the gamma band for beginners groups.

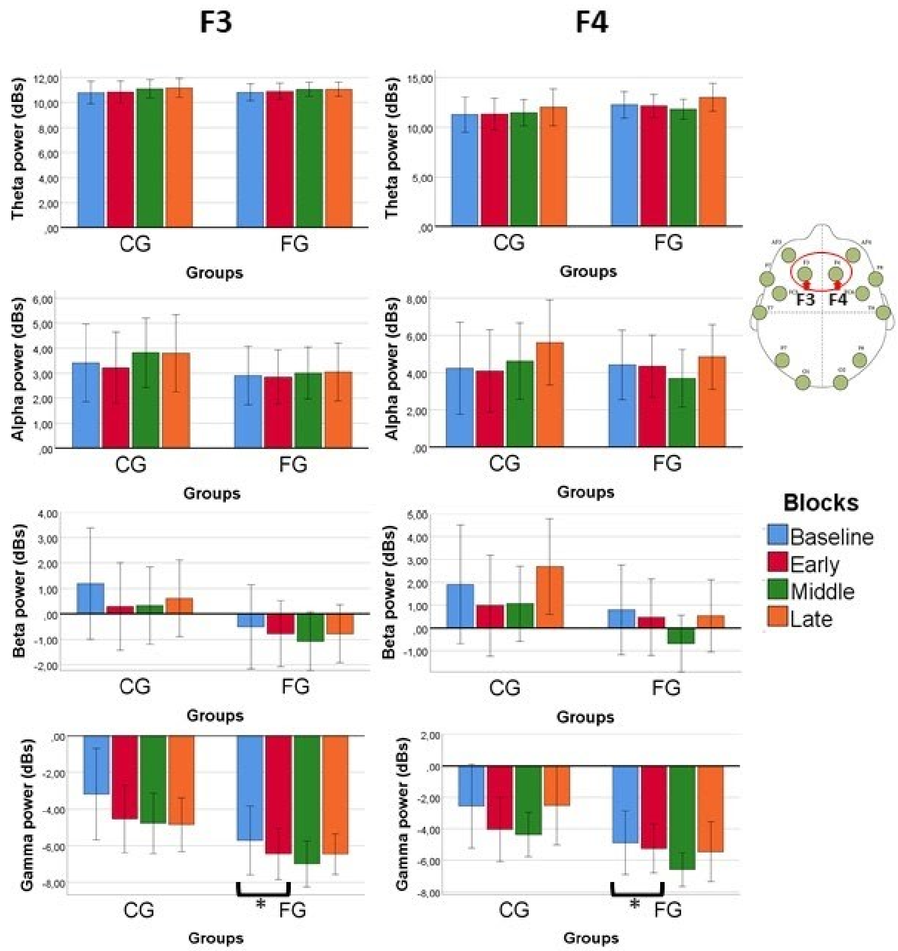

Figure 9: Differences of power between conditions for each electrode and groups of beginners. The FG tended to show significant desynchronizations between blocks not seen for the CG. Significant differences were found at both F3 and F4 between the Baseline and the Early block at the gamma band for the FG.

The visual feedback considered in this study consisted of a sound quality indicator computed using audio descriptors extracted from the audio produced by participants. The feedback was presented offline to participants in the form of a graph where the sound quality of the last trial was shown relative to the previous ones. They could also compare their performance to that of an expert participant who previously did the experiment. We referred to this type of technology as a sound quality visual feedback system (SQVFS).

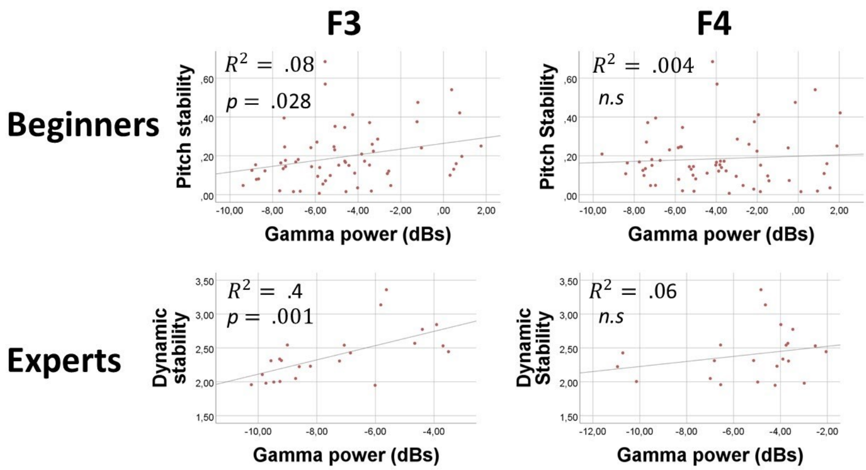

Figure 10: Relationship between power at gamma at each electrode and sound quality descriptors for beginners and experts. Beginners showed a significant correlation between gamma power at the F3 electrode and pitch stability. Experts also showed a significant correlation between gamma power at the F3 electrode and dynamic stability.

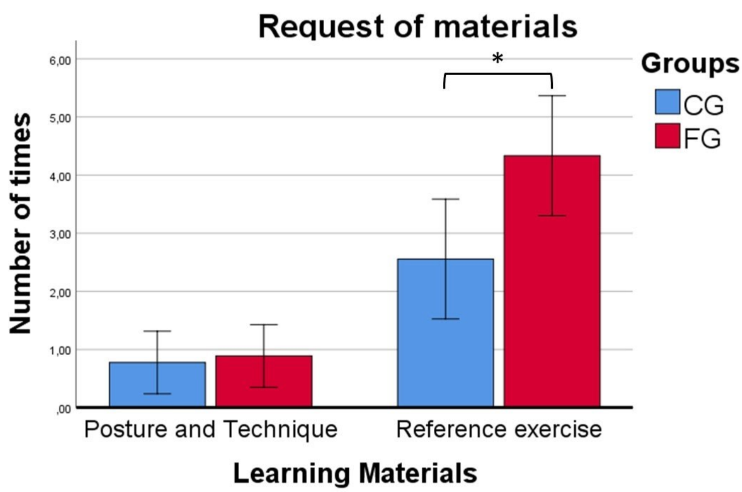

Figure 11: Number of times each group of beginners requested learning materials during the experiment. Participants from the FG tended to request more times the reference exercise video than the CG. No significant differences were found between the amount of time each group requested videos on good posture and technique.

At the end of the session, both groups of participants significantly improved their scores with the exception of the expert group. However, only participants who received feedback from the SQVFS seemed to maintain those improvements in the Late block for pitch stability. They also seemed to improve more in dynamic stability when comparing those results with the baseline. Nonetheless, no significant differences were found at the end of the session regarding the amount of improvement between the Late and the Baseline blocks between beginner groups at each one of the sound quality descriptors.

Regarding the electrophysiological analysis, we found an asymmetry of power between the F3 and F4 electrodes. This asymmetry was bigger for the experts in the alpha, beta, and gamma frequency ranges who showed lower amplitudes in the F3 electrode. This activity may be related to movement-related cortical potentials (MRCP) contralateral to the arm which executes the bow movements (the right arm in this case). Several studies have reported lower amplitudes of MRCP for expert performers compared with novice performers (Di Russo, Pitzalis, et al., 2005, Fattapposta et al., 1996, Hatta et al., 2009, Kita et al., 2001, Wright et al., 2012). MRCP are usually associated with power at the beta range (Tan et al., 2014; Torrecillos et al., 2015), however, in this study, we found similar levels of amplitude difference between alpha and beta.

We also found significant gamma desynchronizations in the F3 electrode through the different blocks of the experiment in beginners that were not found in experts. Those desynchronizations seemed to show some correlation with the amount of improvement in pitch stability, which was the sound descriptor in which participants improved the most throughout the experiment. That is, those trials produced with a greater sound quality were correlated with lower levels of gamma power at the F3 electrode. Similar results were found for the experts but for the dynamic stability descriptor. Those changes in gamma could be related to those found by Gutierrez and Ramírez-Moreno (2016). Gutierrez and Ramírez-Moreno (2016) found desynchronizations at both beta and gamma bands as participants started to learn and achieve proficiency in typewriting in a Colemak keyboard. These results may be interpreted from the temporal binding model which associates gamma band with the role of integrating (binding) information processed in distributed cortical areas (Bhattacharya et al., 2001; Howard et al., 2003; Fitzgibbon et al., 2004). Task complexity demands more cognitive resources, more binding, and thus, gamma-band power enhancement, which may be reduced as the demanded task begins to be automated which could have been the case of both beginners groups.

The FG showed clearer desynchronizations at the gamma band through blocks than the CG and it was found at both F3 and F4 electrodes. This could be related to the apparently better results that the FG obtained in the sound quality results through blocks compared with the CG. However, the sample size was very small and the differences might not be that important. Future research should address this issue with a larger sample size. Also, the limited number of electrodes of the Emotiv Epoc prevents a deeper analysis of the results.
The lack of central electrodes in the Emotiv Epoc makes this device not very appropriate to study this type of motor activity. However, its low cost and easy setup make it a good candidate to be used in educational environments once we can more clearly interpret the results obtained.

Finally, we also found behavioral differences between the FG and the CG. The FG requested more times the reference video of the exercise than the CG. Participants from the FG could clearly see how much their results deviated from an expert performance which could have led them to want to improve themselves and find different ways to continue improving their results. On the other hand, without any kind of feedback or KR, the CG must not have felt any type of motivation to continue improving and they could feel satisfied with the results obtained after reaching a certain threshold.

3.6 Conclusions

In this work, we have studied the effects of an SQVFS in violin beginner students while learning to produce a stable sound using the bow. A group of experts was included in the study as a reference. Experts did not show improvement along with the session, while both groups of beginners did. In particular, only the FG (beginners with SQVF) showed a constant improvement through the blocks of the session while the CG (beginners without SQVF) seemed to stabilize or not maintain their results at the last block. We hypothesize that the SQVF increased the awareness of participants about how far they were from an expert performance, leading them to experiment more with the instrument and getting more involved in the task.

The improvement of participants among blocks seemed to be related to desynchronization in the gamma band in frontal electrodes, especially in the F3. We also found a power asymmetry between the F3 and the F4 electrode in all groups of participants.

Interestingly, experts tended to show less power especially at the alpha and beta range in the F3 electrode which was contralateral to the arm used to control bow movements during the exercise (the right arm).

3.7 References

Antonenko, P., and Niederhauser, D. (2010). The influence of leads on cognitive load and learning in a hypertext environment. *Comput. Hum. Behav.* 26, 140–150. doi: 10.1016/j.chb.2009.10.014

Antonenko, P., Paas, F., Grabner, R., and van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educ. Psychol. Rev.* 22, 425–438. doi: 10.1007/s10648-010-9130-y

Babiloni, C., Del Percio, C., Iacoboni, M., Infarinato, F., Lizio, R., Marzano, N., et al. (2008). Golf putt outcomes are predicted by sensorimotor cerebral EEG rhythms. *J. Physiol.* 586, 131–139. doi: 10.1113/jphysiol.2007.141630

Bazanova, O., and Aftanas, L. (2006). Using individual eeg peculiarities increase neurofeedback efficiency. *Ann. Gen. Psychiatry* 5:S98. doi: 10.1186/1744-859X-5-S1-S98

Bazanova, O., Mernaya, E., and Shtark, M. (2009). Biofeedback in psychomotor training. Electrophysiological basis. *Neurosci. Behav. Physiol.* 39, 437–447. doi: 10.1007/s11055-009-9157-z

Berka, C., Behneman, A., Kintz, N., Johnson, R., and Raphael, G. (2010). Accelerating training using interactive neuro-educational technologies: applications to archery, golf, and rifle marksmanship. *Int. J. Sports Soc* 1, 87–104. doi: 10.18848/2152-7857/CGP/v01i04/54040

Besson, M., Faïta, F., and Requin, J. (1994). Brain waves associated with musical incongruities differ for musicians and non-musicians. *Neurosci. Lett.* 168, 101–105. doi: 10.1016/0304-3940(94)90426-X

Bhattacharya, J., Petsche, H., and Pereda, E. (2001). Long-range synchrony in the γ band: role in music perception. *J. Neurosci.* 21, 6329–6337. doi: 10.1523/JNEUROSCI.21-16-06329.2001

Blanco (2018). *Evaluation of Audio-based Feedback Technologies*. Available online at: https://github.com/adavidBlancoUPF/Evaluation-of-Audio-Based-Feedback-Technologies/

Brandmeyer, A., Timmers, R., Sadakata, M., and Desain, P. (2011). Learning expressive percussion performance under different visual feedback conditions. *Psychol. Res.* 75, 107–121. doi: 10.1007/s00426-010-0291-6

Bundy, O. R. (1987). *Instrumentalists' Perception of Their Performance as Measured by Detection of Pitch and Rhythm Errors Under Live and Recorded Conditions.* Doctoral dissertation, Pennsylvania State University. Dissertation Abstracts International, 48–10A, 2567.

Casares, A. D. B., and Ramírez, R. (2018). *Evaluation of a Sound Quality Visual Feedback for Bow Learning Technique in Violin Beginners: An EEG Study.* Zenodo. Barcelona: Universitat Pompeu Fabra.

Crews, D. J., and Landers, D. M. (1993). Electroencephalographic measures of attentional patterns prior to the golf putt. *Med. Sci. Sports Exerc*. 25, 116–126. doi: 10.1249/00005768-199301000-00016

Dalmazzo, D., Tassani, S., and Ramírez, R. (2018). "A machine learning approach to violin bow technique classification: a comparison between imu and mocap systems," in *Proceedings of the 5th international Workshop on Sensor-Based Activity Recognition and Interaction* (Berlin: ACM), 12. doi: 10.1145/3266157.3266216

De Cheveigne, A., and Kawahara, H. (2002). Yin, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.* 111, 1917–1930. doi: 10.1121/1.1458024

Delorme, A., and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Deutsch, L. M. (2011). *Motion Study of Violin Bow Technique*. Ph.D. thesis, University of california Los Angeles.

Di Russo, F., Pitzalis, S., Aprile, T., & Spinelli, D. (2005). Effect of practice on brain activity: an investigation in top-level rifle shooters. *Medicine and science in sports and exercise*, *37*(9), 1586.

Elwell, J. L., and Grindley, G. C. (1938). The effect of knowledge of results on learning and performance: I. A co-ordinated movement of the two hands. *Brit. J. Psychol.* 29, 39–54. doi: 10.1111/j.2044-8295.1938.tb00899.x

Erbil, N., and Ungan, P. (2007). Changes in the alpha and beta amplitudes of the central EEG during the onset, continuation, and offset of long-duration repetitive hand movements. *Brain Res.* 1169, 44–56. doi: 10.1016/j.brainres.2007.07.014

Fattapposta, F., Amabile, G., Cordischi, M. V., Di Venanzio, D., Foti, A., Pierelli, F., ... & Morrocutti, C. (1996). Long-term practice effects on a new skilled motor learning: an electrophysiological study. *Electroencephalography and clinical Neurophysiology*, *99*(6), 495-507.

Fitzgibbon, S., Pope, K., Mackenzie, L., Clark, C., and Willoughby, J. (2004). Cognitive tasks augment gamma EEG power. *Clin. Neurophysiol.* 115, 1802–1809. doi: 10.1016/j.clinph.2004.03.009

Gentili, R., Bradberry, T., Hatfield, B., and Contreras-Vidal, J. (2008). “A new generation of non-invasive biomarkers of cognitive-motor states with application to smart brain computer interfaces,” in *Proceedings of the 16th European Signal Processing Conference - 2008* (Lausanne).

Gevins, A., and Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theor. Issues Ergon. Sci.* 4, 113–131. doi: 10.1080/14639220210159717

Giraldo, S., Ramirez, R., Waddell, G., and Williamon, A. (2018). "Computational modelling of timbre dimensions for automatic violin tone quality assessment," in *Proceedings of the Timbre is a Many-Splendored Thing Conference* (Montreal, QC).

Gruber, T., and Müller, M. (2005). Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. *Cereb. Cortex* 15, 109–116. doi: 10.1093/cercor/bhh113

Gruzelier, J. (2009). A theory of alpha/theta neurofeedback, creative performance enhancement, long distance functional connectivity and psychological integration. *Cogn. Process.* 10, 101–109. doi: 10.1007/s10339-008-0248-5

Gurtubay, I., Alegre, M., Valencia, M., and Artieda, J. (2006). Cortical gamma activity during auditory tone omission provides evidence for the involvement of oscillatory activity in top-down processing. *Exp. Brain Res.* 175, 463–470. doi: 10.1007/s00221-006-0561-0

Gutierrez, D., and Ramírez-Moreno, M. (2016). Assessing a learning process with functional anova estimators of EEG power spectral densities. *Cogn. Neurodyn.* 10, 175–183. doi: 10.1007/s11571-015-9368-7

Hamond, L. F. (2017). *The Pedagogical Use of Technology-Mediated Feedback in a Higher Education Piano Studio: An Exploratory Action Case Study*. Ph.D. thesis, University College London.

Hatta, A., Nishihira, Y., Higashiura, T., Kim, S. R., & Kaneda, T. (2009). Long-term motor practice induces practice-dependent modulation of movement-related cortical potentials (MRCP) preceding a self-paced non-dominant handgrip movement in kendo players. *Neuroscience Letters*, *459*(3), 105-108.

Haufler, A. J., Spalding, T. W., Santa Maria, D., and Hatfield, B. D. (2000). Neuro-cognitive activity during a self-paced visuospatial task: comparative eeg profiles in marksmen and novice shooters. *Biol. Psychol.* 53, 131–160. doi: 10.1016/S0301-0511(00)00047-8

Hewitt, M. P. (2001). The effects of modeling, self-evaluation, and self-listening on junior high instrumentalists' music performance and practice attitude. *J. Res. Music Educ.* 49, 307–322. doi: 10.2307/3345614

Howard, M. W., Rizzuto, D. S., Caplan, J. B., Madsen, J. R., Lisman, J., Aschenbrenner-Scheibe, R., et al. (2003). Gamma oscillations correlate with working memory load in humans. *Cereb. Cortex* 13, 1369–1374. doi: 10.1093/cercor/bhg084

Hutchins, S. M., and Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. *J. Exp. Psychol.* 141:76. doi: 10.1037/a0025064

Iglewicz, B., and Hoaglin, D. C. (1993). *How to Detect and Handle Outliers, asqc Basic References in Quality control*. Milwaukee, WI: American Society for Quality Control.

Jaušovec, N., and Jaušovec, K. (2004). Differences in induced brain activity during the performance of learning and working-memory tasks related to intelligence. *Brain Cogn.* 54, 65–74. doi: 10.1016/S0278-2626(03)00263-X

Johnson, R. M. G. (2014). *In Touch With the Wild: Exploring Real-time Feedback for Learning to Play the Violin*. PhD thesis, University College London.

Johnson, R., Rogers, Y., van der Linden, J., and Bianchi-Berthouze, N. (2012). "Being in the thick of in-the-wild studies: the challenges and insights of researcher participation," in *CHI '12 CHI Conference on Human Factors in Computing Systems* (ACM), 1135–1144. doi: 10.1145/2207676.2208561

Kepner, C. B. (1986). *The Effect of Performance Familiarity, Listening Condition, and Type of Performance Error on Correctness of Performance Error Detection by 50 High School Instrumentalists as Explained Through a Sensory Blocking*

Theory. Doctoral dissertation, Kent State University, Kent, OH. Dissertation Abstracts International, 47–05A, 1643.

Kerick, S. E., Douglass, L. W., and Hatfield, B. D. (2004). Cerebral cortical adaptations associated with visuomotor practice. *Med. Sci. Sports Exerc*. 36, 118–129. doi: 10.1249/01.MSS.0000106176.31784.D4

Kita, Y., Mori, A., & Nara, M. (2001). Two types of movement-related cortical potentials preceding wrist extension in humans. *Neuroreport*, *12*(10), 2221-2225.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* 29, 169–195. doi: 10.1016/S0165-0173(98)00056-3

Klimesch, W., Doppelmayr, M., Pachinger, T., and Ripper, B. (1997). Brain oscillations and human memory: EEG correlates in the upper alpha and theta band. *Neurosci. Lett.* 238, 9–12. doi: 10.1016/S0304-3940(97)00771-4

Koelsch, S., Schröger, E., and Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport* 10, 1309–1313. doi: 10.1097/00001756-199904260-00029

Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Müller, K., Curio, G., et al. (2007). *Improving Human Performance in a Real Operating Environment Through Real-Time Mental Workload Detection*. Cambridge, MA: MIT Press.

Konczak, J., vander Velden, H., and Jaeger, L. (2009). Learning to play the violin: motor control by freezing, not freeing degrees of freedom. *J. Mot. Behav.* 41, 243–252. doi: 10.3200/JMBR.41.3.243-252

Korg (2018). *Cortosia*. Available online at: https://www.korg.com/us/products/software/cortosia/ (Accessed December 8, 2018).

Landers, D. M., Han, M., Salazar, W., and Petruzzello, S. J. (1994). Effects of learning on electroencephalographic and electrocardiographic patterns in novice archers. *Int. J. Sport Psychol*. 25, 313–330.

Leong, S., and Cheng, L. (2014). Effects of real-time visual feedback on pre-service teachers' singing. *J. Comput. Assist. Learn.* 30, 285–296. doi: 10.1111/jcal.12046

Llimona, Q. (2015). *Yin Pitch Estimation Toolbox*. Available online at: https://github.com/lemonzi/matlab/tree/master/yin

Locke, E. A. (1966). *The Motivational Effects of Knowledge of Results: the Influence of Goal-setting.* Technical report, American Institutes for Research Silver Spring md.

Locke, E. A., and Bryan, J. F. (1966). Cognitive aspects of psychomotor performance: the effects of performance goals on level of performance. *J. Appl. Psychol.* 50:286. doi: 10.1037/h0023550

MATLAB (2010). *version 7.10.0 (R2010a)*. Natick, MA: The MathWorks Inc.

Maurer, U., Brem, S., Liechti, M., Maurizio, S., Michels, L., and Brandeis, D. (2015). Frontal midline theta reflects individual task performance in a working memory task. *Brain Topogr.* 28, 127–134. doi: 10.1007/s10548-014-0361-y

Middlestadt, S. E., and Fishbein, M. (1989). The prevalence of severe musculoskeletal problems among male and female symphony orchestra string players. *Med. Probl. Perform. Art.* 4, 41–48.

Miller, S. (2001). *Literature Review: Workload Measures*. Iowa City, IA: The University of Iowa.

Neubauer, A. C., Fink, A., and Grabner, R. H. (2006). Sensitivity of alpha band erd to individual differences in cognition. *Prog. Brain Res.* 159, 167–178. doi: 10.1016/S0079-6123(06)59011-9

Newell, K. M. (1974). Knowledge of results and motor learning. *J. Motor Behav.* 6, 235–244.

Ng, K., and Nesi, P. (2008). “I-maestro: technology-enhanced learning and teaching for music,” in *NIME* (Genova), 225–228.

Niemann, J., Winker, T., Gerling, J., Landwehrmeyer, B., and Jung, R. (1991). Changes of slow cortical negative DC-potentials during the acquisition of a complex finger motor task. *Exp. Brain Res.* 85, 417–422. doi: 10.1007/BF00229418

Ortega, F. J. M., Giraldo, S. I., and Ramirez, R. (2017). “Phrase-level modeling of expression in violin performances,” in *Proceedings of the 10th International Workshop on Machine Learning and Music, MML 2017* (Barcelona), 49–54.

Paas, F., Renkl, A., and Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educ. Psychol.* 38, 1–4. doi: 10.1207/S15326985EP3801_1

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., and Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport* 12, 169–174. doi: 10.1097/00001756-200101220-00041

Pardue, L. S., Harte, C., and McPherson, A. P. (2015). A low-cost real-time tracking system for violin. *J. New Music Res.* 44, 305–323. doi: 10.1080/09298215.2015.1087575

Perez Carrillo, A., and Wanderley, M. M. (2012). “Learning and extraction of violin instrumental controls from audio signal,” in *MM '12 ACM Multimedia Conference Nara* (Japan: ACM), 25–30.

Raghavachari, S., Lisman, J. E., Tully, M., Madsen, J. R., Bromfield, E., and Kahana, M. J. (2006). Theta oscillations in human cortex during a working-memory task: evidence for local generators. *J. Neurophysiol.* 95, 1630–1638. doi: 10.1152/jn.00409.2005

Raymond, J., Sajid, I., Parkinson, L. A., and Gruzelier, J. H. (2005). Biofeedback and dance performance: a preliminary investigation. *Appl. Psychophysiol. Biofeedback* 30, 65–73. doi: 10.1007/s10484-005-2175-x

Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., and Delannoy, V. (2010). “OpenViBE: an open-source software platform to design, test, and use brain-computer interfaces in real and virtual environments,” in *Presence* (MIT Press), 33–53. Available online at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6797525&isnumber=6797172

Romaní, O., Parra, H., Dabiri, D, TOkuda, H., Hariya, W., Oishi, K., et al. (2015). “A real-time system for measuring sound goodness in instrumental sounds,” in *138th Audio Engineering Society Convention* (Warsaw).

Salazar, W., Landers, D. M., Petruzzello, S. J., Han, M., Crews, D. J., and Kubitz, K. A. (1990). Hemispheric asymmetry, cardiac response, and performance in elite archers. *Res. Q. Exerc. Sport* 61, 351–359. doi: 10.1080/02701367.1990.10607499

Salmoni, A. W., Schmidt, R. A., and Walter, C. B. (1984). Knowledge of results and motor learning: a review and critical reappraisal. *Psychol. Bull.* 95:355. doi: 10.1037/0033-2909.95.3.355

Sassmannshaus, K. (2018). *Violin Master Class*. Available online at: http://www.violinmasterclass.com/en/masterclasses/stance-violin-position (Accessed March 23, 2018).

Saunders, T. C., and Holahan, J. M. (1997). Criteria-specific rating scales in the evaluation of high school instrumental performance. *J. Res. Music Educ.* 45, 259–272. doi: 10.2307/3345585

Schlegel, A. L., and Gregory Springer, D. (2018). Effects of accurate and inaccurate visual feedback on the tuning accuracy of high school and college trombonists. *Int. J. Music Educ.* 36, 394–406. doi: 10.1177/0255761418763914

Schmidt, R. A., Young, D. E., Swinnen, S., and Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition: support for the guidance hypothesis. *J. Exp. Psychol. Learn. Mem. Cogn.* 15:352. doi: 10.1037/0278-7393.15.2.352

Schoonderwaldt, E., and Demoucron, M. (2009). Extraction of bowing parameters from violin performance combining motion capture and sensors. *J. Acoust. Soc. Am.* 126, 2695–2708. doi: 10.1121/1.3227640

Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J., and Miller, L. M. (2008). Music training leads to the development of timbre-specific gamma band activity. *Neuroimage* 41, 113–122. doi: 10.1016/j.neuroimage.2008.01.067

Silvana, M.-S., Nada, P.-J., and Dejan, G. (2008). Simultaneous EEG and EMG biofeedback for peak performance in musicians. *Prilozi* 29, 239–52.

Sokolov, A., Pavlova, M., Lutzenberger, W., and Birbaumer, N. (2004). Reciprocal modulation of neuromagnetic induced gamma activity by attention in the human visual and auditory cortex. *Neuroimage* 22, 521–529. doi: 10.1016/j.neuroimage.2004.01.045

Tan, H., Jenkinson, N., & Brown, P. (2014). Dynamic neural correlates of motor error monitoring and adaptation during trial-to-trial learning. *Journal of Neuroscience*, *34*(16), 5678-5688.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., and Schröger, E. (2005). Pitch discrimination accuracy in musicians vs. nonmusicians: an event-related potential and behavioral study. *Exp. Brain Res.* 161, 1–10. doi: 10.1007/s00221-004-2044-5

Thorpe, C. W. (2002). "Visual feedback of acoustic voice features in voice training," in *Presentation was held at the Australian International Conference on Speech Science and Technology*, 12.

Torrecillos, F., Alayrangues, J., Kilavik, B. E., & Malfait, N. (2015). Distinct modulations in sensorimotor postmovement and foreperiod β-band activities related to error salience processing and sensorimotor adaptation. *Journal of Neuroscience*, *35*(37), 12753-12765.

Trainor, L. J., Shahin, A. J., and Roberts, L. E. (2009). Understanding the benefits of musical training. *Ann. N. Y. Acad. Sci.* 1169, 133–142. doi: 10.1111/j.1749-6632.2009.04589.x

Trowbridge, M. H., and Cason, H. (1932). An experimental study of Thorndike's theory of learning. *J. Gen. Psychol.* 7, 245–260. doi: 10.1080/00221309.1932.9918465

Van Der Linden, J., Schoonderwaldt, E., Bird, J., and Johnson, R. (2011). Musicjacket—combining motion capture and vibrotactile feedback to teach violin bowing. *IEEE Trans. Instrum. Meas.* 60, 104–113. doi: 10.1109/TIM.2010.2065770

Wang, J.-H., Wang, S.-A., Chen, W.-C., Chang, K.-N., and Chen, H.-Y. (2012). "Real-time pitch training system for violin learners," in *Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on* (Melbourne, VIC: IEEE), 163–168.

Welch, G., Himonides, R., Howard, D., and Brereton, J. (2004). "Voxed: technology as a meaningful teaching aid in the singing studio," in *Proceedings of the Conference on Interdisciplinary Musicology*.

Welch, G., Howard, D., and Rush, C. (1989). Real-time visual feedback in the development of vocal pitch accuracy in singing. *Psychol. Music* 17, 146–157. doi: 10.1177/0305735689172005

Wilson, P. H., Lee, K., Callaghan, J., and Thorpe, C. W. (2008). Learning to sing in tune: Does real-time visual feedback help? *J. interdiscipl. Music Stud*. 2.

Winstein, C. J. (1991). Knowledge of results and motor learning—implications for physical therapy. *Phys. Ther.* 71, 140–149.

Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques, 4th Edn*. Cambridge, MA: Morgan Kaufmann.

Wright, D. J., Holmes, P. S., Di Russo, F., Loporto, M., & Smith, D. (2012). Differences in cortical activity related to motor planning between experienced guitarists and non-musicians during guitar playing. *Human movement science*, *31*(3), 567-577.

Zacharias, V., Alfonso, P. C., and Rafael, R (2018). Acquisition of violin instrumental gestures using an infrared depth camera. *Zenodo* doi: 10.5281/zenodo.1422548

Zhukov, K. (2015).Zhukov, K. (2015). Exploring the role of technology in instrumental skill development of Australian higher education music students. Austral. *J. Music Educ.* 66–77.

Blanco, A. D., Tassani, S., & Ramirez, R. (2021). Real-Time Sound and Motion Feedback for Violin Bow Technique Learning: A Controlled, Randomized Trial. *Frontiers in Psychology, 12,* 1268.

4. REAL-TIME SOUND AND MOTION FEEDBACK FOR VIOLIN BOW TECHNIQUE LEARNING: A CONTROLLED, RANDOMIZED TRIAL.

4.1 Abstract

The production of good sound generation in the violin is a complex task that requires coordination and spatiotemporal control of bowing gestures. The use of motion-capture technologies to improve performance or reduce injury risks in the area of kinesiology is becoming widespread. The combination of motion accuracy and sound quality feedback has the potential of becoming an important aid in violin learning. In this study, we evaluate motion-capture and sound-quality analysis technologies developed inside the context of the TELMI, a technology-enhanced music learning project. We analyzed the sound and bow motion of 50 participants with no prior violin experience while learning to produce a stable sound in the violin. Participants were divided into two groups: the experimental group ($N = 24$) received real-time visual feedback both on kinematics and sound quality, while participants in the control group ($N = 26$) practiced without any type of external help. An additional third group of violin experts performed the same task for comparative purposes ($N = 15$). After the practice session, all groups were evaluated in a transfer phase without feedback. At the practice phase, the experimental group improved their bowing kinematics in comparison to the control group, but this was at the expense of impairing the sound quality of their performance. At the retention phase, the experimental group showed better results in sound quality, especially concerning control of sound dynamics. Besides, we found that the expert group improved the stability of their sound while using the technology. All in all, these results emphasize the importance of feedback technologies in learning complex tasks, such as musical instrument learning.

4.2 Introduction

Audio-based and motion capture technologies could provide us with objective measures of student improvement in musical instrument performance. This could allow music teachers to monitor their students' learning process to provide better and personalized learning strategies. This is even more important when we take into account that traditional teaching methods of musical performance movement may not be based on the understanding of its biomechanics components but on the subjective and vague perception of human movement (Brandfonbrener, 2004). Moreover, learning to play an instrument is based on a master-apprentice relationship which consists of weekly lessons, followed by long periods of self-study. According to Welch (1985), this could dissociate the teacher's feedback from student's online proprioceptive and auditory sensations that follow his/her performance.

Using the violin as a case of study, the TELMI Project (Technology Enhanced Learning of Musical Instrument Performance)[1] had the general objectives to design and implement new technologies for music learning and training (based on multi-modal-feedback-technologies, such as audio, image, video, and motion), together with the evaluation of their pedagogical effectiveness. Together with other bowed-string instruments, the violin requires special considerations compared with other instruments. The process of good sound generation in the violin is a complex task that requires coordination and spatiotemporal control of bowing gestures (Schoonderwaldt and Demoucron, 2009). More than 700 practice hours are needed to achieve bowing skills comparable to those of experts according to Konczak and Jaeger (Konczak and Jaeger, 2009). Because pitch control in the violin is continuous, playing with correct intonation becomes a central issue of violin performance (just as it happens with the voice). And, finally, string players also have the highest risks of playing-related musculoskeletal disorders with the shoulder and the neck being the main body parts affected (Fishbein et al., 1988).

In this work, we aim to evaluate some of the technologies developed inside the context of the TELMI project. In particular, we study, through a controlled and randomized experimental design, the effects of real-time augmented feedback in learning bow control within a group of 57 participants with no prior experience playing the violin or any other bow-string instrument. The technologies

[1] telmi.upf.edu.

evaluated in this experiment are capable of offering augmented feedback on bow kinematics (Vamvakousis et al., 2018) as well as on sound quality (Giraldo et al., 2018). The features that are usually considered as important for beginners to take into account when learning when learning to control the bow are related to the kinematics of the up and down movements, as well as with the force and speed exerted on the strings. All those aspects can disrupt the quality of the desired sound coming from the instrument, being the reason why the presence of this type of feedback could also be of great benefit.

4.3 Background

4.3.1 Good Posture and Violin Technique

Some initial tools based on gesture analysis can be found in the i-Maestro project (Ng and Nesi, 2008). Since then, different techniques have been used to study posture and bowing techniques for the violin. For example, **Scho**onderwaldt and Demoucron (2009) extracted bowing parameters from violin expert performance by combining optical motion capture with sensors (see also Schoonderwaldt and Wanderley, 2007**;** Deutsch, 2011). Low-cost methods have also been investigated to track violin performance gestures. For example, through indirect-acquisition-techniques using audio information (Perez Carrillo and Wanderley, 2012), by using resistive fingerboard and optical reflectance sensors placed on the bow stick (Pardue et al., 2015), or, more recently, by the use of an infrared depth camera (Vamvakousis et al., 2018).

However, little has been done to explore the educational potential of these technologies yet. As Visentin et al. (2008) remarked, the similarities of violin performance with other already tested paradigms in the area of kinesiology are important (Hay, 1993). This means that some of the methodologies which are successful in those areas (including the use of tracking systems to evaluate the effects of training) may have the potential to be used to maximize performance or reduce the risk of injury in violin performance. For that purpose, the finding of common patterns of expert performance employing tracking technologies is an essential part of assessing the learning progress in novice players. Recent studies have been done in that direction (Peiper et al., 2003**;** Visentin et al., 2008**;** Konczak and Jaeger, 2009; Verrel et al., 2013; Dalmazzo and Ramírez, 2019; Volta and Volpe, 2019).

One of the first skills a novice violin student has to learn is “straight bowing.” A common mistake by beginner-violin-students is not to keep the bow parallel to the bridge and perpendicular to the strings. “Round bowing,” as it is called, is said to obstruct the quality of the sound as it makes it difficult to control the contact point between the bow and the string. This contact determines the distance between the bow and the bridge, which directly affects sound production. Van Der Linden et al. (2011) presented and evaluated a system specifically designed for that purpose called MusicJacket. MusicJacket is a wearable system that tracks a player's bowing action and provides vibrotactile feedback whenever the player deviates from a target trajectory. After six training sessions, the authors found a general improvement trend in the test group throughout sessions, although no significant results were found in comparison with the control group at the retention test where the technology was absent. The employed sample of participants and the difficulty of the task was probably an important limitation (four per group). Another important limitation is that “straight bowing,” despite being an essential factor for obtaining a good sound, can hardly be considered by itself an indicator of sound improvement on the violin. Taking into account that both sound and gesture are important features to be considered together, new efforts are being made in the direction of finding audio features to characterize sound quality.

4.3.2 Sound Quality Detection

Probably, some of the first attempts to identify descriptors that could be correlated with the quality of the sound can be found in the work of Romaní et al. (2015). Romaní et al. correlated the subjective opinions about sound quality of professional musicians, after listening to single notes recordings of their own instrument, with audio features extracted from the recordings. Those features were extracted using Essentia (Bogdanov et al., 2013). Based on their work, an educational app called Cortosia (Korg, 2018) was implemented to offer visual feedback to music students about the quality of their produced sound. Posteriorly, Giraldo et al. (2018) implemented a real-time feedback system of sound quality by using machine learning models based on different tone examples recorded by a professional violinist.

