TESIS DOCTORAL

MENTAL IMAGE ACUITY IN AUDITORY IMAGERY AND PERCEPTION

(Precisión de las imágenes mentales en imaginería auditiva y percepción)

ANA NAVARRO CEBRIÁN

DIRECTORES

PETR JANATA

(University of California, Davis)

PIO TUDELA

DEPARTAMENTO DE PSICOLOGÍA EXPERIMENTAL Y FISIOLOGÍA DEL COMPORTAMIENTO

UNIVERSIDAD DE GRANADA

ABRIL 2010

Editor: Editorial de la Universidad de Granada
Autor: Ana Navarro Cebrián
D.L.: GR 3221-2010
ISBN: 978-84-693-4360-9

INDEX

EXPERIMENT II

EXPERIMENT III

Electrophysiological Correlates of Accurate Mental Image Formation in Auditory Perception and Imagery Tasks

Chapter IV 167

Effects of perceptual learning on the formation of mental images

Chapter V 195

General Discussion

CHAPTER I,

INTRODUCTION AND RESEARCH PLAN

Introducción

El presente trabajo doctoral tiene como propósito el estudio de los factores implicados en la precisión o calidad de las imágenes mentales auditivas. Con la ayuda de experimentos conductuales y electrofisiológicos se pretende analizar la influencia de los distintos tipos de memoria en la formación de imágenes mentales. Se estudiarán también las diferencias en el procesamiento neuronal subyacente a la imaginería entre los participantes que son capaces de formar buenas imágenes mentales vs. aquellos que forman imágenes menos precisas. Por último se analizarán los cambios conductuales y neuronales subyacentes a la imaginería mental debidos al aprendizaje perceptivo. El paradigma utilizado está basado en un estudio previo (Janata & Paroo, 2006). Se presentan escalas musicales y el participante tiene que decidir si la última nota de la escala está afinada o desafinada. En una condición (percepción, *bottom-up*) el sujeto escucha cada nota de la escala. En la otra condición (imaginería, *top-down*), el sujeto sólo escucha las cuatro primeras notas (que le introducen en el contexto tonal) e imagina el resto de notas que preceden a la última nota (la última nota, *target*, siempre se escucha).

A continuación se presenta evidencia empírica relevante que nos llevará a un mejor entendimiento de las hipótesis que se plantearán en cada uno de los experimentos. Así, en primer lugar se hace una pequeña introducción con el propósito de facilitar la comprensión del estudio de las imágenes mentales auditivas y los estudios de discriminación tonal. Se analizarán después la relación entre imágenes mentales y percepción, y se describirán algunos estudios que ofrecen evidencia empírica sobre el solapamiento cerebral de la imaginería y la percepción. En el siguiente punto se resumirá la literatura existente sobre la influencia de la memoria a largo plazo (influencia del contexto musical o tonal, e influencia de la experiencia musical) en la precisión de las imágenes mentales auditivas (aunque, en este sentido, la mayor parte de la evidencia hasta el momento se da en experimentos de percepción y no de imaginería). Por último, se resumirán los trabajos más relevantes relacionados con el estudio del aprendizaje perceptivo, ya que la última parte de esta tesis estará dirigida al estudio de la influencia del aprendizaje perceptivo en la precisión de las imágenes mentales.

Papel de las imágenes mentales auditivas en una tarea de discriminación tonal

Independientemente de la experiencia musical, algunas personas tienen grandes habilidades para entonar una melodía o distinguir cuando un tono está desafinado, mientras que otras son incapaces de realizar eficazmente este tipo de tareas. ¿Qué es lo que nos lleva a poder reconocer cuando una nota en una melodía o un instrumento está desafinado? ¿Por qué se dan tantas diferencias individuales en este tipo de tareas? Aunque otros factores están involucrados en esta tarea, la memoria humana está encargada de recrear o recuperar las representaciones de los objetos que han sido percibidos previamente. Esta recuperación de representaciones o imágenes mentales será más o menos eficaz dependiendo de procesos atencionales, motivacionales y de memoria.

Tareas de discriminación tonal

Los estudios típicos de discriminación tonal suelen presentar dos tonos (separados o no por otros tonos) y el participante decide si los dos tonos son iguales o por el contrario existe un intervalo tonal entre ambos (Deutsch, 1978).

Un paradigma *oddball* también puede ser utilizado con este propósito, de forma que el sujeto tenga que comparar la frecuencia de estímulos auditivos. Se presentaría, por ejemplo, una secuencia de tonos con la misma frecuencia y el sujeto tendría que responder cuando aparece un tono inesperado con una frecuencia distinta. En el paradigma *oddball*, al contrario que en una tarea típica de discriminación tonal, el tono infrecuente (*target*) suele estar dentro del umbral de discriminación de los sujetos, es decir, suele ser lo suficiente claro para que los sujetos puedan detectarlo. Existen sin embargo algunas variaciones de este paradigma en las que se varía la dificultad del *target*. Por poner un ejemplo, en un estudio con potenciales evocados, Tervaniemi y sus colaboradores (Tervaniemi, Just, Koelsch, Widmann, & Schröger, 2005) utilizaron un paradigma *oddball* en el que presentaron tonos estándar y tonos desafinados o desviados menos frecuentes a sujetos músicos profesionales y a sujetos sin experiencia musical. Los tonos desafinados tenían 0.8%, 2% o 4% de frecuencia (ciclos por segundo) más alta. Los tiempos de reacción y la precisión de los músicos profesionales fue mejor en intervalos de desafinación más pequeños (0.8% y 2%). Además, los datos electrofisiológicos señalaron que el cerebro de los músicos profesionales mostraba esta ventaja frente a los no músicos sólo en los casos en los que se requería prestar atención y contestar a los estímulos auditivos.

En este tipo de tareas *oddball* y en otros paradigmas típicos de discriminación tonal (en los que se presentan dos notas y el sujeto decide si la segunda es igual a la primera), los participantes hacen sus decisiones comparando la huella de memoria sensorial del tono anterior con la percepción del tono siguiente.

En otro tipo de tarea alternativa de discriminación tonal, los sujetos podrían escuchar una melodía familiar y decidir si una de las notas de la secuencia está desafinada (ej., Janata & Paroo, 2006). Al contrario de lo que sucede en el paradigma *oddball,* en esta tarea la memoria sensorial no está implicada directamente ya que cada nota es distinta a la nota anterior. Por tanto, la nota *target* no se compara con la memoria sensorial de la nota anterior sino más bien con la predicción o representación mental de la nota *target,* que se evoca mentalmente al escuchar la nota/notas anteriores. Cada nota de la melodía familiar evocará la representación o predicción de la nota siguiente que le corresponde en la secuencia. Por tanto, para poder decidir si la nota *target* está desafinada, los sujetos (basado en la nota anterior) recuperan automáticamente la representación de ésta de la memoria a largo plazo, y esta representación es comparada con el estímulo *target* entrante.

Por último, algunos estudios (Herholz, Lappe, Knief, & Pantev, 2008) han presentado una tarea de imaginería en la que los sujetos deciden si una nota está desafinada o no tras haber imaginado las notas que le preceden en la melodía o escala musical. En este caso la nota *target* no se compara ni con la huella de memoria sensorial ni con la predicción elaborada en la memoria a largo plazo basada en la nota anterior, sino más bien con una representación o imagen mental pura (que también se recupera de la memoria a largo plazo).

Formación de imágenes mentales

El estudio del aprendizaje y la memoria es básico para entender la capacidad para formar imágenes mentales. Concretamente se ha comparado el proceso de la imaginería auditiva con la memoria de trabajo, y el modelo de memoria de trabajo de Baddeley y Hitch (1974) ofrece un buen ejemplo para entender el funcionamiento cerebral subyacente a la formación de imágenes mentales. En este modelo, un ejecutivo central asigna los recursos necesarios de procesamiento para que otros sistemas, denominados esclavos, realicen sus funciones. Uno de los sistemas esclavos, el bucle fonológico, que está a su vez formado por dos componentes, un almacén temporal de información fonológica y un proceso encargado de repetir la información para que ésta no se pierda, podría considerarse como la sede para la formación de imágenes (Baddeley & Logie, 1992). Baddeley y Logie (1992) señalan que no sólo los estímulos auditivos provenientes del exterior tienen su sede en el almacén fonológico. Repetir una melodía mentalmente o reconocer una nota desafinada, por ejemplo, requieren también de un almacén temporal de información acústica, con lo que argumentan que éste almacén fonológico pueda ser el escenario donde los estímulos auditivos y los estímulos generados internamente (imágenes mentales) interactúan. En cuanto a los estímulos musicales, Baddeley y Logie (1992) añaden que aunque no se compartan los procesos articulatorios del lenguaje, los mismos mecanismos parecen estar involucrados en el procesamiento de memoria de secuencias de tonos.

La memoria de trabajo está ampliamente distribuida en la corteza prefrontal (Fuster, 2001) y como consecuencia éste área parece ser crucial para la formación de imágenes mentales (Frith & Dolan, 1996). El proceso de formar imágenes mentales sin embargo implica otras áreas cerebrales. Dependiendo de la modalidad sensorial, áreas secundarias de la corteza visual o auditiva estarán activas también mientras los participantes están recreando imágenes visuales o auditivas. La correcta interacción entre la corteza prefrontal y la corteza posterior es por tanto primordial para el funcionamiento de la imaginería mental (Halpern & Zatorre, 1999).

Imágenes o representaciones mentales

Como ya se indicó anteriormente, en una tarea en la que se escucha una melodía familiar o una escala musical, cada nota de la secuencia evocará una predicción de la nota siguiente. Esta predicción equivale a una representación, que al igual que ocurre en una tarea de imaginería, es recuperada de la memoria a largo plazo. En este sentido, desde nuestro punto de vista, la única diferencia entre estas dos tareas sería el mayor o menor grado de ayuda sensorial a la hora de recuperar las representaciones de la memoria a largo plazo.

Además, como se resume en el siguiente punto, existe evidencia suficiente que apoya la similitud entre percepción e imaginería en general, y en música en particular (Zatorre, 1999). Hemos considerado como hipótesis la similitud de estas representaciones / imágenes mentales, y el lector podrá ver que en este trabajo denominamos "imágenes mentales" a los fenómenos resultantes de ambas tareas (imaginería y percepción). De esta forma, las imágenes mentales podrían ser evocadas por dos vías distintas. Cuando escuchamos una serie de notas musicales o una melodía conocida, la presentación de una nota va a evocar de forma automática, de abajo a arriba (*bottom-up*), la imagen mental de la siguiente nota en la secuencia. Por el contrario, cuando no se presentan estímulos exteriores y estamos imaginando una melodía, son las propias imágenes mentales las que llevan a la evocación de las imágenes o representaciones de las siguientes notas en la secuencia (*top-down*), con la consiguiente activación de la corteza auditiva.

En ambas tareas, tras la recuperación de imágenes desde la memoria a largo plazo y el mantenimiento de estas en la memoria de trabajo, las imágenes podrán ser comparadas con la información auditiva proveniente del exterior (la última nota de la escala, *target*). Esta comparación se lleva a cabo gracias a la interacción del lóbulo prefrontal y la activación neuronal evocada por los estímulos sensoriales en áreas corticales posteriores. Cuando la memoria de trabajo interactúa con el ambiente, no sólo trabaja con las representaciones internas, sino que también facilita la activación de áreas corticales posteriores bajo el control selectivo de la corteza prefrontal (Fuster, 1995). Por tanto, en ambas tareas, áreas corticales posteriores serán activadas por la estimulación externa y la influencia de la corteza prefrontal. En la dirección opuesta, para la recuperación de representaciones desde la memoria a largo plazo, la activación neuronal de la corteza auditiva podrá facilitar en parte la activación de neuronas de la corteza prefrontal (Fuster, 2001).

Percepción vs. Imágenes mentales

Hipótesis de la equivalencia funcional

El tema de la relación percepción e imaginería mental lleva tratándose desde hace varias décadas. Una de las líneas de investigación ha buscado responder a la pregunta de la existencia de una equivalencia funcional entre estas dos experiencias. La idea de la equivalencia funcional confirmaría que la imaginería mental comparte al menos algunos de los procesos de la percepción, en cuyo caso la ejecución en ambas tareas sería parecida. Algunas teorías han sugerido que las imágenes son esquemas construidos durante la percepción con lo que se espera que los patrones de repuesta causados por los "estímulos" imaginados sean similares a aquellos producidos por los estímulos percibidos (Neisser, 1976).

Existe mucha investigación que apoya esta idea de la equivalencia funcional sobretodo en la modalidad visual (Finke, 1980; Finke y Shepard, 1986). En un estudio conductual, por ejemplo, Segal y Fusella (1970) combinaron las modalidades visual y auditiva. Los investigadores pidieron a un grupo de participantes que formasen una imagen visual o auditiva y les presentaron después un estímulo (visual o auditivo) mientras ellos mantenían presente su imagen mental. Encontraron que las imágenes auditivas causaban más interferencia cuando se presentaban estímulos auditivos, mientras que las imágenes visuales interferían más con los estímulos visuales.

Estímulos auditivos musicales han sido utilizados también en estudios pioneros para probar la relación percepción e imaginería. En una serie de experimentos llevados a cabo por Hubbard y Stoecking (1988), los sujetos tenían que imaginar una nota musical o un acorde que estuviese un tono por encima de una nota indicada. Tras imaginar la nota o el acorde, se les presentaba un tono que podía ser igual a lo que ellos habían imaginado, un tono diferente pero armónicamente relacionado, o un tono diferente y poco relacionado armónicamente. Encontraron que las imágenes mentales facilitaban la percepción ya que los sujetos eran más precisos cuando el tono era igual o relacionado armónicamente al tono de la nota imaginada que cuando el tono estaba menos relacionado armónicamente.

Estudio del solapamiento neuronal entre imaginería y percepción

Entre las líneas de investigación más recientes que defienden la idea de la similitud entre imaginería y percepción, una de las preguntas que se han formulado es si ambos fenómenos comparten los mismos sustratos neuronales. Estudios realizados con resonancia magnética funcional (RMf) y tomografía por emisión de positrones (PET) han encontrado que las imágenes mentales pueden activar áreas propias de la percepción en ausencia de estimulación externa (Bunzeck, Wuestenberg, Lutz, Heinze, & Jancke, 2005; Zatorre, 1999; Zatorre & Halpern, 2005; Zatorre, Halpern, Perry, Meyer, & Evans, 1996). Zatorre y sus colaboradores (1996), por ejemplo, utilizaron la técnica del PET y presentaron tres tareas a un grupo de participantes no músicos. En una condición control los sujetos veían un par de palabras en cada condición y tenían que juzgar cuál era más larga. En la condición de "percepción" también veían un par de palabras y escuchaban al mismo tiempo una canción familiar donde estas dos palabras aparecían (en la letra de la canción). Los sujetos tenían que juzgar en esta ocasión si la primera palabra aparecía en la melodía en un tono más alto que la segunda palabra o al contrario. Por último, se presentó una condición de "imaginería mental" donde los sujetos veían dos palabras en la pantalla y se les indicaba que imaginasen la canción donde las dos palabras aparecían (no se presentaba ningún estímulo auditivo en este caso). En esta condición los sujetos debían decidir igualmente la altura tonal de las palabras en la canción "imaginada". Los datos mostraron que las tareas de percepción e imaginería llevaron a patrones similares comparadas con la condición control. Ambas condiciones parecen estar asociadas con una mayor actividad en la corteza auditiva secundaria y los lóbulos frontales además de otras áreas como el lóbulo parietal izquierdo y el área motora suplementaria. La hipótesis que plantea el que áreas propias de percepción auditiva son activadas en ausencia de estimulación externa debido a imágenes mentales, ha sido apoyada por otras investigaciones más recientes tanto con PET (Halpern & Zatorre, 1999) como con RMf (Halpern, Zatorre, Bouffard, & Johnson, 2004; Kraemer, Macrae, Green, & Kelley, 2005).

Aunque actualmente existe bastante acuerdo en cuanto a la activación de la corteza auditiva en tareas de imaginería mental auditiva, se desconoce bastante los distintos roles que puedan tener la corteza primaria y la secundaria. En la modalidad visual se ha encontrado alguna evidencia que prueba que la corteza primaria está también involucrado en algunas tareas que requieren imaginería mental. Este hecho está todavía por probar en la modalidad auditiva y más específicamente, con estimulación musical (Zatorre & Halpern, 2005).

Más evidencia sobre la posibilidad de un solapamiento neuronal de la imaginería y la percepción viene dada por hechos que demuestran una activación de la corteza auditiva en ausencia de estimulación. En general, la estimulación eléctrica de la corteza auditiva parece reactivar huellas de memoria auditiva que lleva a experiencias o ilusiones auditivas (Penfield & Perot, 1963). Otro dato que nos pueden ayudar a comprender el funcionamiento de la imaginería mental auditiva viene aportado por el estudio de las alucinaciones auditivas en esquizofrenia (Smith, Reisberg, & Wilson, 1992). Las imágenes mentales y las alucinaciones se han considerando en psiquiatría dentro del mismo continuo y ha habido quien incluso ha considerado las alucinaciones como resultantes de una imaginería anormal (Mitnz & Alpert, 1972). En 1981, Launay y Slade (1981) validaron una escala (Launay-Slade Hallucination Scale, LSHS) para medir la disposición de una persona a tener alucinaciones. Estos autores encontraron una correlación significativa entre tendencia alucinatoria y experiencia de imágenes mentales vívidas no patológicas, es decir, aquellas personas que consideran que tienen una buena imaginación o que tienden a soñar despiertos y además lo hacen de forma vívida. Por otro lado, estudios más recientes realizados con técnicas de neuroimagen han aportado evidencia sobre la aparición de alucinaciones auditivas correlacionado con la activación de áreas relacionadas con la percepción auditiva (Lennox, Park, Jones, & Morris, 1999; McGuire, et al., 1996; Shergill, Brammer, Williams, Murray, & McGuire, 2000).

Además de los estudios citados anteriormente realizados con técnicas de neuroimagen, la técnica de los potenciales evocados ha llevado también a importantes avances en el estudio de la imaginería mental. Componentes tempranos como el Mismatch Negativity (MMN) u otros más tardíos como la P300, que pueden aparecer tras la presentación de un estímulo menos frecuente con relación a los estímulos precedentes (paradigma *oddball*), han sido encontrados en tareas en las que los participantes tienen que formar imágenes mentales.

En un estudio con magnetoencefalografía, Yumoto y colaboradores (Yumoto, et al., 2005) presentaron letras japonesas visualmente. La lectura silenciosa de estímulos visuales lleva a la producción de imágenes mentales auditivas que a su vez pueden interactuar con la estimulación presentada auditivamente. De esta forma, los autores detectaron un componente de localización y latencia iguales a las del MMN cuando los sujetos identificaban desajustes (*mismatches)* entre los estímulos visuales y una secuencia aleatoria de sílabas presentadas auditivamente.

En cuanto a procesamiento musical, Schön y Besson (2005) encontraron componentes más tardíos, como por ejemplo el P3a, evocados por imaginería mental auditiva. En este experimento con potenciales evocados, se presentó estimulación visual (lectura de notación musical) que evocó imágenes mentales auditivas. En esta tarea de imaginería (leer música), los participantes mostraron un patrón psicofisiológico similar a cuando se escucha una melodía, cuando la notación musical presentada era tonal, y de igual forma, cuando se presentaba una terminación atonal, se halló, por ejemplo, un incremento de la P3a.

En un estudio reciente realizado con magnetoencefalografía (Herholz, et al., 2008), a un grupo de sujetos músicos profesionales y a otro grupo sin experiencia musical se les pidió escuchar el inicio de una melodía que les era familiar y continuar imaginando el resto de la melodía. Se les presentó el último tono de la melodía que podía ser correcto (afinado) o estar desafinado. Los autores encontraron un *mismatch* producido por las notas incorrectas o desafinadas basado en las imágenes mentales que precedían a la última nota. Encontraron por otro lado que el componente evocado por esta tarea estaba localizado en la corteza auditiva secundaria. Los autores proponen que éste área está en parte encargada de la detección de la trasgresión de predicciones o expectativas auditivas (*mismatch*). Además afirman que éste área puede recibir *inputs* en forma de una huella de memoria clásica, así como también *inputs* de materiales auditivos "imaginados". Este estudio también nos ofrece un claro ejemplo de los efectos de la memoria a largo plazo en las imágenes mentales (del que hablamos en el siguiente punto), ya que el efecto de la *mismatch negativity* se encontró únicamente en el grupo de sujetos músicos profesionales. Los autores explican que en este grupo de sujetos se da una mejor imaginería mental comparado con el grupo de no músicos, lo cual se puede deber a unas representaciones más precisas junto con unas capacidades de procesamiento mejores por parte de los sujetos músicos.