By using some of the previous audio descriptors, such as *dynamic stability* or *pitch stability*, in a previous study we implemented an offline sound quality visual feedback system (SQVFS) (Blanco and Ramirez, 2019). It was evaluated in an experiment with both expert violinists and non-violinists. The use of an expert group allowed us to posteriorly replicate the validity of those descriptors to differentiate between beginners and experts. The descriptors also demonstrated their value as a reference for tracking the improvement of the participants throughout the session. Furthermore, receiving feedback from the SQVFS allowed the test group to stay engaged and improve their scores at the end of the session compared with the control group who stabilized results after the first block of trials.

4.3.3 Recent Views on Motor Learning

Motivational and social factors are known to influence learning but also motor learning in general (Locke, 1966; Wulf and Lewthwaite, 2016). Regarding music learning, Demorest and Pfordresher (2015) stated that it can be difficult for music students to develop their singing abilities if singing was viewed as a fixed characteristic (like a "talent") rather than a temporary condition that could be improved. Even more, it is well-known from a large list of studies in motor learning (and learning in general) that making efforts in changing this kind of conceptions of ability (as a fixed capacity vs. being amenable to change with practice) can enhance motor learning (Dweck and Leggett, 1988; Jourden et al., 1991; Mangels et al., 2006; Blackwell et al., 2007; Wulf and Lewthwaite, 2009). According to Wulf and Lewthwaite (2016), this is possible due to the enhancement of expectancies which can influence working memory, long-term memory, and attentional capture (Zanto et al., 2010; Shomstein and Johnson, 2013; Jiao et al., 2015).

4.4 Aims of the Study

In this study, we aimed to evaluate in an experimental setup different modalities of SkyNote, a novel tool designed to offer feedback in real-time to violin players. We designed an experiment with both professional violinists and beginners with little or no musical experience to evaluate both the effects of real-time visual motion capture feedback on “straight bowing” (as Van Der Linden et al., 2011) combined with the effects of real-time sound quality feedback. We expected that the evaluation of both indicators would offer us a wide picture of the effects and the impact these technologies can have on learning. Participant's skills were first evaluated in a *Baseline* condition which was followed by an *Acquisition* condition where one group of participants received real-time feedback from SkyNote while a control group just received oral instructions. Finally, participants took part in a *Transfer* condition to study the retention effects.

In general terms, in this study we seek to answer the following questions:

1. Does real-time visual feedback improve the bowing technique and sound stability in violin beginner students?

2. Is this improvement retained after removing the real-time feedback?

We decided to include an expert group in the analysis. If some of the computed descriptors allow us to differentiate between beginners and experts we will consider them potential descriptors of violin performance. What is more, if throughout the session the beginner's results of those potential descriptors resemble those of an expert, we will consider that the participant has *improved* his/her results in those specific variables. As already shown in previous research (Romaní et al., 2015; Blanco and Ramirez, 2019), we expected that variables, such as *dynamic stability* or *pitch stability* would be potential descriptors of the quality of the generated sound. We also expected descriptors, such as *bow skewness* (i.e., how straight is the bow during the performance) could be a potential descriptor of violin performance as has already been used in previous studies (Van Der Linden et al., 2011).

We decided to deliver in different conditions the feedback related to sound quality from the feedback related to bow kinematics. That is, participants from the feedback group took part in two different conditions, each one biasing the focus of their attention on a particular modality by offering sound feedback or motion feedback. Participants from the control group also participated in two different conditions, but instead of receiving feedback, they were explicitly asked to focus their attention on a particular modality when performing the required exercise. Previous studies which evaluated the effects of real-time feedback have shown that although a pattern of worsening results appeared at the moment of receiving feedback, it was compensated with higher improvements at the Transfer conditions (Welch et al., 1989; Wilson et al., 2005; Paney and Tharp, 2019). The reasons attributed to these events are usually related to an increase in cognitive load at the time of receiving the feedback. We expected to find a similar trend with our participants.

We asked participants at the end of the experiment to fill a questionnaire with questions regarding their satisfaction with the technology together with which were the most common problems they faced when using it.

4.5 Materials and Methods

4.5.1. Participants

Fifty-seven participants with no prior violin playing experience were recruited from the university campus to participate in an experiment in which they were told they would receive a free violin lesson. In addition, 15 expert violinists with at least 7 years of experience [EG; eight women, seven men; mean age: 32.4 (10.06); mean years experience: 18.6 (5.53)] were recruited from both the university campus and different music schools and conservatories in Barcelona. Participants provided their written consent and procedures were approved by the Conservatoires UK Research Ethics committee on 04/04/2017, following the guidelines of the British Psychological Society. Participants also filled a questionnaire about their musical skills, main instrument, and years of music training. Beginner participants were randomly split into two different experimental groups: the Feedback Group [FG; 15 female, 14 male; mean age: 29.915 (4.88)] and the Control Group [CG; 19 female, nine male; mean age: 28.91 (7.5)]. All participants reported having received 1 year or less of formal training in a musical instrument [mean: 0.06 (0.23) years]. The study was carried out in one recording studio located in the Information and

Communication Technologies Engineering (ETIC) department of the Universitat Pompeu Fabra, Barcelona.

4.5.2 Experimental Procedure

4.5.2.1 FG and CG

Before starting the experiment both groups of beginners took part in a practice session. In that practice session, they were instructed on violin technique, bow position, stance, and bow grip through the Youtube video which explained some of the most important concepts to realize the required full bow exercises correctly (see section 2.3 for more details). A full bow exercise consisted of the alternation of two up and down bowing movements using the full length of the bow with the goal of producing a stable and clear sound. Participants could play while watching the video and explore creating sound with the violin, they could also rewatch different parts of the video while practicing. Participants were informed about the main variables we will use to evaluate their performance: *bow skewness* (bowing parallel to the bridge), *contact point* (measured as bow-bridge distance), *inclination* (taking care of not playing the other strings during the movement with the bow), *pitch stability* (related to avoiding scratchy sounds), and *dynamic stability* (trying to maintain the energy of the sound stable during the whole exercise, even during up-to-down or down-to-up changes). They also were encouraged to explore how pitch changes when they displace their finger down the fingerboard (the sound produced has a higher pitch) or when they displace it further away (produces a lower pitch). The duration of this practice session was around 16 min (6 min video + 10 min practice).

The experiment consisted of three blocks: *Baseline* (10 trials), *Acquisition* (35 trials), and *Transfer* (10 trials). In each trial, participants had to locate in the fingerboard of the D string the location of the five different musical target notes that were displayed triggering the reference synthetic sound (RSS) of the system, which was a pure tone at the chosen frequency. Then, while centered in front of the Kinect camera, they were asked to perform a full bow exercise taking into account what they learned in the practice session and while the system recorded their sound and motion descriptors (see section 2.4 for more details about the system used). We also collected their pitch deviation from the RSS. However, results related to intonation and pitch matching skills will be reported in an accompanying paper.

Both the *Baseline* and *Transfer* blocks were equal for all the participants. They consisted of five sub-blocks of two trials each (10 trials in total) where participants had to perform a full bow exercise in each one of them. The Acquisition block however differed between the groups although the total number of trials remained the same. It consisted of five sub-blocks with six trials per block that were performed under different conditions. The first two trials of each sub-block were performed under the *Normal Instrument Condition* (NIC) that consisted of two normal full bow exercises as those performed in the *Baseline* and *Transfer* blocks. The third and fourth trials were performed in a row in the Kinematic Instrument Condition (KIC) and it was different for each group of beginners. While performing the full bow exercises, the FG received real-time visual feedback (RTVF) on kinematics allowing them to correct their bow movements when they were not parallel to the bridge (i.e., improving bow skewness) or maintaining stable other important variables, such as bow-bridge distance or inclination. On the other hand, the CG was asked to perform full bow exercises as usual but placing special attention to the demanded kinematic variables and not paying so much attention to the produced sound. Finally, the fifth and last trial of the sub-block was called the *Sound Quality Instrument Condition* (SQIC) and it was also different for each group of beginners. The FG received RTVF on sound quality while performing two more full bow exercises allowing them to see in real-time the score of the descriptors *pitch stability* and *dynamic stability*. On the other hand, the CG was asked again, to perform the full bow exercises in a row and to pay attention to the quality of their sound and to the demanded sound quality variables. Between the NIC and the KIC there was a condition called the *Pitch Instrument Condition* (PIC). In that condition participants had the option to correct their previous decision regarding pitch after receiving different types of augmented feedback. Based on the assumption that pitch-matching skills should not interfere with bowing technique in the violin, details regarding the different types of feedback studied to improve intonation will be discussed in an accompanying article. In Figure 1 we can see a summary of the different blocks of the experiment.

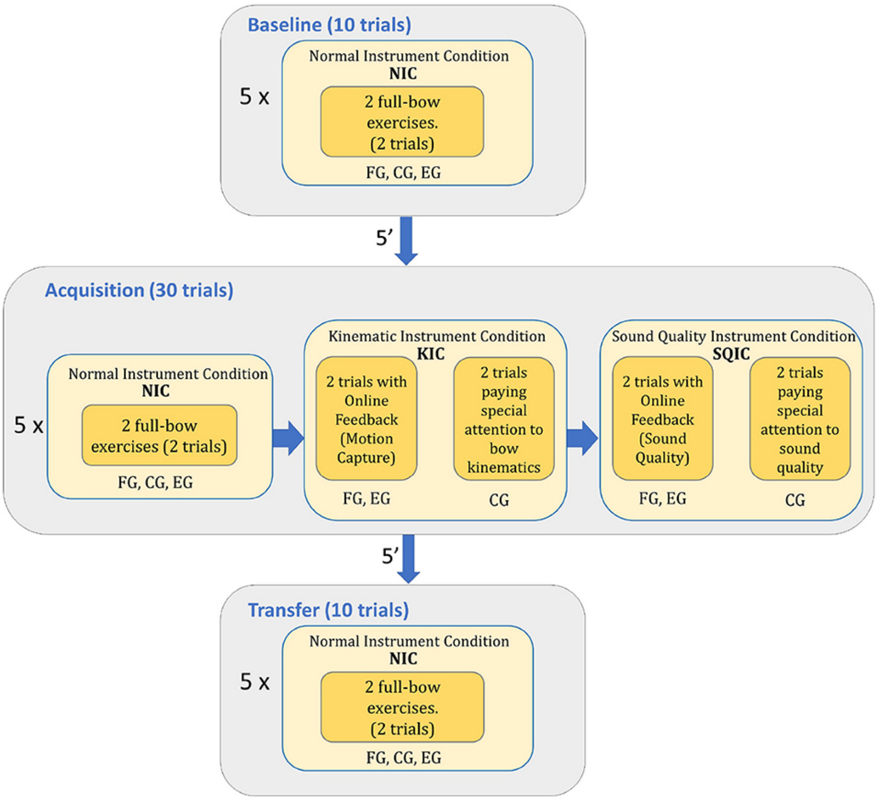

Figure 1. Diagram with the different blocks of the experiment and the different conditions each group of participants went through.

After the *Baseline* block and before the *Acquisition* block participants rewatched the instructional video and remained about the main variables that will be used to evaluate their performance. In addition, the real-time feedback was presented to the FG who received special instruction for its interpretation. On average, between one block and the other, participants rested around 5 min. The duration of the experiment tended to last between 1 and 1 h and a half. At the end of the *Transfer* block, those groups of participants who did not receive RTVF from the software (i.e., the CG) could experiment and practice freely with SkyNote and explore each one of the different feedback modalities the rest of the groups used (pitch, kinematic, and sound quality). After the experiment, all groups of participants answered a questionnaire giving their opinion regarding the technology seen.

4.5.2.1 EG

Before starting the experiment the EG watched the last part of the instructional video which contained a visual example of how to perform the exercise to make sure they understood the task. They were also informed about the main variables that would be used to evaluate their performance. Like those in the FG, the EG also received the same feedback in both KIC and SQIC. Finally, the EG also answered the same questionnaire giving their opinion regarding the technology seen.

4.5.3 Learning Materials

Basic information about violin playing techniques like stance, violin position, bow position, and grip was delivered to the beginner participants through one didactic Youtube video of a professional violinist before the experiment[2]. The video covers some aspects, such as *contact point*. The *contact point* is the point on the string where the bow force is applied, and needs to be located between the bridge and the fingerboard for good sound results. Thus, participants should maintain a constant contact point during the exercise. The video also covers the relation between speed and force, i.e., if you displace more force on the string you should move the bow faster to avoid “scratchy” sounds in the violin, otherwise if you displace less force you should move the bow slower to avoid “whistling” sounds). At the end of the explanation, there is a visual example of how to perform full bow exercises (alternation of up and down movements using the full length of the bow) focusing attention on bowing parallel to the bridge and how to move the wrist of the right hand to achieve a straight bow movement. The duration of the video is about 6 min. The EG visualized only the last part of the explanation to make sure they understood the task.

[2] https://youtu.be/mUz8fIc1FaY.

4.5.4 Providing Visual Feedback With SkyNote

The system we used to deliver real-time feedback to participants, SkyNote, is one of the main outcomes of the TELMI Project (Ramirez et al., 2018). SkyNote is an integrated system that combines different technologies for real-time feedback on pitch, intonation, dynamics (Mayor et al., 2009), motion capture (Vamvakousis et al., 2018), and tone quality (Giraldo et al., 2018). This feedback can be displayed in customized widgets or directly on the musical score, allowing for real-time experimentation and overall performance evaluation. However, for this experiment, we presented feedback of a single performance aspect at a time.

Figure 2 shows the display used for the real-time feedback used for tone quality. Several descriptors, such as "Pitch Stability" and "Dynamic Stability" appear represented in a spider chart delivering online feedback about the score of each one of the descriptors used (for more details see Giraldo et al., 2018).

The system can also monitor specific aspects of the bowing technique when a motion-tracking device is attached (i.e., a Microsoft Kinect) and some markers are placed on the bow and the violin (see Figure 3). Some of these aspects include bow tilt, speed, weight, contact point, inclination, and direction. In Figure 4 you can see the online display on kinematics used for the experiment. For more details see Vamvakousis et al. (2018).

We used an omnidirectional condenser microphone (Behringer, 2013) mounted on a stand to record the audio during the session. One NUC computer to run SkyNote, and two screens: one to deliver feedback to the participant and the other one for the experimenter. The feedback screen could be locked or unlocked by the experimenter based on the condition or group to which the participant belongs.

4.5.5 Questionnaires

Using a questionnaire developed inside the context of the TELMI project we collected some of the views of the participants after the experiment in 27 questions. The questionnaire is available online[3].

[3] https://www.survio.com/survey/d/N2C1Q8P3E9Y4H4A5I.

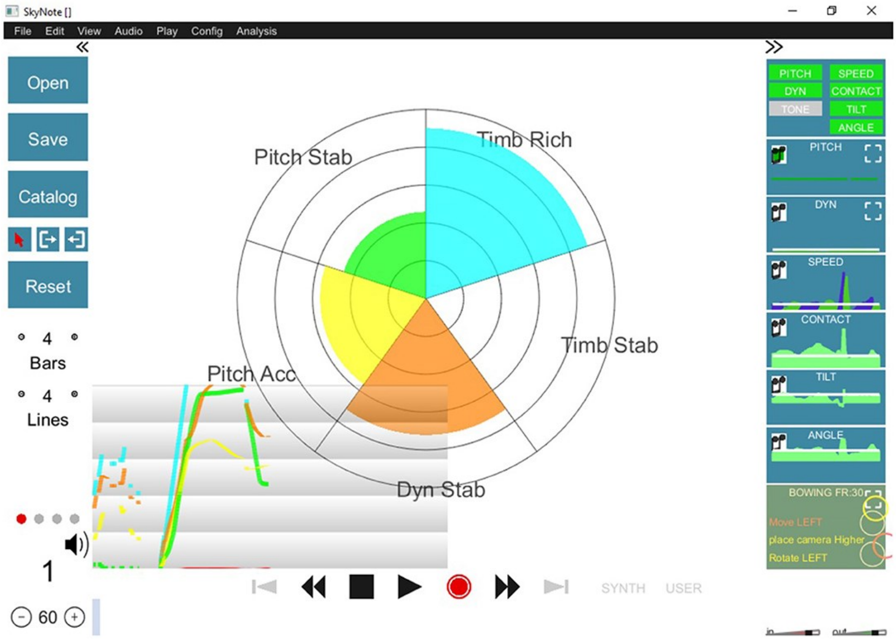

Figure 2. Visual display of the tool used to offer real-time sound quality feedback to participants. Each portion of the spider chart represents a different sound feature while its amplitude represents how close the participant was to the ideal sound.

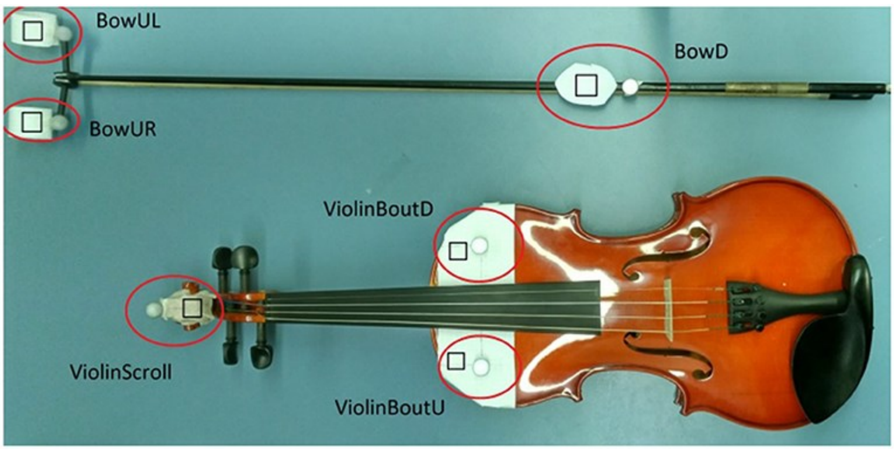

Figure 3. These markers, when placed in the bow and the violin, allow SkyNote to track the bow movement of the participant.

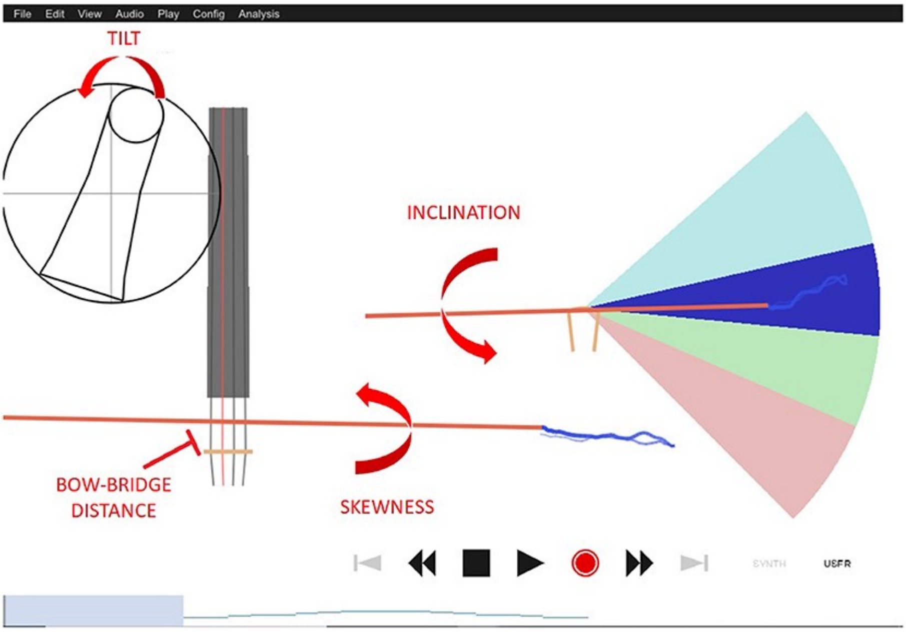

Figure 4. Visual display of the tool used to offer real-time feedback on motion and kinematics to participants. From the display, we can infer some of the descriptors that will be computed later on. The value of skewness, for example, is close to 0 when the bow remains perpendicular to the strings as is the case in this figure.

Questions related to the usability of the technology were ignored as, in this experiment, participants did not operate the tool (but the experimenter). The questionnaire could be separated into four different sections of questions: questions related to satisfaction with the technology, perception of their own performance, effectiveness of the augmented feedback delivered and problems found with augmented feedback.

The *Satisfaction* questions of the questionnaire had the following form:

• To what degree this tool (from 1 *not satisfied at all*, to 5 *very satisfied*)

□ #2…help you learn more quickly?
□ #3…improve your performance?
□ #4…increase your productivity?
□ #5…increase the effectiveness of your practice?
□ #6…make practicing easier?

□ #7…useful?

• How likely are you to (from 1 *not at all likely*, to 5 *very likely*)
□ #25…continue using this tool
□ #26…recommend this tool to others

Questions #16 and #17 were related to the effectiveness of each one of the technologies used.

• Rate the Technology (from 1 *not effective*, to 5 *very effective*)
□ #16…Timber Stability
□ #17…Kinect and motion detection

Question #18 was related to the *perception of their own performance*.

• #18 What do you think has improved more during the session? Select one answer:
□ Pitch-Matching
□ Timber
□ Motion and Kinematics
□ Others.

Questions 20 to 24 were related to the problems found with the augmented feedback. Questions were presented in the form of statements. Answers were from 1 *Strongly Disagree* to 5 *Strongly Agree*:

• #20 Feedback too fast to follow
• #21 Too much feedback information
• #22 Feedback difficult to understand
• #23 Cannot play while watching the feedback.

4.6 Sound and Motion Analysis

All the data was processed in Matlab (MATLAB, 2010), analyzed in Weka (Frank et al., 2016), and in SPSS (IBM Corp, 2011). All the raw data, wav files and statistics for each participant are freely available from Zenodo (Blanco et al., 2020**,** 2021).

4.6.1 Kinematic Features

Figure 4 shows an example of some of the parameters extracted from the exercise of each participant.

• Position: Refers to the distance between the contact point to the frog computed as the euclidean distance.
• Velocity: The derivative of bow position.
• Bow-bridge distance: Distance between the contact point and the bridge.
• Inclination: The first euler angle (roll) of the bow rigid object in the violin coordinate system.
• Tilt: The second euler (pitch) angle of the bow rigid object in the violin coordinate system.
• Skewness: The third euler angle (yaw) of the bow rigid object in the violin coordinate system.
• Bow-violin distance: the distance between the bow and the violin itself.

Each feature was extracted with a sampling rate of 86.13 samples/s. The skewness angle, as defined here, has a value of 0 when the bow is completely perpendicular to the strings. For each trial, we computed the *bow skewness* descriptor as the mean absolute error of the skewness angle referenced to zero (see Equation 1).

$$bowSkewness = \frac{1}{N}\sum_{i=1}^{N}|angle - 0| \qquad (1)$$

Where *angle* is the third euler angle measured and 0 the reference. N is the number of samples in a trial.

4.6.2 Sound Quality Features

Sound quality features were extracted in the same manner as in Blanco and Ramirez (2019). We used the Yin algorithm (Llimona, 2015) to extract the fundamental frequency (f0), instantaneous power, and aperiodicity from the audio signal of each trial using a window size of 33 ms with a hop size of 0.7 ms. The quality of the sound recorded in one trial was assessed through sound descriptors, such as *dynamic stability* (Equation 2) or *pitch stability* (Equation 3) by computing the standard deviation of both f0 and power, respectively throughout the trial

(Romaní et al., 2015). Equations (2) and (3) provide a formal description of these descriptors.

$$dynamicStability = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Pi-\mu)^2} \quad (2)$$

$$pitchStability = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f0i-\mu)^2} \quad (3)$$

Where N is the number of samples in a trial. *Pi* is instantaneous power in Db. *f0i* is the instantaneous fundamental frequency in Hz and μ is the mean value of, respectively, the power (Equation 2) or the fundamental frequency (Equation 3) calculated over the trial. Note that in this definition of the descriptors lower values indicate *more* stability while higher values indicate *less* stability.

4.6.3 Statistical Analysis

We performed two different analyses of the data using SPSS. One for the kinematic results and another one for the sound quality results.

Because we wanted to evaluate the importance of some of the kinematic descriptors extracted to differentiate between beginners and experts, we performed a 3 × 5 mixed-design for each analysis with Group (FG, CG, and EG) as between-subject factors and Condition (*Baseline, Acquisition-NIC, Acquisition-KIC, Acquisition-SQIC*, and *Transfer*) as the within-subject factor. For the kinematic analysis, the mixed-design was univariate with the results of *bow skewness* for each condition while for the sound quality analysis it was multivariate with the results of *dynamic stability* and *pitch stability* for each condition. *Post-hoc* tests using the Tukey method for multiple comparisons with Bonferroni correction were performed between the groups of participants. The descriptors that showed significant differences between the experts and both groups of beginners were considered as good evaluators of performance.

To study the impact of real-time feedback in our beginner's groups we needed to look for possible interactions between both beginner's groups and conditions for those variables previously considered. For that purpose, we performed a 2 × 5 mixed-design with Group (this time only FG and CG) as between-subject factor and Condition as the within-subject factor. For the descriptors that showed a significant interaction between Condition and Group, a posterior simple main effect analysis was performed on each group to find out which conditions were causing the interaction. Pairwise comparisons tests were performed between the conditions using the Bonferroni correction. Finally, to compare the effect of training with SkyNote in the amount of improvement, we performed three independent *t*-tests of the relative difference between the *Transfer* and the *Baseline* for each one of the descriptors applying Bonferroni correction.

Before running the analysis we discarded all the participants who declared to be left-handed (two from the CG and four from the FG) together with one participant from the FG who declared having already received violin lessons as a child. Given that we found deviations due to bad Kinect camera tracking not related to the actual performance of participants (and thus other modalities were not affected), we decided to separately perform the outlier analysis for each modality. We also removed four participants from the CG and three more from the FG in the kinematic analysis because they were labeled as outliers (values bigger than three interquartile ranges). Finally, we removed one participant from the CG and one from the EG in the sound quality analysis for the same reason. After removing the outliers all the data passed the assumptions of normality required to perform the tests. All the results presented in the following sections were Greenhouse-Geisser corrected.

4.7 Results

4.7.1 Analysis of Differences Between Experts and Beginners

Multivariate tests of within-subject effects for *dynamic stability* and *pitch stability* showed significant results in Conditions ($p < 0.0001$) and interaction between Conditions*Group ($p < 0.0001$). Results for *dynamic stability* and *pitch stability* were lower for the EG compared with both FG and CG regardless of conditions (see Figure 5B). That is, the sound of the experts was more stable during the exercise. Tests of Between-Subjects Effects showed significant results for Group both at *dynamic stability* and *pitch stability* ($p < 0.0001$ in both). *Post-hoc* tests showed significant differences between the EG and both the CG and FG in the two descriptors ($p < 0.0001$ in all the tests). Thus, we also considered *pitch stability* and *dynamic stability* as good evaluators of performance and proceeded with the analysis.

Univariate tests of within-subject effects for *bow skewness* showed significant results for Condition ($p < 0.0001$) and interaction between Condition*Group ($p < 0.0001$). Results for *bow skewness* were lower for the EG compared with both FG and CG regardless of conditions (see Figure 5A). Tests of between-subjects effects showed significant results for Group ($p < 0.0001$). That is, their bow was straighter during the exercise. *Post-hoc* tests showed significant results between the EG and the FG ($p = 0.001$) and CG ($p < 0.0001$). Thus, we considered *bow skewness* as a good evaluator of performance and proceeded with the analysis.

4.7.2 Kinematic Analysis

Univariate tests of within-subject effects for the beginner's groups showed significant results for Condition ($p < 0.0001$) and an interaction Condition*Group ($p < 0.007$). *Post-hoc* tests did not show significant differences between CG and FG. Simple main effect analysis revealed significant results for Condition in both the univariate tests of within-subject effects for the FG and CG ($p < 0.0001$ and $p = 0.003$, respectively). Both groups improved on average their results after the *Baseline* (see Figure 5A). The biggest improvements for the FG were found at the *Acquisition-NIC* (14.5% of improvement over the *Baseline*), *Acquisition-KIC* (31% of improvement), and at the *Acquisition-SQIC* (27.8% of improvement). The biggest improvements for the CG were found at the *Acquisition-SQIC* (23% of improvement). Pairwise comparison tests between conditions revealed significant differences between the *Baseline* and the *Acquisition-NIC*, the *Acquisition-KIC* and the *Acquisition-SQIC* in the FG ($p = 0.002$, $p < 0.0001$ and $p = 0.001$, respectively), while in the CG we only found differences between the *Baseline* and the *Acquisition-SQIC* ($p = 0.019$).

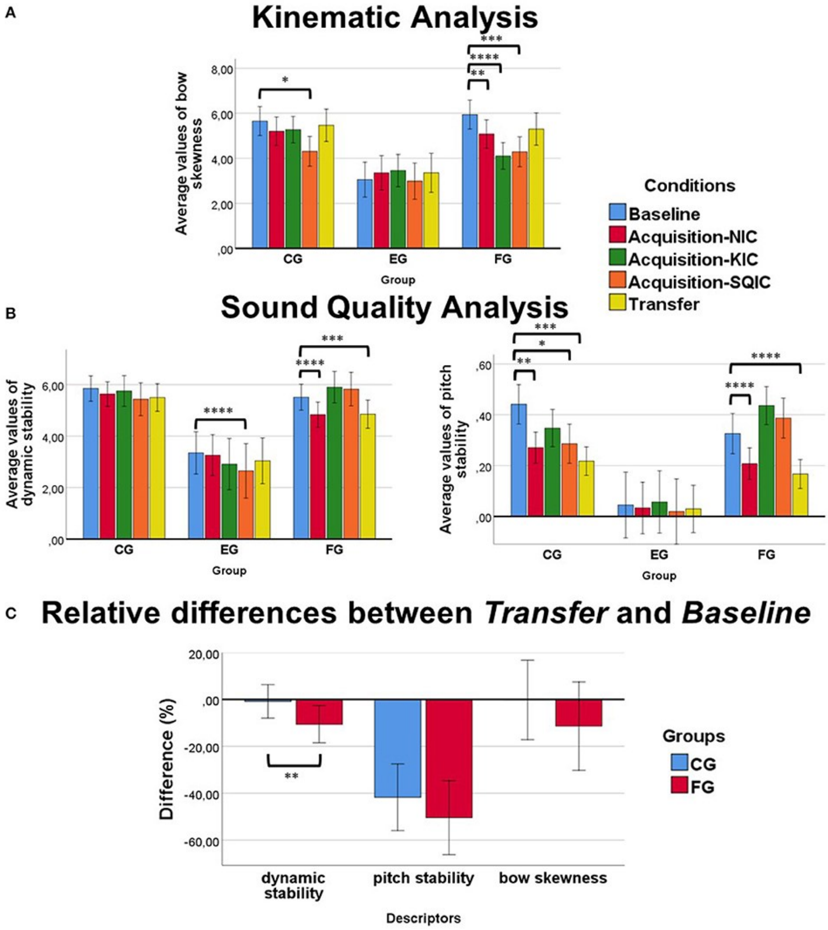

Figure 5. **(A)** Kinematic analysis. Bow skewness results: the FG improved significantly their results compared with the Baseline at the moment of receiving online feedback on bow motion (i.e., at the Acquisition-KIC) and in the rest of the conditions from the Acquisition phase. The CG improved their results only at the Acquisition-SQIC. **(B)** Sound Quality Analysis. dynamic stability results (left): although results for the FG tended to get worse at the moment of receiving online feedback, those results were transferred to conditions without feedback (Acquisition-NIC and Transfer). No significant improvements were found for the CG. pitch stability results (right): both groups of beginners (control and feedback) improved their results in pitch stability at the Acquisition-NIC and Transfer. The FG tended to get worse results when receiving online feedback. **(C)** Relative differences between Baseline and Transfer: The FG seemed to improve on average, more than the CG in all the descriptors. However, only significant results between groups were found at dynamic stability. *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$.

4.7.3 Sound Quality Analysis

Multivariate tests of within-subject effects showed significant results for Condition ($p < 0.0001$ in both descriptors) and an interaction Condition*Group ($p = 0.002$ for *dynamic stability* and $p < 0.0001$ for *pitch stability*). *Post-hoc* tests did not show significant differences between CG and FG. Simple main effect analysis revealed significant results for Conditions in the univariate tests of within-subject effects in *dynamic stability* and *pitch stability* for the FG ($p < 0.0001$ in all the tests). Significant results were found only for *pitch stability* in the CG ($p < 0.0001$). The CG improved their results in *pitch stability* after the *Baseline* (see Figure 5B, left figure). Pairwise comparison tests revealed significant differences in *pitch stability* between the *Baseline* and the *Acquisition-NIC*, the *Acquisition-SQIC*, and the *Transfer* block in the CG ($p = 0.005$, $p = 0.015$, and $p = 0.001$, respectively). No significant results were found between the *Baseline* and the *Acquisition-KIC*. A similar but less pronounced trend was observed for their results in *dynamic stability* although they did not reach significance. On the other hand, the FG seemed to improve their results in *pitch stability* at the *Acquisition-NIC* and at the *Transfer* condition but worsened its results at both the *Acquisition-KIC* and the *Acquisition-SQIC*, i.e., when receiving RTVF. This trend was similar for *dynamic stability* although less pronounced (see Figure 5B, right figure). Significant results in the FG for *pitch stability* were only found between the *Baseline*, the *Acquisition-NIC*, and the *Transfer* condition ($p < 0.0001$ in both conditions). Additionally, the FG showed significant differences in *dynamic stability* between the *Baseline* and the *Acquisition-NIC* and the *Transfer* conditions ($p < 0.0001$ and $p = 0.005$, respectively).

Interestingly, simple main effect analysis also revealed significant results for Conditions in the univariate tests of within-subject effects in *dynamic stability*. The EG also seemed to improve their results in *dynamic stability* after the *Baseline* but especially in the *Acquisition-SQIC*. Pairwise comparisons showed that the EG showed significant results between the *Baseline* and the *Acquisition-SQIC* condition ($p < 0.0001$) and close to significance between the *Baseline* and the *Transfer* condition ($p = 0.07$).

4.7.4. Effect of Training on Performance Improvement and Correlations

The FG obtained on average better results than the CG when comparing the *Transfer* with the *Baseline* condition (see Figure 5C). In *bow skewness*, the FG improved 5.5% more. In *dynamic stability*, they improved 10.3% more and in *pitch stability* a 8.1% more. However, only significant results between groups of participants were found for the *dynamic stability* descriptor ($p = 0.003$).

No significant correlations were found between the average value of *bow skewness* variable for each participant at the *Baseline* and *Transfer* phase with any of the two different sound descriptors used to evaluate the sound quality (two-tailed Pearson's correlation).

4.7.5 Questionnaires

In this section, we offer different results for the four different parts of the questionnaire participants answered. Two participants from the EG were removed from the analysis since they belonged to the project.

4.7.5.1 Satisfaction

After adding up the answers of all the participants we got a "Satisfaction Score" which goes from 8 (in case all the answers were 1) to 40 (in case all the answers were 5). A Univariate Analysis of Variance was performed on the data with *Satisfaction* as the dependent variable and *Group* (FG, CG, and EG) as fixed-factor. The average satisfaction with the technology was similar for the three different groups [CG: 34.611 (1.015); FG: 33.517 (1.131); EG: 33.769 (1.689)]. No significant differences were found in the tests of Between-Subjects Effects at *Group*. See Figure 6A.

4.7.5.2 Participant's Perception of Their Own Performance

In question #18 we asked participants their beliefs regarding what has improved more during the session. Only a small but similar number of participants from the FG and CG (11.1 and 7.7%, respectively) considered that *timber* was the feature that improved more during the session (see Figure 6B). The main differences were found in *motion and kinematics* were a smaller number of participants from the FG compared with the CG (around 17% less) considered it as the most improved feature.

4.7.5.3 How Effective Is Each Technology

In general, the majority of participants rated both technologies as effective or very effective, even the expert group. Motion capture feedback tended to be more valued than sound quality feedback by all the different groups being the expert group the more optimistic with it. 75% of the experts considered the technology to be “very effective” for learning and 25% of them as “effective” (see Figure 6C).

4.7.5.4 Problems With Feedback

We found that a relatively constant number of participants (around 20 and 30% from both FG and CG) agreed with the statements “Feedback too fast to follow” and “Too much feedback information” (see Figure 6D). Also, around 10% of participants in both FG and CG agreed with “Feedback difficult to understand.” The expert group tended in almost equal parts to disagree with the statements or to maintain a neutral position.

A clear division is found in the statement “Cannot play while watching the feedback.” More than half of the participants of the FG agreed with that statement vs. 20% of participants of the CG and 0% of the EG.