Todos estos datos aportan más evidencia sobre la función de la corteza auditiva en la imaginería mental. Al igual que en los estudios de Penfield en los que la estimulación de las posibles huellas de memoria lleva a experiencias auditivas, parece que estas mismas neuronas de la corteza auditiva son activadas en una tarea de imaginería tan común como imaginar una melodía. En este caso, la estimulación eléctrica es sustituida por la función de la corteza prefrontal. En este sentido, diversas áreas de la corteza prefrontal juegan un papel básico en la activación de la corteza posterior para la evocación de imágenes mentales.

A pesar de toda esta evidencia en cuanto a la similitud de la imaginería y la percepción, deben existir también diferencias entre estos dos fenómenos que merecen igual atención. Sin embargo, el estudio de la imaginería mental se ha basado en su equivalencia con la percepción, y es escasa la investigación que intenta buscar las diferencias entre ambos (Chambers & Reisberg, 1985). Como indican estos autores (Chambers & Reisberg, 1985), una diferencia sería el hecho de que la percepción puede requerir de una interpretación del objeto externo percibido. Al contrario, ya que las imágenes mentales son símbolos de objetos percibidos, la imaginería no necesita de un proceso interpretativo. Esto a su vez llevará a distintos procesamientos y resultados en ambos casos y más investigación en esta línea sería muy interesante.

Exposición del paradigma utilizado en nuestro estudio

Al contrario que en la percepción, el estudio de la imaginería se hace difícil ya que son procesos internos de los que además no conocemos el comienzo exacto (*onset*) tan necesario, por ejemplo, en un estudio de potenciales evocados. Los psicológos han elaborado ingeniosos experimentos conductuales mediante los que se puede medir el resultado proveniente de una tarea de imaginería. Un ejemplo podría ser el estudio de la precisión con la que somos capaces de evocar imágenes mentales (Janata & Paroo, 2006.) De esta forma, podemos estudiar la relación entre ambos fenómenos (percepción e imaginería) mediante el análisis de las diferencias o semejanzas de precisión de la ejecución en estas tareas. Concretamente, cuando se escucha o se imagina una melodía con el propósito de saber si la última nota está desafinada (Janata & Paroo, 2006), los sujetos formarán imágenes o representaciones más o menos precisas, que ayudarán a realizar la tarea mejor o peor. Así, asumimos que la mayor o menor facilidad para formar imágenes mentales estará explicada por la ejecución en la tarea (saber si la última nota está desafinada).

Janata y Paroo (2006) llevaron a cabo una serie de experimentos conductuales en los que analizaron la precisión de las imágenes mentales y de las predicciones o expectaciones producidas por estímulos auditivos. Presentaron escalas musicales ascendentes de ocho tonos y, o bien se escuchaban los ocho tonos (escala completa), o se escuchaban varios tonos al principio de la escala y los sujetos tenían que seguir imaginando el resto de tonos hasta el final de la escala. Presentaron de forma aleatoria dos tipos de tareas. En una tarea los sujetos tenían que decidir si la última nota (independientemente de si habían escuchado o imaginado las notas anteriores) estaba afinada, acorde con las notas o imágenes anteriores. En la otra tarea, los sujetos tenían que decidir si la última nota había aparecido en el momento en que tenía que aparecer acorde con las notas o imágenes precedentes, es decir, si era temporalmente correcta. De este modo, los autores analizaron cómo de eficaces eran las imágenes mentales de los participantes en dos dimensiones distintas de las imágenes mentales auditivas, el tono (la altura) y el tiempo. Los autores concluyeron que las imágenes mentales eran tan eficaces en la tarea de imaginería como en la tarea perceptiva cuando los sujetos tenían que juzgar la altura tonal de la nota. En una melodía familiar o una escala musical, la percepción de cada nota de la escala lleva a la creación de una imagen mental o una predicción sobre cuál será la próxima nota en la escala. De este modo, cuando la siguiente nota es escuchada, el sujeto puede comparar el resultado de la percepción de la nota y decidir si esta tiene o no iguales dimensiones que su imagen mental. En una tarea en la que se requiere imaginar una sucesión de notas, bien sea una escala o una melodía, la imagen mental de cada una de las notas llevará a la imagen de la siguiente nota en la secuencia. Janata y Paroo (2006) concluyeron que las imágenes formadas en una tarea de imaginería tenían igual funcionalidad que las imágenes formadas en una tarea de percepción. Sin embargo esto fue cierto únicamente en el caso en el que se requería hacer juicios con respecto a la altura tonal de las notas (la afinación). Cuando los sujetos tenían que decidir sobre si la última nota fue presentada en el momento correcto se encontraron diferencias significativas entre las dos tareas. Las imágenes formadas en la tarea de imaginería parecían ser en este caso mucho menos funcionales o eficaces que las imágenes formadas en la tarea de percepción (cuando se escuchaban todas las notas en la escala). Parece claro a partir de estos datos que la dimensión temporal de las imágenes mentales auditivas no sigue iguales reglas. Por otro lado, resulta muy interesante la variabilidad entre sujetos encontrada a la hora de formar imágenes mentales en ambas tareas (percepción e imaginería). Como era de esperar, aquellos sujetos con más experiencia musical tuvieron más facilidades a la hora de decidir si la última nota estaba o no desafinada en ambas tareas, imaginería y percepción. Sin embargo, no existe una buena explicación para el hecho de que, con igual número de años de entrenamiento musical, algunos sujetos tengan muchas habilidades en este tipo de tareas y otros sujetos sean incapaces de llevarlas a cabo.

Los datos presentados hasta el momento nos han permitido concluir que existe una estrecha relación entre la percepción y la imaginería mental, y ambas tareas pueden llegar a ser funcionalmente equivalentes. Sin embargo, como se ha indicado anteriormente, en un mismo sujeto puede haber diferencias en la calidad de las representaciones en una y otra tarea (imaginería vs. percepción) y ésta diferencia en la calidad o precisión de las imágenes no ha sido suficientemente investigada. El estudio de los factores influyentes en la calidad de las imágenes mentales es por tanto de los principales propósitos de esta tesis. Utilizando el paradigma elaborado por Janata y Paroo (2006) con varias modificaciones, estudiaremos factores influyentes como la memoria a largo plazo (experiencia musical e influencia del contexto tonal), o la posible influencia de un entrenamiento o aprendizaje perceptivo.

A continuación, se analizará la investigación más influyente sobre lo que consideramos los fenómenos que más afectan a la creación de imágenes mentales eficaces (precisas). Por un lado, se estudiará la influencia del contexto y la experiencia con los estímulos (memoria a largo plazo). Después hablaremos de la influencia del entrenamiento con los estímulos o el aprendizaje perceptivo (*perceptual learning*). Se hablará sobre todo de la influencia que estos factores puedan tener sobre tareas de percepción pero no de imaginería, ya que hasta el momento no ha habido mucha investigación de este tipo con imágenes mentales. Más concretamente se analizarán estudios que examinan la discriminación tonal.

Influencia de la memoria a largo plazo. Contexto tonal y experiencia musical

Influencia del contexto

La percepción de objetos o estímulos individuales en la música y el lenguaje va a depender a su vez de la relación de éstos con los demás objetos que le rodean. Así, en el caso del lenguaje, las palabras están unidas entre ellas y estas relaciones están regidas por reglas sintácticas y semánticas. De igual modo, la armonía ofrece una base estructural para dar orden a los estímulos musicales.

Este sistema de reglas, tanto en el lenguaje como en la música, es aprendido de forma implícita por la mera exposición a los estímulos. De esta forma, un sujeto sin experiencia musical formal puede ser capaz de detectar si un músico de una banda está desafinando con respecto al grupo. De igual forma, el sistema de reglas del lenguaje es aprendido implícitamente y las incongruencias lingüísticas son detectadas por la mayoría de las personas. En el caso del lenguaje, existe evidencia sobre los efectos del *priming* semántico en el procesamiento de las palabras. Por ejemplo, Stanovich y West (1979) encontraron que había una facilitación en el procesamiento de las palabras *target* cuando el contexto previo a la palabra era semánticamente congruente con ésta.

En el procesamiento musical, al igual que ocurre en el lenguaje, este conocimiento implícito de la estructura armónica va a producir un efecto de *priming*. De esta forma, cuando escuchamos la primera parte de una melodía, cada nota de la melodía va a llevar a que se creen predicciones sobre cuáles son las notas que siguen. En concreto, en el sistema musical occidental este sistema de reglas armónicas se puede explicar mediante las relaciones en una escala cromática. La escala cromática está compuesta por 12 tonos que a su vez pueden combinarse y formar varias secuencias de 7 tonos o notas musicales. Cada una de estas secuencias es una escala de una tonalidad distinta, las llamadas escalas diatónicas. Estas escalas tienen siete tonos distintos cada una, y el primer tono, llamado la tónica, puede repetirse en una octava más alta. Los siete tonos de la escala, a su vez, tienen distintos status dentro de la jerarquía. Los tonos primero, I (tónica, *tonic*), cuarto, IV, y quinto, V, son más frecuentes que el resto en una melodía y por tanto más esperados o predecibles debido a un efecto *priming* (Krumhansl, 1990). Por otro lado, el séptimo grado o tono de la escala diatónica, llamado sensible (*leading-tone* en inglés) es uno de los tonos menos frecuentes y por tanto menos esperados de la escala. Igual que en el lenguaje, este efecto *priming* ha sido probado con estímulos musicales, e incluso un tono o acorde aislado puede llevar a la activación de este sistema de reglas armónicas y hacer que tonos más relacionados armónicamente sean procesados más rápida y/o eficazmente (Bharucha & Stoeckig, 1986; Janata & Reisberg, 1988). Bigand y Pineau (1997), por ejemplo, presentaron secuencias de ocho acordes y manipularon los seis primeros acordes de modo que las predicciones para el último acorde eran distintas en cada condición. Incluso sin variar el séptimo acorde (el anterior al *target*), los autores encontraron que los sujetos eran más rápidos y eficaces juzgando la entonación del *target* cuando los seis primeros acordes estaban armónicamente relacionados con el *target*. Interpretaron tales resultados como que éste efecto *priming* en la música depende de estructuras armónicas a un nivel mayor (global) ya que no ocurren únicamente de un acorde al siguiente. Tillman, Bigand y Pineau (1998) encontraron resultados similares global y localmente (en dos acordes juntos). Los sujetos fueron más eficaces y rápidos cuando el acorde *target* estaba más relacionado armónicamente a ambos, el contexto global anterior y el contexto local (el acorde precedente) y esta facilitación disminuía si solamente se daba una relación local o una relación global.

El efecto *priming* en contextos musicales ha sido estudiado con la técnica de los potenciales evocados. Por ejemplo, un componente tardío como el LPC (*late positive component*) puede detectarse en un contexto musical cuando se presentan notas incongruentes armónicamente. Besson y Faïta (1995) encontraron que la amplitud de este componente era mayor cuando la última nota de la melodía era incongruente con el contexto (no creaba una sensación de terminación melódica) y además violaba las reglas de la estructura tonal, es decir, no pertenecía a la tonalidad. Cuando la última nota de la melodía era incongruente con el contexto y por tanto no creaba una sensación de terminación o cierre, pero sin embargo pertenecía a la tonalidad (respetando la estructura tonal), se encontró un LPC de menor amplitud. Brattico, Näätänen y Tervaniemi (2001) encontraron un patrón de resultados similares enfocándose en el componente de la *mismatch negativity* (MMN). Los sujetos escuchaban grupos de 5 notas pasivamente (mientras leían un libro), y se en ocasiones se presentaba un tono menos frecuente o incongruente con el contexto. Encontraron que la MMN era más amplia en una condición en la que la nota desafinada entraba dentro de un contexto armónico familiar (la nota cambiaba de modalidad[1], mayor vs. menor) que en una condición en la que las frecuencias de los cinco tonos del grupo estaban determinadas aritméticamente (contexto no familiar). A su vez, la MMN resultó más amplia en el contexto no familiar que la MMN evocada por la aparición de

---

[1] Los modos se refieren a las distintas formas de escalas diatónicas. Dentro de la tonalidad se dan dos formas, modos mayores y menores. Los modos mayores le dan a la melodía un carácter más alegre, mientras que los menores reflejan un carácter más oscuro o triste.

una nota infrecuente dentro de un grupo de notas iguales (condición no-contexto). Los autores concluyen que el contexto armónico puede facilitar la detección de notas desafinadas a un nivel preatencional.

Se han encontrado también efectos *priming* debido al contexto tonal en una tarea de imaginería. En uno de los experimentos del estudio citado anteriormente por Janata y Paroo (2006), los sujetos tenían más facilidades para decidir si la nota *target* estaba desafinada cuando ésta nota era la tónica que cuando era la sensible (séptima nota en una escala diatónica). La tónica es una de las notas más repetidas o frecuentes dentro de una pieza musical y por tanto existe una mayor expectación hacia ésta. Por el contrario, la séptima nota de la escala es una de las menos importantes en la jerarquía tonal y por lo que no se encontró un efecto *priming* en esta condición.

En el capítulo 2 de este trabajo se pretende en parte, replicar estos resultados encontrados en Janata y Paroo (2006). Se estudiará la influencia del contexto tonal en la formación de las imágenes mentales, comparando la calidad de las imágenes o representaciones para la tónica y para la séptima nota (*leading tone*). De la misma manera, estudiaremos también la influencia de la experiencia musical en la calidad de las imágenes mentales, y la interacción de estos factores. En el capítulo 3, se estudiarán estos mismos procesos mediante la técnica de los potenciales evocados. A continuación se presenta un resumen de la literatura que estudia las diferencias entre músicos profesionales y sujetos sin experiencia musical en distintos tipos de tareas de percepción e imaginería.

Experiencia Musical

En una tarea de discriminación tonal, la experiencia musical va a ser uno de los factores que más influyan en la ejecución de los sujetos. Aunque, como ya se ha comentado, incluso los sujetos sin experiencia musical poseen un conocimiento aprendido de forma implícita que les permite percibir una estructura musical compleja de forma parecida a un sujeto con experiencia musical, el entrenamiento musical dota a la persona con más armas que facilitan, por ejemplo, la producción de imágenes mentales de tonos más precisas o acertadas.

Dependiendo del tipo de tarea encontraremos más o menos similitudes entre músicos y no músicos. Así, algunos autores han llegado a afirmar que la relación tonal entre notas es algo aprendido de forma implícita y se espera que sujetos no músicos tengan un rendimiento más parecido al de sujetos con experiencia en tareas que requieren de esta habilidad (Attneave & Olson, 1971; W. Jay Dowling, 1978; W. J. Dowling & Diane, 1971). Por otro lado, reconocer la altura tonal de una nota aislada, independiente del contexto musical o de otras notas, es una labor que requiere de más experiencia (Bachem, 1954). Sujetos sin entrenamiento musical, por consiguiente, tendrán un rendimiento peor en tareas que requieran juzgar la altura tonal (frecuencia absoluta) de una nota ya que este tipo de sujetos no suele hacer uso de este tipo de información musical.

Existen varios estudios con potenciales evocados que muestran diferencias de procesamiento entre sujetos músicos y no músicos. Por ejemplo, en el estudio de Besson y Faïta (1995), los efectos de *priming* difirieron menos entre músicos y no músicos cuando las melodías eran familiares. Sin embargo, cuando las melodías presentadas no eran familiares para los sujetos, únicamente el grupo de sujetos músicos parece verse beneficiados por un efecto *priming*. De igual modo, en el estudio de Brattico y colaboradores (2001), la latencia de la MMN es menor para el grupo con experiencia musical en las condiciones "familiar" y "no familiar". Los autores concluyen que en un contexto tonal estructurado, el sistema parece responder más rápidamente a una incongruencia cuando ha habido una experiencia musical previa a largo plazo. En otro estudio con potenciales evocados (Koelsch, Schröger, & Tervaniemi, 1999) se presentaron acordes mayores y tonos simples en una condición atencional y una preatencional. Se encontró una activación fisiológica (MMN) más fuerte en el grupo de violinistas profesionales incluso a un nivel preatencional lo cual pareció beneficiar a este grupo para poder detectar pequeñas diferencias entre tonos que el grupo de no músicos no pudo discriminar. Un estudio posterior (Tervaniemi, et al., 2005) generalizó estos datos a músicos profesionales de distintos instrumentos (no sólo violín) y además analizaron otros componentes, el N2b, la P3a y la P3b, en una condición preatencional y atencional. Los autores concluyeron que la diferencia entre grupos (músicos vs. no músicos) parece ser más atencional que preatencional, ya que mientras que la amplitud de los componentes N2b y P3 durante la condición atencional es mayor en el grupo de sujetos músicos profesionales, no se encontraron diferencias significativas entre grupos en la condición de escucha pasiva (preatencional). En esta última condición los grupos de músicos y no músicos no diferenciaron en la amplitud de la MMN o de la P3a.

Por último, cabe nombrar en este trabajo que los efectos de la experiencia musical han sido también estudiados en el contexto de la imaginería mental auditiva. Janata y Paroo (2006) encontraron correlaciones significativas entre el número de años de entrenamiento musical y la ejecución de los sujetos en una tarea de discriminación tonal en las dos condiciones, cuando los sujetos escuchaban la escala completa (percepción) y cuando imaginaban parte de ella (imaginería). Aleman y colaboradores (Aleman, Nieuwenstein, Böcker, & Haan, 2000) buscaron la posible influencia que la experiencia musical pudiera tener en una tarea de imaginería mental auditiva. Los sujetos fueron evaluados en dos tareas de imaginería mental auditiva y una tarea de imaginería visual. En una tarea auditiva los sujetos tenían que comparar mentalmente los tonos (la altura tonal) de palabras que aparecen en canciones familiares. En la restante tarea auditiva, los sujetos comparaban mentalmente características acústicas de sonidos cotidianos (ej., silbido de un pájaro). Por último, en la tarea visual, se les pedía comparar mentalmente distintas formas de objetos. Mientras que el grupo con experiencia musical parecía tener más facilidades no sólo en la tarea de imaginería musical sino también en la segunda tarea de imaginería auditiva, no se encontraron diferencias significativas con respecto a la ejecución de la imaginería visual. Los autores concluyen que, debido a que la corteza auditiva es sede de la percepción y la imaginería, el entrenamiento de la corteza auditiva tras años de experiencia musical llevará a un beneficio de la imaginería mental auditiva en general, y no únicamente de aquella que trata con estímulos musicales.

Aprendizaje perceptivo

Los efectos del entrenamiento con estímulos musicales en la imaginería musical no han sido estudiados todavía (Halpern y Zatorre, 2005). El capítulo 4 de este trabajo se centrará en el estudio de la influencia del aprendizaje perceptivo (en música) en la calidad de las imágenes mentales musicales mediante la técnica de los potenciales evocados.

El concepto de aprendizaje perceptivo se refiere a aquellos cambios que se producen en el sistema tras un periodo de entrenamiento con los estímulos. Tales cambios son generalmente positivos o benefician al sistema. En la modalidad visual, por ejemplo, el entrenamiento en una tarea de percepción visual lleva a cambios en la retina relacionados con mejoras en la ejecución en la tarea (Karni & Sagi, 1993). Dependiendo del tipo de tarea presentada, el aprendizaje perceptivo puede requerir de múltiples sesiones hasta empezar a observar los primeros cambios. Estos autores (Karni & Sagi, 1993) por ejemplo, encontraron cambios rápidos en la primera sesión de entrenamiento en la tarea y después hubo un periodo de latencia en el que no hubo mejora apreciable. Se observó poca mejoría incluso hasta después de ocho horas de entrenamiento. Tras ese periodo de latencia se encontraron grandes mejoras que serían permanentes hasta al menos tres años después.

La técnica de los potenciales evocados ha sido utilizada para estudiar los cambios neuronales en la corteza auditiva tras un entrenamiento en una tarea auditiva. Tremblay y cols. (Tremblay, Kraus, McGee, Ponton, & Otis, 2001) entrenaron a un grupo de 7 sujetos a apreciar la diferencia entre dos variantes de la sílaba /ba/ durante 10 sesiones. Tras las sesiones de entrenamiento los sujetos aprendieron a diferenciar los dos sonidos y esto se vio reflejado en los componentes electrofisiológicos evocados. Tras el entrenamiento, la presentación de las sílabas producían un incremento en la N1 y la P2 y una disminución de la amplitud de la P1.

En cuanto al aprendizaje perceptivo en música, uno de los hechos más estudiados ha sido la especificidad vs. generalización del aprendizaje. Específicamente, en experimentos de aprendizaje tonal se ha estudiado la posible generalización del aprendizaje perceptivo en una determinada frecuencia a otras frecuencias. Demany y Semal (2002) realizaron dos experimentos en los que los sujetos aprendían a discriminar en múltiples sesiones de entrenamiento pequeños cambios tonales. Los autores concluyeron que el aprendizaje en discriminación tonal es específico a la frecuencia y al oído en los que se entrena, y ese entrenamiento no se puede transferir al oído opuesto o a otras frecuencias. Del mismo modo, en el segundo experimento, los sujetos fueron entrenados a distinguir cambios en tonos complejos y probaron más tarde que éste aprendizaje no era transferido a sonidos puros. A partir de esos datos concluyeron que el aprendizaje para la discriminación de tonos es específico del timbre (o el espectro).

Se ha especulado que el aprendizaje que puede ser transferido a otros estímulos es más bien un aprendizaje procedural de la tarea o algún aspecto del procedimiento, mientras que aquel aprendizaje relacionado con las características de los estímulos no será transferido para el procesamiento de otros estímulos similares (Robinson & Summerfield, 1996). Sin embargo, otros estudios de discriminación tonal han encontrado generalización a otras frecuencias (Delhommeau, Micheyl, & Jouvent, 2005; Irvine, Martin, Klimkeit, & Smith, 2000) y han mostrado que aunque el aprendizaje pueda ser parcialmente específico a la frecuencia en que se entrena, este aprendizaje facilita el aprendizaje en otras frecuencias, de ese modo, cuando a los sujetos se les presenta frecuencias nuevas, se produce un aprendizaje mucho más rápido que el aprendizaje producido con la primera frecuencia, ya que los sujetos podrían haber aprendido, por ejemplo, a qué características específicas del estímulo deben atender, lo cual podría beneficiar el proceso de aprendizaje (Delhommeau, et al., 2005).

Por otro lado, resulta interesante el estudio del curso temporal que siguen los cambios neuronales producidos por el aprendizaje perceptivo. Se ha distinguido entre cambios neuronales rápidos y lentos. Mientras que algunos estudios han mostrado cambios producidos dentro de la misma sesión (Poggio, Fahle, & Edelman, 1992), la mayoría de los estudios en aprendizaje perceptivo defienden que no se produce aprendizaje hasta después de muchas sesiones, por ejemplo, ocho horas de entrenamiento (Karni & Sagi, 1993), o incluso varios días (Schoups, Vogels, & Orban, 1995). Gilbert (1994) hace una revisión sobre los cambios neuronales rápidos producidos tras un breve entrenamiento y explica que estos cambios podrían deberse a modulaciones en el campo receptivo de neuronas de la corteza primaria. Por otro lado, el sueño REM parece tener un efecto añadido sobre las consecuencias del aprendizaje perceptivo y parece ser necesario para que el aprendizaje se exprese (Gilbert, 1994) reflejando la consolidación de la información en la memoria a largo plazo (cambios neuronales lentos). Estos cambios más lentos tras días de aprendizaje o tras un periodo de sueño una vez que se ha dado un entrenamiento perceptivo parecen estar más relacionados con una reorganización de las representaciones corticales. Atienza y colaboradores (Atienza, Cantero, & Dominguez-Marin, 2002) presentaron un pretest y 5 post-tests (justo después de que se diera la sesión de entrenamiento, 12, 24, 36 y 48 horas después) en los que los sujetos escuchaban pasivamente dos patrones de 5 sonidos. Tras el pretest, presentaron una sesión de entrenamiento en la que los sujetos aprendieron a discriminar los patrones de sonidos. Los autores analizaron los cambios electrofisiológicos producidos debido al entrenamiento y hallaron aumentos en la amplitud de la *mismatch negativity* y la P2, no sólo en la sesión tras el entrenamiento, sino también en sesiones posteriores, demostrando que existen cambios neuronales rápidos y más lentos subyacentes al aprendizaje perceptivo. Por otro lado, existe también evidencia de que estos cambios neuronales más lentos son permanentes, y los beneficios encontrados después de varias sesiones de entrenamiento permanecen visible incluso tras varios años después (Kerni & Sagi, 1993).

El papel que el entrenamiento perceptivo en música tiene en la imaginería musical está todavía por estudiar (Zatorre & Halpern, 2005), y en general, no existe mucha investigación sobre los efectos del aprendizaje perceptivo en las imágenes mentales. Al contrario, un estudio reciente ha señalado las mejoras en una tarea de percepción visual tras darse un entrenamiento en la tarea con imaginería (Tartaglia, Bamert, Mast, & Herzog, 2009). Los sujetos realizaron un pretest y un post-test en los que tenían que decidir si la línea central de una bisección (tres líneas) se encontraba ligeramente más hacia la izquierda o hacia la derecha. Durante el entrenamiento con imaginería los sujetos fueron presentados únicamente con las líneas exteriores y se les pedía que imaginasen la línea central ligeramente hacia la izquierda o ligeramente hacia la derecha. Los autores hallaron que las repetitivas imágenes mentales de tales estímulos visuales fue suficiente para encontrar mejoras similares a las del aprendizaje perceptivo. Este último hallazgo además presenta más evidencia sobre el posible solapamiento neuronal de la percepción y la imaginería.

Objetivos de la investigación

Durante los siguientes 3 capítulos se presentan experimentos conductuales y electrofisiológicos para comparar la precisión de las representaciones mentales auditivas cuando estas están formadas en una tarea de arriba a abajo (*top-down*, imaginería) o de abajo a arriba (*bottom-up*, percepción), y el estudio de los factores implicados en la precisión de estas representaciones.

En el segundo capítulo de este trabajo se analizará la contribución que los distintos tipos de memoria (sensorial, memoria a corto plazo y memoria a largo plazo) tienen en la precisión de las imágenes / representaciones mentales auditivas formadas en una tarea *top-down* y en una tarea *bottom-up.* En una serie de tres experimentos conductuales se evalúa cómo de eficaces son las representaciones mentales de la información tonal dependiendo de varios factores, por ejemplo, dependiendo de la relación tonal de las representaciones mentales con el contexto musical (memoria implícita -experimento 1-), o de la repetición de los estímulos musicales o el entrenamiento (memoria a corto plazo -experimentos 2 y 3-). De este modo, se espera que las imágenes formadas en una tarea de imaginería tengan una precisión más parecida a aquellas representaciones formadas en una tarea de percepción cuando el contexto musical esté más relacionado con éstas, o cuando se de un entrenamiento con los estímulos. También se estudiará la correlación existente entre las tareas de percepción e imaginería presentadas y una tarea típica de discriminación tonal donde las representaciones mentales se basan en la memoria sensorial (experimento 3). Se sugiere que las imágenes mentales formadas en una tarea de percepción (*bottom-up*) y aquellas formadas en una tarea de imaginería (*top-down*) están influenciadas por estos distintos sistemas de memoria en un grado diferente. Nuestra hipótesis sugiere que la precisión de las imágenes formadas en una tarea de imaginería estará más influenciada por la experiencia con el contexto tonal (memoria a largo plazo) mientras que la precisión de las imágenes formadas *bottom-up* en una tarea de percepción tendrá una correlación más fuerte con aquellas representaciones basadas en huellas de memoria sensorial.

En el capítulo tercero, se utiliza la técnica de los potenciales evocados para estudiar los correlatos neuronales subyacentes a la formación de las imágenes mentales. En este experimento, se estudian las diferencias conductuales y psicofisiológicas entre dos grupos, aquellos participantes con imágenes mentales precisas y aquellos incapaces de formar imágenes de forma eficaz. En primer lugar, se espera que las representaciones mentales formadas en una tarea de imaginería sean de igual calidad a las huellas de memoria sensorial hasta el punto de evocar respuestas neuronales similares (N100 y P300).

Además, se espera que las imágenes o representaciones más relacionadas con el contexto tonal tengan una mayor calidad que aquellas menos relacionadas con el contexto y como consecuencia se encuentren diferencias psicofisiológicas en los procesos evocados por la nota *target* en estas dos condiciones.

En el capítulo cuarto se estudian los efectos del aprendizaje perceptivo en las imágenes mentales. Como sugieren Zatorre y Halpern (2005), existe mucha investigación en cuanto a la imaginería musical, sobretodo en lo que concierne a la activación de la corteza auditiva en ausencia de estimulación musical, sin embargo, quedan muchas cuestiones por responder, como por ejemplo los efectos del entrenamiento en la calidad de las imágenes mentales. En el capítulo cuarto, con ayuda de la técnica de potenciales evocados, se estudian los cambios neuronales en una tarea de imaginería musical debidos al aprendizaje en una tarea de percepción musical. Presentamos un paradigma similar que en los capítulos anteriores en el que los sujetos aprenden a discriminar el tono de la última nota de una escala. Los sujetos reciben entrenamiento en una tarea de percepción, y mediante un pretest y dos post-tests se analiza si este aprendizaje perceptivo ha llevado a cambios neuronales que afecten también la ejecución de los sujetos en una tarea de imaginería. Si, como muestra la evidencia, la imaginería mental y la percepción comparten substratos neuronales, el entrenamiento perceptivo en una tarea de discriminación tonal llevará a una mejora en una tarea en la que se requiera la formación de imágenes mentales de los mismos tonos.

Finalmente, en el último capítulo se resumen los resultados encontrados en cada uno de los experimentos y se discuten conforme a los datos y la evidencia de la que se ha hablado en el capítulo primero.

General aims and research plan

Pitch discrimination acuity has typically been studied with tasks in which two tones are presented and the subject decides whether the tones are the same or different in pitch (Demany, 1985; (Micheyl, Delhommeau, Perrot, & Oxenham, 2006). In these studies, pitch judgments are based on the sensory memory trace of the tones. Janata and Paroo (2006) presented a paradigme in which subjects had to either imagine (imagery condition) or listen (cued-attention condition) to a musical scale and decide whether the last note was in tune or out of tune, or whether the last note was correct in time. In both cases, judgments are made based on images or representations evoked by the preceding images (imagery) or notes (cued-attention). When people listen to the complete scale, the sensory information will lead to the formation expectations about the incoming information (bottom-up). This task has been referred to as a cue-attention task because every note in the sequence serve as a predictive cue and directs attention towards the following pitch height in the sequence. On the other hand, when a familiar sequence of notes is mentally imagined (top-down), mental images will be evoked by the preceding images in the sequence.

Pitch discrimination acuity has been proven to be influenced by different forms of short-term and long-term memory. Musical experience together with the effects of training (short-term memory) have been analyzed in several studies (Micheyl, et al., 2006; Spiegel, 1984). Micheyl et al. (2006), for example, found that non-musicians' thresholds were more than 6 times larger (less accurate) than those of musicians. In a second experiment they found that it took from 4 to 8 hours of training for the acuity in the non-musicians group to become as good as that of the musicians group in the first experiment.

The implicit learning of the tonal structure (long-term memory) has also been shown to improve the acuity of pitch representations. The different tones in a diatonic scale are hierarchically structured and some tones (tonic) are more frequent than others (leading tone), meaning there will be tones that are more expected than others due to a priming effect (Krumhansl, 1990). As a result, in a pitch discrimination task, those tones more primed by the tonal context (tonic) are processed more quickly and /or accurately (Bharucha & Stoeckig, 1986; Janata & Reisberg, 1988).

There are only a few studies that consider the effects of memory on an imagery task. Regarding musical experience, for example, (Aleman, et al., 2000) looked at the effects of this factor on auditory imagery and found that more years of musical experience benefits not only auditory images in a music task, but also other kinds of auditory imagery (mentally comparing characteristics of environmental sounds). Herholz et al. (2008) presented incomplete melodies and subjects had to imagine the notes that were not played and decide whether a last (played) note was a correct or incorrect (in pitch) continuation of the (imagined) melody. They compared musicians and non-musicians by means of magnetoencephalography and found that an MMN (mismatch negativity) component after the presentation of an incorrect continuation was only appreciated in the group of musicians as a result of a stronger imagery process in this group.

Janata and Paroo (2006) also presented evidence of the influence of memory on pitch image acuity due to both the effects of musical experience and the influence of the tonal context. They found that those images formed for the tonic, the most primed note in the scale, were more accurate than those formed for the leading tone, one of the least primed note in the scale. They also found positive correlations between the number of years of musical experience and the acuity of pitch images.

The main goal of the research for this dissertation is to compare the acuity of mental images when these are formed under imagery as compared to when they are created in a perception (cued-attention) task, and to study different memory variables that influence the accuracy of images. In chapter 2, the relation of long-term, short-term and sensory memory to the acuity of auditory images is considered in three behavioral experiments. We hypothesize that those images formed under an imagery task will be more influenced by the long-term memory variables manipulated, while images formed in a cued-attention task will be more related to a sensory memory task. We use the paradigm presented in Janata and Paroo (2006). In that study, however, the same tonal key was used throughout each of the experiments, which means that the stimuli presented were repeated in every trial. The repetition of the same scale in every trial may have led to training on recognition of the absolute pitch of the last (target) tone without having to make use of the tonal context information provided by the first heard notes in the scale, or even without having to imagine the remaining notes in the "imagery" scale. In the study presented in chapter 2, the repetition of the tonal key is controlled in two of the three experiments in order to find out whether the results found in Janata and Paroo could be partly influenced by the repetition of the stimuli.

In the first experiment, we study the influences of the tonal context on image acuity. Subjects performed two tasks, a perception (bottom-up) task, and an imagery (top-down) task, in one block in which the target is the tonic (primed by the tonal context) and another block in which the target is the leading tone (less primed by the context).

In the second experiment, we look at whether there are any short-term memory effects as a result of the training with the absolute pitch of the target (due to the repetition of the same stimuli). In this sense, two more blocks are presented in which subjects perform the same task, but the scales vary their key from trial to trial.

In the third experiment, an additional "two-tone" pitch discrimination task is presented. In this task judgments are based on sensory memory traces. This additional task is presented in order to study whether the acuity results in this task are related in a different degree to those images formed in the top-down and those formed in the bottom-up tasks.

In chapter 3, pitch acuity thresholds are calculated in two tasks (imagery and cued-attention) while electrophysiological activity is recorded in order to study the neural activity underlying the formation of mental images. Two groups of participants are formed a posteriori based on their acuity thresholds and the physiological differences between the two groups are analyzed. We first hypothesize that images formed in the imagery task can be as accurate as those formed under a perception task and lead to similar neural responses after the subjects listen to the target note (P3a). Moreover, we also look at the possible ERP effects due to the influence of the tonal context. In this experiment subjects had to randomly listen to scales in which the target was the tonic or the leading tone. We expect stronger ERP effects evoked by the tonic due to its relation to the tonal context.

In chapter 4, the effects of perceptual learning on imagery are studied by means of ERPs. The role of training in the development of mental imagery remains to be understood (Zatorre & Halpern, 2005). Also, because the literature indicates a neural overlapping between imagery and perception in the auditory cortex, we expect that training to discriminate slight pitch changes in a perception task will contribute to the creation of more accurate analogous pitch images in mind. A recent behavioral study has demonstrated that an improvement similar to that of perceptual learning can be achieved by training with mental images instead of external stimuli (Tartaglia, et al., 2009). Subjects had to decide whether the central line in a bisection was offset to the right or the left. After a training session in which they were presented with the lateral lines and had to imagine the central offset line, the authors found an improvement in the actual perceptual task. In the experiment presented in chapter 4 the opposite direction is taken and we expect to find an improvement in the imagery task when subjects are trained in a perception task.

Referencias

Aleman, A., Nieuwenstein, M. R., Böcker, K. B. E., & Haan, E. H. F. (2000). Music training and mental imagery ability. *Neuropsychologia, 38*, 1664-1668.

Atienza, M., Cantero, J. L., & Dominguez-Marin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learning and Memory, 9*, 138-150.

Attneave, F., & Olson, R. K. (1971). Pitch as a Medium - New Approach to Psychophysical Scaling. *American Journal of Psychology, 84*(2), 147-&.

Bachem, A. (1954). The factors in relative and absolute pitch determination. *Journal of the Acoustical Society of America, 26*, 751-753.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 47-89). New York: Academic Press.

Baddeley, A. D., & Logie, R. (1992). Auditory imagery and working memory. In D. Reisberg (Ed.), *Auditory Imagery*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

Besson, M., & Faïta, F. (1995). An Event-Related Potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance, 21*(6), 1278-1296.

Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance,* *12*(4), 403-410.

Bigand, E., & Pineau, M. (1997). Global context effects on musical expectancy. *Perception and Psychophysics, 59*, 1098-1107.

Brattico, E., Näätänen, R., & Tervaniemi, M. (2001). Context effects on pitch perception in musicians and non-musicians: Evidence from ERP recordins. *Music Perception, 19*, 199-222.

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H., & Jancke, L. (2005). Scanning silence: mental imagery of complex sounds. *Neuroimage, 26*, 1119-1127.

Chambers, D., & Reisberg, D. (1985). Can mental images be ambiguous? . *Journal of Experimental Psychology: Human Perception and Performance, 11*, 317-328.

Delhommeau, K., Micheyl, C., & Jouvent, R. (2005). Generalization of frequency discrimination learning across frequencies and ears: implications for underlying neural mechanisms in humans. *Journal of Assoc. Res. Otolaryngol., 6*, 171-179.

Demany, L., & Semal, C. (2002). Learning to perceive pitch differences. *Journal of the Acoustical Society of America, 111*, 1377-1388.

Deutsch, D. (1978). Delayed pitch comparisons and the principle of proximity. *Perception and Psychophysics, 23*(3), 227-230.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85*(4), 341-354.

Dowling, W. J., & Diane, S. F. (1971). Contour, Interval, and Pitch Recognition in Memory for Melodies. *The Journal of the Acoustical Society of America, 49*(2B), 524-531.

Frith, C., & Dolan, R. (1996). The role of the prefrontal cortex in higher cognitive functions. *Cognitive Brain Research, 5*, 175-181.

Fuster, J. (1995). *Memory in the Cerebral Cortex - An Empirical Approach to Neural Networks in the Human and Nonhuman Primate*. Cambridge, MA.

Fuster, J. (2001). The Prefrontal Cortex -An update: Time is of the essence. *Neuron, 30*, 319-333.

Gilbert, C. D. (1994). Early perceptual learning. *Proceedings of the National Academy of Sciences, 91*, 1195-1197.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex, 9*, 697-704.

Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia, 42*, 1281-1292.

Herholz, S. C., Lappe, C., Knief, A., & Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. *European Journal of Neuroscience, 28*, 2352-2360.

Hubbard, T. L., & Stoeckig, K. (1988). Musical imagery: Generation of tones and chords. *Journal of Experimental Psychology: Learning, Memory & Cognition, 14*, 656-667.

Irvine, D. R., Martin, R. L., Klimkeit, E., & Smith, R. (2000). Specificity of perceptual learning in a frequency discrimination task. *Journal of Acoustic Society of America, 108*, 2964-2968.

Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. *Perception and Psychophysics, 68*, 829-844.

Janata, P., & Reisberg, D. (1988). Response-time measures as a means of exploring tonal hierarchies. *Music & Perception, 6*, 161-172.

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature, 365*, 250-252.

Koelsch, S., Schröger, E., & Tervaniemi, M. (1999). Superior attentive and pre-attentive auditory processing in musicians. *NeuroReport, 10*, 1309-1313.

Kraemer, D. J., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: sound of silence activates auditory cortex. *Nature, 434*(7030), 158-158.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Launay, G., & Slade, P. D. (1981). The measurement of hallucinatory predisposition in male and female prisoners. *Personality and Individual Differences, 2*, 221-234.

Lennox, B. R., Park, S. B., Jones, P. B., & Morris, P. G. (1999). Spatial and temporal mapping of neural activity associated with auditory hallucinations *Lancet, 353*, 644.

McGuire, P. K., Silbersweig, D. A., Wright, I., Murray, R. M., Frackowiak, R. S., & Frith, C. (1996). The neural correlates of inner speech and auditory verbal imagery in schizophrenia: relationship to auditory verbal hallucinations. *British Journal of Psychiatry, 169*, 148-159.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*, 36-47.

Mitnz, S., & Alpert, M. (1972). Imagery vividness, reality testing and schizophrenic hallucinations. *Journal of Abnormal Psychology, 79*, 310-316.

Neisser, U. (1976). *Cognition and reality.* San Francisco.

Penfield, W., & Perot, P. (1963). The brain's record of auditory and visual experience: A final summary and discussion. *Brain, 86*, 595-696.

Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity *Science, 256*, 1018-1021.

Robinson, K., & Summerfield, Q. A. (1996). Adult auditory learning and training. *Ear and Hearing, 17*(3), 51-65.

Schön, D., & Besson, M. (2005). Visually induced auditory expectancy in music reading: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 17*(4), 694-705.

Schoups, A. A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: Retinotomy, orientation specifity and monocularity. *Journal of Physiology, 483*, 797-810.

Segal, S., & Fusella, V. (1970). influence of imaged pictures and sounds on detection of visual and auditory signals. *Journal of Experimental Psychology, 83*, 458-464.

Shergill, S. S., Brammer, M. J., Williams, S. C., Murray, R. M., & McGuire, P. K. (2000). Mapping auditory hallucinations in schizophrenia using functional magnetic resonance imaging. *Archives of General Psychiatric, 57*(11), 1033-1038.

Smith, J. D., Reisberg, D., & Wilson, M. (1992). Subvocalization and auditory imagery: Interactions between the inner ear and inner voice. In D. Reisberg (Ed.), *Auditory Imagery* (pp. 95-119). Hillsdale, NJ: Erlbaum Associates.