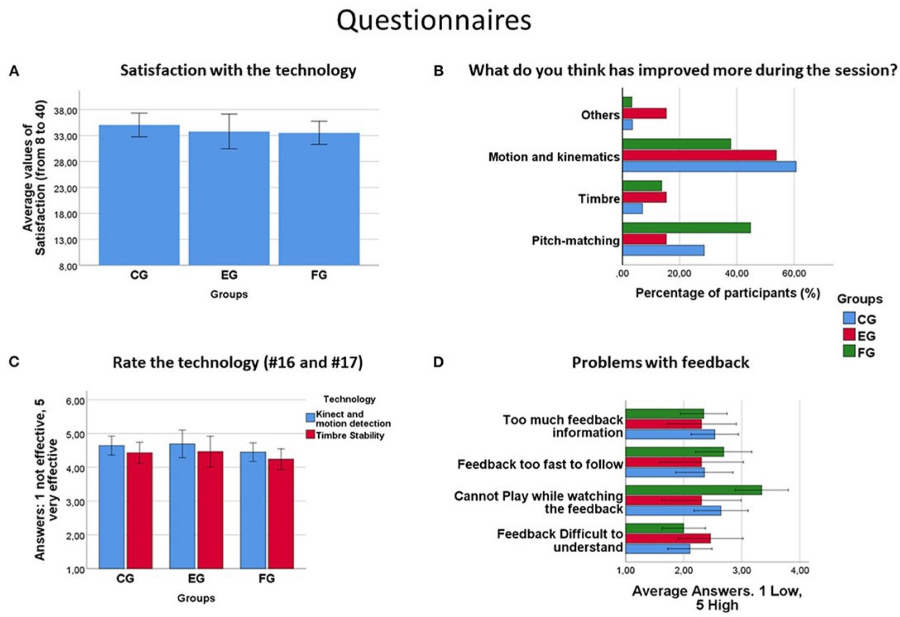

Figure 6. (A) Satisfaction with the technology. No significant differences were found, and the average values were similar for the three groups. **(B)** Answers regarding the perception of the participant's own performance separated by groups. Just a very small percentage of participants of each group considered that timber stability was the feature that improved more during the session. **(C)** Rate the technology. Effectiveness of each technology according to participants. Both technologies tended to be highly valued by the participants. However, motion capture feedback tended to be slightly more valued. Also, experts tended to rate the effectiveness of each technology better than beginners. **(D)** Problems with feedback. Unlike some participants from the FG and CG, experts did not seem to have problems with feedback. More than half of the participants from the FG especially agreed with the fact that it was hard for them to play while watching the feedback (statement #23).

4.8 Discussion

In this study, we have evaluated the use of RTVF of sound and motion capture technologies by comparing a group of participants practicing with such feedback vs. a group of participants practicing without it. Both groups were composed of beginner violin players. We also asked a group of expert violin players to perform the same tasks for comparison purposes. We replicated some of the results from Blanco and Ramirez (2019) and confirmed the usefulness of the proposed audio descriptors (*dynamic stability* and *pitch stability*) to both differentiate the expert performance from the beginner performance and to track the learning process of participants. Just as *pitch stability* and *dynamic stability* are able to differentiate a beginner from an expert, we have found that differences in *bow skewness* can also differentiate between the two groups.

Regarding the effect of sound quality feedback, both beginner groups improved significantly on *pitch stability* obtaining results close to those of the experts in the *Acquisition-SQIC* condition. No improvement in their results were seen for the expert group nor an effect of the technology in their outcomes. However, the presence of technology seemed to affect the beginner group which used it. Unlike the CG, who learned without RTVF and just focusing on practicing each skill separately, the FG did not show significant improvements in *pitch stability* while using the RTVF technology (neither with kinematic feedback nor with sound quality feedback). This effect may be related to the related distraction that a visual real-time feedback technology can impose in learning the violin, especially with beginners as made explicit by Pardue et al. (2015). Evidence in favor of this hypothesis can be seen in the answers to the questionnaires. More than half of the participants who received RTVF (55.2%) answered “Agree” or “Strongly Agree” in equal proportions to the statement which said “Cannot play while watching the feedback.”

RTVF of sound quality was particularly useful for learning to maintain a stable loudness level through the audio descriptor of *dynamic stability*. Although the CG received the same instructions that the FG about the parameters of the sound that will be considered to evaluate their performance, the CG's results in *dynamic stability* did not improve significantly throughout the session. The FG's results, on the contrary, improved significantly in the *Acquisition-NIC* and in the *Transfer* condition for *dynamic stability*. Again, the fact that the results of *dynamic stability* were worse at the *Acquisition-KIC* and at the *Acquisition-SQIC* may be related to the distracting effect of the RTVF. However, despite the distraction, feedback on *dynamic stability* allowed participants to consider it during their learning as evidenced by their improvement at both the *Acquisition-NIC* and *Transfer* condition. This effect coincides with previous results using RTVF for pitch accuracy in singing voice melody production where the results of performance tend to decay while using the technology but improve at later post-tests scores (Welch et al., 1989; Wilson et al., 2005; Paney and Tharp, 2019). This suggests that, besides the increase of cognitive load that RTVF may impose on participants by worsening their performance while receiving feedback, it should not be considered a damaging factor. As Sherwood and Lee (2003) already pointed out, not only do movements need to be practiced but also the cognitive decision-making processes underlying skilled behavior need practice as well. Despite distracting players' attention, RTVF of sound quality could make explicit performance errors that could be going unnoticed otherwise.

Interestingly, the experts improved their performance significantly in *dynamic stability* while using the RTVF at the *Acquisition-SQIC* suggesting that the technology was not distracting them as much as the beginners. Their ability and their strong formed schemas supposedly would allow them to allocate more cognitive resources to the interpretation of the feedback without disrupting their performance. Again, this was also reflected in the questionnaires where no participant in the EG agreed with the statement “Cannot play while watching the feedback.” Besides, the EG's also seemed to improve more in that descriptor in the rest of the conditions to the point of giving results very close to significance in the *Transfer* condition.

In terms of RTVF of kinematic movements, although participants were told to control three different kinematic variables (*bow skewness, inclination*, and *bow-bridge distance*) for this study we decided to focus only on *bow skewness* which, as already pointed out before, seemed to be a reliable estimator to differentiate a beginner's performance from that of an expert. Both groups of beginners seemed to improve their results after the *Baseline*. The FG improved significantly their results in *bow skewness* at the whole *Acquisition* phase, even in those conditions where the feedback was not present. However, that improvement was not transferred at the *Transfer* condition. On the other hand, results from the CG only improved significantly in the *Acquisition-SQIC*.
Unlike the CG, the RTVF of kinematic movements improved significantly the performance of the FG in the *Acquisition-KIC*.

Contrary to previous results with RTVF of sound quality, kinematic feedback seemed to improve the results of participants using it. The reason why performance on *bow skewness* did not decay while receiving feedback was probably due to the nature of the type of feedback itself. The sound quality visual feedback used did not offer information about how to improve the generated sound. On the other hand, kinematic feedback offered real-time information about the movement of the bow allowing participants to immediately correct their bow movements. This distinction between types of feedback is similar to the one we find in visuomotor rotation paradigms between *reward feedback* and *sensory feedback* (Krakauer et al., 2019). Literature in adaptation paradigms reports how each type of feedback could lead to differences in behavior and retention of the learned movements (Izawa and Shadmehr, 2011; Shmuelof et al., 2012; Nikooyan and Ahmed, 2015). This is something that could strongly influence participant behavior and should be taken into account at the moment of designing and evaluating feedback technologies. On the other hand, the fact that performance on *pitch stability* decayed at the *Acquisition-KIC* while *bow skewness* improved in the FG suggests that participants were trying to play with straight bowing at the expense of the quality of the sound. However, it is important to note that for the CG, the performance in terms of *pitch stability* became worse while trying to keep the bow straight but their results in *bow skewness* did not improve significantly as those of the FG.

Both the CG and FG improved their results in *bow skewness* during the *Acquisition-SQIC*. The reasons, however, varied for each group. The CG was able to significantly improve their results both on *bow skewness* and *pitch stability* at the same time. It is possible that by suggesting them to focus only on the quality of the sound, they engaged in an external locus of focus which guided more precisely their arm movements. As Wulf and Lewthwaite (2016) suggested, the external locus of focus prevented learners from interfering with the automatic control processes of their motor system. That could be also the reason why participants from the CG did not improve their results when focusing their attention on their movements. By asking them to focalize their attention on their movements, we would be promoting an internal locus of focus interfering with their automatic control processes. The FG, as mentioned previously, did not improve their results in sound quality during the *Acquisition-SQIC* presumably due to feedback distraction. However, the fact that the FG not only maintained good results at *bow skewness*, but that those results were bigger than for the CG (5% more of improvement) may suggest temporary retention of the kinematic movements needed for straight bowing from the *Acquisition-KIC* to the *Acquisition-SQIC*. The order of the conditions could also have influenced the observed behavior, also in the CG. However, the fact that the FG was able to maintain good results at *bow skewness* during the *Acquisition-NIC* tells us that there was indeed retention at least in the short term that was transferred to the rest of the conditions of the *Acquisition* phase. Moreover, the improvement in *bow skewness* in the *Acquisition-NIC* was accompanied by a significant improvement in both sound quality descriptors. The FG was the only group that showed improvement in all the descriptors at the same time. This suggests that the FG learned how to incorporate together the different feedback received at the *Acquisition-KIC* and at the *Acquisition-SQIC*.

Questionnaires also allowed us to have a broader view of the opinion of participants about the technology. All groups of participants rated both technologies as effective or very effective for learning, especially the EG. In general, motion capture technology tended to be rated as more effective than sound quality feedback. A larger number of participants considered that "Motion and Kinematics" improved more than "Timber" during the session. This contrasts with the obtained outcomes of the experiment where no group retained the levels of straight bowing that they reached at least in one of the three conditions of the *Acquisition* phase. It could be hypothesized that participants from the FG thought that the quality of their sound was not improving because at the time of receiving the feedback they were not receiving a positive one (as inferred from the results in *pitch stability* and *dynamic stability* at the SQIC). At the same time, they improved their bowing movements while using feedback, possibly due to the type of feedback that allowed them to know how to correct their movement. However, the fact that the CG also showed similar results and similar answers in the questionnaires may suggest that straight bowing is a difficult skill to self-assess for those who lack the appropriate metacognitive skills about his/her own level of performance. It also may be unreasonable to expect that learning to bow correctly can be improved in a single session. As Van Der Linden et al. (2011) found, it is even complicated to maintain and retain some of the improvements made during six training sessions. Our results match Linden et al. results by showing how feedback was helping participants to improve their movement. However, although we have seen how this improvement in *bow skewness* came at the cost of disregarding the sound quality of the performance at the moment of receiving feedback, we have also shown how it was retained in conditions where feedback was not present and, accompanied by improvement in sound quality.

SkyNote has been applied at the Royal College of Music with high-level violin students. The results of using SkyNote as well as how the technology can be implemented in teaching and learning practice at a higher education institution will be discussed in an accompanying paper.

4.9 Conclusions

In this work, we have presented and evaluated some of the technologies developed during the TELMI project. We have designed an experimental setup where complete beginners start learning the basics of violin playing, such as the production of a stable and sustained sound. This study extends our previous results (Blanco and Ramirez, 2019) and reaffirms the importance and the impact this kind of technologies may have in the process of learning a musical instrument and evaluating different learning methodologies.

In summary, we can list some of the main findings of this study:

1. We have shown how sound quality and motion-capture descriptors, such as *dynamic stability, pitch stability*, and *bow skewness* may characterize part of the participants' improvement in sound production and bowing technique and may be used to evaluate learning interventions.

2. Although *bow skewness* is usually treated as a precondition for obtaining good sound, the results in this study indicate that, for total beginners, this relation is not straightforward. We have seen how

focusing on the quality of the sound rather than focusing on playing with a straight bow could, in fact, improve straight bowing. This fact could be justified by the choice of an external locus of focus (quality of the sound) rather than an internal one (movement of the arm). However, the order of the conditions could have influenced the results.

3. Real-time kinematic feedback of bow movement influenced differently the participant's performance than the sound quality feedback did. While participants improved their bow movements at the moment of receiving kinematic feedback their results in sound quality got worse. Furthermore, their results in sound quality worsened at the moment of receiving sound quality feedback while their bow movements held up better despite not receiving kinematic feedback. However, when RTVF was removed participants improved in all the descriptors. Again, although the order of conditions could have influenced the results we argue that the type of feedback (and modality) is the main reason for these results. Visual feedback splits attention and can lead to an increase in cognitive load in beginners. This is corroborated by the fact that the expert performance was not influenced by real-time feedback. Even more, real-time feedback improved their performance in *dynamic stability* right at the moment it was received.

4. Finally, we have seen how beginners who received feedback tended to improve more, on average than those who did not in the retention test (*Transfer* condition). However, only significant results were found for *dynamic stability* where the improvement was greater and clearer. Interestingly enough, experts also seemed to slightly improve their performance in *dynamic stability* at the *Transfer* condition. However, that improvement was not statistically significant and we cannot directly infer that feedback was the cause.

Such technologies may help students to avoid bad habits that could occur during their long-periods of self-study, and to increase their motivation and own-expectations toward learning. Furthermore, these technologies can be used to better comprehend and add more clarity to the scarce research in motor learning in music activities. Only by improving the ways we can acquire and track data, and extract and evaluate descriptors from activities, which were previously evaluated based on solely subjective mechanisms, we can objectively gain new insights on how the *body*, understood in its entirety, becomes the subject of learning.

4.10 References

Behringer (2013). *Behringer-C3. Documentation*. Available online at: https://mediadl.musictribe.com/media/sys_master/hcb/h93/8849528520734.pdf (accessed April 19, 2021).

Blackwell, L. S., Trzesniewski, K. H., and Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an intervention. *Child Dev*. 78, 246–263. doi: 10.1111/j.1467-8624.2007.00995.x

Blanco, A., and Ramirez, R. (2019). Evaluation of a sound quality visual feedback system for bow learning technique in violin beginners: an EEG study. *Front. Psychol*. 10:165. doi: 10.3389/fpsyg.2019.00165

Blanco, A. D., Tassani, S., and Ramirez, R. (2020). *Skynote Experiment*. Zenodo. doi: 10.5281/zenodo.4553940

Blanco, A. D., Tassani, S., and Ramirez, R. (2021). *Real-Time Sound and Motion Feedback for Violin Bow Technique Learning: A controlled, Randomized Trial* [Data set]. Zenodo. doi: 10.5281/zenodo.4554201

Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., et al. (2013). "Essentia: an open-source library for sound and music analysis," in *MM'13* (Curitiba), 855–858. doi: 10.1145/2502081.2502229

Brandfonbrener, A. (2004). Healthier music students: can medicina and music prescribe in concert? *Med. Probl. Perform. Art*. 19, 166–165. doi: 10.21091/mppa.2004.1001

Dalmazzo, D., and Ramírez, R. (2019). Bowing gestures classification in violin performance: a machine learning approach. *Front. Psychol*. 10:344. doi: 10.3389/fpsyg.2019.00344

Demorest, S. M., and Pfordresher, P. Q. (2015). Singing accuracy development from k-adult: a comparative study. *Music Percept*. 32, 293–302. doi: 10.1525/mp.2015.32.3.293

Deutsch, L. M. (2011). *Motion study of violin bow technique* (Ph.D. thesis), University of California, Los Angeles, Los Angeles, CA, United States.

Dweck, C. S., and Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychol. Rev*. 95:256. doi: 10.1037/0033-295X.95.2.256

Fishbein, M., Middlestadt, S. E., Ottati, V., Straus, S., and Ellis, A. (1988). Medical problems among icsom musicians: overview of a national survey. *Med. Probl. Perform. Art*. 3, 1–8.

Frank, E., Hall, M., and Witten, I. (2016). *Data Mining: Practical Machine Learning Tools and Techniques, 4th Edn*. Burlington, MA: Morgan Kaufmann.

Giraldo, S., Ramirez, R., Waddell, G., and Williamon, A. (2018). "Computational modelling of timbre dimensions for automatic violin tone quality assessment," in *Proc Timbre is a Many-Splendored Thing Conference*.

Hay, J. (1993). *The Biomechanics of Sports Techniques, 3rd Edn*. Englewood Cliffs, NJ: Prentice-Hall.

IBM Corp. (2011). *IBM SPSS Statistics for Windows, Version 20.0*. Armonk, NY: IBM Corp.

Izawa, J., and Shadmehr, R. (2011). Learning from sensory and reward prediction errors during motor adaptation. *PLoS Comput. Biol*. 7:e1002012. doi: 10.1371/journal.pcbi.1002012

Jiao, J., Du, F., He, X., and Zhang, K. (2015). Social comparison modulates reward-driven attentional capture. *Psychon. Bull. Rev*. 22, 1278–1284. doi: 10.3758/s13423-015-0812-9

Jourden, F. J., Bandura, A., and Banfield, J. T. (1991). The impact of conceptions of ability on self-regulatory factors and motor skill acquisition. *J. Sport Exerc. Psychol*. 13, 213–226. doi: 10.1123/jsep.13.3.213

Konczak, J., vander Velden, H., and Jaeger, L. (2009). Learning to play the violin: motor control by freezing, not freeing degrees of freedom. *J. Motor Behav*. 41, 243–252. doi: 10.3200/JMBR.41.3.243-252

Korg (2018). *Cortosia*. Available online at: https://www.korg.com/us/products/software/cortosia/ (accessed December 8, 2018).

Krakauer, J. W., Hadjiosif, A. M., Xu, J., Wong, A. L., and Haith, A. M. (2019). *Motor Learning*. American Cancer Society.

Llimona, Q. (2015). *Yin Pitch Estimation Toolbox*. Available online at: https://github.com/lemonzi/matlab/tree/master/yin

Locke, E. A. (1966). *The Motivational Effects of Knowledge of Results: The Influence of Goal-Setting*. Technical report. Silver Spring, MD: American Institutes for Research.

Mangels, J. A., Butterfield, B., Lamb, J., Good, C., and Dweck, C. S. (2006). Why do beliefs about intelligence influence learning success? A social cognitive neuroscience model. *Soc. Cogn. Affect. Neurosci*. 1, 75–86. doi: 10.1093/scan/nsl013

MATLAB (2010). *Version 7.10.0 (R2010a)*. Natick, MA: The MathWorks Inc.

Mayor, O., Bonada, J., and Loscos, A. (2009). "Performance analysis and scoring of the singing voice," in *Proceedings of 35th AES International Conference* (London), 1–7.

Ng, K., and Nesi, P. (2008). "i-maestro: Technology-enhanced learning and teaching for music," in *NIME* (Genova), 225–228.

Nikooyan, A. A., and Ahmed, A. A. (2015). Reward feedback accelerates motor learning. *J. Neurophysiol*. 113, 633–646. doi: 10.1152/jn.00032.2014

Paney, A. S., and Tharp, K. L. (2019). The effect of concurrent visual feedback on adult singing accuracy. *Psychol. Music*. 1–11. doi: 10.1177/0305735619854534

Pardue, L. S., Harte, C., and McPherson, A. P. (2015). A low-cost real-time tracking system for violin. *J. New Music Res*. 44, 305–323. doi: 10.1080/09298215.2015.1087575

Peiper, C., Warden, D., and Garnett, G. (2003). “An interface for real-time classification of articulations produced by violin bowing,” in *Proceedings of International Conference on New Interfaces for Musical Expression* (Montreal, QC), 192–196.

Perez Carrillo, A., and Wanderley, M. M. (2012). “Learning and extraction of violin instrumental controls from audio signal,” in *Proceedings of the Second International ACM Workshop on Music Information Retrieval With User-Centered and Multimodal Strategies* (Nara: ACM), 25–30. doi: 10.1145/2390848.2390855

Ramirez, R., C. Canepa, S. G., Kolykhalova, K., Mancini, M., Volta, E., Volpe, G., et al. (2018). “Enhancing music learning with smart technologies,” in *Proceedings of 5th International Conference on Movement and Computing (MOCO'18)* (Genova), 4. doi: 10.1145/3212721.3212886

Romaní, O., Parra, H., Dabiri, D, TOkuda, H., Hariya, W., Oishi, K., et al. (2015). “A real-time system for measuring sound goodness in instrumental sounds,” in *138th Audio Engineering Society Convention* (Warsaw).

Schoonderwaldt, E., and Demoucron, M. (2009). Extraction of bowing parameters from violin performance combining motion capture and sensors. *J. Acoust. Soc. Am*. 126, 2695–2708. doi: 10.1121/1.3227640

Schoonderwaldt, E., and Wanderley, M. (2007). “Visualization of bowing gestures for feedback: the Hodgson plot,” in *Proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS07)* (Barcelona), Vol. 2, 65–70.

Sherwood, D. E., and Lee, T. D. (2003). Schema theory: critical review and implications for the role of cognition in a new theory of motor learning. *Res. Q. Exerc. Sport* 74, 376–382. doi: 10.1080/02701367.2003.10609107

Shmuelof, L., Huang, V. S., Haith, A. M., Delnicki, R. J., Mazzoni, P., and Krakauer, J. W. (2012). Overcoming motor “forgetting” through reinforcement of learned actions. *J. Neurosci*. 32, 14617–14621. doi: 10.1523/JNEUROSCI.2184-12.2012

Shomstein, S., and Johnson, J. (2013). Shaping attention with reward: effects of reward on space-and object-based selection. *Psychol. Sci*. 24, 2369–2378. doi: 10.1177/0956797613490743

Vamvakousis, Z., Perez, A., and Ramirez, R. (2018). *Acquisition of Violin Instrumental Gestures Using an Infrared Depth Camera*. Limassol: Zenodo.

Van Der Linden, J., Schoonderwaldt, E., Bird, J., and Johnson, R. (2011). Musicjacket—combining motion capture and vibrotactile feedback to teach violin bowing. *IEEE Trans. Instr. Meas*. 60, 104–113. doi: 10.1109/TIM.2010.2065770

Verrel, J., Pologe, S., Manselle, W., Lindenberger, U., and Woollacott, M. (2013). Coordination of degrees of freedom and stabilization of task variables in a complex motor skill: expertise-related differences in cello bowing. *Exp. Brain Res*. 224, 323–334. doi: 10.1007/s00221-012-3314-2

Visentin, P., Shan, G., and Wasiak, E. B. (2008). Informing music teaching and learning using movement analysis technology. *Int. J. Music Educ*. 26, 73–87. doi: 10.1177/0255761407085651

Volta, E., and Volpe, G. (2019). "Automated analysis of postural and movement qualities of violin players," in *2019 International Workshop on Multilater Music Representation and Processing* (Milano). doi: 10.1109/MMRP.2019.00018

Welch, G. F. (1985). Variability of practice and knowledge of results as factors in learning to sing in tune. *Bull. Council Res. Music Educ*. 19, 238–247.

Welch, G. F., Howard, D. M., and Rush, C. (1989). Real-time visual feedback in the development of vocal pitch accuracy in singing. *Psychol. Music* 17, 146–157. doi: 10.1177/0305735689172005

Wilson, P. H., Thorpe, C. W., and Callaghan, J. (2005). "Looking at singing: does real-time visual feedback improve the way we learn to sing," in *Proceedings of APSCOM 2005* (Seoul).

Wulf, G., and Lewthwaite, R. (2009). Conceptions of ability affect motor learning. *J. Motor Behav*. 41, 461–467. doi: 10.3200/35-08-083

Wulf, G., and Lewthwaite, R. (2016). Optimizing performance through intrinsic motivation and attention for learning: the optimal theory of motor learning. *Psychon. Bull. Rev*. 23, 1382–1414. doi: 10.3758/s13423-015-0999-9

Zanto, T. P., Rubens, M. T., Bollinger, J., and Gazzaley, A. (2010). Top-down modulation of visual feature processing: the role of the inferior frontal junction. *Neuroimage* 53, 736–745. doi: 10.1016/j.neuroimage.2010.06.012

Blanco, A. D., Tassani, S., & Ramirez, R. (2021). Effects of Visual and Auditory Feedback in Violin and Singing Voice Pitch Matching Tasks. *Frontiers in Psychology, 12,* 2578.

5. EFFECTS OF VISUAL AND AUDITORY FEEDBACK IN VIOLIN AND SINGING VOICE PITCH-MATCHING TASKS

5.1 Abstract

Auditory-guided vocal learning is a mechanism that operates both in humans and other animal species making us capable to imitate arbitrary sounds. Both auditory memories and auditory feedback interact to guide vocal learning. This may explain why it is easier for humans to imitate the pitch of a human voice than the pitch of a synthesized sound. In this study, we compared the effects of two different feedback modalities in learning pitch-matching abilities using a synthesized pure tone in 47 participants with no prior music experience. Participants were divided into three groups: a feedback group (N = 15) receiving real-time visual feedback of their pitch as well as knowledge of results; an equal-timbre group (N = 17) receiving additional auditory feedback of the target note with a similar timbre to the instrument being used (i.e., violin or human voice); and a control group (N = 15) practicing without any feedback or knowledge of results. An additional fourth group of violin experts performed the same task for comparative purposes (N = 15). All groups were posteriorly evaluated in a transfer phase. Both experimental groups (i.e., the feedback and equal-timbre groups) improved their intonation abilities with the synthesized sound after receiving feedback. Participants from the equal-timber group seemed as capable as the feedback group of producing the required pitch with the voice after listening to the human voice, but not with the violin (although they also showed improvement). In addition, only participants receiving real-time visual feedback learned and retained in the transfer phase the mapping between the synthesized pitch and its correspondence with the produced vocal or violin pitch. It is suggested that both the effect of an objective external reward, together with the experience of exploring the pitch space with their instrument in an explicit manner, helped participants to understand how to control their pitch production, strengthening their schemas and favoring retention.

5.2 Introduction

The use of technology (e.g mobile phones, computers, the internet) is widely used for a large number of different purposes related to music education (Zhukov, 2015). Some of the more demanded music learning apps are related to music theory, sight-reading, ear-training, and vocal training. Vocal training apps tend to offer real-time visual feedback of the performed pitch. However, despite the wide use of music education apps, such technologies are rarely employed in music schools where technology is usually restricted to audio/video recording and playback (Ramirez et al., 2015).

This research is part of TELMI (Technology Enhanced Learning of Musical Instrument Performance) a larger H2020 European project. In a previous study, we evaluated the effectiveness of augmented feedback in violin learning. In particular, we studied the effects of augmented feedback on pitch, motion-kinematics, and sound quality during the learning process of participants with no prior music experience \cite{R5}. In the present study, we investigate the results of different types of feedback on intonation learning in singing and violin playing.

5.3 Background

5.3.1 Real-Time Visual Feedback for Improving Intonation

Being able to play or sing in tune is an essential skill for most music students. That is probably the reason why the majority of the scientific literature about the effects of feedback in music learning has focused on intonation learning. Back to the beginning of the twentieth century, researchers from the University of Iowa developed a system to measure the pitch performed by participants and displayed it on a screen in realtime, allowing the participants to correct their performance instantaneously. They named their system Tonoscope (Seashore, 1902). Soon, a new generation of researchers started to study music performance and music learning using objective measures of sound such as frequency, intensity, and duration (for a review see Seashore, 1940). Some experiments attempted to show how training the ear with the visual feedback using the Tonoscope could result in a rapid improvement in pitch intonation and a transfer effect to new tones with different pitches (Seashore and Jenner, 1910; Knock, 1922; Brennan, 1926). Despite some methodological deficiencies (e.g., lack of control groups), these studies represent one of the first attempts to answer questions still relevant today regarding the use of feedback in music learning.

The Seashore's tonoscope was already available in the market in 1915. However, its use did not transcend outside the academic field. Some more recent approaches to characterize singing intonation skills were proposed by Welch with his schema theory of singing (Welch, 1985). From Welch's perspective, singing skills require external right/wrong feedback [also called knowledge of results (KR)] at the beginning of the learning process. The immediacy of this external feedback or concurrent KR is hypothesized to result in a more effective way of learning. Also, and in concordance with Schmidt's schema theory (Schmidt, 1975), the variability of practice may also be able to improve singing skills. Welch (1984) found how both real-time visual feedback with KR and variability of practice seemed to be an effective way to improve pitch-matching skills in children. These results were later replicated for melody production where, despite the fact that the participants worsened their accuracy at the time of receiving feedback, their results improved considerably in retention tests (Welch et al., 1989). Most importantly, realtime visual feedback without KR (that is, without right/wrong feedback) did not improve participants' performance in pitchmatching tasks (Welch, 1984). Similar results were recently found by Hutchins and Peretz (2012) in an experiment involving adult participants. This seems to evidence the importance of reward errors and objective measures for learning to sing.

Many singing apps have been proposed but few studies have attempted to evaluate their efficacy experimentally or in real learning contexts. For example, Wilson et al. (2005) evaluated with participants from different backgrounds and singing levels whether real-time visual feedback improved intonation in sung melodies more than discrete right/wrong feedback. They found that beginners benefit more from pitch real-time visual feedback than advanced singers, just as Welch hypothesized. They also found that participants' results tended to worsen at the moment of receiving feedback. Recently, Paney and Tharp (2019) evaluated the effects of real-time visual feedback with KR after 10 weeks of melody-singing training without finding significant differences with the control group. Some remarkable insights from that study come from the fact that participants using visual feedback tended to obtain better results than the control group although both groups improved. However, the removal of concurrent feedback led to a decay in performance for the experimental group whose retention scores were similar to those obtained by the control group. This drops the possibility that this type of feedback for improving singing skills could create dependency in the long term. On the other hand, both groups received KR at the end of each trial in the form of a score reflecting the overall accuracy of the trial. The lack of a control group without KR makes it hard to interpret if the improvement seen in participants could be related to KR or practice by itself.

In a recent study, Pardue and McPherson (2019) evaluated, both separately and combined, real-time auditory and visual feedback in violin intonation during four real-world violin lessons with beginners (adults and children). The real-time auditory feedback consisted of the pitch-corrected audio of a participant's playing to the nearest allowed pitch in the selected key (inspired by the tradition of students playing along with teachers). No statistical differences were found between each type of feedback, the combination of both nor the absence of feedback. However, their intonation was evaluated while using the technology and not in transfer or retention tests, also all the participants went through the different conditions instead of being separated into groups. Qualitative analysis and interviews of the participants seemed to point in the direction that the main problem of visual feedback was that it required visual attention. This could be the reason why some previous studies which evaluated the effects of real-time visual feedback in melodic production found a pattern of worsening results at the moment of receiving feedback (Welch, 1984; Wilson et al., 2005). On the other hand, some participants mentioned that the main problem with auditory feedback was that it did not provide information about in which direction errors should be corrected.

Most participants seemed to prefer the combination of both types of feedback.

5.3.2 Aural Feedback

Few studies have addressed the effectiveness of auditory feedback although their results remain contradictory. For example, Pfordresher and Brown (2007) found that hearing a synthesized voice concurrently with the singing of participants led to a detrimental effect on the absolute accuracy of poor-pitch singers but had a positive effect on good singers. On the other hand, Wise and Sloboda (2008) found that auditory feedback improved the performances of both "tone-deaf" and "non-tone-deaf" groups when singing familiar songs accompanied by the piano. Finally, Wang et al. (2012) found that the influence of accompanying auditory feedback (a synthesized piano) in song-singing tasks was negative. However, its effect was seen as positive for moderately poor-pitch singers in pitch-matching tasks.

One limitation of previous studies is that they did not study the possible effects of auditory feedback after removing it. Previous studies show that the effect of real-time visual feedback tends to worsen participants' performance in melody production despite improving it in retention conditions (Welch, 1984; Wilson et al., 2005). This effect could also occur with the use of concurrent auditory feedback. On the other hand, and more importantly, participants may be unable to use auditory feedback as KR. One of the reasons we would expect auditory feedback to improve singing skills is because it could be used by participants to recognize that they were not in tune. With the exception of Pardue and McPherson (2019), a common feature of almost all the previous studies was the use of synthesized sounds as auditory feedback. Hutchins and Peretz (2012) suggested that one of the main reasons for poor-pitch singing in their participants was due to a pitch-translation problem. The pitch-translation problem states that participants may not be able to "translate" the pitch from the timbre of the synthesizer to the timbre of their voice. This also leads us to reinterpret the studies that have used realtime visual feedback to improve singing skills: since many of them used synthesizers as reference tones to imitate, it could be argued that what they were really evaluating was the ability of the real-time visual feedback to help participants learn to translate the pitch of the synthesizer's timbre to the timbre of their voice.

Hutchins et al. recorded participants' voices singing different tones and asked them to do a self-matching task. They found improved results when matching their own voice (presumably due to timbral-similarity) but still worse results than when they used a knob-controller. Interestingly, experienced musicians who took part in the experiment were not able to distinguish voice tones differing by 30 cents (compared with the fact that they were able to distinguish synthesized tones with a difference of fewer than 10 cents). According to the authors, these results were due to a "vocal generosity effect." The vocal generosity effect, which was addressed and confirmed in a posterior study (Hutchins et al., 2012), states that a higher degree of mistuning is necessary for listeners, both musicians, and non-musicians, to decide that sung tones were out-of-tune compared with the timbre of other instruments.

Recent work has addressed the effects of self-matching accuracy in melodies (Pfordresher and Mantell, 2014). In their first experiment, they found that participants were more accurate in imitating recorded melodies previously produced by themselves than recorded melodies produced by other participants. In their second experiment, they synthesized the pitch-time trajectories of the recorded melodies using a voicelike tone finding that the self-matching effect was independent of timbre. This self-advantage was also bigger for poor-pitch singers than accurate singers. According to the authors, poorpitch singing is caused by a deficit of inverse modeling during vocal imitation where vocal-pitch patterns of participants are limited to the kinds of patterns they have produced in the past. However, the absolute error scores in the second experiment doubled those of the first experiment. Similar results were also found in previous research suggesting an important humanvoice advantage in pitch imitation (Mantell and Pfordresher, 2013).

Humans, like some other animals (e.g., dolphins, whales, and birds), have the capacity to imitate arbitrary sounds through what has been commonly called auditory-guided vocal learning (Brown et al., 2004; Buccino et al., 2004; Fitch, 2006). Recent views, however, consider vocal learning to be separate phenomena from vocal imitation (Mercado III et al., 2014). Buccino et al. (2004) suggested that, when a motor action is coded in the mirror neuron system, it can be transferred to recombination of the viewed movements to replicate it. Thus, any action already presents in the mirror neuron system could be immediately replicated. Considering the significant exposure to human voices from the birth of any individual, human voices should be easier to replicate than other sounds and not only because of timbral similarity. Actually, Hutchins et al. found that participants from the self-matching task spent less time and required fewer trials than participants from the rest of the tasks. This could mean that participants managed to produce the required tone without hardly any effort. Also, the initial errors in the self-matching condition were much lower than in the slider condition. This implies that, in the slider condition, participants had to start from an almost arbitrary location of the pitch space letting auditory feedback guide their movement to the target note. That is, they did not develop a memory of the location where each pitch had to be found in the slider. However, in the self-matching condition participants seemed to be able to produce the required pitch without the need of starting from any arbitrary location. Both timbral cues and motor imagery may allow participants to recognize pitch due to an implicit/instrument-specific absolute pitch (Pfordresher and Halpern, 2013; Gelding et al., 2015; Reymore and Hansen, 2020).

5.4 Aims

In this study, we aim to evaluate in an experimental setup different modalities of feedback for learning to improve intonation in both the violin and the voice. Complete beginners with no musical experience took part in an experiment where they had to learn to maintain a stable sound with the violin while, additionally, were engaged in a pitch-matching task with their voices or the violin to study the effects of real-time pitch tracking and auditory feedback for this particular type of intonation exercise.

Inspired by the work of Hutchins and Peretz (2012), we designed a new experiment where, instead of using a slider, participants used a real instrument whose results will be compared to those of the voice. Beginners had to learn to translate the pitch from a synthesized pure tone used as a reference to a violin or their voice tone. Participants received help in the form of different types of feedback to improve their intonation skills which were posteriorly evaluated in a retention block. Beginners were randomly distributed into groups: the Control Group (CG) did not receive any type of help to improve their intonation abilities; the Feedback Group (FG) received realtime visual feedback with KR, and the Equal-Timbre Group (ETG) received similar timbre auditory feedback. By studying the retention effects of both the FG and ETG groups we expected to isolate the effects of "external reward" in learning pitch-matching abilities while comparing the effects of auditory feedback in the form of timbre-similarity in different instruments. Finally, we also created an Expert Group (EG) formed by expert violinists. In general terms, in this study we seek to answer the following questions:

1. Does real-time visual feedback improve participant's pitchmatching abilities with a synthesized tone for both violin and singing voice?

2. How does timbre-similarity affect pitch-matching abilities in violin and singing voice?

3. Does timbre-similarity help participants learn how to translate the pitch from a synthesized sound to that of their voice or instrument?

4. How do real-time visual feedback and timbre-similarity affect participants' retention scores?

We expected that real-time visual feedback would positively impact the results of the FG as has already been shown in previous research. However, previous research tended to measure improvements only in pitch accuracy (that is, the error in cents from the desired notes). As in Hutchins and Peretz (2012), we decided to also collect the number of correct notes by considering a note correct if it is within 50 cents of the target pitch.

If timbre-similarity and imitation skills influence the results of pitch-matching tasks, we would expect it would be easier for participants from the ETG to imitate human voice pitches than violin pitches. We would also expect that ETG participants would find voice pitches faster than FG participants.

We followed some of the methodological procedures proposed during the Seattle International Singing Research Symposium (Demorest et al., 2015). All the data used in the current study (raw data, wav files, and statistics) are publicly available in Zenodo (Blanco et al., 2020, 2021b).

5.5 Methods

5.5.1 Participants

Fifty-seven participants with no prior violin playing experience and no musical experience with other instruments (34 female and 23 male) were recruited from the Pompeu Fabra University campus to participate in the study. In addition, 15 expert violinists [EG; 8 women, 7 men; mean age: 32.4 (10.06); mean years experience: 18.6 (5.53)] were recruited from both the university campus and different music schools and conservatories in Barcelona. Participants conceded their written consent and procedures were approved by the Conservatoires UK Research Ethics committee on 04/04/2017, following the guidelines of the British Psychological Society.

Participants filled a questionnaire about their musical skills, main instrument, and years of music training. They also performed a pitch discrimination task (PDT) before (pre) and after (post) the experiment (Musicianbrain, 2021). Those participants who got pitch discrimination thresholds above 18 Hz in both pre and post-tests were asked to realize the Brams Online Test for musical abilities (Peretz et al., 2008). Those participants who got scores below 70% in both the first and third sections of the Brams test were labeled as “possible amusics” and removed from the experiment. We discarded one participant who reported after the experiment being unable to take pleasure in music. She also failed the first block from the Amusia test. Finally, we also discarded all the participants who reported having played a musical instrument for more than 1 year.

Beginner participants were randomly divided into three different experimental groups: the Feedback Group [FG; 9 female, 6 male; mean age: 27.93 (4.33)], the Control Group [CG; 11 female, 4 male; mean age: 27.83 (4.95)], and the Equal Timbre Group [ETG; 10 female, 7 male; mean age: 30.76(8.3)]. The study was carried out in the recording studio located in the Information and Communication Technologies Engineering (ETIC) department of the Universitat Pompeu Fabra, Barcelona.

5.5.2 Materials

5.5.2.1 Learning Materials

Before the experiment, basic information about violin playing techniques like stance, violin position, bow position, and grip was delivered to the beginner participants through a 6-min didactic Youtube video of a professional violinist[4]. The video covered violin technique aspects such as bow-string *contact point*, bow speed-force relationship, and bow angle. The video included an example of how to perform full bow exercises (alternation of up and down movements using the full length of the bow) focusing attention on bowing parallel to the bridge and how to move the wrist of the right hand to achieve a straight bow movement. The experts visualized only the last part of the explanation to make sure they understood the task.

5.5.2.2 Providing Visual Feedback with SkyNote

SkyNote, the system we used to deliver real-time feedback to participants, is one of the main outcomes of the TELMI Project. SkyNote is an integrated system that combines different technologies for real-time feedback on pitch, intonation, dynamics (Mayor et al., 2009), kinematics (Vamvakousis et al., 2018), and tone quality (Giraldo et al., 2019). This feedback can be displayed in customized widgets or directly on the musical score, allowing for real-time experimentation and overall

[4] https://youtu.be/mUz8fIc1FaY.

performance evaluation. For this experiment, we presented feedback of a single performance aspect at a time. In Figure 1, you can see the display used for the real-time feedback on pitch, intonation, and dynamics. The target note appears in the yellow bar on the screen while the performed note is represented by a green line (for more details see Mayor et al., 2009). Five different musical notes were reproduced triggering a reference synthetic sound which consisted of a pure tone at the following frequencies: D#4(311.13 Hz), E4(329.63 Hz), F4(349.23 Hz), F#4(369.99 Hz), G4(392.00 Hz) for most female participants and an octave below for most male participants. Different octaves were chosen if needed to fit the vocal range of participants independently of their gender.

We used a condenser microphone (Behringer-C3, 2013) to record the audio during the session, a NUC computer to run SkyNote, and two screens, one to deliver feedback to the participant and the other one for the experimenter.

5.5.3 Experimental Procedure

Before starting the experiment all groups of participants took part in a practice session monitored by the experimenter. In that practice session, they were instructed on violin technique, bow position, stance, and bow grip through a Youtube video which explained some of the most important concepts required to perform the full bow exercises correctly together with audiovisual examples. A full bow exercise consisted of the alternation of two up and down bowing movements using the full length of the bow with the goal of producing a stable and clear sound. Participants could play while watching the video and explore creating sound with the violin. They could also rewatch different parts of the video while practicing. Participants were informed orally by the experimenter about the violin technique aspects to take into account. These variables were also explained in the Youtube video: *bow skewness* (bowing parallel to the bridge), *contact point* (measured as bow-bridge distance), *inclination* (taking care of not playing the other strings during the movement with the bow), *pitch stability* (related to avoiding scratchy sounds), and *dynamic stability* (trying to maintain the energy of the sound stable during the whole exercise, even during uptodown or down-to-up changes). More details about how these variables were computed can be found in Blanco et al. (2021a). They were also encouraged to explore how the produced pitch changes when they move their finger down the fingerboard. The experimenter verified that all participants were able to perform this task correctly before continuing with the experiment. The duration of this practice session was around 16 min (6 min video + 10 min practice). In order, to find the vocal range for each participant, they were asked to perform some singing warm-up exercises such as sustaining a single comfortable pitch for several seconds. Participants were also asked to make a sweep from the lowest note they could produce to the highest one and another sweep from their highest note to their lower note to ensure that their range covered the space of all the target notes.

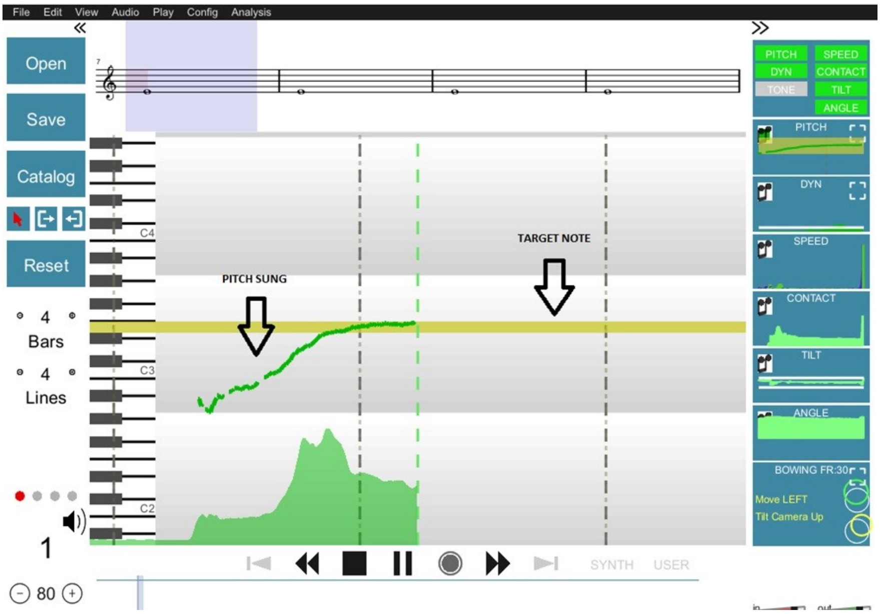

Figure 1: Visual display of the tool used to offer real-time feedback of pitch production. The target note, which in this case is an E3, is represented on the screen as a yellow bar. The produced pitch by the participant is drawn in green at the center of the screen at the moment of the production and displaced to the left of the screen over time.

The experiment consisted of three blocks: *Baseline*, *Acquisition*, and *Transfer*. The *Baseline* was equal for all groups of participants and consisted of a pitch-matching exercise with five different notes alternating between the *violin* and the *voice* condition. At the beginning of the *Baseline* block, one of five target notes was produced with the reference synthetic sound for five seconds (synth-matching task) while participants were not allowed to play or sing. In the *violin* condition, participants had to locate in the fingerboard of the D string the target note by displacing their index finger across the fingerboard while producing sound with the bow. Participants could start from any location on the fingerboard. Once the note was located, they were asked to perform a full bow exercise on that specific location. The note was then reproduced again for five seconds giving participants the possibility of changing their decision. Whether they decided to change or not, they had to perform another full bow exercise. In the *voice* condition participants repeated the same procedure described above but using their voice. The *voice* and *violin* conditions were alternated in random order for each note. That is, sometimes starting with the voice and sometimes starting with the violin.

After the *Baseline* block and before the *Acquisition* block, participants rewatched the instructional video and remained about the main variables that will be used to evaluate their performance. They were also instructed about the procedure of the *Acquisition* block. Real-time visual feedback was presented and explained to the FG. Participants rested around 5 min in between blocks.

As in the *Baseline* block, the *Acquisition* block also consisted in a synth-matching exercise with the same five notes presented in the *Baseline* and an alternation between the violin and the voice. First, participants tried to match the corresponding pitch in a synth-matching task with two attempts just like in the *Baseline*. This was called the *Acquisition pre-Aid* condition. After the second attempt, the *Acquisition Aid* condition started. The *Acquisition Aid* condition differed between the three different groups of beginners. Visual feedback was provided to the FG on how far their performed note was from the target note. Using the feedback participants in the FG could modify their performed notes and hear the reference synthetic sound as many times as they needed. After that, feedback was removed and participants performed a synth-matching task for the same note (*Acquisition post-Aid*). On the other hand, in their *Acquisition Aid* condition, the ETG was able to modify the performed note with matchedtimbre auditory feedback of the corresponding target note. Participants were allowed to request both the recordings and the reference synthetic sounds as many times as they needed (either because they were satisfied or decided to give up). CG participants only had the option to hear the reference synthetic sound and change their performed note, as many times as needed. Following this, both the ETG and CG repeated a last synthmatching task for the same note (*Acquisition post-Aid*). Finally, the experts received real-time visual feedback in the *Acquisition Aid* condition like the FG and performed a last synth-matching task. The reason for that is because we wanted them to fill a questionnaire with their opinion about SkyNote at the end of the experiment (see Blanco et al., 2021a for the answers to the questionnaires). As in the rest of the conditions, participants were not allowed to play or sing during any type of sound reproduction in the *Acquisition-Aid* condition. Summarizing, we can divide the *Acquisition* block into three different conditions: the *Acquisition pre-Aid*, the *Acquisition Aid*, and the *Acquisition post-Aid* (see Figure 2).

After the *Acquisition* block participants rested for 5 min. Then, the *Transfer* condition started. The *Transfer* condition was the same as the *Baseline* condition but with a different order of notes and alternations between the violin and the voice condition (see Figure 2). The order of the notes and alternations between the violin and voice condition was randomized as in the other conditions.

After the *Acquisition post-Aid* condition with the violin half of the participants chosen randomly performed one full bow exercise while receiving real-time feedback about sound quality and one more full bow exercise while receiving real-time feedback on bow kinematics. The other half were also asked to perform the two full bow exercises in a row. In the first one, they were explicitly asked to pay attention to the sound quality feedback and in the second one to the kinematic feedback. These results are presented in Blanco et al. (2021a).

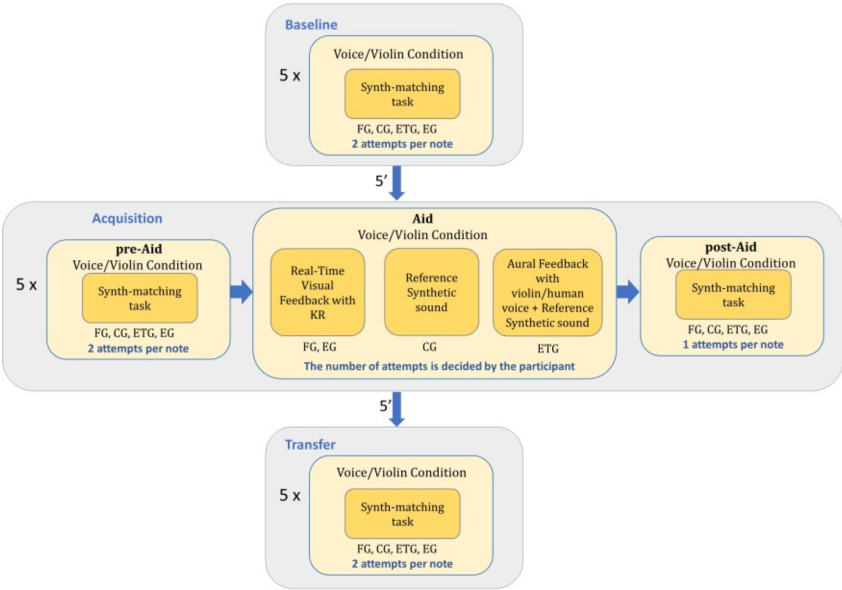

Figure 2: Diagram with the different blocks of the experiment and the different conditions each group of participants went through.

5.6 Intonation Analysis

The Tony software (Mauch et al., 2015) was used to extract information from pitch accuracy from the raw audio of violin and voice exercises. However, it was necessary to visually inspect all the events to ensure the correct operation of the pitch detection algorithm. This data was posteriorly processed in Matlab (MATLAB, 2010) and analyzed in Spss (IBM, 2011).

5.6.1 Pitch Detection

The audios for each condition were recorded in a .wav file by the software at a sample rate of 44,100 Hz. The Tony software was used to extract the pitch of all performed (violin and singing voice) notes (Mauch et al., 2015). Tony is based on the pYIN method for automatic pitch estimation and note tracking (Mauch and Dixon, 2014) together with custom methods for interactive reestimation. It outputs discrete notes on a continuous pitch scale based on the Viterbi-decoding of an independent Hidden Markov Model. This method is particularly robust to small and short pitch variations. If one variation is big and long enough, like in the possible case of one participant accidentally hitting another string, the Tony software considers that a pitch transition occurred, and returns two different pitch estimations separated in time. A visual inspection of all the events was, thus, also necessary to ensure correct pitch extraction.

The pitch performed by each participant was converted to cents using the target note as a reference. To avoid octave errors, those sung or played pitches with a value >+600 or lower than -600 cents were recomputed to a different octave. Finally, we computed the absolute value of the errors. We also considered the number of correct notes, that is, those pitches sung with an error of <50 cents (half semitone).

5.6.2 Violin Technique Analysis

Before starting with the intonation analysis we evaluated whether violin technique could have exerted an influence on the pitchmatching skills of our beginner participants. For that purpose, we looked for possible correlations between beginners' average absolute pitch errors and their technique (both in terms of sound quality and gestures). For sound quality we computed two descriptors which have been proven to be useful in previous research: *dynamic stability* and *pitch stability* (Romaní et al., 2015; Blanco and Ramirez, 2019; Giraldo et al., 2019; Blanco et al., 2021a). We also evaluated the participants' gestural technique using one kinematic descriptor: *bow skewness*. This descriptor represents the angle of the bow with respect to the violin bridge (the closer to zero the better) which is considered to be a common prerequisite to achieve a good sound. More information about how those descriptors were computed can be found in Blanco et al. (2021a). We also evaluated for possible differences between groups in their performance across blocks with one mixed-design 3 × 4 with Group (CG, ETG, FG) as between-subject factor, and Condition (*Baseline*, *Acquisition pre-Aid*, *Acquisition postAid*, and *Transfer*) as within-subject factor. Finally, to ensure that the amount of improvement was not significantly bigger for one group than for the others, we performed three independent t-tests of the relative difference between the Transfer and the Baseline for each one of the descriptors applying Bonferroni correction.

5.6.3 Behavioral Analysis at the Baseline

We evaluated the behavior of beginner participants in the *Baseline* condition and compared it with the experts. To ensure that participants were trying to match the target pitches we compared their average error in cents over the five n otes. Both for the violin and the voice. We also verified if there was a correlation between the frequency of the target notes and the frequency of the produced notes. This helped us to evaluate whether a target pitch was higher than the previous one, the direction of the produced pitch was also higher compared with the previous one. We also evaluated the possibility that some notes could be more difficult to match than others. We performed two 2×2 repeated measures analyses with Instrument (violin, voice) and Note as within-subject factors for beginners and experts. Posteriorly, we evaluated if participants tended to correct their errors in the correct direction in their second attempt.

We performed two more 2×2 repeated measures analysis with Instrument (violin, voice) and Attempts (first and second attempt).

Finally, we also studied if there was any significant trend to flat or sharp notes in the direction of the errors of both beginners and experts when playing violin or singing. For that purpose, we performed four one-sample *t*-tests.

5.6.4 Analysis of the Effects of Feedback

Finally, we performed four more different analyses of the data. One for the error in cents, another one for the number of correct notes, another one for the time in seconds they spent in *Acquisition post-Aid* and finally, another one for the number of times participants from the ETG and CG requested auditory feedback in *Acquisition Aid*.

To study the impact of the different types of feedback in each modality we performed for each analysis one 4x (4 × 2) mixed-design with Group (CG, ETG, FG, and the experts) as between-subject factor, and Condition (*Baseline*, *Acquisition pre-Aid*, *Acquisition post-Aid*, and *Transfer*) and Instrument (violin and voice) as within-subject factors. *Post-hoc* tests using the Tukey method for multiple comparisons were performed between the groups of participants to compare their results. We also performed a 4 × 2 mixed-design with Group as betweensubject factor and Instrument as the within-subject factor for the analysis of duration and a 2 × 2 mixed-design with Group (ETG and CG) and Instrument for the number of feedback requests. For those analyses that showed a significant interaction between Condition and Group, a posterior simple main effect analysis was performed on each group to find out which conditions were causing the interaction. Pairwise comparisons tests were performed between the conditions using the Bonferroni correction.

Finally, we removed two participants from the FG and two more from the expert group because they were labeled as outliers. After removing the outliers all the data passed the assumptions of normality required to perform the tests. All the results presented in the following sections were Greenhouse-Geisser corrected.

5.7 Results

5.7.1 Sound Quality and Bow Technique with the Violin

All groups of participants experienced improvements in violin technique in all the measured descriptors through the different blocks of the experiment. The mixed-analysis showed a significant effect of Condition for *pitch stability*, $F(1.97) = 16.27$, $p < 0.0001$, $\eta 2 = 0.34$, for *dynamic stability*, $F(2.79) = 5.17$, $p = 0.003$, $\eta 2 = 0.14$, and for *bow skewness* $F(2.21) = 5.10$, $p = 0.007$, $\eta 2 = 0.14$. No significant Condition*Group interaction was found. No significant differences were found in the relative amount of improvement at the end of the session between groups. Finally, we did not found significant correlations between the absolute error in cents beginners made with the violin in each condition with the value of each one of the three descriptors used to measure violin technique (*pitch stability*, *dynamic stability*, *bow skewness*).

5.7.2 Behavioral Results at the Baseline

We found significant correlations at the *Baseline* condition between the tone that beginners produced and the target tones of the experiment when using the voice, $R2 = 0.318$, $p < 0.0001$ and the violin, $R2 = 0.47$, $p < 0.0001$ (see Figure 3A). As expected, the experts showed stronger correlations between their produced tone and the target tones of the experiment for the voice, $R2 = 0.98$, $p < 0.0001$ and for the violin, $R2 = 0.99$, $p < 0.0001$. Beginners showed an absolute average error of 150 cents (SD = 12.56) with the violin. Errors for the voice tended to be bigger than for the violin (see Figure 3B). On average beginners showed an error with the voice of 271 cents (SD = 23.83). The mixed analysis for beginners showed a significant effect of Instrument (voice > violin), $F(1) = 20.37$, $p < 0.0001$, $\eta2 = 0.35$. No significant effects of Note nor an Instrument*Note interaction were found. Experts showed an absolute average error of 18 cents (SD = 1.59) with the violin and an absolute average error of 26.82 cents (SD = 2.98) for the voice. The mixed analysis for experts showed a significant effect of Instrument (voice > violin), $F(1) = 5.93$, $p < 0.032$, $\eta2 = 0.33$. No significant effects of Note nor an Instrument*Note interaction were found.
Beginners tended to improve their accuracy in the second attempt when compared with the accuracy of their first attempt by 11.88 cents (SD = 5.63). The repeated measures analysis showed a significant effect of Instrument (voice > violin), $F(1) = 18.73$, $p < 0.0001$, $\eta2 = 0.35$, and of Attempts (second < first), $F(1.00) = 4.43$, $p < 0.041$, $\eta2 = 0.09$. No significant Instrument*Attempts interaction was found. Experts did not show any significant differences between Attempts neither at Instrument.

Finally, beginners did not show any tendency toward sharp or flat errors neither in the violin nor in the voice as revealed by the one-sample *t*-tests. On the other hand, experts did show a flat trend both with the violin, $t(12) = -7.5$, $p < 0.0001$, and with the voice, $t(12) = -6.75$, $p < 0.0001$. On average, the error in cents of the experts while using the violin was -16 cents (SD = 7.69) and -22 cents (SD = 12.15) when using the voice.

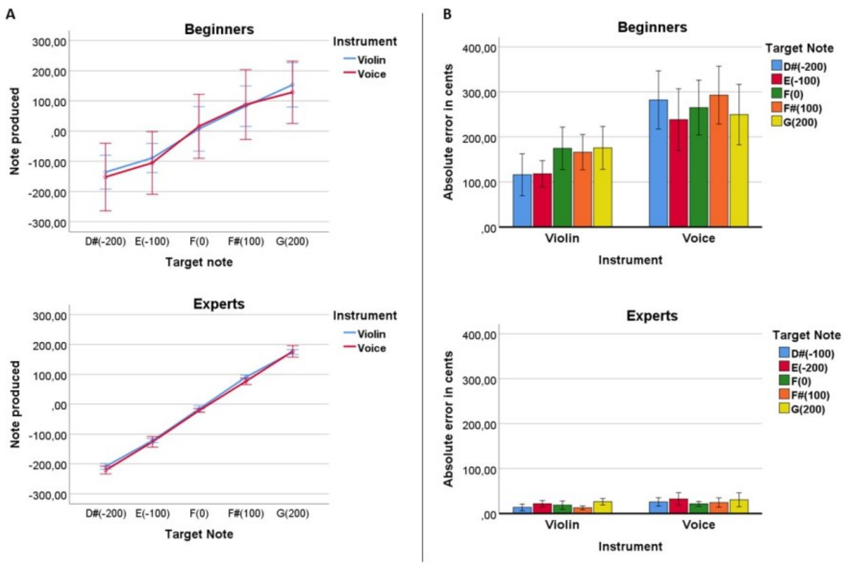

Figure 3: A: Comparison between the target and produced note of beginners and experts both with the violin and with the voice. **B**. Absolute error in cents for each target note. Errors produced with the voice were significantly bigger than errors produced with the violin both for beginners and experts. No significant differences were found between the error produced at each note.}

5.7.3 Pitch Matching Across Blocks

Results showed that all groups of participants showed larger errors for the voice than for the violin. As expected, the error in cents of the experts was on average much lower (M = 18.39, SD = 18.18 cents with the violin and M = 34.46, SD = 29.16 cents with the voice) than the rest of the groups of beginners (M = 135, SD = 17 cents with the violin and M = 217.33, SD = 27.32 cents with the voice). See Figure 4 for a summary of the results. *Posthoc* tests showed significant differences between the experts with the rest of the groups (experts < CG, $p < 0.0001$; experts < ETG, $p < 0.0001$; experts < FG, $p = 0.005$). Also, results from the FG differed significantly from both the ETG and the CG (FG < ETG, $p = 0.006$; FG < CG, $p < 0.0001$).

The FG improved their results through the different conditions at both the violin and the voice (see Figure 4). The ETG showed different behavior for the voice than for the violin at the *Acquisition post-Aid*. We found that, on average, the error with the voice decreased 120.80 cents (SD = 131.56) compared to the *Baseline* when the ETG received aid in the form of a human voice. That decrease was not seen in the violin condition. The Univariate tests of within-subject effects for error in cents showed a significant effect of Instrument (voice > violin), $F(1) = 16.20$, $p < 0.0001$, $\eta 2 = 0.23$, and an Instrument*Group interaction, $F(3) = 2.93$, $p = 0.041$, $\eta 2 = 0.14$. Also significant effects of Condition, $F(2.51) = 20.6$, $p < 0.0001$, $\eta 2 = 0.28$, and a Condition*Group interaction, $F(7.54) = 7.59$, $p < 0.0001$, $\eta 2 = 0.29$. We also found an Instrument*Condition*Group, $F(6.34) = 3.03$, $p < 0.008$, $\eta 2 = 0.144$. We did not found an Instrument*Condition interaction.

The repeated measures for each group revealed a significant effect of Condition in the univariate tests of within-subject effects for the ETG, $F(2.72) = 4.17$, $p = 0.019$, $\eta 2 = 0.20$, and a Condition*Instrument interaction, $F(2.53) = 6.69$, $p = 0.001$, $\eta 2 = 0.28$. Pairwise comparisons tests showed significant results for the ETG between the *Baseline* and the *Acquisition post-Aid* conditions for voice (*Baseline* > *Acquisition post-Aid*, $p = 0.008$). The CG showed only a significant effect of Instrument (voice > violin), $F(1) = 12.03$, $p = 0.004$, $\eta 2 = 0.48$. Finally, the FG showed a significant effect of Condition, $F(1.9) = 18.73$, $p < 0.0001$, $\eta 2 = 0.748$. Pairwise comparisons for the violin showed significant results for the FG between both the *Baseline* and the *Acquisition post-Aid* and *Transfer* conditions (*Baseline* > *Acquisition postAid*; *Baseline* > *Transfer*, $p < 0.0001$ for both tests). Pairwise comparisons for the voice showed significant results for the FG between the *Baseline* and the rest of the conditions: the *Acquisition pre-Aid*, *Acquisition post-Aid*, and *Transfer* (*Baseline* > *Acquisition pre-Aid*, $p = 0.048$; *Baseline* > *Acquisition post-Aid*, $p < 0.0001$; *Baseline* < *Transfer*, $p = 0.045$).

The FG produced better results than the ETG at the *Acquisition post-Aid* condition. The FG showed an average error of 14.73 cents (SD = 2) for the violin and an average error of

20.45 cents (SD = 2.27) for the voice. On the other hand, the ETG showed an average error of 169.24 cents (SD = 31) for the violin and an average error of 145.36 cents (SD = 32.33) for the voice. Independent samples *t*-test showed significant effects at the *Acquisition post-Aid* condition between the FG and ETG for the violin (FG < ETG), $t(29) = 4.17$, $p < 0.0001$, and for the voice (FG < ETG), $t(29) = 3.26$, $p = 0.003$. No significant effects were found in a paired samples t-test between the voice and violin condition for the ETG.

The FG also improved their results more than the ETG at the *Transfer* condition in relation to the *Baseline* (*Transfer - Baseline*). The FG showed an average improvement of 81.17 cents (SD = 51.9) for the violin and an average improvement of 105.31 cents (SD = 168.66) for the voice. On the other hand, the ETG showed an average improvement of 24 cents (SD = 123.4) for the violin and no improvement for the voice (M = -11, SD = 15). Independent samples *t*-test showed significant differences at the degree of improvement between the FG and ETG for the violin (FG < ETG), $t(29) = -3.29$, $p < 0.003$, and for the voice (FG < ETG), $t(29) = -2.41$, $p = 0.022$.

5.7.4 Correct Notes Across Blocks

Despite the main differences found in accuracy between the voice and the violin across blocks, we did not find big differences regarding the number of correct notes. On average, beginners made in the *Baseline* an average number of 1.02 correct notes (SD = 1.23) with the voice and an average number of 1.08 correct notes (SD = 1.04) with the violin. On the other hand, experts made an average number of 4.06 correct notes (SD = 1.48) with the voice and an average number of 4.53 correct notes (SD = 1.3) with the violin. We did not find significant effects of Instrument or Instrument*Group or Instrument*Condition interaction.

The main difference compared with accuracy results across was seen in participants of the ETG. We found an improvement in the number of correct notes at the *Acquisition post-Aid* not

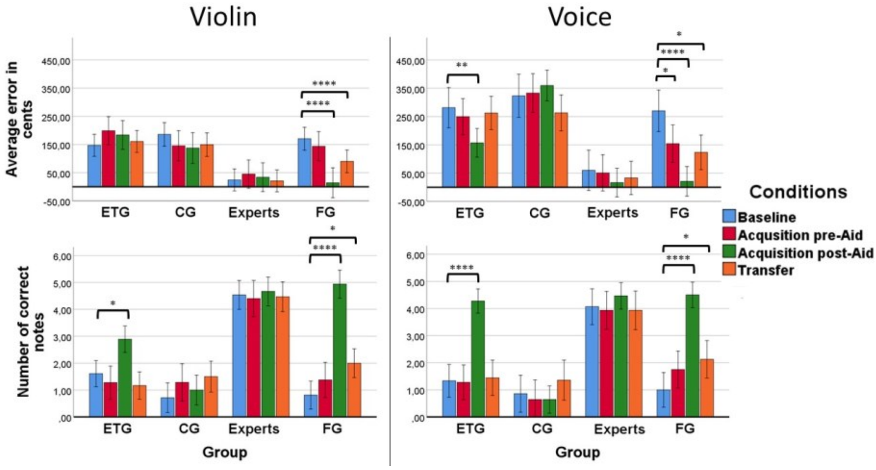

Figure 4: (Upper left) Results of pitch-matching accuracy in the violin condition for each group of participants. Only the FG improved significantly their results compared with the *Baseline* at both the *Acquisition post-Aid* and *Transfer*. (Upper right) Results of pitch-matching accuracy in the voice condition for each group of participants. The FG improved significantly their results at the *Acquisition pre-Aid*, *Acquisition post-Aid*, and *Transfer* conditions. The ETG improved significantly their results at the *Acquisition post-Aid*. However, that improvement was not retained to the rest of the following conditions. (Inferior left) Results of the number of correct notes in the violin condition for each group of participants. The FG improved significantly their results compared with the *Baseline* at both the *Acquisition post-Aid* and *Transfer*. The ETG improved significantly their results only at the *Acquisition post-Aid*. (Inferior right) Results of the number of correct notes in the voice condition for each group of participants. As in the violin condition, the FG improved significantly their results compared with the *Baseline* at both the *Acquisition post-Aid* and *Transfer* while the ETG improved significantly their results only at the *Acquisition post-Aid*. (*$p <= 0.05$, **$p <= 0.01$, ***$p <= 0.001$, ****$p <= 0.0001$).

only for the voice (M = 4.28, SD = 0.24) but also for the violin (M = 2.83, SD = 0.33). The improvement for singing voice, however, was still bigger than the improvement for violin and resembled those of the experts (M = 4.46, SD = 0.18) or the FG (M = 4.38, SD = 0.26) at the moment of receiving real-time visual feedback. Univariate tests of within-subjects effects for correct notes showed a significant effect of Condition, $F(2.38) = 56.51$, $p < 0.0001$, $\eta 2 = 0.51$, a Condition*Group interaction, $F(7.15) = 18.24$, $p < 0.0001$, $\eta 2 = 0.50$, and an Instrument*Condition*Group interaction, $F(8.49) = 3.30$, $p = 0.001$, $\eta 2 = 0.155$. The repeated measures for each group revealed a significant effect of Condition in the univariate test of within-subject effects for the ETG, $F(1.85) = 54.58$, $p < 0.0001$, $\eta 2 = 0.76$, and interaction between Condition*Instrument, $F(2.32) = 7.46$, $p = 0.001$, $\eta 2 = 0.355$. Pairwise comparisons tests showed significant results between the *Baseline* and the *Acquisition post-Aid* conditions for singing voice (*Baseline* < *Acquisition post-Aid*, $p < 0.0001$) and for violin (*Baseline* < *Acquisition post-Aid*, $p = 0.038$). The FG showed a significant effect of Condition, $F(2.11) = 63.28$, $p < 0.0001$, $\eta 2 = 0.841$. Pairwise comparisons for the violin showed significant results between both the *Baseline* and the *Acquisition post-Aid* and *Transfer* conditions (*Baseline* < *Acquisition post-Aid*, $p < 0.0001$; *Baseline* < *Transfer*, $p = 0.036$). We also found a significant effect of Instrument for the experts (violin < voice), $F(1) = 6.09$, $p = 0.03$, $\eta 2 = 0.337$.

The FG improved their results more than the ETG at the *Acquisition post-Aid* condition for the violin condition (difference: M = 2.09, SD = 0.39) although their results in singing voice were similar (difference: M = 0.1, SD = 0.36). Independent samples *t*-test showed significant effects between the FG and ETG for the violin at the *Acquisition post-Aid* (FG > ETG), $t(29) = -5.34$, $p < 0.0001$. No significant differences were found for the voice. The ETG improved more in the *Acquisition post-Aid* with the voice than with the violin (difference: M = 1.4, SD = 1.75). A paired samples t-test showed significant results for the ETG between singing voice and violin in the *Acquisition post-Aid* (voice > violin), $t(17) = -3.48$, $p = 0.003$).

The FG improved their results more than the ETG at the *Transfer* condition in relation to the *Baseline* (*Transfer* - *Baseline*). The FG showed an average improvement of 1.38 correct notes (SD = 1.5) for the violin and an average improvement of 0.92 correct notes (SD = 1.03) for the voice. On the other hand, the ETG showed no improvement for the violin (M = -0.38, SD = 1.57) and an average improvement of 0.16 correct notes (SD = 1.09) for the voice. Independent samples *t*-test only showed significant differences at the degree of improvement between the FG and ETG for the violin (FG > ETG), $t(29) = -2.31$, $p = 0.028$.

5.7.5 Duration of Acquisition Aid

Results for duration showed how participants from the ETG tended, on average, to spend more time trying to match the target notes with the violin than with the voice in the *Acquisition Aid* condition (M = 11.31, SD = 6 s more, see Figure 5A). On the contrary, the experts seemed to spend slightly more time with the voice than with the violin (M = 1.6, SD = 2.25 s more). Univariate tests of within-subject effects showed a significant effect of Instrument (violin > voice), $F(1) = 13.31$, $p = 0.001$, $\eta2 = 0.172$, and an Instrument*Group interaction, $F(3) = 7.25$, $p < 0.0001$, $\eta2 = 0.254$. Simple main effect analysis revealed a significant effect of Instrument in the univariate tests of withinsubject effects for the experts (voice>violin), $F(1) = 18.25$, $p = 0.013$, $\eta2 = 0.36$, and the ETG (violin > voice), $F(1) = 67.29$, $p < 0.0001$, $\eta2 = 0.77$. No significant effect of Instrument was found for the FG or the CG.

Participants from the ETG tended to spend less time with the voice than participants from the FG (M = 9.4, SD = 9.29 s less). Those big differences were not seen in the violin condition where the ETG spent only 1.61 (SD = 9.13) seconds less than the FG. Independent samples *t*-test showed significant effects between the FG and ETG for the voice $t(36) = 2.93$, $p = 0.004$. No significant differences were found for the violin.

Results for the number of requests at the *Acquisition Aid* condition showed how participants from the ETG tended, on average, to request more times the auditory feedback at the violin condition than at the voice condition (M = 0.92, SD = 0.51 times more, see Figure 5B). On the other hand, participants from the CG barely requested more feedback from the synthesizer either at the violin or at the voice condition. Univariate tests of within-subject effects showed a significant effect of Instrument (violin > voice), $F(1) = 26.67$, $p < 0.0001$, $\eta2 = 0.44$, and an Instrument*Group interaction, $F(1) = 34.21$, $p < 0.0001$, $\eta2 = 0.50$. Simple main effect analysis revealed a significant effect of Instrument in the univariate tests of within-subject effects for the ETG (violin>voice), $F(1) = 52.54$, $p < 0.0001$, $\eta2 = 0.71$. Finally, no significant correlations were found between the time spent or the number of attempts in *Acquisition Aid* with the error in cents at the *Transfer* condition. Neither for the violin nor for the voice.

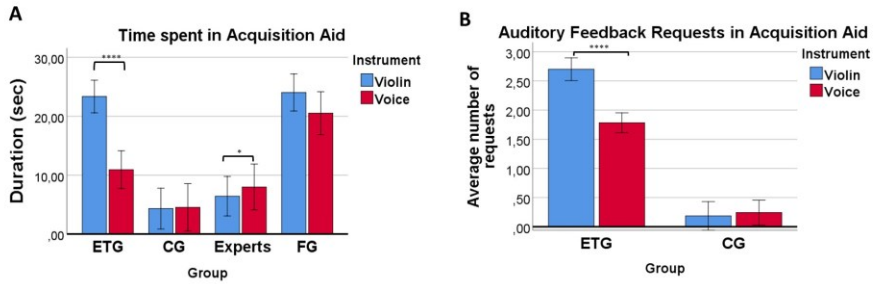

Figure 5: A: Duration of the \textit{Acquisition Aid} for each group. Participants from the FG and ETG spent more time than the control group trying to match the required pitch probably because of the aid they were receiving. The ETG spent significantly less time trying to match pitches with the voice than with the violin. **B:** Number of times participants requested auditory feedback. Participants from the ETG tended to request more times the auditory feedback at the violin condition than at the voice condition. On the other hand, participants from the CG tended to not request more feedback from the synthesizer.

5.8 Discussion

In this experiment, we have evaluated the use of different feedback modalities for learning intonation skills. We have compared real-time visual feedback and auditory feedback with a similar timbre for both violin and singing voice.

First of all, we needed to ensure that beginners were capable of controlling the pitch with the violin and, in any case, that their inability to do so was not an impediment to find and reproduce the demanded note in the fingerboard. Through the use of sound quality and bow kinematics descriptors validated in previous studies (Romaní et al., 2015; Giraldo et al., 2019; Blanco et al., 2021a), we investigated correlations between beginners' violin technique and intonation accuracy at the different conditions. No correlation was found. We also found that all groups of beginners improved their violin technique throughout the experiment in a similar way regardless to which group they belonged. Finally, we also found that beginners obtained lower absolute intonation errors in cents with the violin than with the voice. Even though beginners spent more time before the experiment learning violin technique than voice technique they learned the minimum necessary to control pitch production and realize sweeps through the fingerboard (which was the same technique required for their voice). That seems to reject the idea that differences in violin and singing voice technical skills may have influenced the intonation results.