Spiegel, M. F. (1984). Performance on Frequency-Discrimination Tasks by Musicians And. *Journal of the Acoustical Society of America, 76*(6), 1690-1695.

Stanovich, K. E., & West, R. F. (1979). Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory and Cognition, 7*, 77-85.

Tartaglia, E. M., Bamert, L., Mast, F. W., & Herzog, M. H. (2009). Human Perceptual Learning by Mental Imagery. *Current Biology, 19*, 2081-2085.

Tervaniemi, M., Just, V., Koelsch, s., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Experimental Brain Research, 161*, 1-10.

Tillmann, B., Bigand, E., & Pineau, M. (1998). Effects of global and local contexts on harmonic expectancy. *Music Perception, 16*, 99-118.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training

Ear and hear, 22, 79-90.

Yumoto, M., Uno, A., Itoh, K., Karino, S., Saitoh, O., Kaneko, Y., et al. (2005). Audiovisual phonological mismatch produces early negativity in auditory cortex. *NeuroReport, 16*, 803-806.

Zatorre, R. J. (1999). Brain imaging studies of musical perception and musical imagery. *Journal of New Music Research, 28*(3), 229-236.

Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: Musical imagery and auditory cortex *Neuron, 47*, 9-12.

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience, 8*, 29-46.

CHAPTER II,

INFLUENCES OF MULTIPLE MEMORY SYSTEMS ON AUDITORY MENTAL IMAGE ACUITY

Published as Navarro Cebrian, A., & Janata, P. (in press). Influences of multiple memory systems on auditory mental image acuity. *Journal of the Acoustical Society of America*.

Abstract

The influence of different memory systems and associated attentional processes on the acuity of auditory images, formed for the purpose of making intonation judgments, was examined across three experiments using three different task types (cued-attention, imagery, and two-tone discrimination). In Experiment 1 the influence of implicit long-term memory for musical scale structure was manipulated by varying the scale degree (leading tone versus tonic) of the probe note about which a judgment had to be made. In Experiments 2 and 3 the ability of short-term absolute pitch knowledge to develop was manipulated by presenting blocks of trials in the same key or in seven different keys. The acuity of auditory images depended on all of these manipulations. Within individual listeners, thresholds in the two-tone discrimination and cued-attention conditions were closely related. In many listeners, cued-attention thresholds were similar to thresholds in the imagery condition, and depended on the amount of training individual listeners had in playing a musical instrument. The results indicate that mental images formed at a sensory/ cognitive interface for the purpose of making perceptual decisions are highly malleable.

Introduction

Mental representations (images) of auditory information arise in a variety of behavioral contexts, and must, in principle, arise via one of two routes: either as a result of “bottom-up” sensory processing of an external sound or as a result of “top-down” processes in which the auditory object (sound) is anticipated or imagined. These representations necessarily interact in many common situations. For example, when listening to a familiar melody we are able to anticipate upcoming notes, and we readily discern if their pitches or timbres have unexpectedly changed. Of interest is how these features of auditory mental representations are influenced by mnemonic and attentional processes. Specifically, how does the acuity of a mental image depend on the short-term and long-term memory systems and attentional processes that are invoked to create and maintain the image?

The property of auditory objects that has received the most attention is pitch, and it is this aspect of auditory images that we focus on also. The motivation for our study arose from the recognition that across published studies, the properties of auditory images are examined using diverse paradigms. Such diversity makes it difficult to determine whether and how the mental processes underlying the different paradigms influence a common pitch representation (Figure 1). The acuity of pitch representations of pure tones and harmonic complexes is typically studied using 2-alternative forced choice (2AFC) discrimination tasks (two-note discrimination) in which one sound is compared with a second sound that follows after a brief interval (Demany, 1985; Micheyl, Delhommeau, Perrot, & Oxenham, 2006; Spiegel, 1984; Wier, 1977).

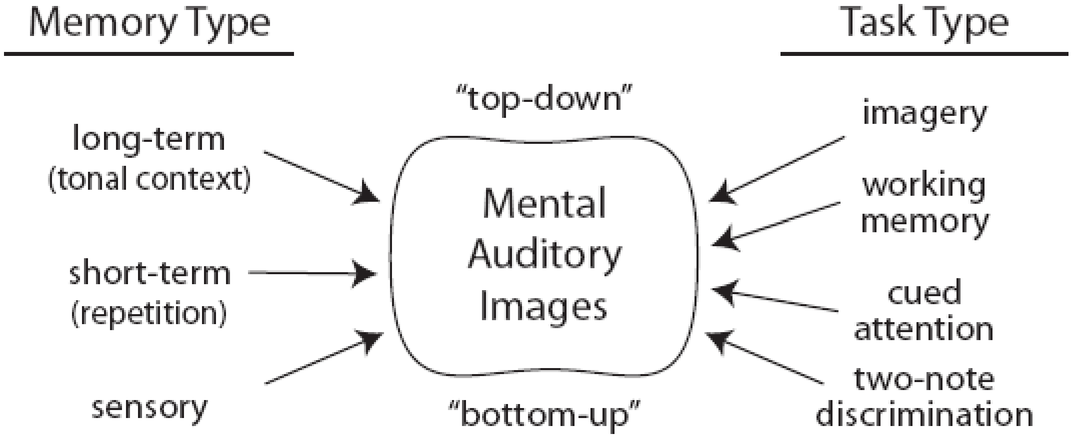


***Figure 1**. Auditory images are studied in many different types of tasks. Image acuity may be influenced by different types of mnemonic processes implicitly engaged by the various task requirements.*

Discrimination tasks are the most common and perhaps most precise behavioral tool available to psychologists to probe the mental representations arising via a "bottom-up" sensory route, and they establish the fidelity of an auditory image within a short term sensory store (Cowan, 1984). By contrast, a task in which a participant is cued to imagine one or more pitches that are then compared with a target pitch or melody might be regarded as a measure of the acuity of a representation arising via a "top-down" process (Crowder, 1989; A.R. Halpern, 1988).

It is likely that the process of imagining a sequence of notes relies on the "phonological loop" as characterized extensively in verbal working memory tasks

(Baddeley & Logie, 1992), given that task manipulations that impact the maintenance of items in the phonological loop, such as articulatory suppression, have been shown to also influence imagined items in auditory imagery tasks (Smith, Wilson, & Reisberg, 1995). Working memory tasks emphasize the maintenance of images in a short-term memory store, but the items in the (phonological) store may arrive there via a "bottom-up" sensory route, e.g. hearing a note or sequence of notes, or from long-term memory stores, e.g. recalling and maintaining a familiar melody in mind. Thus, the acuity of an auditory image in working memory may depend on the source of the image. While factors that disrupt the maintenance of images in working memory have been thoroughly investigated, to our knowledge the acuity of images deriving from different sources has not been examined.

Other types of tasks may utilize both sensory and long-term information, and it is difficult to place them squarely within a sensory memory or working memory framework. For example, what we call a "cued-attention" task is one in which a context serves as a predictive sensory cue that immediately precedes a target. For example in visuospatial attention tasks such as the Posner cueing paradigm, attention can be summoned to a specific spatial location by presenting a cue at the location of a subsequent target or via a central cue, such as an arrow, that indicates the location to which attention should be oriented (Posner, Snyder, & Davidson, 1980). Similarly, the notes in a highly familiar musical stimulus such as a melody or scale serve to orient attention to a specific pitch at a specific moment in time (Jones, 1981; Jones, Moynihan, MacKenzie, & Puente, 2002). In the case of an ascending scale, the mental image of each successive pitch cannot be based on a representation of that pitch in sensory memory because it has not been heard yet. Even though information about the next pitch must come from a longer-term memory store, there is typically no need to maintain the upcoming pitch in working memory over an extended delay period.

Given the different sources of auditory objects in discrimination, cued-attention, working memory and imagery tasks, one might be tempted to conclude that the different tasks are probing the maintenance and acuity of inherently different objects. Indeed, the nature of the similarities and differences between auditory mental representations generated by different means has generated considerable discussion (Reisberg, 1992). Nevertheless, auditory images, be they formed in laboratory target detection tasks or in more naturalistic environmental, musical, or linguistic contexts, must somehow interact with representations of the sensory input when sensory input is present, raising questions about the locus of such interactions and the acuity of the images at the site of interaction. A likely substrate for these interactions is the auditory cortex, as suggested by neuroimaging studies that found activation of the auditory cortex by imagery tasks (Bunzeck, Wuestenberg, Lutz, Heinze, & Jancke, 2005; Halpern & Zatorre, 1999; Kraemer, Macrae, Green, & Kelley, 2005; Zatorre, Halpern, Perry, Meyer, & Evans, 1996), and an event-related potential (ERP) study that found emitted potentials that mimic the N100 component of the auditory evoked potential (Janata, 2001). The N100 is an ERP component with generators in the auditory cortex

(Pantev, et al., 1995; Verkindt, Bertrand, Perrin, Echallier, & Pernier, 1995) that is known to be modulated by attentional demands (Näätänen, 1990, 1992; Naatanen & Picton, 1987). The locus of the interaction may be distributed and also encompass lateral prefrontal areas, such as Broca's area, as these areas are part of a common network for the generation and maintenance of verbal and tonal information (Brown, Martinez, & Lawrence, 2006; Koelsch, et al., 2009). Ventrolateral prefrontal areas respond more strongly when musical target events are unexpected, given a preceding context (Koelsch, et al., 2002; Maess, Koelsch, Gunter, & Friederici, 2001; Tillmann, Janata, Birk, & Bharucha, 2003), indicating a role for the lateral prefrontal cortex in comparing representations of sensory input with established contextual representation.

Across three experiments, we compared the relative influence of long-term and short-term memory processes on the acuity of mental images generated for the purpose of intonation judgments in three types of tasks. We elaborated on previous experiments that examined pitch acuity in the context of the cognitive operations underlying cued-attention and imagery tasks (Janata & Paroo, 2006). The basic task consists of an intonation judgment about the last note of a scale after listening to or imagining the preceding notes in the scale (Figure 2). In the cued-attention conditions all of the notes leading up to the target note are heard, whereas in the imagery condition, several notes preceding the final note are to be imagined. Even though each imagined note might be regarded as a predictive cue for the subsequent note, attention-orienting systems in such a situation do not have the benefit of predictive sensory information from one event to the next as they do in the cued-attention condition.

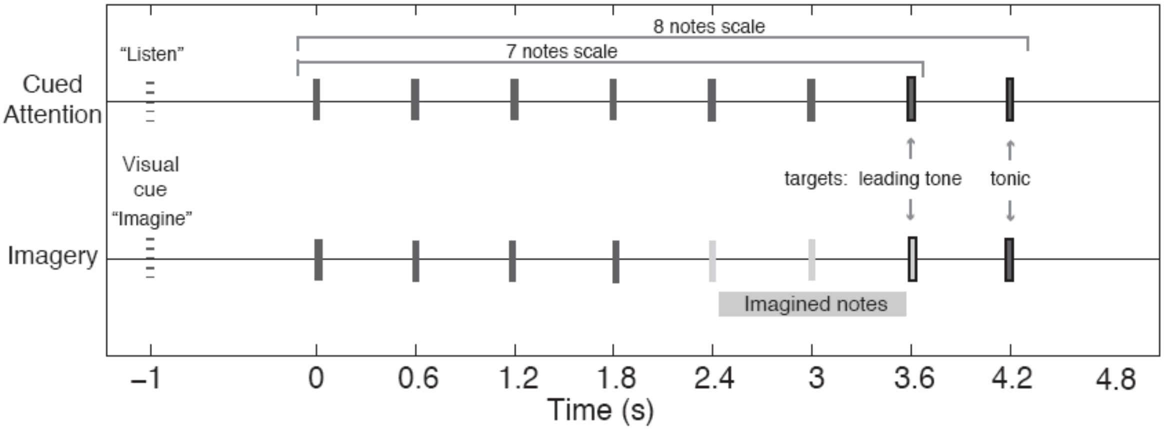


Figure 2. ***Schematic diagram of the top-down (imagery) and bottom-up (cued-attention) tasks used in the three experiments. The dashed vertical bars at -1 s indicate the onset of a visual cue that designated the task type. Solid vertical bars denote the onsets of scale tones. Light gray bars indicate imagined notes.***

Janata and Paroo (2006) found that pitch image acuity was comparable under cued-attention and imagery conditions when the stimuli remain physically identical across trials, although image acuity was slightly worse in the imagery condition and was additionally influenced by contextual and experiential factors. An effect of tonal (musical) context was observed in which the acuity for the most typical/stable note of a key (the tonic) was better than acuity for one of the least stable notes in the key (the leading tone), particularly when subjects performed a context membership task in which no imagery instructions were given. Image acuity was also better for participants who had played a musical instrument for a greater number of years.

Even though each imagined note might be regarded as a predictive cue for the subsequent note, attention-orienting systems in such a situation do not have the benefit of predictive sensory information from one event to the next as they do in the cued-attention condition. Janata and Paroo (2006) found that pitch image acuity was comparable under cued-attention and imagery conditions when the stimuli remain physically identical across trials, although image acuity was slightly worse in the imagery condition and was additionally influenced by contextual and experiential factors. An effect of tonal (musical) context was observed in which the acuity for the most typical/stable note of a key (the tonic) was better than acuity for one of the least stable notes in the key (the leading tone), particularly when subjects performed a context membership task in which no imagery instructions were given. Image acuity was also better for participants who had played a musical instrument for a greater number of years.

Although the experiments of Janata and Paroo (2006) demonstrated that mental images for pitch could attain similar acuity across different cognitive manipulations, it was possible that the physically invariant context and in-tune target that was heard on every trial allowed participants to develop absolute representations for the target pitch in short-term memory (Demany & Semal, 2002; Irvine, Martin, Klimkeit, & Stmith, 2000), thereby facilitating their performance in imagery trials. In other words, participants could have made judgments about the mistuning of terminal notes in the imagery conditions, without having to truly imagine the preceding missing notes in the scale in order to arrive at the final note. This possibility encouraged us to think about the influences of different memory systems and their interactions on image acuity, and led to the present set of experiments. The new experiments sought to replicate some of the previous observations as well as examine the relative influences of short-term versus long-term memory processes on pitch image acuity.

Across the three experiments we manipulated two variables to examine the influence of two forms of memory on image acuity. The first variable examined the influence of long-term implicit memory for tonal knowledge in music. The second manipulated the short-term memory for the target by either facilitating or hindering the build-up of absolute target representations. Keeping the key constant across blocks of trials was expected to facilitate image acuity, whereas varying the key from trial to trial was expected to hinder accurate image formation. Finally, we also studied the relative influence of sensory memory on image acuity by comparing image acuity under cued-attention and imagery conditions with image acuity in a two-tone pitch discrimination task in which the sensory trace of a target is compared directly with the sensory trace formed immediately prior. We expected that the acuity of images formed under influence of other sensory memory traces (cued-attention condition) would be more similar to the image acuity based on sensory memory, while the acuity of images formed under an imagery task (where no sensory information is presented and greater attentional resources are required to form accurate images) would be more influenced by long-term memory. In addition, given evidence that mental image acuity is correlated with the amount of musical training (Janata & Paroo, 2006; Micheyl, et al., 2006), we sought to determine the degree to which image acuity and musical training were related across the different task types.

Experiment I

Overview

We addressed two main issues in this experiment. First, we sought to replicate the effect of tonal knowledge on image acuity observed in Janata and Paroo (2006). Tonal knowledge is an example of implicit long-term memory that likely derives from extensive exposure to the statistical regularities of Western tonal music (Krumhansl, 1990; Tillmann, Bharucha, & Bigand, 2000). Even inexperienced participants possess an implicit knowledge of the harmonic structure of music, and this implicit memory influences the expectancies of the tones or chords that follow a priming context (Bigand, Madurell, Tillmann, & Pineau, 1999). With regard to the notes that make up major keys and minor keys, tonal knowledge manifests itself as a hierarchy of stability or goodness-of-fit, in which the first scale degree (the tonic) is considered the most stable note in the key, whereas the leading tone is among the least stable (Krumhansl, 1990). In this experiment, we expected tonal context to influence moment-to-moment expectations in our cued-attention condition and lead to more accurate images for the tonic compared to the leading tone. We expected the tonal context effect to be enhanced in the imagery condition, which demanded greater reliance on tonal knowledge.

A significant difference between the present experiment and the previous one (Janata & Paroo, 2006) was the use of seven keys (tonal contexts) instead of only one. This manipulation was introduced to emphasize an imagery strategy on each trial and interfere with an implicit buildup of a memory trace for a specific target pitch across trials. It was possible that in the previous study participants could simply resort to retrieving an absolute pitch memory of the target pitch on each trial in order to perform the task, rather than relying on forming a sequence of mental images in order to form an accurate image of the terminal pitch. Given that one of the target pitches in each trial was always in-tune, perceptual learning of the target pitch would prove effective for making finer discrimination judgments. Disrupting that process by changing the key on each trial should increase the efficacy of an imagery strategy in which each note leading up to the terminal note is imagined relative to the image of the preceding note based on knowledge of an ascending major scale.

Method

Participants Eighteen University of California, Davis students (age: M= 19.39 SD=1.24; 11 females) participated in exchange for partial course credit. The experiment was carried out with the informed consent of the participants in accordance with the guidelines of the Institutional Review Board at the University of California, Davis.

Stimuli Stimuli consisted of ascending major scales terminating on either the seventh scale degree (leading tone) or eighth scale degree (tonic). The notes were synthesized in MATLAB from their fundamental frequency and the next seven higher harmonics as described in Janata and Paroo (2006). A 5-ms linear ramp was applied to the envelope at the beginning and end of each note. The duration of each note was 250 ms and the onset asynchrony between successive notes was 600 ms. Seven different keys were used in this experiment. The starting note frequencies were: B2 (123.47 Hz), C3 (130.81 Hz), Db3 (138.59 Hz), D3 (146.83 Hz), Eb3 (155.56 Hz), E3 (164.81 Hz), F3 (174.61 Hz).

Figure 2 illustrates the basic task conditions. Stimuli in the *cued-attention* condition consisted of all of the notes leading up to and including the final target note (either the leading tone or tonic). Stimuli in the *imagery* condition consisted of the first four notes of the ascending scale, silence, and then a target note with an onset at the time that the leading tone or tonic would have occurred. The target note was either in-tune or out-of-tune. Pitch deviations for the target note stepped from -100 to +100 cents in 2.5 cent intervals.

Procedure The participants were seated in a darkened, sound-attenuated room. The stimuli were presented using Sony Professional headphones (MDR-7506) and the volume was adjusted at 70 dB. The task was implemented using Presentation software (Neurobehavioral Systems). Participants first underwent training with two different blocks of trials. First, to familiarize themselves with the stimuli, they listened to several examples of the cued-attention and imagery conditions, and both in-tune and out-of-tune targets. Second, in order to further familiarize themselves with the task, they performed a block of 12 trials simulating the actual experiment, though in contrast to the actual experimental trials, they received feedback after every trial.

Each trial consisted of two scales separated by one of ten different tonal masks. The tonal masks were random sequences of four notes in the same timbre as the scales of 250 ms of duration. One of the scales was the standard scale in which the last note (seventh or eighth) was in tune. The other was a probe scale in which the last note was out of tune. One second before the first scale started, a cue word appeared on the screen to indicate the trial type. The word was either “Listen” or “Imagine” and remained on the screen throughout the trial. In the cued-attention trials, the participants heard all notes leading up to the target note in both scales of the trial. In the imagery trials, the participants heard only the first four notes and the last one (leading tone or tonic) of each scale of the trial, and were instructed to imagine the notes that were not played. The target note was played at the time at which the last note should have occurred (3600 or 4200 ms, for the seventh or eighth note respectively).

The task of the participants was to either listen to or imagine the scales and press “F” or “J” on the keyboard to indicate in which of the two scales the terminal note was out of tune ("F" for the first scale with the note out of tune and "J" for the second). Imagery and cued-attention trials were interleaved randomly across trials and two thresholds (sharp and flat deviations) were estimated for each condition. There were two different blocks in the experiment. In one block, the scales terminated in the tonic (eighth note) and in the other block the scales terminated in the leading tone (seventh note). The order of the presentation of the blocks was randomized. Participants were instructed to refrain from making any rhythmic movements or vocalizations.

In order to minimize possible effects of fatigue and waning attention on the estimates of image acuity, thresholds were determined using the ZEST algorithm (Marvit, Florentine, & Buus, 2003). ZEST converges more quickly on threshold estimates than a traditional adaptive staircase procedure. In ZEST, a probability density function represents the probable distribution of the listener's threshold, and it is modified after every trial using Bayes' theorem given the participant's response. The starting deviation was ± 75 cents. The approximate duration of the experiment was 40 minutes. After the task, the participants answered several questions about their musical experience, their perception of the difficulty of the task and some other aspects of their mental images.