Participants in this study produced an intonation error with the voice greater than the found in previous studies evaluating pitch-matching skills (Dalla Bella et al., 2007; Pfordresher and Brown, 2007; Wise and Sloboda, 2008; Pfordresher et al., 2010; Hutchins and Peretz, 2012; Berkowska and Dalla Bella, 2013). Hutchins and Peretz (2012), for example, reported an average singing absolute intonation error of 129 cents for non-musicians while in this study the absolute error was 217 cents. We believe that the choice of a pure tone synthesized sound used in this study may have caused a larger error, while the voice-like synthesizer sound used in the literature may reduce the intonation errors. The fact that the error with the violin was less than with the voice could be due to a possible greater timbral similarity of the violin with the synthesizer. Expert violinists also tended to be more accurate with the violin than with the voice, however, this could be due to increased experience with the instrument. On the other hand, Hutchins and Peretz (2012) found that the vocal tones tended to be matched with less accuracy than the tones produced with the slider, which could also have influenced the results of the experts. In Hutchins and Peretz (2012), musicians showed an average error of 2 cents with the slider and 17 cents with the voice while in this experiment experts showed an average error of 18 cents with the violin and 26.8 cents with the voice.

Although the accuracy with the violin was higher than with the voice, we did not find differences between the number of correct notes produced by each modality that oscillated between one and two correct notes (out of five) in the Baseline. Even assuming that the participants gave the same note in each attempt we could find similar results if that note was located between two of the target notes. To make sure that the participants were trying to hit the notes and not producing the same in each attempt, we calculated if there was some kind of correlation between the target notes and the produced notes. We found significant correlations for both the violin and the voice. Furthermore, we did not find significant differences in the absolute error of each note neither with the violin nor with the voice. This suggests that the frequency of the note was not a factor influencing the intonation accuracy. Interestingly, experts tended to show a trend to make flat errors both with the violin and the voice that was not seen in beginners. This contrasts with the results reported by Hutchins and Peretz (2012) where results with the slider showed no trend to flat or sharp error neither in musicians and non-musicians whereas results with the voice showed a trend to flat errors in both groups. Interestingly, Pfordresher and Brown (2007) did not report any prevailing tendency toward flat or sharp singing among poorpitch singers. Although our participants might not be considered poor-pitch singers, the difficulty of the task (matching a puretone) may have led to similar behaviors. In relation to the experts, we hypothesize that pure tones could be perceived slightly flat compared with the timbre of the violin, although we cannot offer evidence of this fact.

Once the behavior of the participants in the Baseline condition had been studied, we proceeded to study their behavior in the rest of the blocks of the experiment and what possible influences the feedback received could exert on them.
Both FG and ETG improved significantly their results with both the violin and the voice when received visual or aural feedback at the *Acquisition post-Aid*. However, the FG was the only group that showed retention at both the *Acquisition preAid* and the *Transfer* conditions at both modalities. As expected, neither the CG nor the experts improved among the session (in the case of the experts because their errors were minimal). Despite the average error in cents at the *Acquisition postAid* for the voice was larger in the ETG than in the FG, the number of correct notes did not differ between groups (around four correct notes, which is also the average of correct notes of the experts). We can also see an improvement in the number of correct notes at the *Acquisition post-Aid* for the violin in the ETG compared with the rest of the conditions (around

three correct notes out of five), although not as high as with the voice.

The fact that the CG participants were not able to match the notes, even after having the opportunity to try and listen to the synthesized pitch as many times as they wanted, confirms that the lack of trials was not the reason for the poor results beginners showed in their singing abilities. Both ETG and FG were able to match a similar number of notes when receiving help, either in the form of auditory feedback or in the form of visual feedback. However, only the group of participants who received visual feedback seemed to retain their results both at the *Acquisition pre-Aid* and the *Transfer* condition.

The FG spent more time at the *Acquisition Aid* trying to match the pitch than the rest of the groups in the voice condition. This seems to be confounded with the degree of improvement. However, no correlation between the duration or the number of attempts with the produced error at the *Transfer* condition was found neither for the violin nor the voice. A similar effect has been already reported in Hutchins and Peretz (2012). They found that participants tended to spend more time and make more attempts with the slider than with the voice. To determine whether the advantage of the slider condition compared with the voice condition was due to the number of attempts, they required participants to make a minimum number of voice responses comparable to those of the slider. No changes in their voice responses were found across their attempts. As Hutchins and Peretz (2012) suggested, the reason why participants tended to make few responses and spent less time with the voice is probably because participants determined that further responses would not aid their accuracy.

Welch (1984) and Hutchins and Peretz (2012) showed how online visual feedback of the pitch does not improve by itself the pitch-matching results of participants. It seems that there is the need for objective information on the screen regarding how far or close is the produced pitch from the correct result, just as in linear positioning tasks, where subjects had to move an object toward a target out of sight with no time limit. However, our participants were not "tone deaf " and some of the hypothesized reasons why they may not be able to pitch-match our synthesized tone were because of their inability to translate from one timbre to another (Hutchins and Peretz, 2012). By offering them aural feedback of a similar timbre to the one produced, we expected to help them to establish the parameters of the translation by themselves, without the need for an unmistakable sign of "correct" or "incorrect." The fact that on average, participants from the ETG were able to produce a higher number of correct notes at the *Acquisition post-Aid*, highlights that their difficulties were in part based on a pitch-translation problem. Nonetheless, this technique did not seem to help participants to retain the new mapping.

As we mentioned before, timbral similarity had stronger effects for the voice than for the violin. It is possible that participants were more able to be in tune with a human voice due to implicit imitation skills (Buccino et al., 2004; Christiner and Reiterer, 2013). The fact that participants in the ETG were capable of finding the correct pitch in less time than the FG during the *Acquisition Aid* and requested feedback a lesser number of times suggests this was the case. For some participants of the ETG, finding the pitch after hearing the vocal sounds was an almost automatic task done without effort while, for the FG, they had to explore sweeping their voice through the screen until the objective pitch was matched.

Humans are capable to imitate arbitrary sounds thanks to similar mechanisms that operate in other animal species through auditory-guided vocal learning/imitation (Brown et al., 2004; Fitch, 2006). It is known that both auditory memories and aural feedback interact to guide vocal imitation which probably explains why it was easier for participants to imitate human voices than violin sounds. Auditory-guided imitation thus, although helped participants to improve in their task, did not seem to help participants to establish and retain the new mapping between timbers. It is possible that both the objective visual measures and the experience of exploring the pitch space with their voice in an explicit manner, helped participants to understand how they got where they wanted to go, strengthening the schema and favoring retention. Another possibility is that although participants from the ETG were able to perform the task correctly in the *Acquisition post-Aid*, the level of confidence of their chosen answers could have been less as those of the FG, impairing them to learn the mapping between one timbre and the other. Future research could address this issue by offering binary feedback (right/wrong) to some participants in the equal-timbre group after their chosen answers. For now, as already pointed out by Welch (1984), reward seems almost indispensable when learning to translate from one timbre to another.

5.9 Conclusions

In summary, we can list some of the main findings of this study:

1. We found that auditory feedback in the form of timbresimilarity helped more to sing in tune with the voice rather than with the violin. Participants from the ETG also spent less time choosing their answers than participants from the FG at the *Acquisition pre-Aid*. We suggest that implicit imitation skills, above timbre-similarity, may also play an important role in matching the desired pitch.

2. Participants from the ETG were not able to retain their results for the rest of the conditions where auditory feedback was removed.

3. We have revalidated the importance of real-time visual feedback and KR for learning intonation. Participants from the FG were the only ones which improved their results significantly at the *Transfer* condition both with the violin and the voice.

5.10 References

Behringer-C3 (2013). Behringer-C3. Available online at: https://www.behringer.com/behringer/product?modelCode=P0262

Berkowska, M., and Dalla Bella, S. (2013). Uncovering phenotypes of poorpitch singing: the sung performance battery (SPB). *Front. Psychol*. 4:714. doi: 10.3389/fpsyg.2013.00714

Blanco, A., and Ramirez, R. (2019). Evaluation of a sound quality visual feedback system for bow learning technique in violin beginners: an EEG study. *Front. Psychol*. 10:165. doi: 10.3389/fpsyg.2019.00165

Angel David Blanco, Simone Tassani, & Rafael Ramirez. (2020). SkyNote Experiment. Zenodo. http://doi.org/10.5281/zenodo.4553940

Angel David Blanco, Simone Tassani, & Rafael Ramirez. (2021). Effects of Visual and Auditory Feedback in Violin and Singing Voice Pitch Matching Tasks [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4630144

Blanco, A. D., Tassani, S., and Ramirez, R. (2021a). Real-time sound and motion feedback for violin bow technique learning: a controlled, randomized trial. *Front. Psychol*. 12:648479. doi: 10.3389/fpsyg.2021.648479

Brennan, F. (1926). A report of three singing tests given on the tonoscope. *Psychol. Monogr*. 36, 249–262. doi: 10.1037/h0093227

Brown, S., Martinez, M., Hodges, D., Fox, P., and Parsons, L. (2004). The song system of the human brain. *Cogn. Brain Res*. 20, 363–375. doi: 10.1016/j.cogbrainres.2004.03.016

Buccino, G., Binkofski, F., and Riggio, L. (2004). The mirror neuron system and action recognition. *Brain Lang*. 89, 370–376. doi: 10.1016/S0093-934X(03)00356-0

Christiner, M., and Reiterer, S. (2013). Song and speech: examining the link between singing talent and speech imitation ability. *Front. Psychol*. 4:874. doi: 10.3389/fpsyg.2013.00874

Dalla Bella, S., Giguére, J.-F., and Peretz, I. (2007). Singing proficiency in the general population. *J. Acoust. Soc. Am*. 121, 1182–1189. doi: 10.1121/1.2427111

Demorest, S. M., Pfordresher, P. Q., Bella, S. D., Hutchins, S., Loui, P., Rutkowski, J., and Welch, G. F. (2015). Methodological perspectives on singing accuracy: an introduction to the special issue on singing accuracy (part 2). *Music Percept*. 32, 266–271. doi: 10.1525/mp.2015.32.3.266

Fitch, W. (2006). The biology and evolution of music. A comparative perspective. *Cognition* 100, 173–215. doi: 0.1016/j.cognition.2005.11.009

Gelding, R., Thompson, W., and Johnson, B. W. (2015). Auditory imagery and the poor-pitch singer. *PLoS ONE* 10:e0121809. doi: 10.1371/journal.pone.0121809

Giraldo, S., Waddell, F., Nou, I., Ortega, A., Mayor, O., Perez, A., et al. (2019). Automatic assessment of tone quality in violin music performance Front. *Psychol*. 10:334. doi: 10.3389/fpsyg.2019.00334

Hutchins, S., and Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. *J. Exp. Psychol*. 141:76. doi: 10.1037/a0025064

Hutchins, S., Roquet, C., and Peretz, I. (2012). The vocal generosity effect: how bad can your singing be? *Music Percept*. 30, 147–159. doi: 10.1525/mp.2012.30.2.147

IBM Corporation (2011). *IBM SPSS Statistics for Windows, Version 20.0*. rmonk, NY: IBM Corporation.

Knock, C. (1922). Visual training of the pitch of the voice. *Psychol. Monogr*. 31, 102–127. doi: 10.1037/h0093174

Mantell, J. T., and Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition* 127, 177–202. doi: 10.1016/j.cognition.2012.12.008

MATLAB (2010). *MATLAB Version 7.10.0 (R2010a)*. Natick, MA: The MathWorks Inc.

Mauch, M., Cannam, C., Bittner, R., Fazekas, G., Salamon, J., Dai, J., et al. (2015). "Computer-aided melody note transcription using the tony software: accuracy and efficiency," in *Proceedings of the First International Conference on Technologies for Music Notation and Representation*.

Mauch, M., and Dixon, S. (2014). "pyin: A fundamental frequency estimator using probabilistic threshold distributions," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 659–663. doi: 10.1109/ICASSP.2014.6853678

Mayor, O., Bonada, J., and Loscos, A. (2009). "Performance analysis and scoring of the singing voice," in *Proceedings of the 35th AES International Conference* (London), 1–7.

Mercado III, E., Mantell, J. T., and Pfordresher, P. Q. (2014). Imitating sounds: a cognitive approach to understanding vocal imitation. *Compar. Cogn. Behav. Rev*. 9, 17–74. doi: 10.3819/ccbr.2014.90002

Musicianbrain (2021). Available online at:
http://musicianbrain.com/pitchtest/
(accessed March 22, 2021).

Paney, A., and Tharp, K. (2019). The effect of concurrent visual feedback on adult sing accuracy. *Psychol. Music*. doi: 10.1177/0305735619854534

Pardue, L., and McPherson, A. (2019). Real-time aural and visual feedback for improving violin intonation. *Front. Psychol*. 10:627. doi: 10.3389/fpsyg.2019.00627

Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., et al. (2008). On-line identification of congenital amusia. *Music Percept*. 25, 331–343. doi: 10.1525/mp.2008.25.4.331

Pfordresher, P., and Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness". *Mus. Percept*. 25, 95–115. doi: 10.1525/mp.2007.25.2.95

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., and Liotti, M. (2010). Imprecise singing is widespread. *J. Acoust. Soc. Am*. 128, 2182–2190.doi: 10.1121/1.3478782

Pfordresher, P. Q., and Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychon. Bull. Rev*. 20, 747–753. doi: 10.3758/s13423-013-0401-8

Pfordresher, P. Q., and Mantell, J. T. (2014). Singing with yourself: evidence for an inverse modeling account of poor-pitch singing. *Cogn. Psychol*. 70, 31–57. doi: 10.1016/j.cogpsych.2013.12.005

Ramirez, R., Canepa, C., Ghisio, S., Kolykhalova, K., Mancini, M., Volta, E., et al. (2015). "Enhancing music learning with smart technologies," in *Proceedings of the 5th International Conference on Movement and Computing (MOCO'18)*, 49:4.

Reymore, L., and Hansen, N. C. (2020). A theory of instrument-specific absolute pitch. *Front. Psychol*. 11:2801. doi: 10.3389/fpsyg.2020.560877

Romaní, O., Parra, H., Dabiri, D., Tokuda, H., Hariya, W., Oishi, K., et al. (2015). "A real-time system for measuring sound goodness in instrumental sounds," in *138th Audio Engineering Society Convention*.

Schmidt, R. (1975). A schema theory of discrete motor skill learning. *Psychol. Rev*. 82:225. doi: 10.1037/h0076770

Seashore, C. (1902). A voice tonoscope. *Univ. Iowa Stud. Psychol*. 30, 18–29.

Seashore, C. (1940). The psychology of music. *Mus. Educ. J*. 30–33. doi: 10.2307/3385750

Seashore, C., and Jenner, E. (1910). Training the voice by the aid of the eye in singing. *J. Educ. Psychol*. 1, 311–320. doi: 10.1037/h0075088

Vamvakousis, Z., Perez, A., and Ramirez, R. (2018). "Acquisition of violin instrumental gestures using an infrared depth camera," in *15th Sound and Music Computing Conference Sonic Crossings, 2018*, eds A. Georgaki and A. Andreopoulou (Limassol: Cyprus University of Technology).

6. ONLINE TONE MANIPULATION IN VIOLIN PERFORMANCE

6.1 Abstract

During the performance, musicians need to monitor information such as proprioceptive feedback of the ongoing movements and relate that to the planned movement. Supposedly, whenever there is a mismatch between actual and predicted consequences an error signal is generated leading, if necessary, to corrective movements. Previous research investigated neural error-related processes of music production in pianists by manipulating feedback of correct notes. One of the advantages of using a bow-string instrument is that feedback manipulated notes allows online corrections allowing us to study processes of correction in music production. In this study 15 bow-string players with more than 7 years of experience were asked to a) listen to a reference melody of four notes b) play that melody with the violin (Action Condition, AC), and c) listen passively to the replay of their own performance (Replayed Condition, RC). Randomly, the auditory feedback of one of the four notes of the melody was manipulated in the AC by lowering or lifting the pitch of the tone by half semitone. We found that mistuned notes elicited an f-ERN followed by a P3a and P3b. In addition, between the f-ERN and the P3a we found a parietal negativity, the N340, previously reported in visual-tracking tasks. It is suggested that compensating manipulations recruited parietal areas specifically related to auditory-motor integration which are thought to play a key role in the evaluation of sensory-prediction errors and the following movement adjustments to the motor command together with the cerebellum.

6.2 Introduction

Playing music is a complex skill that requires a complex interaction between motor, auditory, and somatosensory systems. Imitating a simple melody requires processes related to movement planning (which movements have to be realized and in which order), movement execution at the correct tempo, and constant monitoring of the movements that allow fine and continuous adjustments in response to possible errors produced. The strong coupling between the auditory and motor systems required for musical activity gives auditory feedback an important role to establish the required associations between action and perception and thus, enable musical learning (Lappe et al., 2018; Pfordresher & Chow, 2019; Pfordresher et al, 2011; Stewart et al., 2013; Brown and Palmer, 2012; Chen, Raw, & Watkins, 2012; Baumann et al., 2007; see Nunes-Silva et al., 2020 for a review).

Auditory feedback seems to be less important for expert musicians than for beginners in instruments such as the piano (Bishop, Bailes & Dean, 2013; Finney & Palmer, 2003; Highben & Palmer, 2004; Repp, 1999; Finney, 1999; Pfordresher, 2005, 2008). This suggests that musicians have strong associations between actions and their expected auditory consequences, allowing anticipatory imagery to compensate for the lack of feedback information (Bishop et al., 2013; Keller, Dalla Bella, & Koch, 2010). Unlike the piano and other keyed or fretted instruments like the guitar, pitch control in the violin is continuous and, thus, movements have to be much more precise. This makes music production and intonation monitoring with those instruments much more dependent on auditory feedback (Chen et al., 2008, 2013) as it already happens with the voice (Kleber et al. 2017).

This work aims to study the electrophysiological correlates of error monitoring, correction, and self-generation with string instruments in violin and cello players. For that purpose, we have designed a setup that allows us to manipulate the pitch of random notes from short melodies played with the violin and collect the electrophysiological responses of right and wrong feedback both in active conditions (during the performance) and in replayed conditions (listening to themselves).

6.3 Background

6.3.1 Effects of self-generation in N1 amplitude

It is thought that when hearing self-produced sounds an efference copy of the motor command is transformed into a prediction of auditory feedback and sent to the auditory cortex supposedly suppressing N1 responses of the temporal lobe (Ford et al., 2001, Horvath, 2015; Hughes, Desantis & Waszak, 2013). This N1 reduction, however, becomes weak when auditory feedback is altered during speaking (Heinks-Maldonado et al., 2005; Jones et al., 2013, Scheerer et al., 2014; Scheerer and Jones, 2018) supporting the existence of a forward model modulating cortical responses to self-generated sounds.

Katahira et al., (2008) addressed this issue in musical performance with the piano. Unlike speech, proficiency in music performance differs between subjects. In their study, they predicted that musically trained subjects would be able to predict their own errors by comparing them with actual auditory feedback, while non-trained subjects would lack this ability. They found in both experiments how the amplitude of N1 tended to be higher for manipulated feedback than for congruent feedback as was predicted if a corollary discharge for musical performance was created. Interestingly, this effect of manipulated feedback was significant only in the trained group in the first experiment. However, they did not find differences in the amplitude of N1 between the active and the passive condition. Probably, as the authors recognize, because both participants and the physical condition of the stimuli were not the same.

6.3.2 Error-related negativities (f-ERN and r-ERN)

Previous research in online note manipulation during music performance found a prominent frontocentral negativity peaking around 200 and 250ms for manipulated notes that was not present in non manipulated and correct notes (Maidhof et al, 2009; Katahira et al, 2008; Mathias et al, 2016; Loehr et al, 2013). This negativity was maximal in frontocentral electrodes and it is thought to reflect the f-ERN (Maidhof et al, 2009, Hajcak, Moser, Holroyd, & Simons, 2007; Hajcak, Holroyd, Moser, & Simons, 2005; Miltner, Braun, & Coles, 1997; Krigolson, Pierce, Holroyd & Tanaka, 2008; Krigolson & Holroyd, 2007; Palidis, Cashaback, Gribble, 2018). The f-ERN was also found in conditions where participants listened to the replay of their performance errors (Maidhof et al, 2009; Herrojo-Ruiz et al., 2009). Although with a lower amplitude, presumably due to stronger expectancies build up during performance (Maidhof et al, 2009).

However, not all performance errors elicit an f-ERN. Speeded response tasks usually elicit a negativity around 50 and 100ms just after the production of an error (for a review see Gehring, Liu, Orr & Carp, 2012). This negativity is usually called a response-ERN (rERN). Although it is thought that the f-ERN and the rERN may share contributions from a common generator in the Anterior Cingulate Cortex (ACC), the main difference between them according to the Reinforcement-Learning-ERN theory is that the rERN is elicited by an efference copy of the motor command (internal feedback) whereas the f-ERN by external feedback. That's the reason why the rERN is supposed to occur more quickly than the f-ERN (Gehring, Liu, Orr & Carp, 2012). What is more, relatively recent research found that the ERN could even anticipate the onset of an incorrect note already 100ms before in the performance of expert pianists (Maidhof, Rieger, Prinz & Koelsch, 2009; Maidhof, Pitkäniemi, Tervaniemi (2013); Strübing, Ruiz, Jabusch & Altenmüller, 2011; Herrojo-Ruiz, Jabusch & Altenmuller (2009)). This adds support to the idea that the r-ERN relies on a comparison between the predicted outcome of action with the actual action goal. Interestingly, no f-ERN was reported in either one of those studies following incorrect responses presumably because they were self-performed errors and not manipulated ones. According to Heldmann et al (2008), when there is internal self-monitoring information about errors additional feedback information about the error becomes redundant, which is reflected in a lack of f-ERN.

6.3.3 The P300

The P300 is a positivity that usually follows the f-ERN (Falkenstein et al. 1995, for a review, see Overbeek et al, 2005; Polich, 2007). The P300 can be interpreted as two different components: an earlier one with a frontal distribution (P3a) and a more posterior one that also occurs later (P3b) (Arbel & Donchin, 2009; Ruchsow et al., 2005b; van Veen & Carter, 2002). The P3a, as the ERN, seems to be unrelated to error awareness (Endrass et al., 2007). Adding the fact that has a similar topography with the ERN, it has led to the suggestion that it may have a neural generator also in the medial frontal cortex (Herrmann et al., 2004; van Veen & Carter, 2002) or that both the ERN and the P3a are both parts of a single oscillatory potential at the Theta band (4-7Hz). On the other hand, the P3b (or parietal P300) is hypothesized of being involved in adapting response strategies following an error, and with the updating of an internal model of the environment (Krigolson et al. 2008; Palidis et al. 2018, Donchin and Coles 1998, Overbeek et al., 2005). One clear distinction between these two components can be seen in the previously mentioned study of Maidhof et al. (2009). They found the P3b to be larger when participants were asked to detect errors in the passive condition (task-relevant) than when they were not (task-irrelevant).

6.3.4 Reward and sensory-prediction errors

However, contrary to previous research which studied error monitoring processes using discrete response tasks, the violin is fretless and allows online corrections of errors. Kieffaber et al. (2016) used a non-discrete task by making participants move a cursor from a starting location to a target location. They found as expected an r-ERN following the initiation of an erroneous response. In addition, they found that theta/alpha power in frontal sites was clearly related to the corrective action. These results may suggest that frontocentral theta power may be distinguished from the ERN as a separated component contrary to some previous evidence (Bernat et al., 2005; Cavanagh et al., 2009; Gehring & Willoughby, 2004; Hall et al., 2007; Trujillo & Allen, 2007).

Krigolson et al (2007, 2008) distinguished between two different kinds of motor errors that depend on external feedback to be elicited: outcome errors and target errors. An analogous distinction that the recent research in motor learning uses between reward and sensory prediction errors (SPE) (Krakauer et al, 2019). It is thought in general terms that reward errors are evaluated within the medial-frontal cortex involving the ACC and the basal ganglia (Krigolson and Holroyd 2006). Reward errors represent the failure to achieve the desired movement goal while sensory-prediction errors represent a discrepancy between the actual motor command and the appropriate motor command. Sensory-prediction errors involve in general terms the posterior parietal cortex (PPC), which is thought to play a key role in the evaluation of sensory-prediction errors and the following movement adjustments to the motor command, and the cerebellum (Culham et al. 2003; Desmurget et al. 1999, 2001; Desmurget and Grafton 2000; Diedrichsen et al. 2005).

Krigolson and Holroyd (2005, 2007) recorded the ERPs from participants engaged in a tracking task to investigate the neural correlates of each type of error. They found that these types of tasks elicited an ERN suggesting that the frontal system was evaluating reward errors. However, they also found a parietal negativity peaking later than the frontal negativity around 362 ms. According to the authors, high-level error information (reward errors) was communicated to the PPC once evaluated by the medial-frontal cortex for the adaptive modification of behavior. This negativity coincides both in time and location with that reported by Leuthold and Jentzsch (2002) associated with motor reprogramming of a movement that had already commenced in a response priming task. Krigolson et al (2008) added additional support for the hypothesis that the medial frontal cortex was sensitive to outcome errors but not by target errors by showing no ERN effects following sensory-prediction errors during performance of a manual aiming task. Palidis et al (2018) were also able to design a visuomotor rotation task to isolate reward-based learning from sensory error-based learning. In that experiment, they found that the FRN was elicited specifically by reward feedback but not sensory error feedback. However, a parietal P300 was present in both types of errors as well as in both previously mentioned studies. Recent neuroimaging studies have offered support for the role of the PPC for adaptive modification of behavior in music production with a string instrument. Segado et al. (2018) found overlapping brain activity between cello and singing in auditory and dorsal-motor regions. In a posterior study (Segado et al., 2021), participants were required either to ignore or compensate for pitch-shifting manipulations of their produced tone. They found that compensating manipulations recruited parietal areas specifically related to auditory-motor integration, in particular the intraparietal sulcus (IPS) and the supramarginal gyrus (SMG).

6.3.5 Movement-related cortical potentials

Movement-related cortical potentials (MRCP) are also an important tool in studying the processes of motor learning (Masaki and Sommer, 2012). Recent studies have reported movement-related cortical potentials of smaller amplitude in experienced guitarists compared with nonmusicians when playing scales with the guitar

(Wright et al., 2012), an effect already found in performers of different sports backgrounds (DiRusso, et al., 2005; Fattapposta et al., 1996; Hatta et al., 2009; Kita et al., 2001) although with less ecological tasks such as simple button pressing to investigate clay target shooting (DiRusso, et al., 2005) and pistol shooting (Fattapposta et al., 1996).

Recently, Tan et al. (2014a,b) and Torrecillos et al (2015) studied the effect of perturbed movements in a force field task in modulations of the post-movement Beta-rebound and correction Beta-enhancement after perturbed trials with different degrees of perturbation. MRCPs are typically composed of a foreperiod Beta enhancement occurring before the onset of a movement, a movement Beta-suppression following the onset of the movement, and a beta-rebound following the end of the movement. Torrecillos et al (2015) designed two experiments where participants had to reach a visual target in a force field. They found how unpredictable changes in the strength of the applied force modulated the post-movement Beta rebound by attenuating it as Tan et al (2014) reported. This Beta rebound seemed to be independent of the online correction of the movement and, as found in their second experiment, was insensitive for both goal or sensorimotor errors, suggesting that post-movement Beta-rebound reflects salience processing independent of sensorimotor adaptation.

6.4 Aims

In this current experiment, we aimed to extend previous research on error monitoring in piano melody-playing to string instruments like the violin. Unlike the piano, the violin is a fretless instrument where musical notes are placed in a continuum resembling more the singing voice. As we have seen, the use of discrete tasks in error monitoring has limitations that can only be avoided with other kinds of tasks that allow online error correction. Although singing is probably one of the most interesting candidates to study music monitoring it has certain limitations to study feedback manipulation due to bone conduction of self-performed feedback. However, the violin and bow-string instruments also have their limitations for these kinds of tasks. First of all, even being electric violins, it is hard to mute them completely to not produce any external sound that could add to the manipulated feedback of headphones. Even more, because western violin players tend to hold the violin with their chin, the vibration of the instrument when being played is carried to the participant through bone conduction interfering again with the manipulated feedback from the headphones. In this study, we have designed a setup that allows us to manipulate the notes of the violin while maintaining a naturalistic environment for music performance. This setup is also suited for other string instrument players like cellists and can be a potential tool to study the electrophysiological correlates of learning intonation skills in future research. To our knowledge, no previous research has studied the electrophysiological correlates of melody-playing in a string instrument.

Both violinists and cellists participated in the experiment where they listened to a reference melody, played that reference melody with the violin, and listened to the replay of their performance. Randomly chosen, one of the participant's melody notes was online manipulated during their performance by raising it half a tone up or half a tone down. The inter-onset interval between notes of the melody was 2 seconds. Previous experiments with the piano (except for Katahira et al. 2008) have used smaller inter-onset intervals between notes around 125ms. Those higher tempos made it impossible to compare amplitude modulations of the N1 component between active and passive conditions. We expected to find a self-generation effect between the passive and active condition in the amplitude of the N1 component as has been previously reported (Ford et al., 2001). We also expected to find motor-evoked potentials in the active condition related both to the left hand (finger placement in the fingerboard) and the right arm (bow movement) together with their respective post-movement beta rebound and their possible modulations after note manipulations (Tan et al. 2014; Torrecillos et al 2015)

Participants were required to online correct any perceived error during their performance, whether they felt it was self-generated or externally manipulated. Because of the non-discrete nature of the violin, errors could be of varying magnitude. We hypothesized that higher errors would involve more corrective movements than lower errors implying, in turn, greater awareness of the error. We hypothesized to find an f-ERN after manipulated notes in the active and passive condition. However, considering the bigger dependency on auditory feedback that string-instruments demand and the big inter-onset interval used, we did not expect to find an r-ERN after self-generated errors as in previous studies but also an f-ERN. We also expected to find lower f-ERN amplitudes in the passive condition as Maidhof et al. (2009) reported and a lack of P3b considering that our passive condition was task-irrelevant. Finally, we expected to find mid-frontal theta power after the initiation of errors and related to the onset of corrective movements as previous research has reported (Kieffaber et al., 2016).

6.5 Methods

6.5.1 Participants

Fifteen adult right-handed participants (7 females; age 28.06 (SD = 7.86)) with no self-reported history of neurological, psychiatric, or hearing impairment and with normal or corrected-to-normal visual acuity were recruited from the university campus and different music academies and conservatories of Barcelona to participate in the experiment. All participants had, at least, 6 years' experience playing a bowed string instrument: 7 violinists with 14.14 (7.81) years' experience; 7 cellists with 14.57 (8.58) years' experience; 1 participant who played both plus the Chinese Erhu, with 25 years' experience. Participants conceded their written consent before their participation and after all procedures were explained to them. Procedures were approved by the Conservatoires UK Research Ethics committee, following the guidelines of the British Psychological Society and in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). The study was carried out in the EEG labs located in the Department of Clinical Psychology and Psychobiology of the University of Barcelona (Brainlab).

6.5.2 Stimuli

Individual violin notes were created using a real violin sample sound retrieved from the webpage of the Philharmonia Orchestra (www.philarmonia.co.uk). The violin sound was recorded with a "piano" dynamic and played "con sordino" (with mute) in the G3 (196.59Hz) note of the 4th string with a duration of 1.5s. The same sound was used to generate the rest of the notes that composed our stimuli set using Audacity (https://www.audacityteam.org/). First, the DC offset of the sound was removed and the maximum amplitude was normalized to -1 dB. Then, the pitch of the sound was shifted to generate different notes, in an equal-tempered scale, that are possible to be played in the 4th string of a violin without changing the position of the hand: A3 (220.671Hz), Bb3 (233.8Hz), B3 (247.65Hz), C4 (262.39Hz) and D4 (294.441Hz). These sounds were then arranged to create short melodies of 4 different notes (2s onset to onset; thus, from note offset to subsequent note onset, 500ms silent gap) using custom-made Matlab (R2020a; Mathworks) scripts. All melodies started in G3 (hereinafter referred to as open-string note). The following three notes (hereinafter referred to as target notes) were arranged according to either the Ionian mode (A3, B3, C4, D4) or the Aeolian mode (A3, Bb3, C4, D4). Target notes were chosen randomly without repetition, resulting in a total of 24 possible different melodies per modal scale. Hereinafter, we will refer to these short melodies as reference melodies, and the sounds composing them as reference sounds (see Procedures below).

The remaining auditory stimuli consisted of the self-generated sounds created by each participant when playing the violin to replicate the reference melodies. These sounds were recorded at a sampling rate of 44.1 kHz by a Shure dynamic microphone (model: PPPER) located pointing at the front of the violin's case, using Max/MSP 7 software. We refer to these sounds as self-generated sounds when delivered to the participant in real-time while playing, and as replayed sounds when delivered to the participant immediately after replicating the reference melody, as a replay of the preceding performance. All sounds (reference, self-generated, and replayed) were delivered binaurally through passive noise-reduction in-ear earphones (Sony MDR-EX110LP) at the maximum comfortable level chosen by the user when listening to the reference melodies (approx. 50-75dB range for reference and 45-65dB range for played-replayed).

During playing, one note out of the three target notes in a melody, chosen at random in each melody, was shifted by 50 cents (half semitone) above or below its pitch using a real-time pitch-shifting algorithm (see Supplementary Materials). We will refer to these target notes as manipulated notes; to the target note immediately following a manipulated note (when the manipulation occurred before the last target note in the melody), as the post-manipulated note; and to the rest of the target notes as non-manipulated notes. Replayed sounds reproduced exactly what the participant heard during playing, thus preserving pitch manipulations.

Visual stimuli consisted of scores (~22x2 degrees) displayed at the center of the screen on a white background with the word 'Score' above them (Arial; ~3x1 degrees). Scores were presented with a G-clef, in the key signature of C-Major (i.e., no flats and no sharps) and with a 4/4 time signature. The notes were distributed between two bars with two half-notes in each of them. Participants received visual instructions displayed at the bottom of the screen consisting of instruction messages: "Listen to the scale. Mode: Jonico", "Listen to the scale. Mode: Eolico", "Go!" and "Listen" (Arial; ~3x1 degrees). Finally, between the score and the visual instructions, a visual countdown was displayed at certain moments of the experiment with the text "Get ready in:". The countdown went from 3 to 1 and was displayed in red together with a red arrow that moved over each of the last three notes of the score.

6.5.3 Procedures

Participants sat on an armless chair in an electrically shielded, sound-attenuated EEG recording chamber, holding a muted violin (standard Stradivarius; primavera ultra rubber mute) in vertical position between their legs as if it were a cello (see Fig.1). The position and the mute were adopted to reduce the amplitude of the sound generated by the violin during playing that would reach the participant's ears, as well as eliminating the possibility to hear the violin by bone conduction. Since the participants only played on the G string, a piece of cotton was introduced between the D, A, and E strings and the fingerboard to mute any resonances. Protective earmuffs 3M PELTOR Optime III covered the participant's ears to provide an extra layer of sound attenuation, besides that furnished by the earbuds of the noise-reduction earphones. All these measures ensured that, while playing, the participant would only hear the self-generated sounds as delivered by the earphones, whose loudness was already set to the maximum self-chosen comfortable level, so that pitch-shifting manipulations would not be heard mixed with the direct, non-manipulated, violin acoustic sound.

A computer screen (28 inches) to deliver instructions was placed at a distance of 1.20 m from the participant. A keyboard was placed on the floor right next to the participants' bare feet to enable them to start an experimental block by pressing the spacebar with their toes, minimizing ample arm movements that would be produced if the hand was displaced from the keyboard to the violin and vice versa.
The experiment was composed of 24 blocks of 6 trials each (144 trials in total; each trial lasted 26 s). A block consisted on the following sequence of events (illustrated in Fig.1): 1) the participant pressed the spacebar on the keyboard with his/her right foot toes to start the block when desired; 2) Immediately after, the score of a short melody appeared on the screen with visual instructions on the top ("Listen to the scale. Mode: Jonico" or "Listen to the scale. Mode: Eolico") and the sounds composing the reference melody were delivered through the earphones (*Reference Condition*); synchronously with each sound, a visual countdown composed of numbers from 3 to 1 (with the instruction "Get ready in: ") appeared on the screen together with a red arrow that pointed to the reference note being played; 3) 0.5 seconds after the offset of the last reference note, the instruction "GO!" appeared on screen; from this moment, the countdown and the red arrow disappeared and the participant disposed of 9 seconds to play with the violin the previously heard reference melody (*Active Condition*) respecting its pitch, tempo and note duration as best as possible; 4) after those 9 seconds, regardless of whether the participant ended his/her performance, the instruction "Listen" appeared on the screen and the replay of the recorded performance was reproduced (*Replayed Condition*); 5) 9 seconds after the offset of the replay, a new trial started (step 2); 6) After 6 trials the screen displayed a message informing that the block had finished and that the participant could start a new block by pressing the spacebar on the keyboard when desired.