Data analysis The data were analyzed using SPSS 15. The dependent variable was the width of the mental images, defined as the difference between the positive (sharp) and the negative (flat) deviation thresholds. Data were evaluated using an ANOVA with Task Condition (cued-attention, imagery) and Probe Position (tonic, leading tone) as within-participants factors. Post-hoc t-tests were used for multiple comparisons using Bonferroni criterion for correction and an alpha of .05. In the leading tone block we expected that the offset of the images in the imagery condition would be shifted upward toward the tonic if the participants were looking for the best possible note that fits with the established key. Therefore, we performed an additional analysis using another property of the mental images as the dependent variable: the offset of the center of the mental image tuning curve, defined as the average of the positive and negative thresholds. A one-sample t-test was performed for the imagery condition in the leading tone group, using offset as the dependent variable.

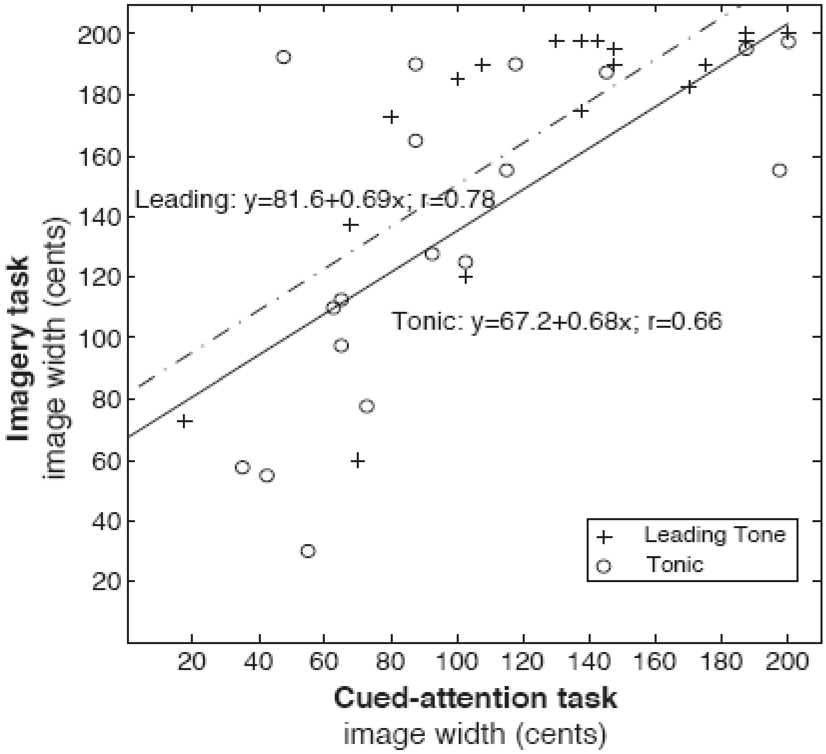
Table 1. Mean image widths (in cents) in Experiments 1–3

Experiment	Number of keys	Probe	Cued-attention	Imagery
Experiment 1	7 keys	Tonic	98.75 (12.39)	134.44 (12.88)
	7 keys	Leading tone	128.19 (11.54)	170 (10.26)
Experiment 2	1 key	Tonic	99.63 (11.26)	116.63 (10.44)
		Leading tone	122.38 (12.89)	160.25 (10.5)
	7 keys	Tonic	118.88 (10.01)	131.25 (10.73)
		Leading tone	130.38 (11.83)	175.15 (10.48)
Experiment 3	1 key	Tonic	89.21 (8.57)	114.44 (9.63)
	7 keys	Tonic	96.309 (6.6)	142.42 (7.53)

Note. Standard errors of the means are indicated in parentheses. Pitch values are given in cents.

Results

The means and standard deviations for the average image widths are shown in Table 1. There was a main effect of Task in that the width of the mental images was significantly greater in the imagery conditions than in the cued-attention conditions, $F(1,17) = 28.34$, $p < .0001$. We observed a significant effect of Probe Position, $F(1,17) = 19.54$, $p < .0001$, in which acuity was better for the tonic than for the leading tone. The Task X Probe Position interaction was not significant, F(1,17) = .581. In the leading tone condition, the offset in the imagery condition was not significantly different from 0, *t*(1,17) = 1.26, p < .222, indicating that mental images of the leading tone were not consistently biased toward the tonic.



<u>Figure 3.</u> Relationship of tuning thresholds obtained in cued-attention and imagery tasks in individual listeners judging either the intonation of the tonic or the leading tone in Experiment 1.

Figure 3 illustrates the variability in image acuity across participants and the significant correlations (Tonic: *r*(18) = .66, p<0.003; Leading tone: (18) = .78, p<0.0001) between image acuity in the cued-attention and imagery conditions for both the tonic and leading tone targets. The more accurate participants' images were in the cued-attention condition, the better they were in the imagery condition. It is important to note that some participants reached the maximal deviation limit of one semitone for either their sharp or flat thresholds. Thus their image widths may be underestimated.

Discussion

The observation of better image acuity in the cued-attention condition than in the imagery condition corroborates the previous report of Janata and Paroo (2006) in which Probe Position (tonic versus leading tone) was a between-participants manipulation. In contrast to the previous study, however, we observed no significant interaction of Probe Position and Task, and there was an effect of Task even for the tonic. Moreover, image acuity in the present study was generally poorer than in the previous study. The main difference between the previous experiment and the present one was that the present one employed seven different keys instead of one. The repetition of the same musical scale in every trial in the experiments reported in Janata and Paroo (2006) may have led to consolidation of a more accurate mental representation for the in-tune target pitch. In other words, the task may have become more of a target detection task in which one is monitoring a stream of sounds for a specific target. By contrast, the changing of keys from trial to trial in the present experiment reduced the likelihood that participants were judging target events against an invariant absolute representation, and that they were instead more likely to use a strategy of imagining the remaining notes in the scale if they were to arrive at an accurate representation of the target note in every trial.

Although the key changed from trial to trial, the relational structure of the stimuli remained invariant across trials. In other words, the pattern of notes that was heard and imagined remained the same. Although relational information in music, such as the successive steps in a major scale, has been shown to consolidate more easily in memory than absolute information (Deutsch, 1975), most participants appeared unable to make effective use of the relational information present in the ascending major scale, despite its constancy across trials. In the literature, there is some controversy about the consolidation of absolute memory. The use of an absolute frequency seems to require more experience and non-musicians do not usually make use of this information (Bachem, 1954). Some theories assume that categorization is necessary in order to consolidate information in long-term memory (LTM) storage, and data that cannot be coded categorically will be difficult to store in memory for extended periods. In this respect, pitch height might be regarded as a type of information that cannot be coded and some authors suggest that its trace decays in a few seconds (Bachem, 1954). However, other studies have shown that the codification or categorization of the information is not a necessary condition for the consolidation of pitch height information, and tone frequencies have been shown to consolidate in LTM thanks to repetition and rehearsal (Deutsch, 1972, 1975, 1979; Peynircioglu, 1995). In Deutsch

(1975), for example, the pitch of two tones had to be compared when they were separated by other tones. When the pitch of the first tone was repeated in the tones in between subjects' performance improved. Moreover, several studies have demonstrated that non-musicians are also able to remember a song in the key in which they have heard it most often (A. R. Halpern, 1989; A.R.. Halpern, 1992; Levitin, 1994; Schellenberg & Trehub, 2003). In accordance with the latter studies, it seems that participants are able to learn pitch height information when the scales are repeated over and over in the same key, even within the short-term context of this type of experiment.

In agreement with Janata and Paroo (2006), we found that participants' image acuity was better for the tonic than the leading tone, irrespective of task condition. Thus, implicit memory for the tonal context generally influences the acuity of mental images, although we did not find evidence that participants systematically mistuned images of the less stable leading tone toward the more stable tonic, as might be suggested by the typical musical role of the leading tone as a note that precedes, or resolves to, the tonic.

Experiment II

Overview

Taken together, the results of Experiment 1 and those reported in Janata and Paroo (2006) suggest that image acuity suffers when the tonal context (key) changes from trial to trial and participants are inhibited from forming short-term absolute memory representations and must instead rely on a more cognitively demanding imagery strategy in order to arrive at an appropriate mental image to compare against an incoming target. However, some of the observed differences in the experiments may have been driven by differences in participant samples. Therefore, in this experiment we used a within-participants design to compare image acuity between blocks of trials that favored either absolute (1-key) or relational (7-key) representation strategies, in order to better understand the influence of stimulus repetition on the formation of accurate images. We predicted that the acuity of images formed in the imagery condition would be more similar to the acuity of images formed by cued attention when the scales were repeatedly presented in a single key due to the learning effects of repetition.

Method

Participants Twenty students participated in this experiment (age: M = 19.45, SD =1.32; 14 females). None of them participated in the previous experiment. Seven of them had at least 1 year of musical training and the amount of training was 3.29 ± 3.02 (mean ± standard deviation). All the participants were compensated with course credit and the experiment was carried out with the informed consent of the participants in accordance with the guidelines of the Institutional Review Board at the University of California, Davis.

Stimuli The stimuli for the scales in this experiment were the same as the stimuli in Experiment 1. However, half of the trials in this experiment were presented in one single key while the other half were presented in seven different keys. The key used for the 1 key condition fell in the middle of the range of keys used in the seven-key blocks (D-major; first note: 146.83 Hz).

Procedure The paradigm was the same as in Experiment 1 (2AFC), using the ZEST procedure for threshold estimation (Marvit et al., 2003). Three factors of interest, Task Condition (cued-attention, imagery), Probe Type (tonic, leading tone) and Key Number (1 key, 7 keys) constituted a 2x2x2 design with each combination of Probe Position and Key Number forming one of four blocks of trials. The blocks of trials were counterbalanced across participants. Cued-attention and imagery trials were randomly interleaved within blocks. The approximate duration of this experiment was 80 minutes.

Data analysis Image widths were analyzed using an ANOVA with the three within-participants factors described above.

Results

Mean image widths for all conditions are provided in Table 1. There was a significant effect of Probe Position, $F(1,19) = 22.52$, $p < .0001$, with more accurate images for the tonic, and Key Number, $F(1,19) = 11.24$, $p < .003$, indicating more accurate images when only one key was used in a block of trials. As in Experiment 1, image widths were significantly narrower in the cued-attention than the imagery conditions, $F(1,19) = 20.53$, $p < .0001$. There was a significant interaction between Probe Position and Task, $F(1,19) =$

8.78, p < .008. A pairwise comparison showed that the averages for the imagery and cued-attention conditions differed significantly for the leading tone variable (one key condition: t(1,19) = 14.62; p < .001; seven keys condition: t(1,9) = 21.01; p < 0.0001), but not for the tonic (one key condition: t(1,19) = 3.81; p = .066 ; seven keys condition: t(1,19) = 2.7; p = .117). There was no interaction of Key Number and Task condition, F(1,19) = .032, and Probe Position and Key Number (F(1,19)=.837). The 3-way interaction of Key Number, Task Condition and Probe Position was not significant either, F(1,19) = .453.

As in Experiment 1, we observed a great deal of variability between participants. For example, the narrowest image width in the 1-key tonic condition was 10 cents and the worst image width in the same condition was 197.5 cents.

Discussion

As predicted, we found that participants were able to achieve comparable image acuity in cued-attention and imagery conditions, but only when the probe note was the tonic. It would appear that mental images formed in an imagery task have the same effect on performance accuracy as that of auditory expectations but this seems to be true only when the object to be imagined is strongly primed by the context (as is the case for the tonic). This result corroborated the result of Experiment 1 in Janata and Paroo (2006). The significant difference in image acuity between the 1-key and 7-key blocks of trials suggests that as participants listened to the same scale in all trials, they created an absolute representation of the target note. In principle, participants could then perform the task without having to create a sequence of images in their minds in order to arrive at an accurate image against which an incoming sensory input was judged. The observation that image acuity was better overall when probe tones were constant from trial to trial parallels observations of perceptual learning in pitch discrimination tasks in which participants are trained across many trials to discriminate pure tones or harmonic complexes at a constant fundamental frequency (Demany, 1985; Demany & Semal, 2002; Irvine, et al., 2000; Micheyl, et al., 2006). Most importantly, the averages in Table 1 indicate that the difference between the 1-key and 7-key blocks in the cued-attention condition is greater when the probe is the tonic (mean difference = 19.25; t (1,19) = 2.66; p < .015) than when the probe is the leading tone (mean difference = 8; t (1,19) = 1.45; p = .27). This finding shows how image acuity improves due to short-term memory training and this improvement is greater when the target is primed by the context, demonstrating an effect of long-term memory on short-term memory representations.

Once again, the obtained image widths were considerably wider than those reported previously for pure tones or harmonic complexes (Demany, 1985; Demany and Semal, 2002; Irvine et al., 2000; Janata & Paroo, 2006; Micheyl et al., 2006). Several differences between this study and previous studies could account for differences in threshold estimates. One of these is the difference in adaptive threshold estimation procedures. In this experiment the ZEST algorithm was used in order to reduce the number of trials taken to estimate threshold. Although ZEST has been validated using within-participants comparisons across thresholding procedures, including up-down staircase methods (Marvit, et al., 2003), it is nonetheless sensitive to a number of parameters that influence convergence on a threshold as well as the stopping rule. Nonetheless, differences in threshold estimates arising from differences in thresholding procedures do not impact within-procedure comparisons or inferences such as those pertaining to image acuity under 1-key or 7-key conditions.

More interestingly, differences in threshold estimates are likely to arise from task differences. The tasks used here differ from those previously reported in the literature in subtle but important ways. One is that we estimate thresholds for detecting both higher and lower pitch deviations in which the participant does not know on any given trial whether the deviant will be higher or lower than the reference pitch. By contrast, as far as we can tell, in two-tone discrimination judgments the standard pitch is held constant and the deviant is always higher in pitch, thereby allowing for a listening strategy in which attention is directed to one side of the tuning curve surrounding the reference pitch. Perhaps most importantly, in two-tone discrimination tasks participants receive direct sensory information about the target pitch in every comparison interval (Campbell & Small, 1963; Demany & Semal, 2002; Irvine, et al., 2000; Micheyl, et al., 2006), whereas this direct sensory information is absent when a probe tone arrives in both the cued-attention and imagery tasks that are employed here. While the same reference pitch is heard during one of the observation intervals of each trial in our tasks, there is no sensory memory benefit as is the case in two-tone discrimination tasks.

Experiment III

Overview

Given the task differences between our study and previous pitch discrimination studies in which a sensory representation of the pitch to be compared is established immediately prior to the comparison, we sought to determine the degree to which the threshold estimates in our tasks were a consequence of not receiving bottom-up sensory information that can directly serve as the basis for the discrimination judgment. Therefore, participants were presented with a two-tone pitch discrimination task in addition to our other tasks. In doing so, it was also possible to determine whether performance on this sensory memory task correlated with performance in the cued-attention and imagery tasks. In keeping with the observed correlations between performance on the cued-attention and imagery tasks in Experiments 1 and 2, it was expected that those participants with smaller thresholds in the discrimination task would also form more accurate images in the imagery task.

The imagery task was similar to the tasks in Experiments 1 and 2. In this instance, two blocks of trials were presented, one where the scales were presented in the same key and another in which the key was chosen randomly from seven different keys. The target note was always the tonic.

Method

Participants Thirty-four students participated in this experiment (age: M = 19.81, SD = 1.8; 18 female). None of them participated in previous experiments of this study. Eighteen of them had at least 1 year of musical training and the amount of training was 6.33 ± 5.18 (mean ± standard deviation) between them. All of them were compensated with partial course credit. The experiment was carried out with the informed consent of the participants in accordance with the guidelines of the Institutional Review Board at the University of California, Davis.

Stimuli

a. Two-tone discrimination task Stimuli for the discrimination task were constructed from the same notes that were used as target notes in Experiments 1 and 2. The first note of each pair in the trial was one of the seven different target notes in the imagery task: B3 (246.94 Hz), C4 (261.63 Hz), Db4 (277.18 Hz), D4 (293.66 Hz), Eb4 (311.13 Hz), E4 (329.63 Hz), F4 (349.23 Hz). In one of the pairs of the trial, the second note was the same note as the first one. In the other pair, the second note was different. The deviation in this note was from -100 to +100 cents in 2.5-cent intervals. The order of the pairs was randomized.

b. Tuning judgment task The stimuli for the scales in this experiment were the same as the stimuli in Experiment 2. However, in this experiment we presented two blocks of scales. In one block seven different keys were used. In the other block, the key (D-major) remained constant from trial to trial. The target note was the tonic in both blocks.

Procedure

a. Two-tone discrimination task The first task of this experiment contained trials with two pairs of notes separated by one of ten different tonal masks. The participants had to decide in which pair the second note differed from the first. They were first presented with 12 practice trials during which they were given feedback after every trial. After this training, they performed the task with no feedback. In this block of the experiment, the task followed the same structure as Experiments 1 and 2, and thresholds were estimated using the ZEST procedure.

b. Tuning judgment task After the block of 2-note discrimination trials, the participants were trained in an imagery task like that used in Experiments 1 and 2. They listened to two musical scales every trial and had to judge which one had a mistuned eighth note (tonic). During these practice trials, they were given feedback. After the examples, they were presented with two experimental blocks. In one block, the key was held constant across trials. Trials in the other block varied across 7 different keys. The order of blocks was randomized across participants. As in the previous experiments, the participants also answered questions about their musical experience and about their perception of the task. The duration of the experiment was approximately 1 hour.

Data analysis Image widths were analyzed as before using a repeated-measures ANOVA with two within-participants factors, Key Number (1, 7) and Task (cued-attention, imagery). Post-hoc t-tests were used for multiple comparisons using Bonferroni criterion for correction and an alpha of .05. The relationship between performance on the discrimination task and in the cued-attention and imagery conditions was assessed by calculating the correlation across participants of the threshold estimates obtained in the tasks. Mean image widths were also compared for the two-note discrimination and the 1-key cued-attention tasks using a *t*-test for the groups with and without musical experience.

Results

Threshold estimates obtained in the pitch discrimination task and in the cued-attention and imagery conditions were significantly correlated (Figure 4), both when the scales were presented in one key, r(32) = .735, p < .0001 (cued-attention), and r(32) = .567, p < .001 (imagery), and in seven different keys, r(32) = .702, p < .0001 (cued-attention), and r(32) = .572, p < .0001 (imagery). Mean image widths for the cued-attention and imagine conditions are shown in Table 1. Mean width (cents) for the two-tone pitch discrimination task was 76.37 ± 7.82 (mean ± standard error). A repeated-measures ANOVA showed a significant effect of Key Number, F(1,31) = 14.79, p < .001, Task, F(1,31) = 48.69, p < .0001, and an interaction between Key Number and Task, F(1,31) =6.53, p < .016. This pattern of results indicated that images were less accurate in the imagery condition and when the key varied from trial to trial, but disproportionately so for imagery trials in blocks with varying key.

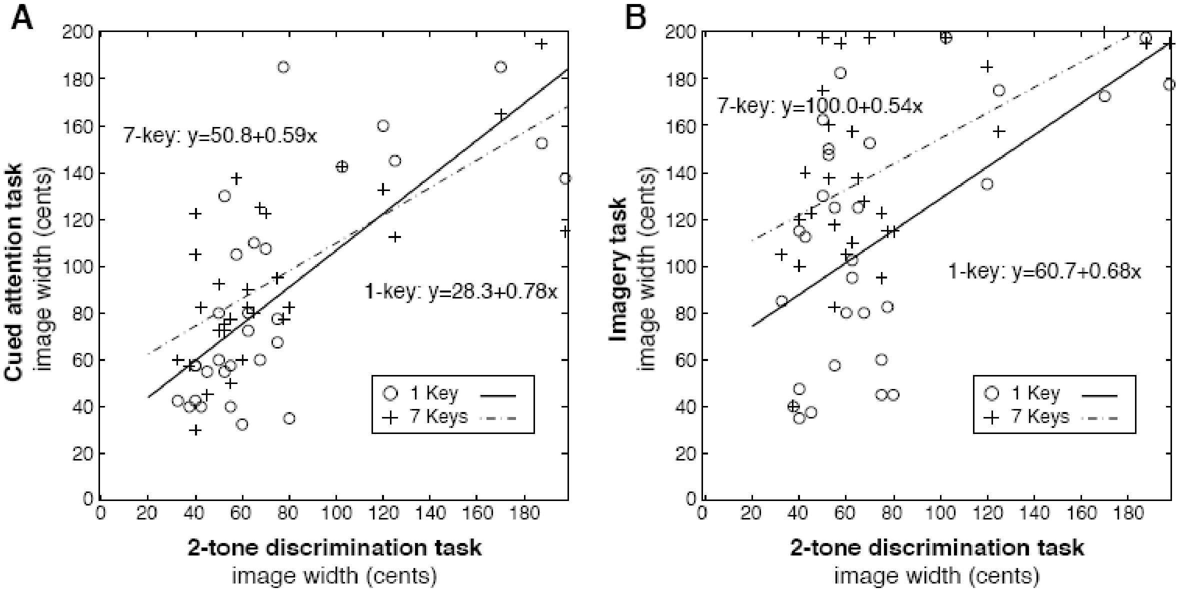


Figure 4. ***Correlations between the two-tone pitch discrimination task in Experiment 3 and image widths in the cued-attention (A) and imagery conditions (B) in 1 and 7 keys.***

The individual differences shown in Figure 4 indicate that for many participants, image acuity in the two-tone discrimination task was comparable to image acuity in the cued-attention conditions, especially during single-key blocks. While a small number of participants were able to maintain comparable image acuity in the two-tone discrimination and imagery conditions when the key was held constant, variability in imagery condition thresholds increased significantly among most participants.

The direct comparison of image widths in the two-note and the cued-attention tasks showed no difference between these two conditions either in the group without any musical training (*t* (1,31) = 1.33; p = .208) or in the group with some musical experience (*t* (1,31) = 1.87; p = .079). Correlations between the years of musical training and image widths were calculated (Figure 5). Significant negative correlations were found between the amount of musical training and image width in all conditions, indicating that image widths decreased with increasing amount of musical training, both when the scales were all presented in the same key, r(32) = -.53, p < .002 (imagery), and r(32) = .-387, p < .029 (cued-attention), and when the key varied across trials, r(32) = -.483, p < .005 (imagery), and r(32) = -.431, p < .014 (cued-attention). Despite the strong relationship between amount of musical training and image acuity, one must note that there were many participants with no musical training whose image acuity was comparable to that of participants with considerably more musical training.

The 18 participants with some music training (amount of training, 6.33 ± 5.18; mean ± standard deviation) and the 14 participants with no music training were further compared using ANOVA. There was a main effect of Task, F(1,31) = 52.17, p<.0001, Key Number, F(1,31) = 13.63, p < .001, Experience, F(1,31) = 11.37, p < .002, and an interaction between Task and Key Number, F(1,31) = 5.86, p < .022. Pairwise comparisons showed that when the scales were presented in the same key, the widths in the imagery and the cued-attention conditions did not differ significantly in the group with musical experience (see Table 2).

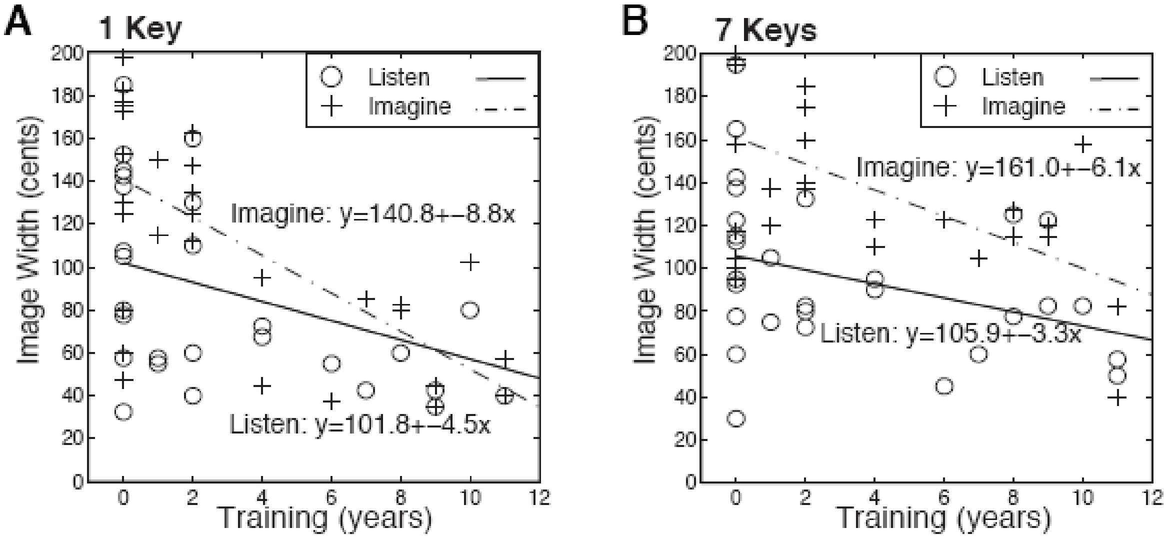


Figure 5. ***Correlations between the amount of musical training and the width of the images in the cued-attention and imagery tasks when the scales were presented either in 1 key (A) or 7 keys (B).***

Discussion

The comparison of image acuity across three tasks indicated that participants were capable of achieving similar acuity in all three tasks, though only one individual illustrated equally good acuity during the imagery conditions when the key of the scale was changing from trial to trial. Overall, image acuity was most similar between the two-tone discrimination task and the cued-attention condition in blocks of trials where the scales were drawn from the same key. These two conditions are the most closely matched in terms of the amount of reinforcement offered for the representation of the target pitch. The results suggest that when a memory trace that has been reinforced by constancy across trials interacts synergistically with contextual cues that facilitate the retrieval of the target pitch at a particular moment in time (e.g. an ascending major scale), the quality of the representation is comparable to that which arises when a sensory trace of the target pitch is instantiated in a two-tone discrimination task. In other words, representations built in memory with help of contextual cues (long-term memory) and repetition (short-term memory) can become as accurate as those representations based on a sensory memory trace (two-note discrimination task).

Table 2. *Mean differences between the two imagery conditions for participants with music experience and nonmusicians*

Music	Number of keys	Imagery	Mean	Mean difference (HRD-IMG)	Sig.
No experience	1 key	Cued-attention	110.21 (11.8)	- 38.9 (11.09)	.001
		Imagery	149.11 (12.56)		
	7 keys	Cued-attention	115.01 (8.99)	- 51.6 (8.2)	<.0001
		Imagery	166.61 (10.02)		
Music experience	1 key	Cued-attention	74.028 (10.54)	- 17.78 (9.48)	.07
		Imagery	91.806 (10.89)		
	7 keys	Cued-attention	83.639 (7.78)	- 42.61 (7.37)	<.0001
		Imagery	126.250 (8.92)		

Note. Standard errors of the means are indicated in parentheses. Pitch values are given in cents.

However, the results also indicate that individuals vary considerably in their ability to generate images of comparable quality across task conditions.

Within our sample there was considerable variation in the amount of participants' vocal or instrumental musical training. Those individuals with more musical training were more readily able to form accurate pitches across a greater variety of task conditions, though when the opportunity to maintain an absolute representation of the target pitch was eliminated (7-key condition) and the task demands for generating images were increased (imagery condition), the quality of the images also suffered on average for this group.

General discussion

Across a series of three experiments we sought to extend our understanding of the factors that influence the acuity of mental auditory representations. As a basis for our experiments, we utilized a pair of cognitive tasks in which the amount of sensory/perceptual support for the formation of auditory images is varied (Janata & Paroo, 2006). The objective of that paper was a comparison of image acuity in two cognitive tasks that have historically been regarded as different (cued-attention versus imagery). In this paper we considered the influences of different forms of memory on the acuity of auditory objects/images that are formed for comparison with the mental representation of an arriving sound. We note that while various traditional paradigms share a requirement of instantiating mental images, representations, or objects, they differ in the memory systems they engage. We examined the relative accuracy of these images as different mnemonic influences were emphasized by different task requirements (Figure 1).

Across the experiments, the influence of memory systems that operate on several timescales was varied, ranging from brief sensory memory traces, to short-term memory traces formed over blocks to trials, to long-term implicit memory for tonal structure in music. We assumed that memory systems would be recruited in accordance with the specific task demands and characteristics. For example, a two-tone discrimination task emphasizes the maintenance of an image in sensory memory for a short period of time so that it can be compared with a second image as it enters sensory memory.

Although the properties of different memory systems have been studied extensively, as have the properties of cognitive processes such as attention, much of the research focus has been on the capacity and temporal properties of the memory systems, and how those properties interact with attention (Cowan, 1995, 2008). However, less attention has been paid to the representations of the items that are being maintained in different memory stores, and how these representations may differ with changing task demands. In his extensive review of the literature on the interactions of memory and attention, Cowan (1995) briefly contends with the issue of what form neural activations would take if generated by sensory stimuli as opposed to mental processes that represent those stimuli in the absence of sensory input. He suggests they should be similar, “although there would be limits to how well the sensory properties of stimuli could be imagined.” (pg. 252) Further along the same page he writes, “As for the proposed similarity (but not identity) of activation generated through stimuli vs. mental events, the recent behavioral evidence, though indirect is fairly convincing.” He proceeds to cite one of the best pieces of evidence to date that mental images assume sufficient detail, in this case timbral detail, to facilitate the processing of a sensory item that is physically congruous with the imagined detail (Crowder, 1989).

One way of treating what may be a broad latent discomfort with the idea that externally and internally derived representations of a stimulus might be the same in terms of activation of some form of short-term memory store, is to allow for the possibility that there exists a common memory store that is capable of maintaining representations of auditory objects with variable acuity at any given moment in time. We would like to suggest that our experiments explore some of the influences on the acuity of representations within such a store. Importantly, we regard the store we are discussing as a substrate for the comparison of images arising directly from sensory input, e.g. a note that was just heard, with images that have found their way to that store, with varying degrees of success (acuity), from any one of several memory sources depending on the task demands. Considerable neurophysiological evidence has amassed that some region within the auditory cortex serves in that capacity (Janata, 2001; Näätänen, 1992; Näätänen & Winkler, 1999).

The store we are describing is likely the “phonological” store that is part of the phonological loop of working memory (Baddeley & Hitch, 1974). Auditory images can find their way to the phonological store either via auditory input, e.g. hearing a word, or by the activation of longer-term representations, e.g. reading a word. Although the properties of the phonological store have largely been examined with regard to the maintenance of verbal material, the possibility that the phonological loop serves as a basis for auditory imagery has been considered (Baddeley & Logie, 1992; Smith, Reisberg, & Wilson, 1992; Smith, et al., 1995). Moreover, verbal and tonal working memory have been shown recently to have highly overlapping neural substrates (Koelsch, et al., 2009). We now consider our results regarding the acuity of pitch representations on different memory timescales and across diverse task demands (Figure 1) in relation to previous work that has been focused on particular tasks or timescales.

Sensory memory

Although the exact definitions of sensory memory vary, it is generally regarded as a store lasting on the order of several hundred milliseconds to several seconds in the case of audition (Cowan, 1995). Importantly, representations of auditory objects in sensory memory might be regarded as the highest fidelity representations available to an individual for the purpose of making discrimination judgments. Accordingly, a two-tone pitch discrimination task, which is essentially a comparison of an item held in sensory memory with an item entering sensory memory, is expected to provide an estimate of the bestattainable acuity for any given individual. Most of what is known about the limits of pitch discrimination derives from such tasks.

Acuity estimates depend further on the spectral characteristics of the tones as well as the long-term and short-term training of participants. One recent study comparing individuals with and without musical training found frequency discrimination thresholds of ~0.2% and ~1.8%, respectively, for pure tones, and slightly better thresholds for complex tones consisting of harmonics 2–5 of a fundamental frequency of 330 Hz (Micheyl et al., 2006). We observed average thresholds of 2.3% in our two-tone discrimination task, with the best participants achieving ~1%. The threshold estimates here were slightly larger, though this could be attributable to several factors, including use of spectrally broader harmonic sounds, a different adaptive thresholding algorithm, the type of headphones used in the experiment, and perhaps most importantly the fact that higher (sharp) and lower (flat) threshold estimates were obtained simultaneously. Thus, attention could not be focused reliably to one side or the other of the reference pitch on each trial. Nonetheless, the two-tone discrimination task provided us with an estimate of pitch acuity when the representation of the auditory object was based directly on sensory input of the target pitch.

In our cued-attention and imagery conditions, the auditory object that was to be compared with the probe item was never based on a sensory memory representation of the probe item. In the cued-attention condition, the mental image of the auditory object could be formed through a series of interactions between sensory representations of the notes of the ascending scale and explicit knowledge of the pitch relations between successive notes of an ascending major scale. Thus, attention could be cued accurately to a particular pitch representation. It was found that images were generally as accurate in the cued-attention task as in the two-tone discrimination task (Figure 4A), indicating that an internally generated (top-down) mental image could be as accurate as a mental image held in sensory memory. This remained true, even when the key changed from trial to trial, indicating that the interaction between sensory input and knowledge of the pitch relations in the ascending scale were sufficient to generate accurate mental images. However, the apparent parity between top-down and bottom-up mental images began to break down as support for forming a specific pitch image deteriorated.

Buildup of short-term absolute memory representations

Overall, the removal of sensory information leading up to the probe note reduced acuity of mental images. This may have been due in part to poor timing of the mental images. Janata and Paroo (2006) found that temporal acuity suffered greatly in the imagery task. If the wrong note is being imagined when the probe arrives, or if the mental image is not formed fully, acuity would be expected to suffer (Jones et al., 2002). However, some listeners remained able to instantiate mental images that were as accurate as the sensory images, but only when the key remained constant from trial to trial (Figure 4B). The preserved image acuity in the 1-key imagery trials suggests that some form of absolute memory for the probe note was used by these listeners. It is important to note that in Experiment 3, with the exception of one listener, all of the listeners who had accurate images in the 1-key imagery condition had less accurate images in the 7-key condition, indicating that their acuity in the 1-key condition was not due to their abilities to imagine successive notes in a sequence alone.

Across the experiments, it was observed that the repetition versus variation of the key significantly improved image acuity. Acuity improvements have been observed in pitch discrimination experiments in which the reference frequency is kept constant (Campbell & Small, 1963; Demany, 1985; Demany & Semal, 2002; Irvine, et al., 2000; Micheyl, et al., 2006; Wright & Fitzgerald, 2004). Improvements in acuity can be separated into general training of pitch discrimination abilities as well as frequency specific effects that reflect perceptual learning (Demany & Semal, 2002; Irvine et al., 2000). Further evidence that a consistently presented tone forms a representation in short-term memory that is readily reactivated after its sensory memory trace decays comes from an electrophysiological study of the mismatch negativity (Cowan, Winkler, Teder, & Naatanen, 1993). The MMN is the difference between the evoked potentials elicited by deviant and standard tones and is regarded as marker of a preattentive deviance detection process located in the auditory cortex (Näätänen, 1992). In conditions in which the standard remained consistent across trials (brief sequences of standards with occasional deviants), the MMN to a deviant in the 2nd position of a sequence was larger than when the standard changed from trial to trial, indicating that a memory trace of the standard persisted across trials. Our findings are consistent with all of these observations, and together with the results of Janata and Paroo (2006), indicate that the short-term memory representations of consistently mapped target tones remain accurate across varying degrees of top-down processing demands. It is worth noting that target tones in these tasks occurred no more frequently than other tones, so in contrast to MMN experiments or even two-tone discrimination experiments, it can be ruled out that the acuity effects were due to the sensory predominance of single tones.

The acuity differences between the 1-key and 7-key conditions also demonstrated that listeners made better use of built-up absolute than relative information, even though the relative information present in the pitch relationships of the ascending major scale remained invariant within the experiments. Relative information in music may be retained more easily in memory than pitch height (Deutsch, 1972, 1975), and even inexperienced participants are able to transpose a known melody to other keys maintaining the same pitch intervals (Attneave & Olson, 1971; Dowling, 1978; Dowling & Diane, 1971). However, in such studies the emphasis is on memory for pitch class categories rather than the tuning of individual notes. Thus, those memory mechanisms that ensure accurate recognition or recall of sequences of notes may not support accurate intonation judgments.

Implicit memory for tonal structure

At the longest memory timescale, we observed effects of implicit memory for tonal structure. The effects of a primed context influencing the processing of a target event are well documented in language at both local (word) and global (sentence or paragraph) levels (Meyer & Schvaneveldt, 1971; Stanovich & West, 1979). These effects have also been replicated with musical stimuli (Tillmann & Bigand, 2002). Even a single tone or chord can prime tonal hierarchy representations such that more and less-related items are processed with differential speed and accuracy (Bharucha & Stoeckig, 1986; Janata & Reisberg, 1988). Here, we examined the role of tonal context on the acuity of mental images and found that the key primed by the ascending scale context consistently led to more accurate images of the tonic than those created for the leading tone. On the other hand, our results account for the reciprocal relationships between long-term and short-term memory systems and specifically the contributions of long-term memory to the retention process of short-term representations. Cowan (1995), for example, in his working memory model, proposes that long-term memory activation is the key for short-term memory storage. The relation between these systems has been broadly studied in language, such as the effect of lexical knowledge on short-term memory of words (Brown & Hulme, 1992), or the contribution of phonological representation to short-term memory span (Hulme, Roodenrys, Brown, & Mercer, 1995). This study shows for the first time the effects that the implicit memory of tonal context has on the consolidation of pitch representations in short-term memory. Our data illustrate that pitch image acuity improves after training and storage in short-term memory, and that improvement is increased dramatically by the activation of long-term memory priming codes.

The influence of musical training and individual differences

We found considerable individual differences in image acuity across all three experiments. Overall, an individual's image acuity covaried across tasks. Experiment 3 was particularly interesting in this regard because it allowed us to observe the departures in image acuity from a baseline established by the two-tone discrimination task. For a number of participants, the image widths in the discrimination task matched their widths in the cued-attention conditions when the key remained constant from trial to trial as well as when it changed, indicating that the sensory input and expectations generated by abstract knowledge of the scale were as effective at instilling an accurate image as when a sensory template was presented. For other participants, the image in the cued attention condition started to suffer when the key changed from trial to trial, suggesting a dissociation between the build-up of an absolute representation of the target and reliance on the relational cues of the scale. The imagery task further amplified these differences. Many participants were able to maintain baseline acuity, provided that the scales were always played in the same key. Thus, these individuals exhibited the ability to form and use short-term memories to attain accurate pitch images. However, others in whom baseline acuity was comparable were unable to form or utilize the short-term representations and showed no better performance in the 1-key trial blocks than in the 7-key blocks. Thus these tasks tap a range of influences on the acuity of pitch images in a store that is used for assessing the intonation of an incoming pitch.

In general, musical experience is related to the neural activity implicated in the detection of pitch changes in both pre-attentional (Koelsch, Schröger, & Tervaniemi, 1999) and attentional tasks (Tervaniemi, Just, Koelsch, Widmann, & Schröger, 2005). The degree of similarity between musicians and non-musicians may vary depending on the kind of task used. For example, in tasks in which the judgment is based on relative pitch information, musical experience may have less influence than when subjects have to make use of pitch height (Dowling, 1978).

The amount of experience playing a musical instrument strongly influenced image acuity in the imagery conditions, with musically trained participants exhibiting better acuity. Previous studies have compared participants with and without musical experience and have shown that musicians generally have better auditory imagery abilities (musical and non-musical), for example in comparing acoustical features in everyday sounds (Aleman, Nieuwenstein, Böcker, & Haan, 2000). The present experiments show how musical experience influences top-down and bottom-up processing to different degrees. We found that musical experience had a greater influence in the "top-down" imagery condition rather than the cued-attention condition. One possible explanation for the better performance of musicians overall and in the imagery condition in particular is a long-term memory (LTM) effect from years of musical training with diatonic scales, as well as active expectancy formation in countless trial and error situations encountered while practicing. The ability to listen to a stimulus with the minds' ear requires that we have first experienced this in perceptual life. For those participants with no musical experience, the idea of hearing or imagining scales over and over may have seemed quite foreign. It is likely that their LTM representations of the successive pitch relationships in an ascending major scale were less developed, the repetition across trials within the experiments notwithstanding.

However, the results in Experiment 3 show how some individuals without musical experience can perform at the same level as musicians. Since this benefit is not due to training, it may be explained by an innate ability to make finer pitch discriminations. Innate factors may take the form of neuroanatomical variability in the arcuate fasciculus (AF), a fiber bundle that links auditory areas in the temporal lobe with lateral prefrontal cortex. Recent studies have shown that pitch processing deficits in tone-deafness are associated with greatly reduced superior AF volume in the right hemisphere and that AF volume is generally related to pitch discrimination ability (Loui, Alsop, & Schlaug, 2009). The morphology of the left AF is correlated with pitch naming acuity in absolute pitch possessors (Oschslin, Imfeld, Loenneker, Meyer, & Jancke, 2009). Moreover, within the auditory cortex, the relative morphology of Heschl's gyrus in the two hemispheres predicts the manner in which individuals perceive the pitch of complex harmonic sounds, and also depends on musical training (Schneider, et al., 2005).

Given the heterogeneity in tuning thresholds across listeners as well as their susceptibility to task conditions and training, it is interesting to consider the implications of these thresholds for music perception and the enjoyment of music. Specifically, do broader tuning curves accommodate greater degrees of mistuning and render the average listener more forgiving of poor intonation? Most of our listeners had thresholds that safely fell within the semitone range, although on average they were surprisingly poor. This suggests that their categorical perception of pitch class would remain intact and accurate across some amount of mistuning, and that they would remain relatively impervious to the fact that a note had not been played perfectly in tune.