All reference melodies in a block belonged either to the Ionian or to the Aeolian modes, alternating between blocks (first block modal scale counterbalanced across participants). All possible melodies were covered in the experiment (each melody appeared 3 times throughout the experiment: 144/(24 possible melodies x 2 modes) = 3). The experiment lasted 62 min, without pauses (total experimental duration, including EEG recording preparation and pauses, about 3h). Participants could rest between blocks as desired, and a communication channel with the experimenter via CCTV remained open during the whole experiment to report any problem or discomfort.

Participants performed three training blocks, identical to the experimental blocks, supervised by the experimenters before the EEG recording preparation, to adapt to the demanded task. Participants were told that some notes would be externally manipulated and were required to correct any perceived error, either if it was self-generated or externally manipulated.

6.5.4 Audio Analysis

The audio of each experimental block was recorded in mono to a stereo channel (audio channel 1) and stored as a stereo .wav file. We used Tony Software (Mauch et al., 2015) to export the onset, pitch, and duration of each note from the recorded audio files. We also exported the onset, offset, pitch, and duration of the sounds produced by corrective movements. Tony is based on the pYIN method for automatic pitch estimation and note tracking (Mauch and Dixon, 2014) together with custom methods for interactive reestimation. It outputs discrete notes on a continuous pitch scale based on Viterbi-decoding of an independent Hidden Markov Model. After using Tony, we performed a visual inspection of the onsets of all events to ensure the correct operation of the pitch detection algorithm.

In a posterior analysis in Matlab (Matlab, 2010), we labeled each detected onset depending on its condition (Reference, Active, and Replayed), type of note (open-string, non manipulated, manipulated, and post-manipulated), correction (with, without), type of onset (start of the note, onset of correction, offset of correction), and absolute error in cents ([0-15], [15-30], [30-50], [50-70], [>70]).

6.5.5 EEG recording and preprocessing

EEG data were acquired from 67 Ag/AgCl electrodes and digitized at a sampling rate of 500Hz by a Neuroscan 4.4 software and Neuroscan SynAmps RT amplifier (NeuroScan, Compumedics, Charlotte, NC, USA). 62 electrodes were mounted in a nylon cap

(QuickCap 64; Compumedics, Charlotte, NC, USA) following the 10-10 international system; additionally, 2 electrodes were positioned over the left and right mastoids (M1 and M2 respectively), and 3 electrodes to record the electrooculogram (EOG; 1 under the left eye, to assist in blink detection; 2 located at the left and right outer canthi of the eyes, to record lateral eye movements). The reference electrode was located at the tip of the nose and the ground electrode at AFz. All impedances were kept below 5 kΩ during the whole recording session.

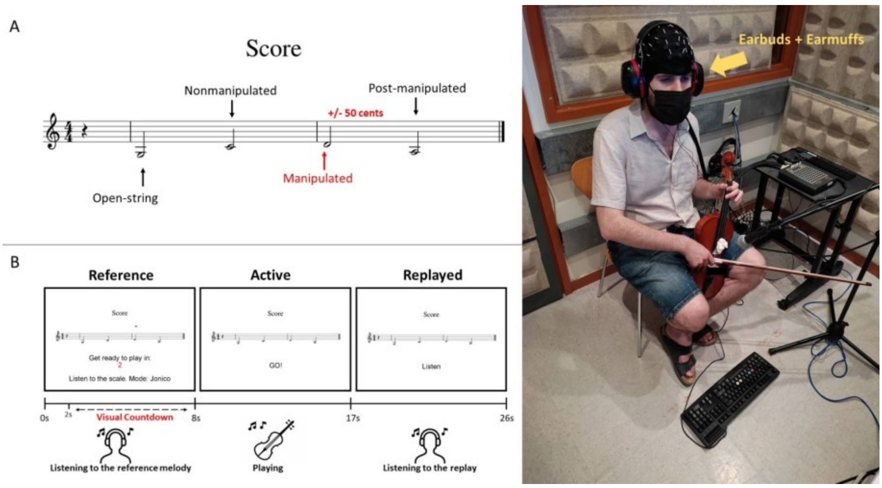

Figure 1. A. One random melody from the experiment. In this specific example, the open-string note (which is always G3) is followed by a nonmanipulated note. The next note, and without the participant's knowledge, is shifted by 50 cents during playing (half a semitone; manipulated note). We refer to the next note as the post-manipulated note. **B.** Conditions. Each trial is composed of three different conditions: in the Reference Condition, participants hear the sounds composing the reference melody; in the Active Condition, participants play with the violin the previously heard reference melody; finally, in the Replayed Condition, participants hear the replay of their own performance in the Active Condition. Participants hear all sounds (reference, self-generated, and replayed) through earphones **C.** During the experiment, participants held the violin between their legs as if it were a cello. Protective earmuffs 3M PELTOR Optime III covered the participant's ears to provide an extra layer of sound attenuation, besides that furnished by the earbuds of the noise-reduction earphones.

EEG recordings and audio files were synchronized by simultaneously sending, at the beginning of each experimental condition, triggers to the EEG signal and audio clicks to audio channel 2 (see *6.11 Supplementary Materials*). This allowed us to find, through interpolation, where the onsets of the triggers in the audio file were located in the EEG recording.

Data preprocessing was performed offline using EEGlab v2021.0 software (Delorme and Makeig, 2004) running on Matlab R2017a. EEG raw data was pass-band filtered from 1 to 50Hz with a windowed sinc FIR filter (Hamming window) and re-referenced to the average of all channels. Then, EEG data were visually inspected and those periods contaminated from noise and non-stereotyped muscle artifacts were rejected. Independent component analysis (ICA) decomposition was applied to remove from the signal ocular, heart rate, and muscular components identified on the basis of their scalp topography and time course (Jung et al., 2000). Finally, EOG channels were removed.

6.5.5.1 Event-related Potentials (ERPs) processing

Preprocessed EEG data were low-pass filtered at 25Hz with a windowed sinc FIR filter (Hamming window) and epoched from -600 to 800ms time-locked to each auditory stimulus onset (from all conditions (Reference, Active, and Replayed), types of notes (open-string, non manipulated, manipulated and post-manipulated) and absolute error in cents ([0-15], [15-30], [30-50], [50-70], [>70]) without baseline correction. Epochs from each condition containing improbable data 3 SD above or below the mean probability distribution of values across all epochs were excluded (EEGlab's function pop_jointprob.m).

6.5.5.2 Event-related Spectral Perturbations (ERSPs) processing

Preprocessed EEG data were epoched from -1500ms to 2000ms time-locked to each auditory stimulus onset (from all conditions (Reference, Active, and Replayed), types of notes (open-string, non manipulated, manipulated, and post-manipulated), and absolute error in cents ([0-15], [15-30], [30-50], [50-70], [>70]) without baseline correction. The ERSP was computed with a linearly increasing number of cycles with increasing frequency. The wavelet used in each time window began with a 3-cycle wavelet (Hanningtapered window) and ended with 50% of the number of cycles in the equivalent FFT window at its highest frequency. The ERSP was also computed with an output of 30 frequency bins from 3Hz to 50Hz and 200 output time samples (EEGlab's function *newtimef.m*). Epochs from each condition containing improbable data 3 SD above or below the mean probability distribution of values across all epochs were excluded (EEGlab's function pop_jointprob.m). For illustration purposes, in some figures, ERSPs were baseline corrected subtracting the mean ERSP of the whole epoch for each frequency bin separately.

6.6 Data Analysis

6.6.1 Behavioral Analysis

In the behavioral analysis, we aimed to study the different aspects that could have influenced the performance of participants. That is if tuning error magnitude or the onset of corrective movements depended on whether the notes played were manipulated, nonmanipulated or post-manipulated.

The open-string note did not require corrective movements (no finger placed on the fingerboard) and was expected to stay in tune throughout the experiment. However, because strings tend to lose tension with time, we asked participants to tune the string in case they noticed that they were getting out of tune. To ascertain that the open-string note remained in tune during the experiment, we measured its pitch through the 12 experimental blocks. The error of the note was on average 6.53 cents (SD: 2.9), which is below the normal pitch discrimination threshold for musicians (Hopkins, 2015), and no significant changes were found across blocks (repeated measures ANOVA). To study whether target note mean pitch error differed for each type of note (nonmanipulated, manipulated, post-manipulated) we performed another repeated measures Anova. The error was measured in the notes heard by the participants. Thus, it includes the participants' error plus the external manipulation.

We also evaluated the behavior of participants concerning their corrective movements. We expected that lower errors would be more difficult to perceive than larger ones and therefore would have been less likely to suffer corrections. We performed a 6x3 repeated measures ANOVA with absolute error in cents ([0-15], [15-30], [30-50], [50-70], [>70]) and type of note (non manipulated, manipulated and post-manipulated) as within-factors and the percentage of notes which were followed by a corrective movement as the dependent variable. We also expected that those notes which suffered corrective movements were in the direction of decreasing the perceived error and presented a magnitude of the correction related to the magnitude of the error. We performed two more 6x3 repeated measures ANOVA with absolute error in cents and type of note as within-factors. One with the magnitude of the correction in cents (error after the correction minus the error before the correction) as the dependent variable and another one with the absolute error in cents after the corrective movements as the dependent variable.

Finally, to study the possible effects of the type of note or the magnitude of the error in the timing (onset and offset) of the corrective movements, we performed two more 6x3 repeated measures ANOVA with absolute error in cents and type of note as within-factors and the onset and the offset (in seconds) of the corrective movements as dependent variables.

We decided to set a minimum percentage of notes as a threshold for each type of note and level of error to consider the inclusion of the results of the participant in the analysis. 8 participants did not produce more than 5% of their total notes with an error higher than 70 cents and their results were removed from that specific range. Of these 8 participants, 3 of them also did not produce more than 5% of their total notes between the 50 and 70 range and were thus removed from that range.

6.6.2 Electrophysiological Analysis

6.6.2.1 Auditory responses to violin notes during listening

In order to have an overview of the electrophysiological auditory responses elicited by the individual violin notes embedded in the melody, we first inspected the responses to the notes of the reference melody played by the synthesized violin, which did not contain any tuning errors, averaged across participants. We then inspected the auditory responses to the passive replay of the self-performed notes (replayed condition), selecting only those notes that were tuned (i.e., tuning error < 15 cents). In controlled situations, an experienced violinist can identify differences between pitches in the violin from 7 cents of separation (Hopkins, 2015) up, while nonmusicians can identify differences from 15 cents up. Recent studies have found higher thresholds of up to 15 cents in expert violinists (Blanco et al., 2021). That is the reason why we considered 15 cents to be a reasonable threshold for tuned notes. We expected to find in both conditions (reference and replayed) the classical auditory P1, N1, and P2 ERP components.

6.6.2.2 Self-generation effects on auditory and motor responses to tuned notes

Previous studies have reported modulations of the N1 and P2 auditory components to self-generated compared to externally-generated sounds (Ford et al., 2001, Horvath, 2015; Hughes, Desantis & Waszak, 2013). Thus, we inspected the responses to tuned notes in the active condition and compared them with the tuned notes in the replayed condition. As we expected to find a superposition of auditory and motor responses caused by both bow movement (right arm) and finger placement on the fingerboard (left arm) in the active condition, we performed a P1 to N1 and N1 to P2 peak-to-peak amplitude measurement extracted from the Cz electrode referenced to the average of Mastoids rather than an absolute amplitude measurement, to further limit the influence of possible confounds. The amplitude of the P1 component for each participant was computed as the mean amplitude within a 20 ms window centered on the peak, which was determined by searching for the maximum positive amplitude peak between 0 and 90ms in the grand average of all subjects. The same procedure was applied to the N1 (negative; between 70 and 140ms; 20 ms window) and the P2 (positive; between 130 and 240ms; 40 ms window). Two paired samples t-test were performed between the active tuned and the replayed tuned condition for the P1 to N1 and N1 to P2 amplitudes. We also inspected the influence of possible motor-related potentials related to bow movement or finger placement employing a cluster-based analysis

6.6.2.3 Effects of tuning

To assess the effects of tuning, we compared responses to tuned and mistuned notes in both the active and replayed conditions. We considered a note was mistuned when its error was larger than the threshold of 30 cents. Notes with an error between 15 and 30 cents were removed from the analysis. This threshold was decided concerning behavioral results (see section 6.7.1 Behavioral Results). Initially, we planned to analyze separately nonmanipulated and manipulated mistuned notes. However, a cluster-based analysis showed no significant differences between the ERPs of the two types of notes nor with the ERPs of post-manipulated notes. Neither in active nor replayed conditions (see 6.11 Supplementary Materials). The behavioral results also showed no significant differences between the timing of the onset and offset of the corrective movements and therefore the influence of motor activity should be comparable between the three different types of mistuned notes. Thus, we pooled the responses to all mistuned notes independently of whether the pitch was externally altered or not.

We inspected possible modulations of the beta rebound due to mistuned responses in the active condition in central and frontocentral electrodes. We defined a time window extending from the onset of the note to 1500ms. A paired-samples t-test was performed between mistuned and tuned responses in a 200ms window centered around the maximum peak and at the electrode where this peak was maximum. To assess the influence of other components related to error-monitoring in the replayed condition and active condition and to what extent MRCPs could overlap to those components or interact with them in the active condition we performed a cluster-based analysis (see Section 6.6.2.5 Cluster-based Analysis).

Finally, to see if the components of interest found were modulated by the magnitude of the error we computed a 4x2 repeated measures analysis with Magnitude (15-30, 30-50, 50-70, >70) and Condition (Active, Replayed) as within-subject factors for the amplitude of each component found.

6.6.2.4 Effects of corrective movements

To assess whether the components of interest were modulated by the onset of corrective movements we computed a 3x2 repeated measures analysis with Speed of correction (High (< 250 ms), Normal (250 to 350 ms), Slow (> 350 ms)), and Condition (Active, Replayed) as within-subject factors for each component found.

6.6.2.5 Cluster-based Analysis

Cluster-based analyses were performed using a nonparametric randomization procedure (Maris, 2004; Maris and Oostenveld, 2007). We performed a two dimensional (time, electrode) analysis comparing the activity of the tuned notes from the active condition vs the replayed condition; between mistuned and tuned notes from each condition; and the interaction Tuning x Condition (comparison between conditions after subtracting tuned notes from mistuned notes on each condition), on the ERP amplitudes (from -600 to 600ms) and the power estimates in the theta (4-7Hz) and beta (15-25Hz) frequency bands. We defined the neighboring electrodes using a Delaunay triangulation over a 2D projection of the electrode montage. We also established a minimum of 2 nearby electrodes per cluster. For each comparison, we performed a two-tailed dependent t-test assessed with the nonparametric Montecarlo Method. The p-value was determined by calculating the proportion of 2D samples from 10000 random partitions of the data. Those 2D points exceeding a significance level set to 0.05 were grouped to create the clusters. The sum of the t-statistics within every cluster was used to calculate the cluster-level statistic. The Monte Carlo method was used to assess the significance probability of the clusters. Those values of $p < 0.025$ corrected for two-tailed tests, were considered significant.

6.7 RESULTS

6.7.1 Behavioral

6.7.1.1 Mean error in cents by Type of note

Participants showed an absolute average error of 33.31 cents (SD: 16.21; range: [20.59, 70.45]) when playing non-manipulated notes; of 46.88 cents (SD: 7.29; range: [35.31, 62.01]) when playing manipulated notes; and of 35.31 cents (SD: 13.32; range: [17.5, 66.8]) when playing post-manipulated notes. In Table 1 we offer the mean, range, and standard deviation of some of the most relevant measures of the analysis (see Supplementary Materials). These error differences were significant, $F(1.41) = 23.04$, $p < 0.0001$, $\eta2 = 0.62$. Pairwise comparisons revealed larger errors in manipulated notes as compared to nonmanipulated notes ($p = 0.001$) and post-manipulated notes ($p < 0.0001$). No significant differences were found between non-manipulated and post-manipulated notes.

6.7.1.2 Corrective movements

Participants were instructed to online correct any detected pitch error in their performance. As it would be expected, since it is more difficult to detect small errors than large ones, participants tended to perform more corrective movements on those notes with higher errors than those with lower errors, independently of the type of note (see Figure 2, A). The repeated measures yielded a main effect of absolute error in cents ($F(1.41)=55.47$; $p<0.0001$; effect size $\eta2 = 0.90$). Pairwise comparisons yielded significant results between the 0-15 error range and the rest of error ranges with the exception of the 15-30 error range ([0-15] < [30-50], $p = 0.01$; [0-15] < [50-70], $p = 0.003$; [0-15] < [>70], $p < 0.0001$). No significant differences were found between the 50-70 and >70 range. The 30-50 range was significantly different both with the 15-30 and the 50-70 range ([30-50] > [15-30], $p = 0.015$; [30-50] < [50-70], $p = 0.018$). No significant differences were found between the 30-50 and the >70 range. No main effect was found for the type of note nor interaction.

As it would be expected, we also found that the magnitude of the corrective movements depended on the magnitude of the produced error (see Figure 2, B). Higher produced errors were followed by a higher corrective movement. The repeated measures analysis yielded significant results at the test of within-subject effects of absolute error in cents ($F(1.96)=81.972$; $p<0.0001$; effect size $\eta2 =0.93$). Pairwise comparisons yielded significant results between the 0-15 error range and the 50-70 and the >70 error range ([0-15] < [50-70], $p = 0.003$; [0-15] < [>70], $p < 0.0001$). No significant differences were found between the 0-15, the 15-30 and the 30-50 error range. The >70 range was significantly higher than the rest ([>70] > [50-70]; $p < 0.0001$; [>70] > [30-50]; $p < 0.0001$; [>70] > [15-30]; $p = 0.001$; [>70] > [0-15]; $p < 0.0001$]. No main effects of type of note nor interaction were found.

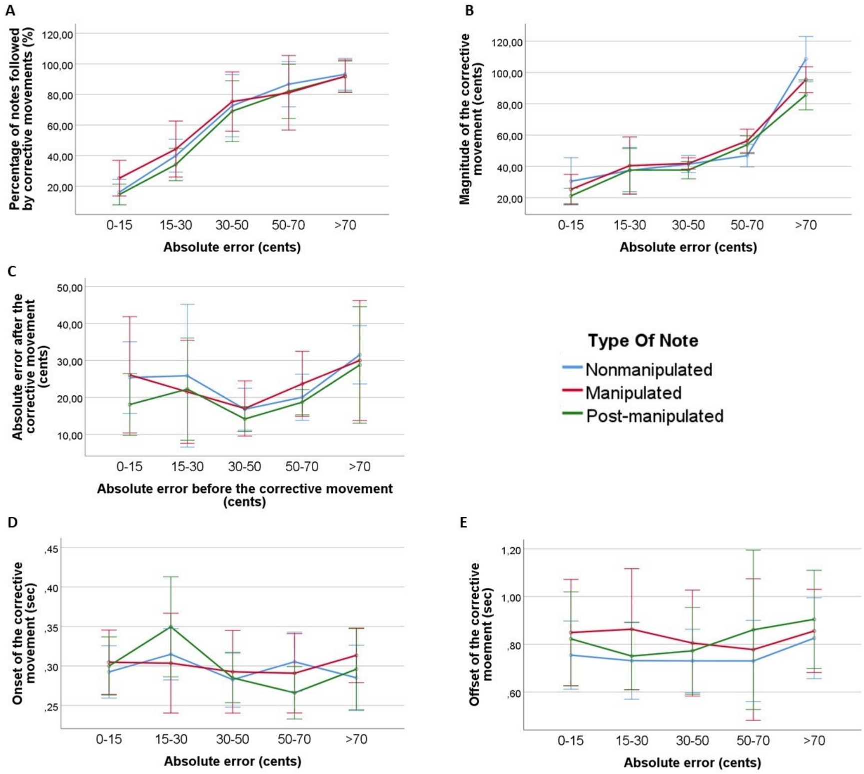

Figure 2: A. Percentage of notes followed by corrective movements for each range of error and type of note. Higher errors had a higher tendency to be followed by corrective movements than lower errors. No significant differences were found between the three types of notes. **B.** Magnitude of the corrective movements for each range of error and type of note. Higher errors tended to be followed by larger corrective movements. No significant differences were found between the three types of notes. **C.** Absolute error after the corrective movements for each range of error and type of note. No significant differences were found between the three types of notes nor the final error after the corrective movements for each range of error. **D.** Onset of corrective movements for each range of error and type of note. No significant differences were found between the three types of notes nor the time onset of the correction for each range of error. **E.** Offset of the corrective movement for each range of error and type of note. No significant differences were found between the three types of notes nor the time offset of the correction for each range of error.

We did not find differences between the three types of notes on the absolute error of the note after the end of the corrective movement. The repeated measures analysis yielded no significant results at the test of within-subject effects (see Figure 2, C). On average, the absolute error after the correction independently of the type of note was 22.66 cents (SD: 3.89; range: [13.13-32.2]).

The onset and offset of corrective movements depended neither on the type of note nor on the magnitude of the produced error (see Figure 2, D, and E). On average, participants initiated their corrective movements at 299ms (SD: 18; range: [256-342]) and ended them at 802ms (SD: 95; range:[569 1035]).

We found that the percentage of corrective movements at the error range between 15 and 30 cents did not show significant differences with the error range between 0 and 15 cents. We also found that the average error after the corrective movements was 22.66 cents (see Table 1). This may lead us to think that the 15-30 cents region could be not considered “out of tune” at least for an important part of our participants. This contrasts with experiments in controlled situations where experienced violinists can identify differences between pitches in the violin from 7 cents of separation (Hopkins, 2015) up, while nonmusicians can identify differences from 15 cents up. It is also possible that the limitations of playing in tempo imposed by the experiment make the participants settle for that value even considering it out of tune. All in all, This leads us to the conservative decision to consider only notes between 0 and 15 cents as “in tune” and those above 30 cents as “out of tune” for posterior analysis.

Also, the behavior of participants did not seem to be influenced by the type of note. Taking into account that we also did not find significant differences between their ERPs for both tuned and mistuned notes (see 6.7.11 Supplementary Materials) we decided to include them all in a single condition (tuned or mistuned) for subsequent analysis.

6.7.2 Electrophysiological

6.7.2.1. Auditory responses to violin notes during listening

In order to have an overview of the electrophysiological auditory responses elicited by the individual violin notes embedded in the melody, we first inspected the responses to the notes of the reference melody, which did not contain any tuning errors, averaged across participants (see Fig. 3). Each trial started with the playback of the reference melody by a synthesized violin. The reference notes elicited the prototypical frontocentrally distributed auditory ERPs: P1 peaking at ~75 ms, N1 peaking at ~120 ms, and P2 peaking at ~200 ms. These auditory components were followed by a frontal negativity (FN) peaking at ~450 ms in FCz. Subsequently, we observed a negative to positive slow drift over frontal electrodes (Frontal drift, FD) covering the interval between successive notes which terminated with the onset of the next note. The time-frequency analysis also showed a typical evoked auditory response on frontocentral electrodes with power in the Theta 4-7Hz frequency range locked to note onset and lasting until about 500ms. Having the electrophysiological responses to notes played by the synthesized violin as a reference, we then inspected the auditory responses to the passive replay of the self-performed notes, selecting only those notes that were performed correctly (tuned notes, i.e., tuning error < 15 cents). The replay of the self-performed notes that were tuned elicited the same pattern of electrophysiological responses as the synthesized notes from the reference melody (see Figure 3), albeit with reduced amplitude, consistent with the softer volume of the replayed notes compared to the reference melody (-25 dB approx).

6.7.2.2 Auditory and motor responses while performing violin notes

Next, we inspected the responses to tuned notes in the active condition. During the performance of the melody, the electrophysiological measurements show a superposition of auditory and motor responses (see Fig. 4). The auditory P1, N1, and P2 components can be identified showing a similar time-course and topography as during listening, as well as the FN following these components. The time period preceding the onset of the note, however, differed markedly between playing and listening to the replay, particularly on frontal electrodes. This difference is due to the superposition of the motor evoked potentials (MEPs) related to the placement of the left finger on the board and the moving of the arc with the right arm in the playing condition, which are absent during listening. A direct contrast between the active and replayed conditions allows us to better isolate these MEPs. The cluster-based permutation tests showed significant differences between the active and replayed conditions over the left and right centroparietal, as well as frontal electrodes in the time period between -130 and 28 ms (see Fig 2, panel A). Specifically, we found two overlapping clusters: one negative frontal cluster between -130 and 28ms (T=-3.226; p=0.0004) and one positive cluster over left and right centroparietal electrodes in the time period between -120 and 24ms (T=2.909; p=0.0006). The scalp maps of this difference (see Fig 4, panel B) strongly suggest a motor origin, with contributions from both the left (right arm) and right (left finger) motor cortices. Nevertheless, the MEP appears larger over the right hemisphere, possibly due to the more ballistic and discrete nature of the finger compared to the arm movement, thus also allowing a better synchronization of the response. The time-frequency analysis also showed a typical evoked auditory response on frontocentral electrodes with power in the Theta 4-7Hz frequency range locked to note onset for both active and replayed conditions (see Fig 4, panel C). The cluster-based analysis did not yield any significant differences between conditions in the Theta frequency range. Additionally, in the active condition, a further motor-related induced response is observed in the beta range (ca. 15-25Hz, Fig 4, panel C) maximal over right centroparietal electrodes (Fig 4, panel D) consisting of a pre-movement event-related desynchronization (ERD) followed by a post-movement beta rebound (event-related synchronization, ERS) after the movement execution (after the start of the played note) and maximum at 706ms. The cluster-based analysis in the Beta frequency range between the active and the replayed condition yielded a positive cluster from the whole length of the epoch and all the electrodes (active > replayed, T = 26.918, p = .00009).

6.7.2.3 Self-generation effects on auditory responses to tuned notes

Thus, we found substantial differences in the responses between the active and replayed conditions, due to the superposition of motor activity in the active condition. It is possible that some differences between the active and replayed conditions might also be explained by a modulation of the auditory responses during playing. However, Single paired t-tests did not show significant differences between the active and replayed conditions neither at N1-P1 nor at N1-P2.

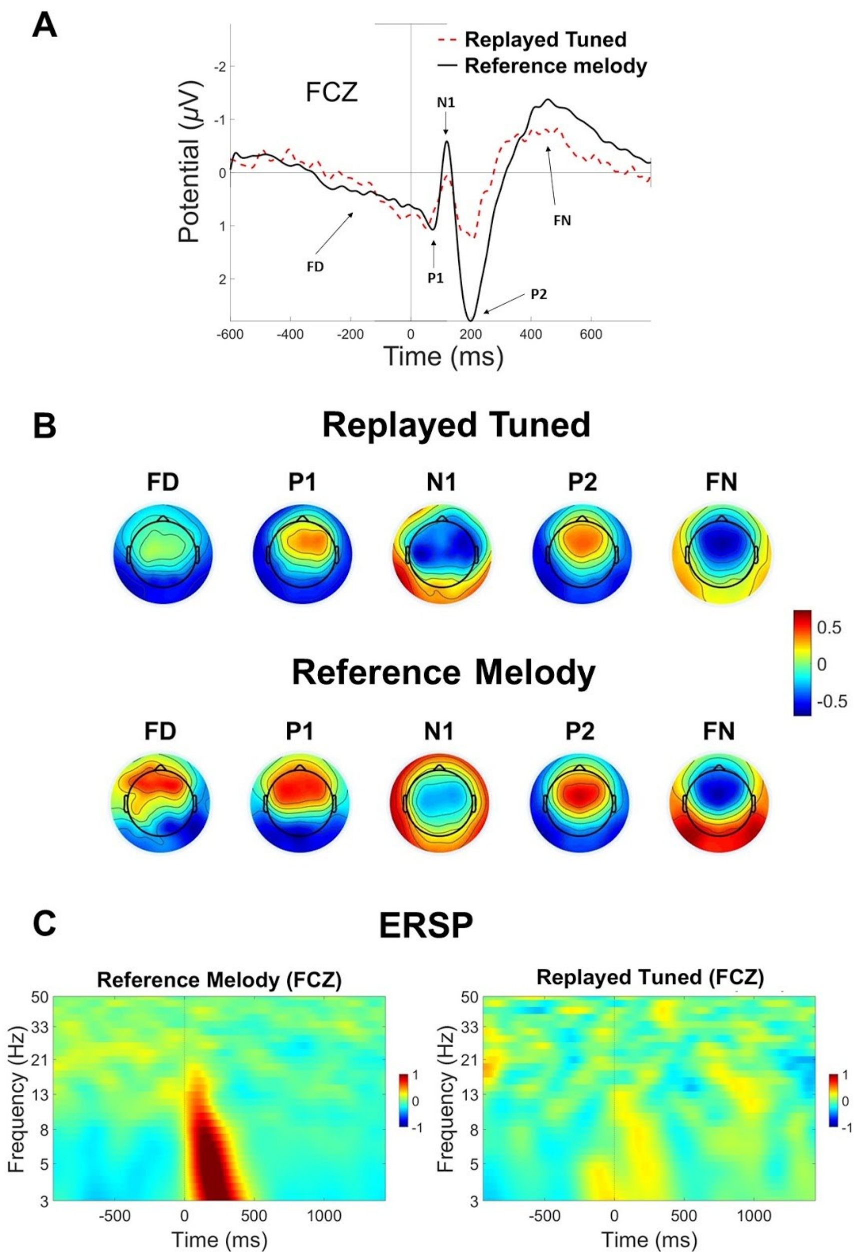

Figure 3. Auditory responses to the notes of the reference melody and tuned notes in the passive replay. A ERPs at FCz elicited to reference notes (solid black line) and replayed tuned notes (dashed red line); **B** Scalp maps for Frontal drift (-200 to 0) P1(70 to 80ms), N1(110 to 130ms), P2(180 to 220ms) and FN(300 to 600ms); C ERSP plots at FCz.

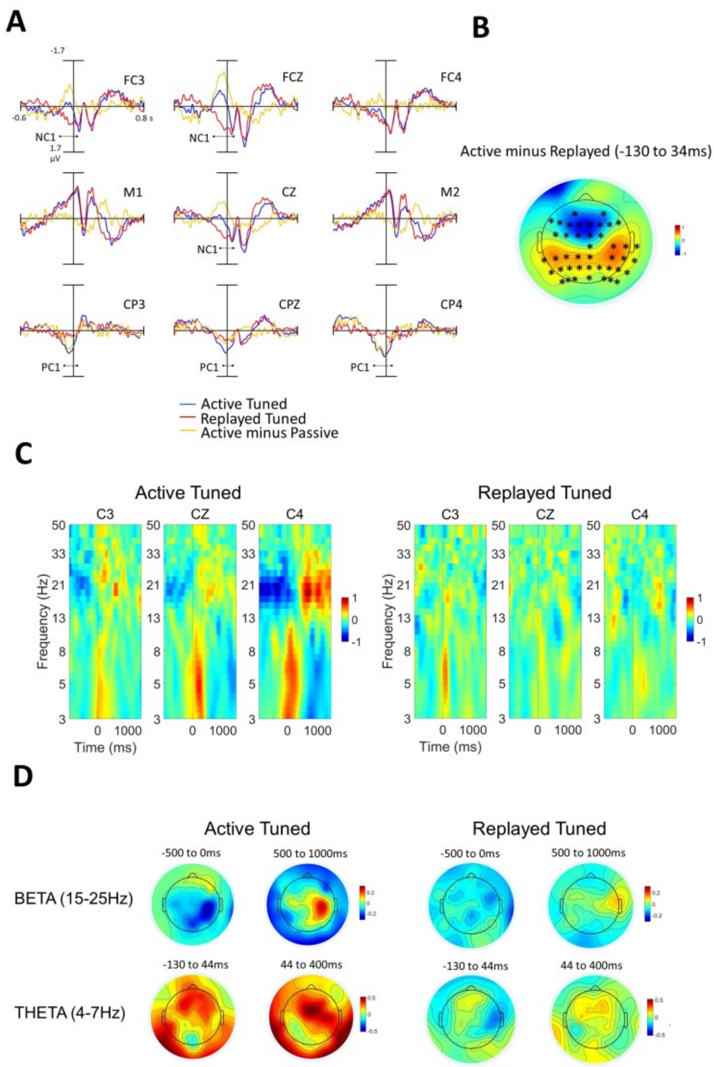

Figure 4. Responses during playing tuned violin notes compared to listening. A ERPs elicited to active (blue) and replayed (red) notes and difference waves (yellow) for a subset of electrodes. The black lines below the ERP at each electrode show the significantly different time-periods according to the cluster analysis. NC1 for Negative Cluster 1 and PC1 for Positive Cluster 1; **B** Motor-evoked potential scalp map (active – replayed difference from -130 to 34 ms). The marks correspond to the significant electrodes in this time period according to cluster analysis; **C** ERSP plots at central electrodes elicited to replayed and active notes; **D** Scalp plots of ERSP in theta and beta frequency ranges.

6.7.2.4. Effects of tuning

In order to assess the effects of tuning, we compared responses to tuned and mistuned notes. As we have already justified (see *6.7.1.1 Effects of tuning* and *6.7.1.2 Corrective movements*), notes were considered mistuned when they deviated more than 30 cents from the pitch of the reference to ensure that the error was clearly perceived. The effects of tuning were analyzed separately during playing and listening (Fig 5) and then compared across the two conditions (Fig 6).

Producing a mistuned note during playing elicited an attenuation of the post-movement beta rebound maximal on C4 (Fig 5, panel A). A paired sample t-test 200ms around the maximum peak of the beta rebound between self-generated tuned and mistuned notes showed significant differences at C4 between 606 and 806ms (active tuned < active mistuned; $t(14) = 3.487$; $p=0.004$). In the ERPs, we observed a frontal negativity peaking at ca. 240 ms (f-ERN) followed by a broad positivity at frontocentral electrodes between ca. 380-600ms (P3). In the topography, we observed how this f-ERN was slightly lateralized to the right hemisphere (Fig 5, panel A). We also observed, between the f-ERN and the P3, a right central and right centro-parietal negativity that started around 200 ms, maintained its amplitude until 380 ms and was maximum at 284 ms at C4 electrode (N-280). This negativity was also followed by a parieto-occipital negativity peaking at 340ms (N-340), maximum at P03 electrodes. The P3 could be decomposed into two components: an earlier component localized in frontal electrodes between 380 and 500ms (P3a) and a more central one between 500 and 600ms (P3b). The cluster-based permutation tests showed that during playing, the ERP to mistuned notes differed significantly from the ERP to tuned notes between 182-580 ms. Specifically, we found a negative cluster in the time period between 180 to 396ms ($T=-4.721$; $p=0.0013$) containing frontal, central, parietal, and occipital electrodes resembling the f-ERN and the N-340. Another negative one between 424 to 562ms ($T=-3.062$; $p=0.006$) containing parietal, occipital, and left temporal electrodes. We also found a positive cluster between 380 to 600ms ($T=4.484$; $p=0.0014$) containing frontal and central electrodes resembling both the P3a and the P3b. The time-frequency analysis comparing tuned and mistuned notes during playing showed an increase in theta power (3-8Hz) for mistuned notes in mid-frontal electrodes between 124-508 ms reaching its greatest amplitude in FCZ at 313 milliseconds. Cluster-based permutation tests in theta power showed a positive cluster between 124 to 508 which covered the whole scalp ($T=2.313$; $p=0.00009$). No significant clusters were found in the beta range.

Similarly to during playing, listening to the replay of a mistuned note also elicited a frontal negativity (f-ERN) followed by a parieto-occipital negativity (N-340) and a frontal positivity in the ERPs (resembling the early component of the P3, the P3a), but no late central positivity (P3b). During listening, the responses to mistuned notes differed significantly from the responses to tuned notes between 206-460 ms at frontal, central, and parietal electrodes (see Fig. 5, panel B). Specifically, we found a negative cluster in the time period between 206 and 300ms ($T=-1.962$; $p=0.0009$) containing frontal, central and parietal electrodes resembling the f-ERN which was followed by a right centro-parietal negativity that ended at 300 ms resembling the N-280. Another one between 332 and 446ms ($T=2.013$; $p=0.0004$) containing parietal, occipital, and left-temporal electrodes resembling the N-340 maximum at Pz. We also found a positive cluster between 346 and 496ms ($T=1.877$; $p=0.0013$) containing frontal and central electrodes resembling the P3a. We also observed an increase in theta power on frontal midline electrodes reaching its greatest amplitude in FCZ at 300 milliseconds (Fig 5, panel B). However, no significant results were yielded by the cluster-based analysis. No significant results were found at Beta power either. The tuning effects appeared less robust during listening than during playing (Fig 6). However, the direct contrast of the tuning effects (mistuned - tuned difference wave) during playing compared to during listening yielded only significant differences between 466 and 560ms at central electrodes ($T=1.083$; $p=0.027$) corresponding to the absent P3b in the replayed condition. Playing a mistuned note also elicited greater midfrontal theta responses than listening to the replay of the mistuned note. Cluster-based analysis in theta showed a positive cluster between 140 and 374ms at frontal and central electrodes ($T=471$; $p=0.017$). Finally, a paired sample t-test in beta power between active (tuned minus mistuned) and replayed (tuned minus mistuned) notes showed significant differences at C4 between 606 and 806ms (active < replayed; $T=3.460$; $p=0.004$).