The multifaceted auditory image

Our findings clearly indicate that numerous factors influence the acuity of pitch images. Although there has been a tendency within psychology and neuroscience to consider mental representations of objects as restricted to specific tasks and perceptual and cognitive processes, thus leading to an implicit differentiation of mental representations and their neural stores, we would like to suggest that it is perhaps equally important to consider a more unified view of auditory mental images in which multiple memory systems and processes shape image representations at a sensory/cognitive interface that supports discrimination of salient acoustic properties like pitch.

References

Aleman, A., Nieuwenstein, M. R., Böcker, K. B. E., & Haan, E. H. F. (2000). Music training and mental imagery ability. *Neuropsychologia, 38*, 1664-1668.

Attneave, F., & Olson, R. K. (1971). Pitch as a Medium - New Approach to Psychophysical Scaling. [Article]. *American Journal of Psychology, 84*(2), 147-&.

Bachem, A. (1954). The factors in relative and absolute pitch determination. *Journal of the Acoustical Society of America, 26*, 751-753.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 47-89). New York: Academic Press.

Baddeley, A. D., & Logie, R. (1992). Auditory imagery and working memory. In D. Reisberg (Ed.), *Auditory Imagery*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

Bigand, E., Madurell, F., Tillmann, B., & Pineau, M. (1999). Effect of global structure and temporal organization on chord processing. *Journal of Experimental Psychology: Human Perception and Performance, 25*(1), 184-197.

Brown, G. D. A., & Hulme, C. (1992). Cognitive psychology and second language processing: The role of short-term memory. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 105-121). Amsterdam: Elsevier.

Brown, S., Martinez, M. J., & Lawrence, M. P. (2006). Music and language side by side in the brain: A PET study of the generation of melodies and sentences. *European Journal of Neuroscience, 23*, 2791-2803.

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H., & Jancke, L. (2005). Scanning silence: mental imagery of complex sounds. *Neuroimage, 26*, 1119-1127.

Campbell, R. A., & Small, A. M. (1963). Effect of practice and feedback on frequency discrimination. *Journal of the Acoustical Society of America, 35*, 1511-1514.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96*, 341-370.

Cowan, N. (1995). *Attention and Memory.* New York: Oxford University Press.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? In W. S. Sossin, J. C. Lacaille & V. F. Castellucci (Eds.), *Essence of Memory* (pp. 323-338). Amsterdam: Elsevier.

Cowan, N., Winkler, I., Teder, W., & Naatanen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology-Learning Memory and Cognition, 19*(4), 909-921.

Crowder, R. G. (1989). Imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 472-478.

Demany, L. (1985). Perceptual learning in frequency discrimination. *Journal of the Acoustical Society of America, 78*, 1118-1120.

Demany, L., & Semal, C. (2002). Learning to perceive pitch differences. *Journal of the Acoustical Society of America, 111*, 1377-1388.

Deutsch, D. (1972). Effect of repetition of standard and comparison tones on recognition memory for pitch. *Journal of Experimental Psychology, 93*, 153-162.

Deutsch, D. (1975). The organization of short-term memory for a single acoustic attribute. In D. Deutsch & J. A. Deutsch (Eds.), *Short Term memory* (pp. 107-152). New York: Academic Press.

Deutsch, D. (1979). Octave generalization and the consolidation of melodic information. *Canadian Journal of Psycholgy, 33*, 201-205.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85*(4), 341-354.

Dowling, W. J., & Diane, S. F. (1971). Contour, Interval, and Pitch Recognition in Memory for Melodies. *The Journal of the Acoustical Society of America, 49*(2B), 524-531.

Halpern, A. R. (1988). Mental scanning in auditory imagery for songs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 434-443.

Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition, 17*(5), 572-581.

Halpern, A. R. (1992). Musical aspects of auditory imagery. In D. Reisberg (Ed.), *Auditory Imagery* (pp. 29-44). Hillsdale, NJ: Erlbaum Associates.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex, 9*(7), 697-704.

Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology, 86*, 527-536.

Irvine, D. R. F., Martin, R. L., Klimkeit, E., & Stmith, R. (2000). Specifity of perceptual learning in a frequency discrimination task. *Journal of the Acoustical Society of America, 108*, 2964-2968.

Janata, P. (2001). Brain electrical activity evoked by mental formation of auditory expectations and images. *Brain Topography, 13*(3), 169-193.

Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. *Perception and Psychophysics, 68*, 829-844.

Jones, M. R. (1981). Music as a stimulus for psychological motion: I. Some determinants of expectancies. *Psychomusicology, 1*, 34-51.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science, 13*(4), 313-319.

Koelsch, S., Gunter, T. C., Cramon, Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *NeuroImage, 17*, 956-966.

Koelsch, S., Schröger, E., & Tervaniemi, M. (1999). Superior attentive and pre-attentive auditory processing in musicians. *NeuroReport, 10*, 1309-1313.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009).

Functional architecture of verbal and tonal working memory: and FMRI study. *Human Brain Mapping, 30*(3), 859-873.

Kraemer, D. J., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: sound of silence activates auditory cortex. *Nature, 434*(7030), 158-158.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception and Psychophysics, 56*, 414-423.

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *Journal of Neuroscience, 29*, 10215-10220.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Brocas' area: an MEG study. *Nature Neuroscience, 4*, 540-545.

Marvit, P., Florentine, M., & Buus, S. (2003). A comparison of psychophysical procedures for level-discrimination thresholds. *Journal of the Acoustical Society of America, 113*(6), 3348-3361.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227-234.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*, 36-

47.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behav Brain Sci, 13*, 201-288.

Näätänen, R. (1992). *Attention and Brain Function*. Hillsdale: Lawrence Erlbaum Associates.

Naatanen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24*(4), 375-425.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin, 125*(6), 826-859.

Oschslin, M. S., Imfeld, A., Loenneker, T., Meyer, M., & Jancke, L. (2009). The plasticity of the superior longitudinal fasciculus as a function of musical expertise: A diffusion tensor imaging study. *Frontiers in Human Neuroscience, 3*, 1-12.

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., et al. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and clinical neurophysiology, 94*, 26-40.

Peynircioglu, Z. F. (1995). Covert rehearsal of tones. *journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 185-192.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*.

Reisberg, D. (1992). *Auditory Imagery*. New Jersery: Lawrence Erlbaum Associates.

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science, 14*, 262-266.

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., et al. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience, 8*, 1241-1247.

Smith, J. D., Reisberg, D., & Wilson, M. (1992). Subvocalization and auditory imagery: Interactions between the inner ear and inner voice. In D. Reisberg (Ed.), *Auditory Imagery* (pp. 95-119). Hillsdale, NJ: Erlbaum Associates.

Smith, J. D., Wilson, M., & Reisberg, D. (1995). The Role of Subvocalization in Auditory Imagery. *Neuropsychologia, 33*(11), 1433-1454.

Spiegel, M. F. (1984). Performance on Frequency-Discrimination Tasks by Musicians And. *Journal of the Acoustical Society of America, 76*(6), 1690-1695.

Stanovich, K. E., & West, R. F. (1979). Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory and Cognition, 7*, 77-85.

Tervaniemi, M., Just, V., Koelsch, s., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Experimental Brain Research, 161*, 1-10.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychol Rev, 107*(4), 885-913.

Tillmann, B., & Bigand, E. (2002). A comparative review of priming effects in language and music. In P. McKevitt, S. O Nualláin & C. Mulvihill (Eds.), *Language, vision and music* (pp. 231-240). Amsterdam: John Benjamins.

Tillmann, B., Janata, P., Birk, J., & Bharucha, J. J. (2003). The costs and benefits of tonal centers for chord processing. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2), 470-482.

Verkindt, C., Bertrand, O., Perrin, F., Echallier, J. F., & Pernier, J. (1995). Tonotopic organization of thehumanauditory cortex: N100 topography and multiple dipole model analysis. *Electroencephalography and clinical neurophysiology, 96*, 143-156.

Wier, C. C. (1977). Frequency Discrimination as a Function of Frequency and Sensation Level. *Journal of the Acoustical Society of America, 61*(1), 178-184.

Wright, B. A., & Fitzgerald, M. B. (2004). Learning and generalization on five basic auditory discrimination tasks as assessed by threshold changes. In D. Pressnitzer, A. de Cheveigne, S. McAdams & L. Collet (Eds.), *Auditory signal processing: Physiology, psychoacoustics, and models* (pp. 510-516). New York: Springer Verlag.

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience, 8*, 29-46.

Chapter III

CHAPTER III,

Electrophysiological Correlates of Accurate Mental Image Formation in Auditory Perception and Imagery Tasks

Published as Navarro Cebrian, A., & Janata, P. (in press). Electrophysiological Correlates of Accurate Mental Image Formation in Auditory Perception and Imagery Tasks. *Brain Research*

Abstract

Event-related potentials (ERPs) were recorded while listeners made intonation judgments about target notes that terminated a sequence of heard notes (bottom-up task) or a sequence of imagined notes (top-down task). We hypothesized that the neural processes underlying the accurate formation and evaluation of mental images would behave similarly in both tasks. In the imagery condition, the amplitude of the N100 component of the auditory evoked potential in response to the target tone was smaller for those listeners who formed more accurate mental images. It was comparable in amplitude to the N100 evoked when all of the notes leading to the target were heard, consistent with a process of habituation of the N100 in the auditory cortex due to the formation of a sequence of mental images. The P3a response – a marker of deviance detection – to mistuned targets was also found in the imagery condition and it was larger for listeners who formed more accurate images. Additionally, the influence of long-term implicit memory for tonal structure of Western music on the acuity of mental images was examined by comparing responses to leading-tone (contextually unstable) and tonic (contextually stable) targets. Images were more accurate for targets that were related more closely to the established tonal context. The results suggest that successful top-down activation of pitch representations activates the same neural processes that underlie the N100 response to perceived notes, and that the engagement of these processes underlies successful detection of mistuning as indexed by the P3a component.

Chapter III

Introduction

Tonality and intonation

When listening to a melody, two types of pitch judgments are possible. One type is a *tonality* judgment in which a listener determines whether a note in the melody belongs to the key that is primed by the melody, or whether one note in the key has been substituted for another in-key note. Such tasks are easy for most people but difficult for those with amusia – commonly referred to as tone deafness (Ayotte, Peretz, & Hyde, 2002). Detection of deviant pitches in a melody is associated with a variety of enhanced components in event-related potentials (ERPs) at both early (< 300 ms) and late (> 300 ms) time points (Besson & Faïta, 1995; Besson & Macar, 1987; Brattico, Winkler, Näätänen, Paavilainen, & Tervaniemi, 2002; Fujioka, Trainor, Ross, Kakigi, & Pantev, 2005; Miranda & Ullman, 2007; Peretz, Brattico, Jarvenpaa, & Tervaniemi, 2009). The presence and size of these potentials is modulated by the amount of individual musical training (Besson & Faïta, 1995; Besson, Faïta, & Requin, 1994; Miranda & Ullman, 2007), and also on innate anatomical factors that appear to underlie amusia (Loui et al., 2009; Peretz et al., 2009).

A second type of judgment is an *intonation* judgment in which the listener must determine whether a note has been mistuned by a fraction of a semitone (the smallest frequency step size between notes in Western tonal music). Although intonation judgments can be made without a musical context, e.g. in two-tone pitch discrimination tasks, musical contexts nonetheless facilitate intonation judgments, perhaps by providing stronger support for the pitch representation that serves as the reference point for the intonation judgment (Warrier & Zatorre, 2002). The ability to make successful tonality and intonation judgments appears to be dissociated in amusics (Peretz et al., 2009), who are more impaired in the former relative to the latter when compared with normal controls. Nonetheless, there is considerable individual variability in the accuracy of intonation judgments and this variability is explained in part by musical training (Janata & Paroo, 2006; Micheyl, Delhommeau, Perrot, & Oxenham, 2006; Navarro-Cebrian & Janata, in press).

The influence of musical training on pitch discrimination has also been studied with ERPs. The N100 and mismatch-negativity (MMN) components elicited by pitch deviations are influenced by musical training (Koelsch, Schröger, & Tervaniemi, 1999; Pantev et al., 1998; Tervaniemi, Rytkönen, Schröger, Ilmoniemi, & Näätänen, 2001). Professional musicians are able to detect pitch changes that are not detectable by non-musicians , and such pitch changes lead to stronger MMN responses in musicians (Koelsch et al., 1999).

Overall, accuracy in pitch judgments may depend on familiarity of the stimuli, musical training, as well as individual differences in the recruitment of cognitive processes that are necessary for the retrieval and maintenance of pitch representations in memory. In this paper we focus on the ERP correlates of intonation judgments as they are influenced by a number of factors, including tonal context.

Mental representations, neural representations, expectations, and images

The objective of this study is to understand the influences of a number of factors on the form of a neural representation of pitch that is generated for the purpose of comparison with the neural representation of the pitch on an incoming sound. In doing so, we bridge a number of paradigms that have been used to examine the structure of mental representations and their neural counterparts. Our operational definition of a mental representation of the pitch associated with an auditory object is the width of the pitch tuning curve obtained by psychophysical methods (Janata & Paroo, 2006; Navarro-Cebrian and Janata, in press). Because this mental representation can be arrived at under different task paradigms, we are faced with the conundrum of how to refer to this mental representation in relation to the names that have been used in the different paradigms.

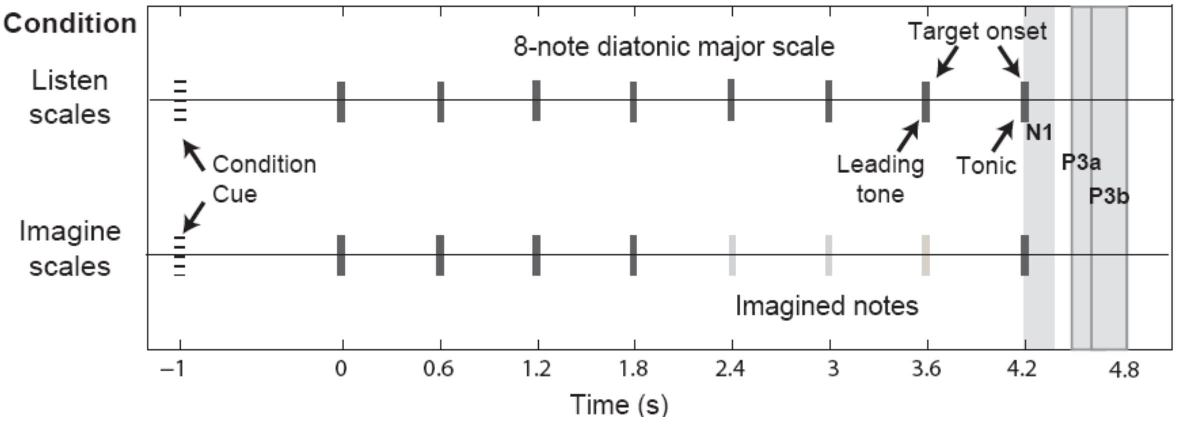
The two most common terms are “expectation” and “image.” Expectations are typically studied in oddball/target detection paradigms and may refer either to a mental/neural representation of a repeating standard stimulus or a specific infrequently occurring target stimulus. Expectations are also studied in music cognition paradigms and are established by virtue of general tonal knowledge or familiarity with a specific piece of music. By contrast, the term “image” is generally associated with imagery paradigms in which an individual must imagine a sequence of notes in order to perform the task at hand.

Given that different paradigms are used to examine the mental and neural representations of “expectations” and “images” it is tempting to assume *de facto* that these constructs are categorically different. However, neuroimaging studies indicate that imagery tasks activate auditory cortical areas (Halpern & Zatorre, 1999; Janata, 2001; Zatorre, Halpern, Perry, Meyer, & Evans, 1996), suggesting that the distinction between the neural representations of "expectations" and "images" is blurry at best. Behavioral evidence indicates that the acuity of a mental representation of pitch is influenced by various forms of memory that are available to support the formation of the mental representation (Navarro-Cebrian and Janata, in press). We use the term "image" synonymously with "representation." We prefer using the term image instead of expectation because the latter term implies that sensory input will occur at some moment in the near future whereas a mental image refers to a representation without such constraints. The term image also implies a representation that is formed and maintained voluntarily, by contrast to a representation that is generated implicitly.

Pitch discrimination, "bottom-up", and "top-down" tasks

Pitch discrimination studies typically use a "bottom-up" task in which subjects compare the pitch of two tones presented one after the other (Demany, 1985; Micheyl et al., 2006; Spiegel, 1984; Wier, 1977). In this type of task, the two mental representations that are compared are that of the incoming sensory stimulus and the sensory memory trace of the immediately preceding stimulus, which is ostensibly the same as incoming sensory stimulus. By contrast, we are interested in the acuity of mental pitch representations in two tasks in which the mental representation/image of an auditory object must be retrieved from longer-term memory stores in order to be compared with an incoming auditory stimulus (Figure 6; Janata & Paroo, 2006; Navarro-Cebrian & Janata, in press). We refer to the first task as a “bottom-up” or “cued-attention” task because the notes in a familiar sequence of notes (melody) leading up to a target note serve as highly predictive cues that direct attention towards a specific pitch height. Even though attention is oriented endogenously toward the expected pitch of the target, and therefore guided by a memory process that specifies the expected pitch, the preceding notes provide strong sensory (“bottom-up”) support for directing attention or expectation accurately toward the target pitch. We refer to the second task, an imagery task, as a “top-down” task because there is a much greater reliance on endogenous processes associated with retrieving the melody’s notes from memory than on sensory input in order to form an accurate mental representation of the target note.

Janata and Paroo (2006) compared these two conditions and showed that mental images measured behaviorally were functionally equivalent for pitch, whereas temporal acuity was compromised in the imagery task. In a recent study (Navarro-Cebrian & Janata, in press) we further manipulated the influence of short and long term memory systems on the acuity of mental images formed in both types of tasks, and found considerable variation across individuals in the utilization of the different sources of support for accurate image formation.



***<u>Figure 6.</u>* Schematic diagram of the stimulus and task conditions. Here, the 8th note in an ascending major scale (the tonic) was the target tone about which an intonation judgment was made. In other trials, intonation judgments were made about the leading tone (the 7th scale degree). The shaded areas following the target tone indicate the time windows used for the analysis of the N1, P3a, and P3b ERP components.**

In the present experiment, we compare the properties of images formed in these two tasks by means of ERPs, and seek to correlate performance measures with ERP components that reflect the representation and processing of auditory objects. We hypothesize that the representation of an auditory event can be influenced by an imagery task that emphasizes the formation of mental images with minimal sensory support, and that the representation will be similar when it is formed under conditions that cue attention to a specific location along the pitch continuum. Initial evidence for this hypothesis comes from a study in which imagined notes, in particular the first of several imagined notes, were found to elicit an emitted potential with a topography corresponding to that of the N100 component elicited in response to heard notes (Janata, 2001). However, that study did not directly examine the acuity of mental images formed under bottom-up and top-down conditions.

Some studies have already applied the ERP technique to the study of pitch deviance detection when the judgment is based solely on pure images (top-down imagery task), but none of them has made a direct comparison of cued-attention (bottom-up) vs. imagery (top-down). For example, Schön and Besson (Schön & Besson, 2005) found that when only visual information was presented (subjects were reading musical notation), a P300 increment was reported following musical rule violations. Also, an MMN increment was found to follow the onset of the incorrect termination of a familiar melody when musical experts imagined the notes preceding the final note (Herholz, Lappe, Knief, & Pantev, 2008).

Tonal contexts

Finally, in the present experiment we also examine the influence of long-term memory on auditory image formation and the consequent performance in the tasks by comparing the ERP components evoked by tonic and leading tone targets. In Western music, the tonic (the first and last notes of a major or minor scale) is the most significant and perceptually expected note, while the leading tone (seventh note of the scale) is one of the least stable notes of the tonal context primed by a scale (Krumhansl, 1990). The tonal structure that governs the relationships of individual notes to a primed key (tonal) context is implicitly learned by exposure to the regularities of Western tonal music (Krumhansl, 1990; Tillmann, Bharucha, & Bigand, 2000). Items that are more closely related to the context are processed more quickly and accurately (Bharucha & Stoeckig, 1986; Tillmann et al., 2000). Given the ERP correlates of processing tonality violations mentioned above, and the ability of a melodic context to influence intonation judgments (Janata & Paroo, 2006; Navarro-Cebrian & Janata, in press; Warrier & Zatorre, 2002), we expected that images formed either by a top-down or a bottom-up task would be more accurate when the target note is the tonic, and that this accuracy would be reflected in ERP components that are modulated by attention and error detection processes.

Method

Participants Twenty-four UC Davis students participated in exchange for course credit. Seventeen of them had at least one year of musical training (*M*=5.25 *SD*=4.61). One participant who did not complete the task and three participants with excessive artifacts in the EEG recordings were excluded from the analyses. None of them had hearing problems and four of them reported having perfect pitch, though this was not verified with a formal test.