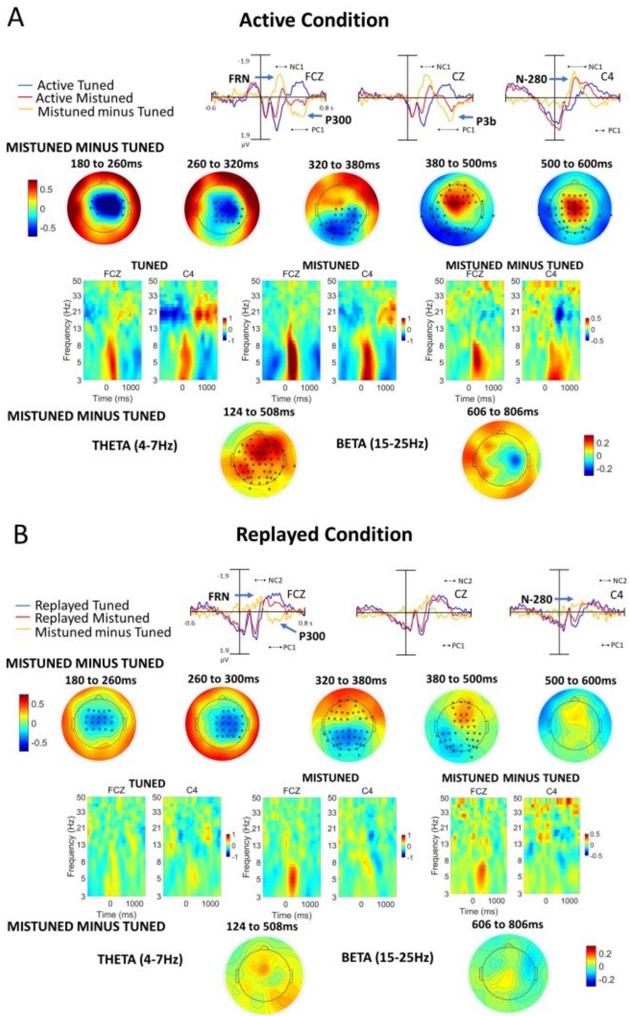

Figure 5. Effects of producing a mistuned note (left) and listening to the replay of a mistuned note (right). Panel A: Active Condition. Panel B: replayed Condition. From up to down: - 1. ERPs elicited to tuned (blue) and mistuned (red) notes and difference waves (yellow) for a subset of electrodes. The lines below the ERP of each electrode show the significantly different time-periods according to the cluster analysis. NC1: Negative Cluster 1. PC1: Positive Cluster 1. NC2: Negative Cluster **2.** Scalp maps (mistuned – tuned difference). The marks correspond to the significant electrodes in this time period according to cluster analysis-ERSP plots of tuned, mistuned and difference on FCz and C4. **3.** Scalp plots of ERSP (mistuned – tuned difference) in theta and beta frequency ranges. The marks correspond to the significant electrodes in this time period according to cluster analysis.

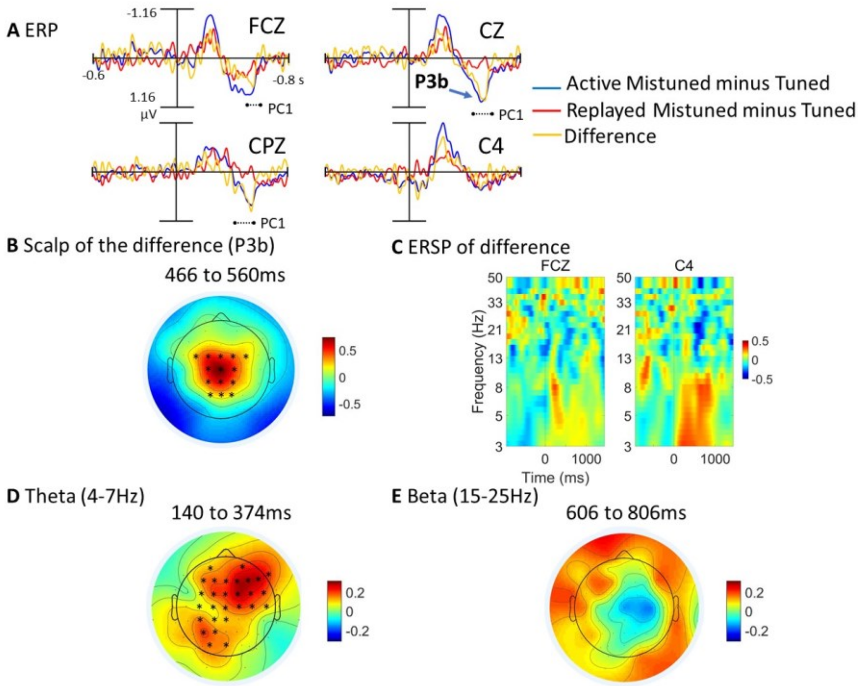

Figure 6. Contrast of tuning effects in the active vs replayed condition. A ERPs Active Mistuned – tuned and replayed Mistuned minus Tuned on electrodes FCz, Cz, CPz C4. The lines below the ERP of each electrode show the significantly different time-periods according to the cluster analysis. PC1: Positive Cluster 1. **B** Scalp maps of the difference for the P3b time-periods (466 to 560ms). The marks correspond to the significant electrodes in this time period according to cluster analysis. **C** TF power plots of difference on FCz and C4. **D and E.** Scalp plots of theta power and beta power. Active minus replayed difference from 140 to 374 ms for theta, and 606 to 806 ms for Beta. The marks correspond to the significant electrodes in this time period according to cluster analysis.

We found that the f-ERN, P3b, N340, and midfrontal theta seemed to be modulated by the magnitude of the error (see Figure 7). The amplitude of the components tended to grow with the error. The repeated measures analysis in the amplitude of the f-ERN (computed as the average value between 200 and 280 ms of the mistuned minus tuned difference) yielded a significant effect of Magnitude, $F(1.87) = 7.43$, $p = .003$, eta squared = .21. A simple main effect analysis yielded a significant effect of Magnitude for the active condition, $F(2) = 3.71$, $p < .037$, eta squared = .485. The repeated measures analysis in the amplitude of the N280 (computed as the average value between 260 and 320), and the P3a (computed as the average value between 360 and 440 ms) did not yield any significant effect The repeated measures analysis in the amplitude of the P3b (computed as the average value between 460 and 560 ms) yielded a significant effect of Magnitude, $F(2) = 4.91$, $p = .015$, eta squared = .260 and of Condition, $F(1) = 20.94$, $p < .0001$, eta squared = .617. Finally, the repeated measures analysis in the N340 yielded a significant effect of Magnitude, $F(2) = 7.74$, $p = .002$, eta squared = .356, and an interaction Condition*Magnitude $F(2) = 3.84$, $p = .034$, eta squared = .215. Simple main effect analysis showed a significant effect of Magnitude in the Active condition, $F(2) = 8.61$, $p = .001$, eta squared = .381. Pairwise comparisons for the active condition showed significant differences between the 30-50 range and the >70 range ($p = .006$). A simple main effect analysis did not yield significant effects of Magnitude in the replayed condition. The repeated measures analysis for the beta rebound yielded significant effects of Condition, $F(1) = 10.95$, $p = .005$, eta squared = .439, but not of Magnitude. Finally, the repeated measures analysis of the midfrontal theta yielded significant effects of Condition $F(1) = 8.71$, $p = .01$, eta squared = .384, and of Magnitude $F(2) = 6.57$, $p = .005$, eta squared = .320. Pairwise comparisons showed significant differences between the 30-50 range and the >70 range ($p = .006$).

6.7.2.5 Corrective movements

To assess whether the onset of corrective movements could have influenced the latency of our components of interest we computed the latency of the maximum peak of each component for the ERPs of notes which were followed by fast, medium, and slow corrections. The latency of the f-ERN was computed as the minimum peak between 0 and 400 ms in the FCZ electrode for each type of corrective movement (Slow [<250 ms], Medium [250 to 350 ms] and Fast [>350 ms], Condition (Active and Replayed) and participant after subtracting the tuned notes from the signal (mistuned minus tuned difference). Similarly for the N280 (minimum value between 200 and 500 ms at C4), the N340 (minimum value between 200 and 500 ms at Pz), the P3a (maximum value between 150 and 450 at FCZ), the P3b (maximum value between 450 and 600 ms at CZ), the midfrontal theta (maximum value between 0 and 600 at FCZ), and beta rebound (minimum value between 500 and 1000 ms). The latency of corrective movements did not seem to exert any influence in the latency of the components in the Active condition.

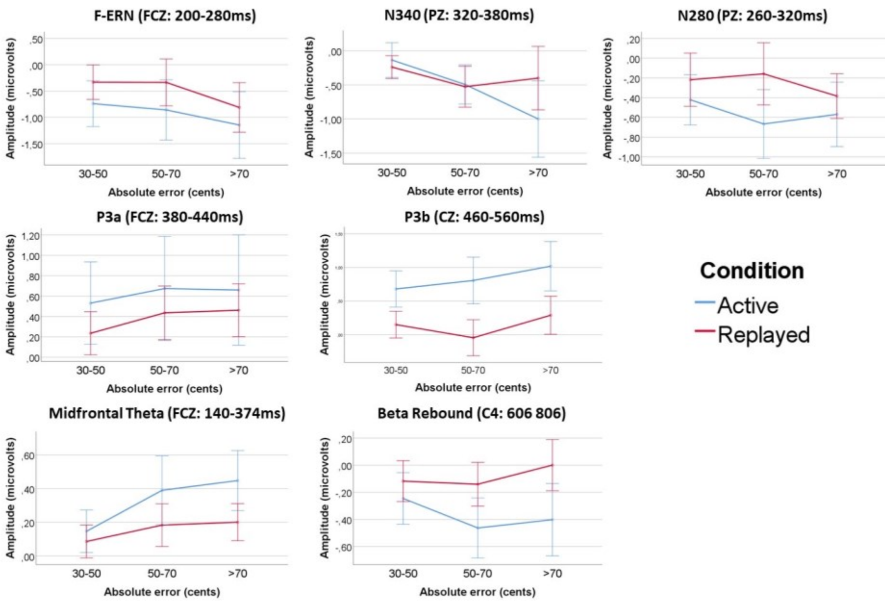

Figure 7. Effects of the magnitude of the error in the amplitude of the components (mistuned minus tuned). We found significant increments between the amplitude of the FRN, P3b, N340 and midfrontal theta components and the magnitude of the produced or heard error.

We found that the latency of corrective movements modulated the amplitude of some of the components analyzed. The amplitude of the f-ERN tended to be higher for fast corrections than for slow corrections. On the contrary, the amplitude of the N280, the N340, and midfrontal theta tended to be higher for slow corrections than fast corrections. These effects were only seen in the active condition. On the other hand, the amplitude of the P3b reflected lower amplitudes for slow corrections than for medium correction in both conditions. Both the amplitude of the P3b and the N300 were significantly bigger for the active condition. The repeated measures analysis for the f-ERN yielded a significant

Condition*Speed of correction interaction, $F(2) = 7.45$, $p = .003$, eta squared = .348. Simple main effect analysis yielded a significant effect of Speed of correction only for the Active condition, $F(2) = 8.78$, $p = .001$, eta squared = .386. Pairwise comparisons yielded significant effects between fast and slow corrections, slow corrections > fast corrections, $p = .008$, and between medium and slow corrections, slow corrections > medium corrections, $p = .046$. The repeated measures analysis for the N300 yielded a significant main effect of Condition, Active < Replayed, $F(1) = 5.94$, $p = .029$, eta squared = .298, and a Condition*Speed of response interaction, $F(2) = 3.8$, $p = .034$, eta squared = .214. Simple main effect analysis yielded only significant effects of Speed of correction for the active condition, $F(2) = 5.57$, $p = .009$, eta squared = .285. Pairwise comparisons yielded significant effects between fast and slow corrections, slow corrections < fast corrections, $p = .018$. The repeated measures analysis for the N340 yielded a significant Condition*Speed of correction interaction, $F(2) = 3.39$, $p = .048$, eta squared = .195. Simple main effect analysis yielded a significant effect of Speed of correction only for the Active condition, $F(2) = 4.96$, $p = .014$, eta squared = .262. The repeated measures analysis for the midfrontal theta yielded a significant effect of Speed of correction, $F(2) = 5.05$, $p = .013$, eta squared = .265. Simple main effect analysis yielded only significant effects of Speed of correction only for the Active condition, $F(2) = 3.42$, $p = .034$, eta squared = .197. Pairwise comparisons yielded significant differences between fast and medium corrections, medium corrections > fast corrections, $p < .0001$. Finally, the repeated measures analysis for the P3b yielded a significant main effect of Condition, Active > Replayed, $F(1) = 18.43$, $p = .001$, eta squared = .568, and a significant main effect of Speed of correction, $F(2) = 4.8$, $p = .016$, eta squared = .566. Pairwise comparison yielded significant differences between slow and medium corrections, medium corrections > slow corrections, $p = .003$.

Although the N280 showed to have its amplitude modulated by the onset of corrective movements it did not show a clear maximum peak as the other components. It seemed to extend longer in time in slow corrections than in fast corrections and seemed to maintain its amplitude until the onset of the corrective movement (see Figure 8).

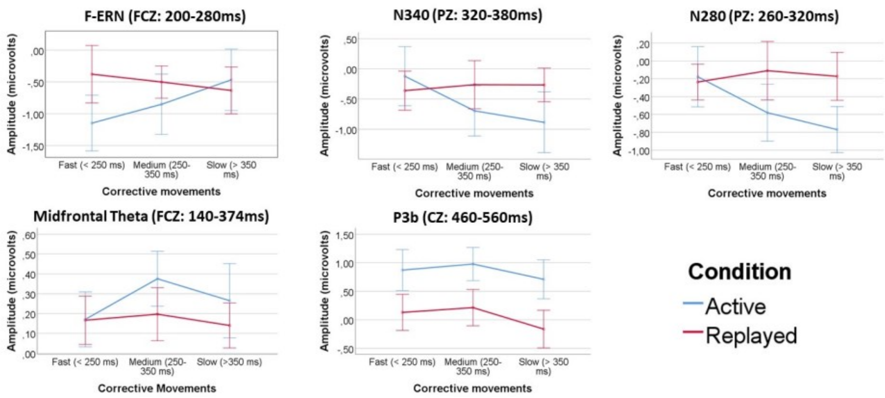

Figure 7. Effects of the latency of corrective movements in the amplitude of the components (mistuned minus tuned). We found that the latency of corrective movements modulated the amplitude of the components in the active condition. Specifically, the amplitude of the f-ERN tended to be higher for fast corrective movements (< 250 ms) than for slow corrective movements (> 350 ms). On the contrary, The N340, N280, and midfrontal theta tended to show higher amplitudes for slow corrective movements than fast ones. The P3b component seemed to show less amplitude after slow corrective movements than after fast ones.

6.8 Discussion

6.8.1 Behavioral

In this work, we have studied the effects of self-generation in a string instrument both with violinists and cellists. We have also studied the effects of note manipulation and online correction in melody production.

String players tended to make a greater error (around 30 cents in nonmanipulated notes) than that found in other experiments with string instruments which reported errors from less than 7 cents (Hopkins, 2015) to 15 cents (Blanco et al., 2021). However, the task in this experiment was more complex than those evaluated in previous works. First, in this work we were evaluating melody production and no single pitch-matching which imposes higher cognitive demands such as planning a sequence of movements, playing them in tempo and constant monitoring of movements to

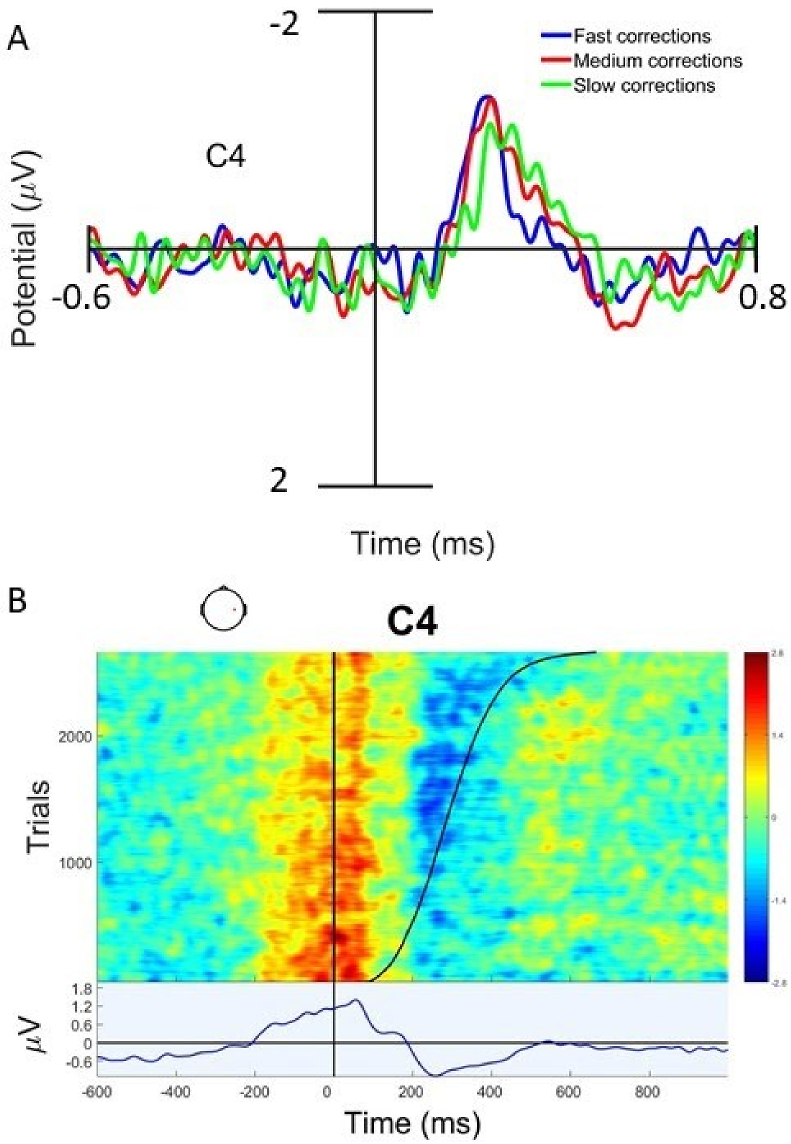

Figure 7. Effects of the latency of corrective movements in the amplitude of the N280. (A) The ERP of the mistuned minus tuned difference at C4 for notes that were followed by slow, medium, and fast corrections. **(B)** Trials of all the participants were sorted with respect to the latency of error correction and smoothed with a 100-trials moving average.

respond and produce online correction of errors. On the other hand, both violinists and cellists were playing in an unusual position they are used to, which, despite the three blocks of practice before the experiment, could have influenced the results. Finally, the large number of manipulated notes present in the task (1 out of 3) could have made it difficult to create the auditory-motor associations corresponding to the new situation.

As expected, the manipulated notes showed an error in cents greater than that of the nonmanipulated notes and also the post-manipulated ones. No differences were found in the level of error of the non-manipulated and post-manipulated notes, suggesting that pitch manipulation did not exert any readaptation effect in our participants. That is, that after a flat or sharp manipulation participants could decide to plan the finger location of the following note in the fingerboard in a more flat or sharp location expecting to find the correct note there. This could have led to bigger amounts of errors in post-manipulated notes, an effect that was not seen in our study.

The participants were instructed to correct all those notes that they perceived as erroneous, regardless of whether it was their own error or a manipulation of the experiment. We did not find any type of significant differences in our participants' behavior in relation to corrective movements between nonmanipulated, manipulated, and post-manipulated notes and seemed independent of the magnitude of the error. The participants started their corrective movements around 300ms and finished them around 788ms. Also, in this study, participants tended to stop their corrective movements around 20 cents of error. This led us to the conservative decision to consider only notes between 0 and 15 cents as "in tune" and those above 30 cents as "out of tune" for posterior analysis.

6.8.2 Error monitoring potentials

Although there is mention in previous literature about possible electrophysiological differences between self-generated errors or errors due to external manipulation (Herrojo-Ruiz et al., 2009, Maidhof et al., 20XX) a cluster-based analysis carried out with both the manipulated, non-manipulated, and post-manipulated notes both tuned and out of tune responses in the active and replayed conditions did not yield any type of significant differences in the resulting ERPs (see *6.7.11 Supplementary Materials*). This led us to include manipulated, nonmanipulated, and post-manipulated in a single set of tuned and out-of-tune notes regardless of whether they had been manipulated or not. As we expected, self-generated errors also elicited an f-ERN as manipulated errors. Taking into account that fretless instruments (e.g. violin) are much more dependent on auditory feedback than fretted instruments (e.g. piano) it may be expected to find that self-generated errors elicited an f-ERN like manipulated notes. In addition, the inter-onset interval between notes in this experiment was 2 seconds which is much longer than those used by Maidhof et al and Herrojo-Ruiz et al. (around 125ms) making the appearance of an r-ERN very unlikely since it is normally reported in speeded response tasks (Maidhof, Rieger, Prinz & Koelsch, 2009; Maidhof, Pitkäniemi, Tervaniemi, 2013; Strübing, Ruiz, Jabusch & Altenmüller, 2011; Herrojo-Ruiz, Jabusch & Altenmuller (2009), for a review see Gehring, Liu, Orr & Carp, 2012).

In both the active and replayed conditions we have found effects of activity related to f-ERN and P300 as previously reported (Maidhof et al, 2009; Herrojo-Ruiz et al., 2009). Despite the fact that on average, the amplitude of the f-ERN seemed greater in the active than in the replayed condition, this difference was not statistically significant as compared with previous studies with the piano (Maidhof et al, 2009). Maidhof et al. argued that the f-ERN is modulated by expectancies and expectancies during performing are stronger than during listening. According to Maidhof et al. in the active conditions, two different kinds of expectancies were violated: the preceding musical context (which is shared in passive conditions) and a specific auditory effect on the basis of their intention, while in the passive condition, it is only the first one which is violated. Taking into account that in this experiment replayed melodies were reproduced immediately after the active condition we should add the expectancy of knowing in advance which note of the melody is going to present the error or the manipulation which is not present in the active condition.

One possible explanation for the previous results may be the activity of other components that overlapped with the f-ERN in both, or in the passive condition. To ensure the correct performance of participants, in this experiment, we decided to use scores. However, this decision could have elicited an imaginary Mismatch Negativity (iMMN) (Yumoto et al., 2005, Katahira et al., 2008). The iMMN is supposed to reflect the discrepancy between the note of the score and the heard sound and is elicited between 150 and 200ms. Future studies should address this issue by removing possible influences in the use of a score when studying the error-monitoring processes of playing a musical instrument.

We found that the f-ERN component seemed to be directly related both to the magnitude of the errors and the latency of corrective actions. Greater errors tended to elicit greater f-ERN amplitudes than lower errors, and notes which were followed by faster corrective movements elicited a greater f-ERN amplitude than notes followed by slower corrective movements. Considering that the major errors may tend to be more unexpected than the minor errors and that, in addition, a greater part of them were the result of external manipulations, it is predictable to expect greater amplitudes of the f-ERN after major errors (Goyer et al., 2008). On the other hand, it is also expected that major errors are more clearly perceived than minor errors, causing earlier corrective responses (Hewig et al., 2011). This would explain the results obtained. However, in the behavioral results of this study we found that although the lowest errors certainly tended to suffer fewer corrections than the highest errors, these results did not tend to be very different from errors greater than 30 cents (which was the threshold chosen for this study), nor did they seem to present differences in the latency of the onset of corrective movements or in their offset. This could lead us to reverse the direction of causality and suggest that the f-ERN could be related to the awareness that an error has been made. Previous studies have shown that the amplitude of the f-ERN can be modulated by cognitive load (Krigolson et al., 2011) suggesting that not only expectancies but also attentive processes could modulate the f-ERN amplitude.

It is thought that the f-ERN, rERN, P3a and midfrontal theta share a common neural generator in the ACC (Bernat et al., 2005; Cavanagh et al., 2009; Gehring & Willoughby, 2004; Hall et al., 2007; Trujillo & Allen, 2007, Gehring, Liu, Orr & Carp, 2012), or even that the rERN and P3a are both parts of a single oscillatory potential at the theta band (Herrmann et al., 2004; van Veen &

Carter, 2002). However, recent research on response correction by moving a mouse cursor reported that while the rERN remained invariant across fast and slow error corrections the latency of midfrontal theta was related to the onset of corrective movements (Kieffaber et al., 2016). In our experiment, the midfrontal theta activity peaked at 313 ms which seemed to coincide with the average onset of corrective movements (~300 ms). However, unlike Kieffaber et al. we found that its latency was not related to the latency of corrective movements and was maintained in the same temporal location. On the other hand, its amplitude, unlike the f-ERN, tended to increase for corrective movements with higher latencies. This effect was also found for other components that may be related to the corrective movements and will be described below.

We have also found two late-positivities after mistuned notes resembling the P3a and the P3b components. The P3a appeared around 380 and 460 ms with a frontal distribution while the P3b between 460 and 560 ms with a central distribution. The P3b seemed absent in the replayed condition, probably due to the fact that the participants were not instructed to detect errors. The latency of these components was greater than that reported by Maidhof et al. (2009) which was, as usual, around 300 ms for the P3a and 400 ms for the P3b (Arbel & Donchin, 2009; Ruchsow et al., 2005b; van Veen & Carter, 2002). We hypothesize that this result is due to both the possibility or existence of corrective movements after a mistuned note. In our experiment, participants tended on average to start their corrective movements around 300 ms, almost 100 ms before the P3a. Recent research on response correction by moving a mouse cursor reported a Pe-like positivity that consistently followed the response correction 100 ms later (Kieffaber et al., 2016). However, we did not find any effect of the latency of the corrective movements in the latency of the P3a and the P3b, although the amplitude of the P3b seemed, in fact, being modulated by the latency of corrective movements. Slow corrective movements tended to elicit lower P3b amplitudes. This is consistent with our previous interpretation that slow corrective movements represented more “unsure” errors than fast corrective movements. While the P3b has been associated with awareness or affective response to an error (Overbeek et al. 2005), the P3a seems to be unrelated to error awareness (Endrass et al., 2007). Consistent with this interpretation we also found that the P3b, but not the P3a, was modulated by the magnitude of the error.

Between the f-ERN and the P3a we found a central-right negativity which extended from 200 to 380 ms with a maximum peak at 280 ms in C4 that we called the N280. Considering that participants started on average their corrective movements at 300 ms which consisted of a displacement of the left-hand finger, it is possible that this negativity could be an MRCP related to the onset of the corrective movements. The lack of a clear and defined peak for this component could be due to the fact that the corrective moves were not time-locked with the onset of the note. In addition, we could also see how its amplitude tended to decrease just after the corrective movement had started. Similar behavior has been reported for the ERN, which has come to assume the function of an "alarm system" that does not turn off until error remediation starts (Burle et al., 2008; Kieffaber et al., 2016). We did not find this behavior in the Replayed condition where the same component appeared although with lower amplitude in slow corrections.

We also found a parietal negativity peaking at 340 ms in Pz that we termed the N340. As with the N280, this component has not tended to be reported in previous studies involving discrete responses (with the exception of Leuthold & Jentszch, 2002). Discrete response tasks usually elicit reward errors, but not sensory-prediction errors (Krakauer et al., 2019; Krigolson and Holroyd, 2006). Krigolson and Holroyd (2005, 2007) reported a parietal negativity following the f-ERN peaking at 360 ms after errors in a visual-tracking task. According to the authors, high-level error information evaluated within the medial frontal cortex (reward errors) was communicated to the PPC for the adaptive modification of behavior (sensory-prediction errors). This view is supported by recent research in cello performance and singing which found that compensating pitch-shifting manipulations recruited parietal areas, in particular, the IPS and the SMG (Segado et al., 2021). We cannot ensure that the neural generator of our N340 was located in parietal areas. However, although its latency was not modulated by corrective movements its amplitude tended to be higher for trials where the latency of corrective movements was higher just as the results found for midfrontal theta and the N280.

In summary, one possible interpretation of these results is that we have found different components that may be related to the functioning of two error monitoring systems. The “fast” system, mediated by the medial frontal cortex, detects the error with a certain degree of certainty (reflected by the amplitude of the f-ERN). If the certainty is very high, it initiates the corrective action and, because the action has already been initiated, the activity of the “slow” system becomes redundant (reflected in a lower midfrontal theta and N340 amplitude). However, when the uncertainty is high (reflected in a lower f-ERN amplitude), other processes start accumulating evidence that an error has been made until the “slow” system reacts (higher midfrontal theta and N340 amplitude) and initiates the corrective movements. Future work should study better the effects of corrective movements in our signal by employing a condition where participants are instructed to ignore errors.

6.8.3 Self-generated effects in N1

Given that we found that the activity of the motor potentials did not seem to interfere with the 3 auditory components, we studied possible effects of self-generation in the relative amplitude of N1 in relation to P1 and P2. We did not find any effect related to the decrease in N1 due to corollary discharge (Ford et al., 2001). The reason why we do not see a reduction in N1 may be due to multiple factors (Horvath, 2015; Hughes, Desantis & Waszak, 2013). In most of the studies where the attenuation effect has been found, self-generated sounds were not relevant for the subjects. Several studies have reported that attention in self-generated sounds generated an enhancement effect of the N1 that could overlap or interact with the suppression effect (Timm et al., 2013; Kok Rahnev, de Lange, 2012). Finally, some studies reported that musicians have more sensitivity to the timbre of their own instrument, reflected in a greater N1 amplitude (Pantev et al., 2001). In our experiment, participants were violinists and cellists and, despite both playing a stringed instrument, the characteristic timbre of the violin could have influenced the results obtained in the N1 amplitude.

6.8.4 Beta Rebound

We found the effects of beta suppression and beta rebound before and after the onset of tuned notes in the active condition. The beta rebound had a maximum peak at 706 ms in tuned notes and was suppressed after mistuned notes as reported (Tan et al. 2014; Torrecillos et al 2015). The beta rebound was independent of online corrections and insensitive to the magnitude of the error. These results are consistent with those reported by Torrecillos et al (2014) suggesting that beta-rebound may reflect salience processing independent of sensorimotor adaptation.

6.9 Conclusions

In this work, we have extended previous studies done on the piano to string instruments. Music production and intonation monitoring with those instruments are much more dependent on auditory feedback and they allow us to better study error monitoring processes in contexts where pitch correction is possible. In addition, we have validated a setup made to be able to collect EEG data during violin performance while introducing pitch manipulations with the possibility of extending it to future research.

6.10 Reference

Baumann, S., Koeneke, S., Schmidt, C. F., Meyer, M., Lutz, K.,Jancke, L. (2007). A network for audio-motor coordination in skilled pianists and non-musicians. Brain Research, 1161(1), 65–78. https://doi.org/10.1016/j.brainres.2007.05.045

Bernat, E. M., Williams, W. J., & Gehring, W. J. (2005). Decomposing ERP time–frequency energy using PCA. *Clinical neurophysiology*, *116*(6), 1314-1334.

Bishop, L., Bailes, F., & Dean, R. T. (2013). Musical expertise and the ability to imagine loudness. *PloS one*, *8*(2), e56052.

Blanco, A. D., Tassani, S., & Ramirez, R. (2021). Effects of Visual and Auditory Feedback in Violin and Singing Voice Pitch Matching Tasks. *Frontiers in Psychology*, *12*, 2578.

Brainard, D. H. (1997) The Psychophysics Toolbox, Spatial Vision 10:433-436. [PDF]

Brown, R. M., Palmer, C. (2012). Auditory-motor learning influences auditory memory for music. Memory and Cognition, 40, 567–578. https://doi.org/10.3758/s13421-011-0177-x

Burle, B., Roger, C., Allain, S., Vidal, F., & Hasbroucq, T. (2008). Error negativity does not reflect conflict: a reappraisal of conflict monitoring and anterior cingulate cortex activity. *Journal of cognitive neuroscience*, *20*(9), 1637-1655.

Cavanagh, J. F., Cohen, M. X., & Allen, J. J. (2009). Prelude to and resolution of an error: EEG phase synchrony reveals cognitive control dynamics during action monitoring. *Journal of Neuroscience*, *29*(1), 98-105.

Chen, J. L., Rae, C., Watkins, K. E. (2012). Learning to play a melody: An fMRI study examining the formation of auditory-motor associations. NeuroImage, 59, 1200 1208. https://doi.org/10.1016/j.neuroimage.2011.08.012

Chen, Jessie, Marjorie Woollacott, Steve Pologe, and George P. Moore. 2013. "Stochastic Aspects of Motor Behavior and Their Dependence on Auditory Feedback in Experienced Cellists." *Frontiers in Human Neuroscience* 7 (July). https://doi.org/10.3389/fnhum.2013.00419.

Culham, J. C., Danckert, S. L., De Souza, J. F., Gati, J. S., Menon, R. S., & Goodale, M. A. (2003). Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Experimental brain research*, *153*(2), 180-189.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in cognitive sciences*, *4*(11), 423-431.

Desmurget, M., Epstein, C. M., Turner, R. S., Prablanc, C., Alexander, G. E., & Grafton, S. T. (1999). Role of the posterior parietal cortex in updating reaching movements to a visual target. *Nature neuroscience*, *2*(6), 563-567.

Desmurget, M., Gréa, H., Grethe, J. S., Prablanc, C., Alexander, G. E., & Grafton, S. T. (2001). Functional anatomy of nonvisual feedback loops during reaching: a positron emission tomography study. *Journal of Neuroscience*, *21*(8), 2919-2928.

Di Russo, F., Pitzalis, S., Aprile, T., & Spinelli, D. (2005). Effect of practice on brain activity: an investigation in top-level rifle shooters. *Medicine and science in sports and exercise*, *37*(9), 1586.

Diedrichsen, J., Verstynen, T., Lehman, S. L., & Ivry, R. B. (2005). Cerebellar involvement in anticipating the consequences of self-produced actions during bimanual movements. *Journal of neurophysiology*, *93*(2), 801-812.

Donchin, E., & Arbel, Y. (2009, July). P300 based brain computer interfaces: a progress report. In *International Conference on Foundations of Augmented Cognition* (pp. 724-731). Springer, Berlin, Heidelberg.
Donchin, E., & Coles, M. G. (1998). Context updating and the P300. *Behavioral and brain sciences*, *21*(1), 152-154.

Endrass, T., Reuter, B., & Kathmann, N. (2007). ERP correlates of conscious error recognition: aware and unaware errors in an antisaccade task. *European Journal of Neuroscience*, *26*(6), 1714-1720.

Falkenstein, M., Koshlykova, N. A., Kiroj, V. N., Hoormann, J., & Hohnsbein, J. (1995). Late ERP components in visual and auditory Go/Nogo tasks. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *96*(1), 36-43.

Fattapposta, F., Amabile, G., Cordischi, M. V., Di Venanzio, D., Foti, A., Pierelli, F., ... & Morrocutti, C. (1996). Long-term practice effects on a new skilled motor learning: an electrophysiological study. *Electroencephalography and clinical Neurophysiology*, *99*(6), 495-507.

Finney, J. (1999). The rights and wrongs of school music: considering the expressivist argument and its existential component. *British Journal of Music Education*, *16*(3), 237-244.

Finney, S., & Palmer, C. (2003). Auditory feedback and memory for music performance: Sound evidence for an encoding effect. *Memory & cognition*, *31*(1), 51-64.

Ford, J. M., Mathalon, D. H., Heinks, T., Kalba, S., Faustman, W. O., & Roth, W. T. (2001). Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *American Journal of Psychiatry*, *158*(12), 2069-2071.

Gehring, W. J., & Willoughby, A. R. (2004). Are all medial frontal negativities created equal? Toward a richer empirical basis for theories of action monitoring. In *Errors, conflicts, and the brain. Current opinions on performance monitoring* (Vol. 14, p. 20).

Gehring, W. J., Liu, Y., Orr, J. M., & Carp, J. (2012). The error-related negativity (ERN/Ne).

Goyer, J. P., Woldorff, M. G., & Huettel, S. A. (2008). Rapid electrophysiological brain responses are influenced by both valence and magnitude of monetary rewards. *Journal of cognitive neuroscience*, *20*(11), 2058-2069.

Hajcak, G., Holroyd, C. B., Moser, J. S., & Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. Psychophysiology, 42(2), 161-170.

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2007). It's worse than you thought: The feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology*, *44*(6), 905-912.

Hall, J. R., Bernat, E. M., & Patrick, C. J. (2007). Externalizing psychopathology and the error-related negativity. *Psychological science*, *18*(4), 326-333.

Hatta, A., Nishihira, Y., Higashiura, T., Kim, S. R., & Kaneda, T. (2009). Long-term motor practice induces practice-dependent modulation of movement-related cortical potentials (MRCP) preceding a self-paced non-dominant handgrip movement in kendo players. *Neuroscience Letters*, *459*(3), 105-108.

Heinks-Maldonado, Theda H., Daniel H. Mathalon, Max Gray, and Judith M. Ford. 2005. "Fine-Tuning of Auditory Cortex during Speech Production." Psychophysiology 42 (2): 180–90. https://doi.org/10.1111/j.1469-8986.2005.00272.x

Heldmann, M., Rüsseler, J., & Münte, T. F. (2008). Internal and external information in error processing. *BMC neuroscience*, *9*(1), 1-8.

Herrmann, M. J., Saathoff, C., Schreppel, T. J., Ehlis, A. C., Scheuerpflug, P., Pauli, P., & Fallgatter, A. J. (2009). The effect of ADHD symptoms on performance monitoring in a non-clinical population. *Psychiatry research*, *169*(2), 144-148.

Hewig, J., Kretschmer, N., Trippe, R. H., Hecht, H., Coles, M. G., Holroyd, C. B., & Miltner, W. H. (2011). Why humans deviate from rational choice. *Psychophysiology*, *48*(4), 507-514.

Highben, Z., & Palmer, C. (2004). Effects of auditory and motor mental practice in memorized piano performance. *Bulletin of the Council for Research in Music Education*, 58-65.

Holroyd, C. B., & Krigolson, O. E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology*, *44*(6), 913-917.

Hopkins, M. T. (2015). Eighth-grade violinists' instrument tuning ability: A comparison of pitch perception and tuning accuracy. *Journal of Research in Music Education*, *63*(3), 349-368.

Horváth, J. (2015). Action-related auditory ERP attenuation: Paradigms and hypotheses. *Brain Research*, *1626*, 54-65.

Hughes, G., Desantis, A., & Waszak, F. (2013). Attenuation of auditory N1 results from identity-specific action-effect prediction. *European Journal of Neuroscience*, *37*(7), 1152-1158.

Jones, Jeffery A., Nichole Scheerer, and Anupreet Tumber. 2013. "The Relationship between Vocal Pitch Feedback Error and Event-Related Brain Potentials." Proceedings of Meetings on Acoustics 19 (1): 060151. https://doi.org/10.1121/1.4800544.

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, *111*(10), 1745-1758.

Katahira, K., Abla, D., Masuda, S., & Okanoya, K. (2008). Feedback-based error monitoring processes during musical performance: An ERP study. *Neuroscience research*, *61*(1), 120-128.

Keller, P. E., Dalla Bella, S., & Koch, I. (2010). Auditory imagery shapes movement timing and kinematics: Evidence from a musical task. *Journal* *of Experimental Psychology: Human Perception and Performance*, *36*(2), 508.

Kieffaber, P. D., Hershaw, J., Sredl, J., & West, R. (2016). Electrophysiological correlates of error initiation and response correction. *NeuroImage*, *128*, 158-166.
Kita, Y., Mori, A., & Nara, M. (2001). Two types of movement-related cortical potentials preceding wrist extension in humans. *Neuroreport*, *12*(10), 2221-2225.

Kleber, B., Friberg, A., Zeitouni, A., & Zatorre, R. (2017). Experience dependent modulation of right anterior insula and sensorimotor regions as a function of noise-masked auditory feedback in singers and nonsingers. *NeuroImage*, *147*, 97-110.

Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., & De Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral cortex*, *22*(9), 2197-2206.

Krakauer, J. W., Hadjiosif, A. M., Xu, J., Wong, A. L., & Haith, A. M. (2019). Motor learning. *Compr Physiol*, *9*(2), 613-663.

Krigolson, O. E., & Holroyd, C. B. (2006). Evidence for hierarchical error processing in the human brain. *Neuroscience*, *137*(1), 13-17.

Krigolson, O. E., Heinekey, H., Kent, C. M., & Handy, T. C. (2012). Cognitive load impacts error evaluation within medial-frontal cortex. *Brain research*, *1430*, 62-67.

Krigolson, O. E., Holroyd, C. B., Van Gyn, G., & Heath, M. (2008). Electroencephalographic correlates of target and outcome errors. *Experimental Brain Research*, *190*(4), 401-411.

Lappe, C., Lappe, M., & Keller, P. E. (2018). The influence of pitch feedback on learning of motor-timing and sequencing: A piano study with novices. *Plos one*, *13*(11), e0207462.

Leuthold, H., & Jentzsch, I. (2002). Spatiotemporal source localisation reveals involvement of medial premotor areas in movement reprogramming. *Experimental brain research*, *144*(2), 178-188.

Maidhof, C., Pitkäniemi, A., & Tervaniemi, M. (2013). Predictive error detection in pianists: a combined ERP and motion capture study. *Frontiers in Human Neuroscience*, *7*, 587.

Maidhof, C., Rieger, M., Prinz, W., & Koelsch, S. (2009). Nobody is perfect: ERP effects prior to performance errors in musicians indicate fast monitoring processes. *PLoS One*, *4*(4), e5032.

Masaki, H., & Sommer, W. (2012). Cognitive neuroscience of motor learning and motor control. *The Journal of Physical Fitness and Sports Medicine*, *1*(3), 369-380.

MATLAB (2010). *MATLAB Version 7.10.0 (R2010a)*. Natick, MA: The MathWorks Inc.

Mauch, M., Cannam, C., Bittner, R., Fazekas, G., Salamon, J., Dai, J., et al. (2015). "Computer-aided melody note transcription using the tony software: accuracy and efficiency," in *Proceedings of the First International Conference on Technologies for Music Notation and Representation*.

Mauch, M., and Dixon, S. (2014). "pyin: A fundamental frequency estimator using probabilistic threshold distributions," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 659–663. doi: 10.1109/ICASSP.2014.6853678

Miltner, W. H., Braun, C. H., & Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a "generic" neural system for error detection. *Journal of cognitive neuroscience*, *9*(6), 788-798.

Moore. 2008. "Pitch and Space Maps of Skilled Cellists: Accuracy, Variability, and Error Correction." *Experimental Brain*

Nunes-Silva M, Janzen TB, Rodrigues RG, Luz AR da. Sensory feedback in music performer–instrument interactions. *Psychology of Music*. June 2020. doi:10.1177/0305735620928397

Overbeek, T. J., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable components of error processing: On the functional significance of the Pe vis-à-vis the ERN/Ne. *Journal of psychophysiology*, *19*(4), 319-329.

Palidis, D. J., Cashaback, J. G., & Gribble, P. L. (2018). Neural Signatures of Reward and Sensory Prediction Error in Motor Learning. *BioRxiv*, 262576.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport*, *12*(1), 169-174.

Pelli, D. G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies, Spatial Vision 10:437-442. [PDF]

Pfordresher, P. Q. (2005). Auditory feedback in music performance: the role of melodic structure and musical skill. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1331.

Pfordresher, P. Q. (2008). Auditory feedback in music performance: The role of transition-based similarity. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 708.

Pfordresher, P. Q., & Chow, K. (2019). A cost of musical training? Sensorimotor flexibility in musical sequence learning. *Psychonomic bulletin & review*, *26*(3), 967-973.

Pfordresher, P. Q., Keller, P. E., Koch, I., Palmer, C., Yildirim, E. (2011). Activation of learned action sequences by auditory feedback. Psychonomic Bulletin and Review, 18, 544
549. https://doi.org/10.3758/s13423-011-0077-x

Pierce, L., Krigolson, O., Tanaka, J., & Holroyd, C. (2008). Reinforcement learning and the acquisition of perceptual expertise in ERPs. *Journal of Vision*, *8*(6), 475-475.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148.

Repp, B. H. (1999). Effects of auditory feedback deprivation on expressive piano performance. *Music Perception*, *16*(4), 409-438. *Research* 188 (4): 493–503. https://doi.org/10.1007/s00221-008-1380-2.

Ruchsow, M., Spitzer, M., Grön, G., Grothe, J., & Kiefer, M. (2005). Error processing and impulsiveness in normals: evidence from event-related potentials. *Cognitive brain research*, *24*(2), 317-325.

Ruiz, M. H., Jabusch, H. C., & Altenmüller, E. (2009). Detecting wrong notes in advance: neuronal correlates of error monitoring in pianists. *Cerebral cortex*, *19*(11), 2625-2639.

Scheerer, N. E., & Jones, J. A. (2018). The role of auditory feedback at vocalization onset and mid-utterance. *Frontiers in psychology*, *9*, 2019.

Scheerer, Nichole E., J. Behich, Hanjun Liu, and Jeffery A. Jones. 2013. “ERP Correlates of the Magnitude of Pitch Errors Detected in the Human Voice.” *Neuroscience* 240 (June): 176–85. https://doi.org/10.1016/j.neuroscience.2013.02.054.

Segado, M., Hollinger, A., Thibodeau, J., Penhune, V., & Zatorre, R. J. (2018). Partially overlapping brain networks for singing and cello playing. *Frontiers in neuroscience*, *12*, 351.

Segado, M., Zatorre, R. J., & Penhune, V. B. (2021). Effector-independent brain network for auditory-motor integration: fMRI evidence from singing and cello playing. *NeuroImage*, *237*, 118128.

Stewart, L., Verdonschot, R. G., Nasralla, P., Lanipekun, J. (2013). Action-perception coupling in pianists: Learned mappings or spatial musical association of response codes (SMARC) effect? Quarterly Journal of Experimental Psychology, 66, 37–50. https://doi.org/10.1080/17470218.2012.687385

Strübing, F., Ruiz, M. H., Jabusch, H. C., & Altenmüller, E. (2011). Error monitoring is altered within musician’s dystonia: evidence from ERP-based studies.

Tan, H., Jenkinson, N., & Brown, P. (2014). Dynamic neural correlates of motor error monitoring and adaptation during trial-to-trial learning. *Journal of Neuroscience*, *34*(16), 5678-5688.

Tan, H., Zavala, B., Pogosyan, A., Ashkan, K., Zrinzo, L., Foltynie, T., ... & Brown, P. (2014). Human subthalamic nucleus in movement error detection and its evaluation during visuomotor adaptation. *Journal of Neuroscience*, *34*(50), 16744-16754.

Timm, J., SanMiguel, I., Saupe, K., & Schröger, E. (2013). The N1-suppression effect for self-initiated sounds is independent of attention. *BMC neuroscience*, *14*(1), 1-11.

Torrecillos, F., Alayrangues, J., Kilavik, B. E., & Malfait, N. (2015). Distinct modulations in sensorimotor postmovement and foreperiod β-band activities related to error salience processing and sensorimotor adaptation. *Journal of Neuroscience*, *35*(37), 12753-12765.

Trujillo, L. T., & Allen, J. J. (2007). Theta EEG dynamics of the error-related negativity. *Clinical Neurophysiology*, *118*(3), 645-668.
Van Veen, V., & Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology & behavior*, *77*(4-5), 477-482.

Wright, D. J., Holmes, P. S., Di Russo, F., Loporto, M., & Smith, D. (2012). Differences in cortical activity related to motor planning between experienced guitarists and non-musicians during guitar playing. *Human movement science*, *31*(3), 567-577.

Yumoto, M., Matsuda, M., Itoh, K., Uno, A., Karino, S., Saitoh, O., ... & Kaga, K. (2005). Auditory imagery mismatch negativity elicited in musicians. *Neuroreport*, *16*(11), 1175-1178.

6.11 Appendix. Supplementary Materials

6.11.1 Online Tone Manipulation System in Max MSP

We developed a system en Max 7 (cite) to record the session, reproduce the reference melodies to the participant, manipulate certain notes of their performance, and replay each trial.

The system needs to read a previously generated text file that contains the information about which melody to reproduce, the number of the note pertaining to the melody which should receive the manipulation, and the direction of the manipulation. The information is expressed in the following form:

'file line number (from 1 to 144 in order)', 'number of the melody to be reproduced', 'onset to manipulate (from 2 to 4)', 'direction of the manipulation (0 is -50 cents, 1 is +50 cents)';

For example, if the first line of the file is "1,6,3,1;" it means that: we are reading the first line of the file, as expressed by the "1" in the first position; the melody to be reproduced is melody number 6 from a previously specified folder which contains all the reference melodies of the experiment in .wav format. The onset/note to be manipulated is number 3 from the melody (it can never be 1 because the first note is the open string note which has to be always in tune). And the direction of the manipulation +50 cents (sharp).

The system receives two independent audio channels (stereo) as input. One audio channel contains the sound recorded by the microphone inside the EEG chamber and the other contains audio clicks that, when detected by the system, controls its behavior. The first time the system detects an audio click, it opens, reads the first line of the text file, and reproduces the corresponding wave file. This corresponds to the Reference Condition of the experiment. The second time the system receives an audio click, it starts recording all the audio coming from the channel of the microphone. It also starts counting all the detected sound onsets from that audio channel. When the number of detected onsets is equal to the onset to be manipulated minus one and the energy of the previous onset has descended below a certain threshold, the system pitch-shifts all the incoming sound +50 or -50 cents until the energy of the onset to be manipulated descends again below the certain threshold (see Figure 1). This corresponds to the Active Condition of the experiment.

Finally, when the system receives the third audio click it stops recording the incoming audio from the channel of the microphone, saves it in a wave file, and reproduces it. This corresponds to the replayed Condition. We can see a flux diagram describing this process in Figure 2.

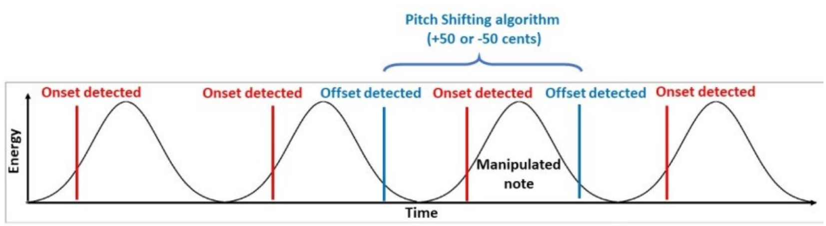

Figure 1. Example of the energy distribution of the Active Condition. Every time the energy reaches a certain threshold the system detects an onset. When the number of detected onsets is equal to the onset to be manipulated minus one and the energy has descended a certain threshold

(offset detected in the figure) the system pitch-shifts all the incoming sound +50 or -50 cents until another offset is detected.

The STIM PC controlled the order and the duration of the conditions in an automated way counting the number of seconds, sending the audio clicks to the PC in charge of the online tone manipulation system (DSP) and the EEG markers to the SCAN PC via parallel port. It also displayed the corresponding scores of the melodies on the screen of participants together with their visual instructions. The code was developed in Matlab (MATLAB, 2010) and with the aid of the Psychtoolbox (Brainard, 1997; Pelli, 1997).

In Figure 3 we can see the interface of the online tone manipulation system developed in Max MSP and a description of its components.

Outside the EEG chamber, where the experiment was carried out, a Behringer audio mixer was used to distribute audio signals among three computers. See Figure 4 for a whole description of the complete setup.

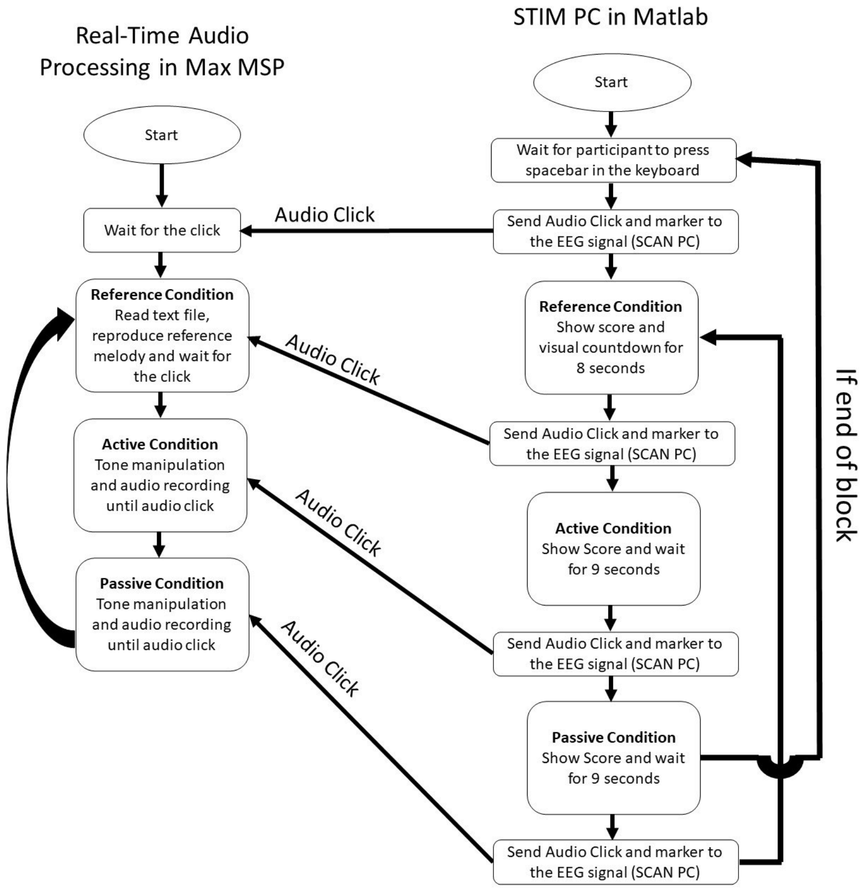

Figure 2. Flux Diagram of the process followed by the online tone manipulation system developed in Max and how it is controlled by the STIM PC in Matlab. The audio clicks were sent via the sound card of the STIM PC and the markers to the SCAN PC via parallel port.

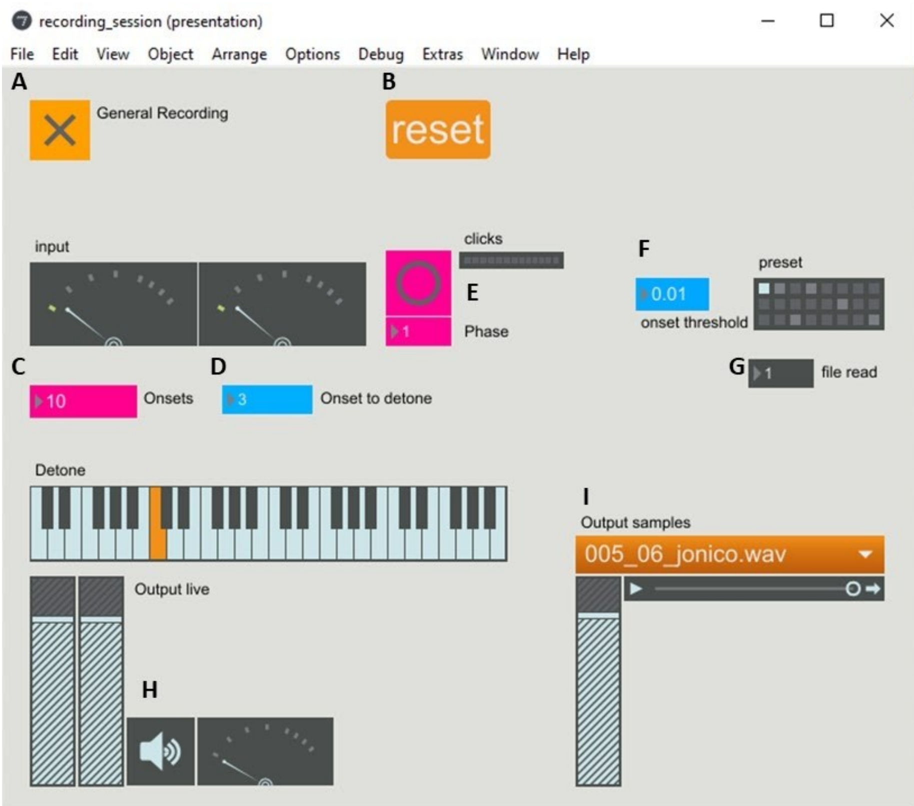

Figure 3: Interface of the Online Tone Manipulation System. A. Button for the general recording of the session. **B.** Reset Button. If pressed, all the parameters return to their default values. **C.** Number of detected onsets in the performance detected during the Active Condition. **D.** Number of the onset that will receive the tone manipulation. This information is extracted from a previously created .csv file. **E.** Audio click detector. If an audio click is detected through its respective channel the system changes the condition of the experiment: 1 is Reference Condition, 2 is Active Condition and 3 is replayed Condition. **F.** Onset threshold. This number adapts the sensitivity of the system to detect an onset from a note played by the participant. **G.** Current .csv file being read. The .csv file contains the information about the melody to be reproduced, the number of the manipulated note, and the direction of the manipulation. H. Output levels of the sound sent to the participant. **I.** The output level of the reproduced reference melody.

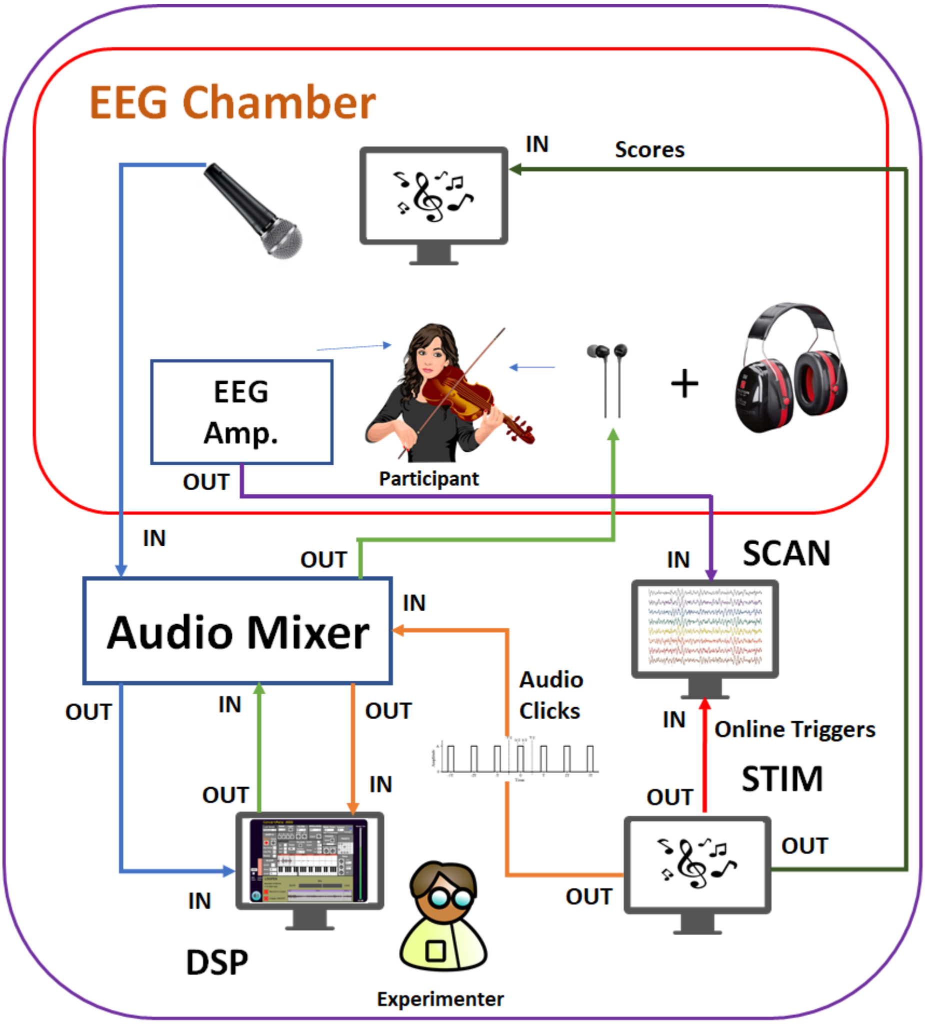

Figure 4. The final setup used for the experiment. The STIM PC controls the operations of the DSP PC via audio clicks. It also sends online triggers ('markers') to the SCAN PC and displays the corresponding scores of the melodies to the melodies of participants. The DSP PC is the one in charge of reproducing the reference melodies to the participant, manipulating certain notes of their performance, and replaying each trial to the participant.

6.11.2 Tables

	Mean	Range	Standard Deviation
Error nonmanipulated (cents)	33.31	20.59-70.45	16.21
Error manipulated (cents)	46.88	35.31-62.01	7.29
Error post manipulated (cents)	35.31	17.5-66.8	13.32
Error after corrective movement ALL (cents)	22.66	13.13-32.20	3.89
Onset corrective movement (ms)	299	256-342	18
Offset corrective movement (ms)	802	560-1035	95

Table 1. Mean, range and standard deviation of some of the most relevant measures extracted from the behavioral analysis.

	Mean	Range	Standard Deviation
Reference Melody	418.733	345-493	37.98
replayed Tuned (0-15)	102	42-158	31.21
replayed Mistuned (ALL) (>30)	146.73	114-238	28.56
Replayed Mistuned (15-30)	75.46	34-98	16.8

Replayed Mistuned (30-50)	63.53	46-82	9.28
Replayed Mistuned (50-70)	41.53	28-67	9.3
Replayed Mistuned (>70)	42.74	20-114	23.92
replayed Mistuned Manipulated	67.2	46-83	10.27
replayed Mistuned Postmanipulated	33.2	16-55	9.08
replayed Mistuned Nonmanipulated	47.4	20-109	22.36
Replayed Fast corrections	55.66	18-129	36.48
Replayed Medium corrections	49.73	16-81	20.58
Replayed Slow corrections	42.86	17-77	21.93
Active Tuned (0-15)	98.8	62-140	26.32
Active Mistuned (ALL) (>30)	140.73	95-231	35.67
Active Mistuned (15-30)	71.2	49-96	14.33
Active Mistuned (30-50)	59.8	50-80	9.22
Active Mistuned (50-70)	41.13	20-65	10.56
Active Mistuned (>70)	43	18-104	26.71
Active Tuned Manipulated	21.8	12-37	7.09
Active Mistuned Manipulated	63.86	41-78	9.7
Active Tuned Postmanipulated	25.73	16-39	7.468
Active Mistuned	67.2	12-56	10.27

Postmanipulated			
Active Tuned Nonmanipulated	53.86	21-82	18.02
Active Mistuned Nonmanipulted	45.8	21-105	22
Active Fast corrections	49.46	17-123	34.77
Active Medium corrections	48.86	22-83	20.32
Active Slow corrections	41.6	15-71	20.2

Table 2. Number of trials included in each averaged note type after rejection. Mean, range and standard deviation for each one of the conditions and sub conditions.

6.11.3 Manipulated, nonmanipulated and post-manipulated notes

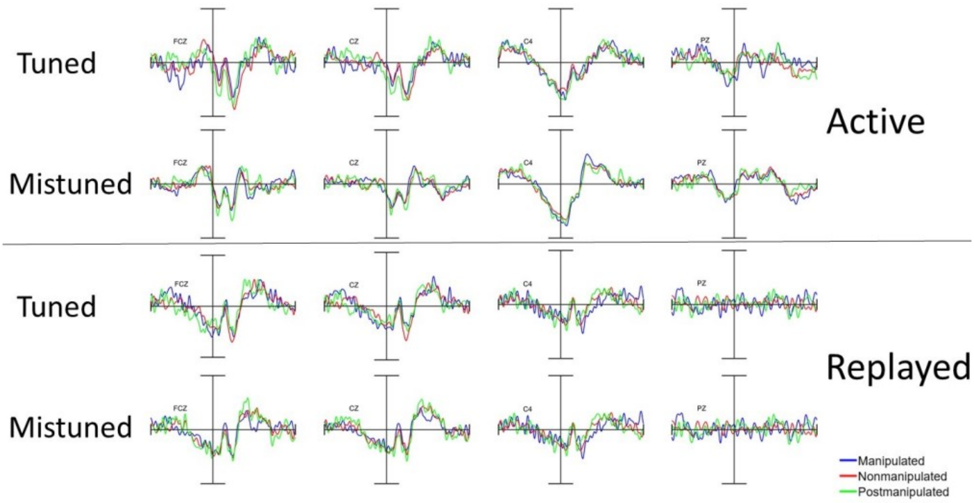

Figure 1. ERPs Active Mistuned, Active Tuned and replayed Mistuned minus Tuned on electrodes FCz, Cz, CPz C4, for manipulated (blue lines), nonmanipulated (red lines), and postmanipulated (green lines). No significant differences were found between the Mistuned and Between the Tuned conditions for each type of note.

7. CONCLUSIONS

7.1 Summary

In this thesis, we have explored the use of both sound and kinematic descriptors to track the process of learning to play the violin. We evaluated the impact of different types of augmented feedback in learning bow control with the violin, and intonation both with the violin and the voice. Finally, we have also studied the electrophysiological correlates of error-monitoring in violin and cello expert performers.

We did a first pilot study where we evaluated the effects of an offline sound quality visual feedback system while participants learned to produce a stable sound with the violin (Blanco and Ramirez, 2019). We also measured their cortical activity at pre-frontal sites using a low-cost EEG system. 18 participants with no prior experience with the violin now any bowed string instruments were randomly distributed between an experimental and a random group. 7 violin experts participants were also recruited for comparative purposes. Both groups of beginners could access instructional videos about violin and bow technique. The experimental group could, in addition, demand a visualization of their scores at each trial allowing them to compare the quality of the sound of their last trial with the previous ones and with the performance of an expert participant. Participants performed 20 trials in total. The sound quality for each trial was measured using audio descriptors such as *dynamic stability* (the standard deviation of the energy of the signal), *pitch stability* (the standard deviation of the power), and *aperiodicity* (the ratio between aperiodic power and total power of the signal) (Romani et al., 2015). We found that those descriptors not only allowed us to differentiate between beginners and experts but also gave us the possibility to track beginner's improvement in sound stability through the different blocks of the experiment. We found that the experimental group behaved differently than the control group during the experiment. They tended to demand the instructional videos more times than the control group, probably in an attempt to improve their previous scores. We also found that the experimental group tended to slightly improve their results at *dynamic stability* while maintaining good results at *pitch stability* until the last block while the control group did not (compared with the Baseline). Regarding the electrophysiological analysis, we found that both beginners and experts showed lower power at alpha, beta, and gamma frequency bands at F3 compared with F4. That desynchronization was found to be higher in expert participants at the Baseline. However, through the experiment, beginners tended to show desynchronizations at the F3 electrode that seemed to be slightly correlated with *pitch stability*. We hypothesized that this activity could be related to movement-related cortical potentials contralateral to the arm controlling the bow and suggest that its amplitude could be related with expertise and task complexity as has been reported previously in the literature (Di Russo, Pitzalis, et al., 2005, Fattapposta et al., 1996, Hatta et al., 2009, Kita et al., 2001, Wright et al., 2012).

Following the pilot study, we designed an experiment where we evaluated a more advanced prototype of the Telmi project, SkyNote, which, in addition to an online sound quality analysis feedback system, also included motion-capture technologies that allowed us to track participant's bow movement and offer kinematic feedback to them. Contrary to the previous experiment, in this one, we added a retention block with the same structure as the Baseline at the end of the experiment where the online feedback was removed. We also collected the opinion of participants about the technology with some questionnaires. 50 participants with no prior experience with the violin nor any bowstring instrument were randomly distributed between an experimental and a control group. 15 violin experts were also recruited for comparative purposes. Participants performed 55 trials in total consisting of full bow movements. However, this time, they were also explicitly encouraged to maintain their bow parallel to the bridge during the movement. In addition, the experimental group received online kinematic and sound quality feedback. We found that using the technology helped improve the experimental group's sound quality at the retention block. However, although kinematic feedback helped participants to improve their bow movements those improvements were not maintained in the experimental group.

Until now, however, we have studied motor learning processes that are very specific to the violin but that may have little to do with more musical aspects such as intonation skills. Taking advantage of the same participants from the previous experiment, we designed a third experiment where we wanted to evaluate the effects of feedback in improving pitch-matching abilities with both the violin and the voice. The participants were separated into three groups: a group that received help in the auditory feedback mode with a timbre similar to that of the instrument used (the Equal-Timbre Group), a group that received help in the form of visual feedback offered by SkyNote (Feedback Group) and a group without any help (Control Group). Both the Equal-Timbre Group (ETG) and the Feedback Group (FG) improved their results after receiving help. Although the ETG got better results for the voice than for the violin, these results were not maintained in the retention block while the FG did. We hypothesized that ETG participants improved more on voice than violin due to implicit imitative abilities. However, both a lack of confidence in the correctness of their answers, as well as a lack of exploring the pitch space in an explicit way, could have been the cause that their results were not maintained in the retention block.

But, what is it that makes a beginner unable to recognize when he is producing an incorrect note in this type of task? In Chapter 6, we reported the results of an experiment with expert violinists and cellists in which we studied the processes of monitoring, detecting, and correcting errors on the violin using electroencephalography techniques. To do this, we have developed a setup that allows us to manipulate the notes played by our participants online by lowering or lifting the pitch of the tone by half semitone and reproduce the performance again to be heard in a passive condition. We found a fronto-central negativity (resembling the f-ERN) after the out-of-tune notes that was accompanied by a right central negativity (N-280) and a parietal negativity (N-340). Finally we found some late positivities from 380 ms resembling the P3a and the P3b. The main differences that we find at those components between the active and passive conditions were due to the amplitude of the P3b. We also found a midfrontal theta activity and a beta rebound that was only present in the active condition. We did not find any effect on N1 activity due to possible self-generation effects. All these signals, with the exception of P3a, were shown to be sensitive to the magnitude of the error made. Regarding the error correction processes, we found that the amplitude of the f-ERN turned out to be higher for fast corrections and lower for slow corrections. On the other hand, the N-340, the N-280, and the midfrontal theta tended to show greater amplitude for slow corrections than fast ones. These effects due to the correction did not alter the amplitude of the same components in the passive condition.

7.2 Future directions

Throughout the different chapters of this thesis, we have seen how with relatively simple methods to implement we can evaluate the first steps of learning complex skills such as playing the violin and learn about the effects of feedback on them. The inclusion of electroencephalography techniques in experiments, or the extension of their use to more ecological contexts to which we are normally accustomed, can offer us a broader perspective on how to interpret the different learning processes that our participants may be subjected to and how the feedback may be affecting them.

These technologies have the potential to become useful tools that students can use to assess the quality of their performance from different points of view and, in turn, even compare it numerically with that of their teachers and peers. It is important to evaluate the impact that their inclusion may have in a more ecological context and with wider time spans than those used in the experiments presented. It would also be valuable to compare them with other types of skills more complex than those used and that would require the inclusion of participants with a more advanced musical level. There are already technologies, that also came out of the Telmi project, that allow us to differentiate between different types of arc movements (Dalmazzo et al., 2020; Dalmazzo et al, 2019), and even technologies that can allow us to study expressive parameters between different types of violin performance (Ortega et al., 2019).

Nor can we end this thesis without mentioning the historical context experienced in this last year of the global pandemic. The proliferation of online classes and courses and the physical separation between students and students have accelerated a process that has been brewing for many years and in which these types of technologies will play an undeniable role. The incorporation of them in these online contexts could allow the availability of important sources of data as we have never had before. All in all, the design and implementation of these technologies would help us to learn to better characterize the different learning processes and answer some of the questions with which we began this thesis such as: what is it that distinguishes the expert performance of the amateur? Only a truly multidisciplinary enterprise involving psychologists, neuroscientists, engineers, and pedagogues can lead us to transcend our understanding of musical learning in all its deepest essences and facets, and lastly, offer light and better answers to all the questions in this thesis that are still open.

7.3 References

Blanco, A. D., & Ramirez, R. (2019). Evaluation of a sound quality visual feedback system for bow learning technique in violin beginners: An EEG study. *Frontiers in psychology*, *10*, 165.

Dalmazzo, D., & Ramírez, R. (2019). Bowing gestures classification in violin performance: a machine learning approach. *Frontiers in psychology*, *10*, 344.

Dalmazzo, D., Waddell, G., & Ramírez, R. (2020). Applying deep learning techniques to estimate patterns of musical gesture. *Frontiers in psychology*, *11*.

Di Russo, F., Pitzalis, S., Aprile, T., & Spinelli, D. (2005). Effect of practice on brain activity: an investigation in top-level rifle shooters. *Medicine and science in sports and exercise*, *37*(9), 1586.

Diedrichsen, J., Verstynen, T., Lehman, S. L., & Ivry, R. B. (2005). Cerebellar involvement in anticipating the consequences of self-produced actions during bimanual movements. *Journal of neurophysiology*, *93*(2), 801-812.

Fattapposta, F., Amabile, G., Cordischi, M. V., Di Venanzio, D., Foti, A., Pierelli, F., ... & Morrocutti, C. (1996). Long-term practice effects on a new skilled motor learning: an electrophysiological study. *Electroencephalography and clinical Neurophysiology*, *99*(6), 495-507.

Hatta, A., Nishihira, Y., Higashiura, T., Kim, S. R., & Kaneda, T. (2009). Long-term motor practice induces practice-dependent modulation of movement-related cortical potentials (MRCP) preceding a self-paced non-dominant handgrip movement in kendo players. *Neuroscience Letters*, *459*(3), 105-108.

Kita, Y., Mori, A., & Nara, M. (2001). Two types of movement-related cortical potentials preceding wrist extension in humans. *Neuroreport*, *12*(10), 2221-2225.

Ortega, F. J., Giraldo, S. I., Perez, A., & Ramírez, R. (2019). Phrase-level modeling of expression in violin performances. *Frontiers in psychology*, *10*, 776.

Romani Picas, O., Parra Rodriguez, H., Dabiri, D., Tokuda, H., Hariya, W., Oishi, K., & Serra, X. (2015, May). A real-time system for measuring sound goodness in instrumental sounds. In *Audio Engineering Society Convention 138*. Audio Engineering Society.

Wright, D. J., Holmes, P. S., Di Russo, F., Loporto, M., & Smith, D. (2012). Differences in cortical activity related to motor planning between experienced guitarists and non-musicians during guitar playing. *Human movement science*, *31*(3), 567-577.