Stimuli Pairs of ascending major scales were used in the experiment. The scales consisted of seven or eight notes and were in 7 different keys. The frequencies of starting notes were: B_2 (123.47 Hz), C_3 (130.81 Hz), D_{b3} (138.59 Hz), D_3 (146.83 Hz), E_{b3} (155.56 Hz), E_3 (164.81 Hz), F_3 (174.61 Hz). The duration of each note was 250 ms and the stimulus onset asynchrony was 600 ms. Each note was synthesized in MATLAB from its fundamental frequency and the next seven higher harmonics as described previously (Janata and Paroo, 2006). A 5-ms linear ramp was applied to the envelope at the beginning and end of each note. The last note in the scale (seventh or eighth) was either in-tune or mistuned. Pitch deviations for mistuned notes could be either higher (sharp) or lower (flat) than the standard note and ranged from –100 to +100 cents in 2.5 cent intervals.

Procedure The participants were seated in a comfortable chair in a dimly illuminated, sound-attenuated room. The stimuli were presented using headphones and the volume was adjusted at ~70 dB. The task was presented using Presentation software, running under Windows XP. Participants first underwent training in which they performed a block of twelve trials simulating the actual experiment. They received feedback after every trial. Each trial consisted of two scales separated by one of ten different tonal masks. The tonal masks were random sequences of four 250 ms notes in the same timbre as the scales. One of the scales in the trial was the standard scale in which all notes were in tune. The other was a probe scale in which the last note (seventh or eighth) note was out of tune.

Trials were drawn from one of two task conditions (Figure 6). On each trial, a cue word appeared on the middle of the screen one second before the first scale started to indicate the trial type. The word was either “Listen” or “Imagine” and remained on the screen throughout the trial. Participants were asked to keep their eyes fixed on the words and not to blink until the end of the trial. In the cued-attention (Listen) trials, participants heard every note of the scale. In the imagery (Imagine) trials, participants heard only the first four notes and the last note of the scale. They were instructed to imagine the (two or three) notes that were not played. The target note was played at the time that one of the last notes should have occurred if all the notes had been played (3600 or 4200 ms, seventh or eighth note respectively). The use of the seventh or the eighth note as the target was randomized throughout the experiment. Given this uncertainty on which degree of the scale (leading tone or tonic) the target was going to fall, we expected that subjects would be more likely to imagine the notes that were not played in order to perform the task accurately. Depending on which scale degree was the target (seventh or eighth), they had to imagine either two or three notes in the imagery condition.

The task of the participants was to either listen to or imagine the scales and judge, pressing “F” or “J” on the keyboard, which of the two scales contained the mistuned note. The order of the target scale was varied randomly across trials. Participants were instructed to tap the space bar with their right index finger along with every listened and imagined note in order to better keep the tempo in the imagery condition. Tap latencies were also recorded. Participants were instructed not to move or make any vocalizations along with the scales, except for the taps accompanying the stimuli. Imagine and Listen trials were ordered randomly across trials.

The experiment started with the estimation of the participant's pitch thresholds in the cued-attention and imagery conditions. These trials were recorded and included in the ERP analyses. Two pitch discrimination thresholds (sharp and flat deviations) were estimated for both cued-attention and imagery conditions. Pitch thresholds were determined using the ZEST procedure (Marvit, Florentine, & Buus, 2003) with the starting deviation set at ± 75 cents. ZEST converges more quickly on threshold estimates than a traditional adaptive staircase procedure, reducing the effects of fatigue. In ZEST, a probability density function represents the probable distribution of the listener's threshold, and it is modified after every trial using Bayes' theorem given the participant's response.

After the subjects converged on their thresholds they were given more trials of both cued-attention and imagery conditions with deviations ranging between their threshold level and 100 cents, in order to have an adequate number of trials for averaging. All trials were included for the analysis (trials during the estimation of thresholds and trials presented thereafter). Note that because the degree of mistuning during the mistuned trials depended on the listeners' thresholds, poor performers were presented with larger physical deviants during those trials on average than were good performers. In principle, the larger physical deviants could result in ERP differences, e.g. in the N100, between the groups of performers. However, no such effects were observed in the cued-attention task, the task most likely to show such an effect.

In addition to the main task, they were also presented with a synchronization/continuation tapping task in which they heard the same tone (same frequency; D_3, 146.83 Hz) eight times or they heard four tones, a space equivalent to 3 tones and a last tone (simulating the musical scales in the cued-attention and the imagery condition). In the tapping-only conditions they were required to tap along with the tones keeping the tempo. Once the tones finished, they were required to continue tapping, following the same tempo, until they heard an auditory cue indicating that they could stop tapping. The time that they continued tapping, without hearing anything, was equivalent to the time of a scale. This last condition was presented in order to record the activity elicited by tapping along with a monotonic isochronous sequence and the activity elicited tapping without listening to anything. The tapping-only trials were randomly interleaved between the other conditions. An average of 15 trials were presented through the experiment (every trial containing a condition in which they tapped along with constant-frequency tones was immediately followed by another condition in which they continued tapping but they did not hear anything). At the beginning of these trials, the cue word that appeared 1 second before the onset of the first note was "Tap" instead of "Listen" or "Imagine”.

At the end of the experiment subjects were asked about their musical background. They also answered several questions regarding the difficulty of the task and the strategies used.

Statistical analysis The dependent variable was the width of the mental images, defined as the difference between the positive (sharp mistuning) and the negative (flat mistuning) thresholds. Listeners were divided into two groups (good/bad performers) based on the their behavioral thresholds. For each listener, an average image width was calculated by averaging the image widths from the imagery (top-down) and cued-attention (bottom-up) conditions. Across listeners, the median average image widths served as a cut-off point to form the groups. An ANOVA was performed with Task (imagery vs. cued-attention) and Tonal Context (leading tone vs. tonic) as within-subjects variables and Group (good/bad performers) as a between-subjects factor. Post-hoc t-tests were used for multiple comparisons using the Bonferroni criterion for correction and an alpha of .05. An additional ANOVA analysis, with Tonal Context as within-subjects factor and Groups as between-subjects factor, was performed to analyze the differences for the response latencies.

EEG recording and analyses EEG data were collected from 128 channels at a sampling rate of 512 Hz. Vertical and horizontal eye movements were recorded using EOG electrodes. EEG data analysis was performed using Matlab scripts based on functions from the EEGLAB toolbox. First, spherical spline interpolation was used to correct bad channels for 13 subjects. For each of those subjects, fewer than 4% of all channels were defined as bad and replaced. The data were filtered offline to remove frequencies above 50 Hz and below 1.5 Hz. Portions of continuous data with excessive artifact were rejected after visual inspection.

Independent components (ICs) were estimated for individual subjects using the filtered but un-segmented EEG data in order to identify eye-movement and muscular artifacts. ICA decomposition of the EEG provides spatially fixed and temporally independent components of the underlying processes. Analyses of the data relied on the independent component analysis functions of EEGLAB. 128 ICA components were identified for each subject. As a preprocessing stage for each subject, IC scalp topographies, time courses, and spectral characteristics were inspected visually to identify eye-movement and muscular artifacts (Jung et al., 1998). Artifactual ICs were marked and their estimated contribution to the original EEG waveform was excluded from subsequent analyses. In addition to a small number of components related to blinks and eye movements, most of the rejected components were related to anomalous amplitude deviations in single channels. These components were readily identified by plotting all of the component time series as a 2D color image. Aberrant components saturated the colormap, and their time series often showed slow drifts until a peak in an extreme value. Trials containing these deviations could have been rejected but due to a large number of channels, ICA components rejection seemed a better option. As recommended in the EEGLAB tutorial manual, a second ICA analysis was calculated after eliminating the most salient artifacts. An average of 54 components were rejected for each subject. These typically reflected aberrant activity within individual channels. On average, 74 components (range: 59-98) were retained in each dataset.

Data were separated into epochs time-locked to the onsets of eight different types of scales: standard scales (with seven or eight notes) for both the cued-attention and the imagery conditions, and deviant scales in which either the 7th or the 8th note was mistuned, (again for both the cued-attention and the imagery conditions). To obtain comparable number of trials in the standard and deviant conditions, scales with sharp and flat deviations were pooled together. The final number of standard and deviated scales used for analyses was approximately the same. ERPs were averaged for each condition. 6.55 s epochs were created starting 1 s before the onset of the first note in the scale. The time window spanning the onset of the first note and the onset of the third note (0 to 1200 ms) was used as the baseline. This window was chosen as the baseline because it encompasses many perceptual, cognitive, and motor processes, and the demands of the different conditions are most similar at this time. The period preceding the onset of the target notes was not used as a baseline because of likely differences between the cued-attention and imagery conditions in this time window.

N100

The average maximum amplitude between 80 to 120 ms following the onset of the target was calculated across a fronto-central set of electrodes and used as a measure of N1 amplitude. The activity location was examined by plotting the scalp maps of the participants' average for several time windows (Figure 7). 16 channels were selected for analyses (see Figure 8).

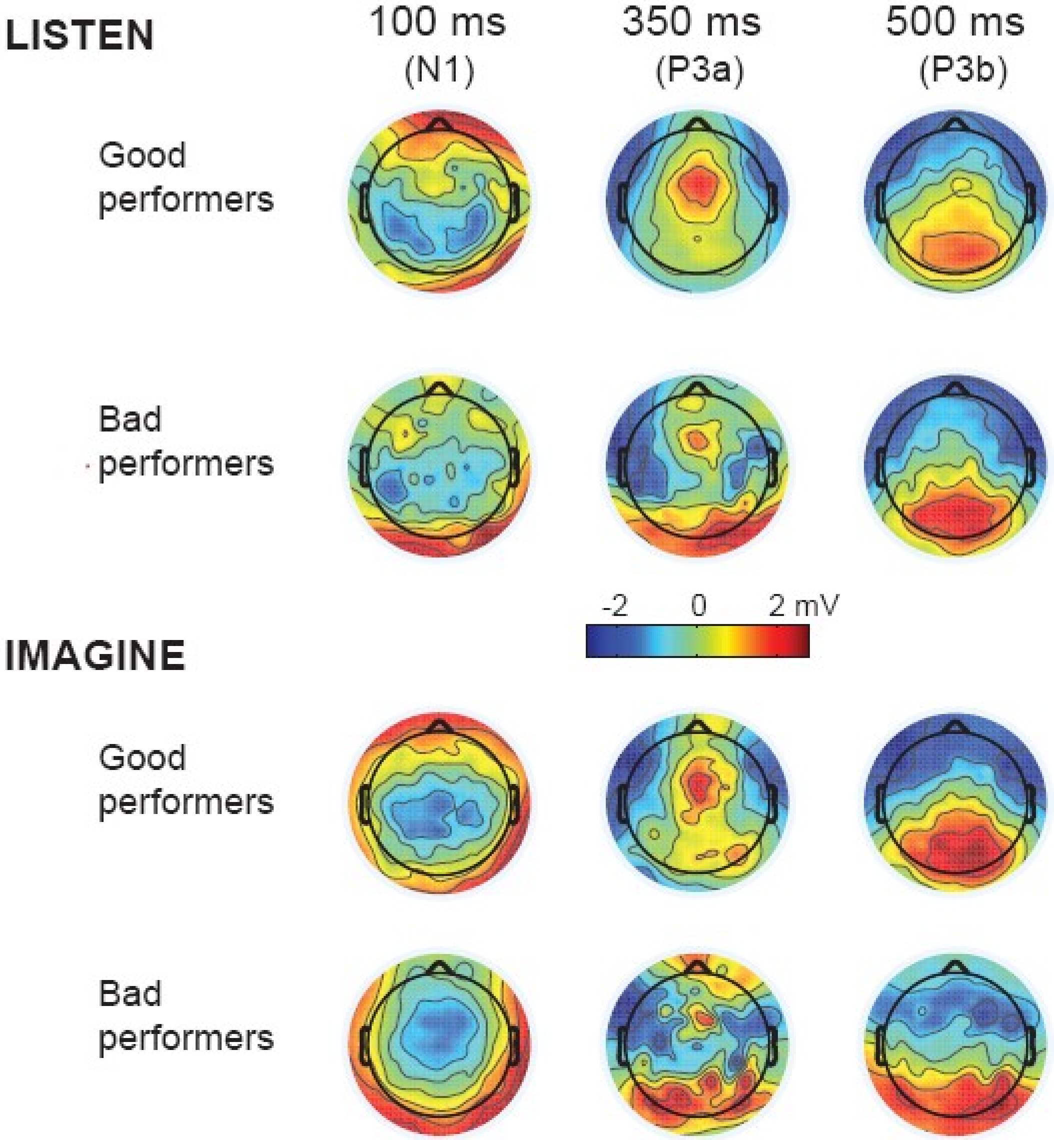


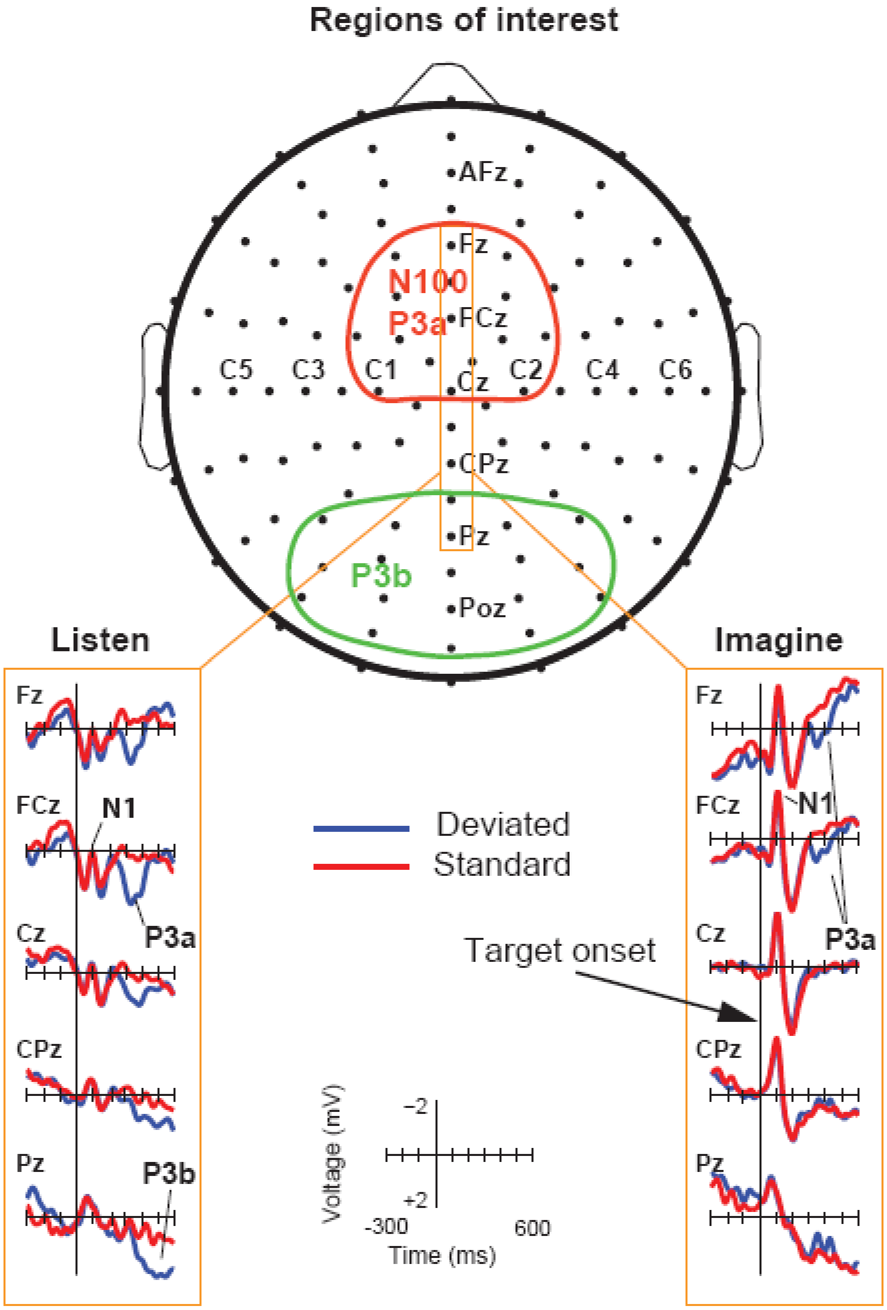
Figure 7. ***Scalp voltage distributions at 100 ms, 350 ms, and 500 ms after target onset for the cued-attention (upper figure) and the imagery (bottom figure) conditions for both good and bad performers.***

P300

Time windows used for analysis of the P3a and P3b components were 300–400 ms and 450– 600 ms respectively. The same fronto-central region of interest (ROI) that was used for the N100 was used for the P3a, whereas a parietal ROI was used for the P3b (Figure 8) after examining the topographies for the groups of good and bad performers (Figure 7).

For the tapping condition data were separated into epochs time-locked to the onset of the first tone. Therefore, three types of epochs were obtained, "listen tapping", "imagine tapping" (in which three tones before the last one were not played), and "silent tapping" (in which they continued tapping keeping the tempo but no tones were presented). As in the main task, a baseline corresponding to the first two tones (from 0 to 1200 ms) was used for baseline correction. Later, these epochs were separated into 400 ms epochs time-locked to the onset of every single tap.

Statistical analysis For the N100, an ANOVA was performed with Deviation (standard vs. deviant note), Task (cued-attention vs. imagery condition), and Tonal Context (leading tone vs. tonic) as within-subjects variables, and Group of performers (good vs. bad performers) as a between-subjects variable.



<u>Figure 8.</u> Electrode sites included in the analyses of the N100, P3a, and P3b effects, and ERP responses for the target (from 300 ms before the target) in representative electrodes for the cued-attention (left) and the imagery (right) conditions.

For the P3a and P3b components, the difference between the deviant and the standard waveform was analyzed in an ANOVA with Tonal Context and Task as within-subjects variables, and Group of performers as between-subjects variable. Post-hoc comparisons were made with the least significant difference test (Fisher, LSD).

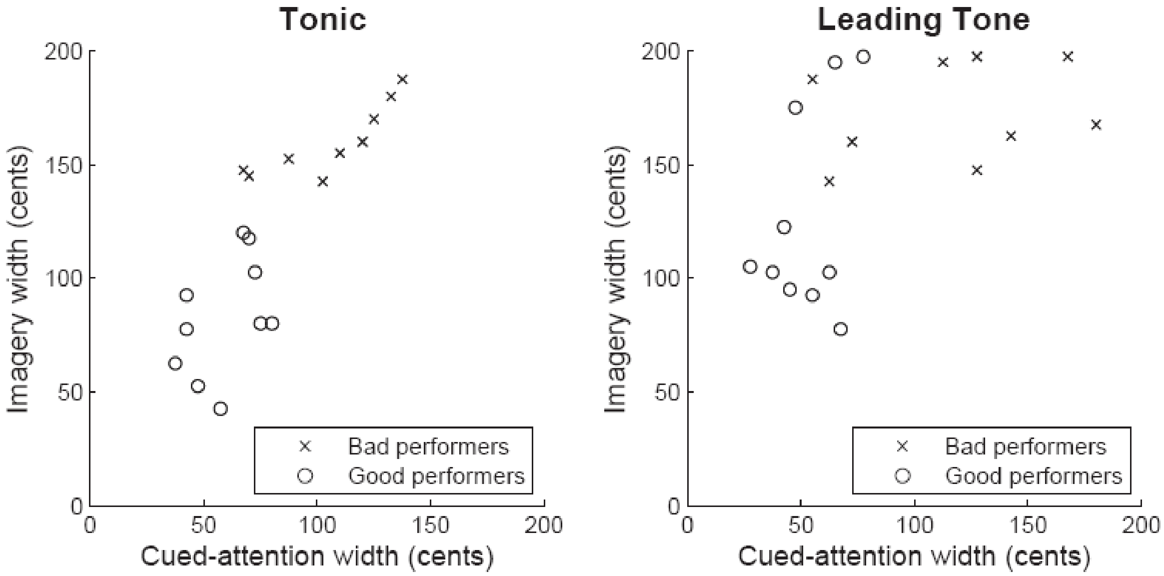


Figure 9. ***Scatter plots showing the image widths (differences between sharp and flat thresholds) for the imagery and the cued-attention conditions when the target was the tonic (left panel) or leading tone (right panel). Listeners were designated as good (o) or bad (x) performers for the ERP analysis based on a median split of the average cued-attention and imagery thresholds.***

Results

Behavioral results

Participants were divided using a median-split into two groups based on their performance in the cued-attention and imagery conditions. The median average image width for the cued-attention and the imagery conditions that was used to separate the groups was 107.5 cents. Average musical experience differed significantly for the groups ($t(1,18) = 5.1$, $p < .009$). The mean music experience for the good performers was 7.8 years and for the bad performers was 2.7 years. However, some of the good performers had no formal musical training, and some of the bad performers had more years of musical training (12 y) than the average of the good performers. Figure 9 shows the relationship between image widths in the imagery and the cued-attention conditions for both groups. There was a significant positive correlation between the width of images in the cued-attention and imagery conditions for both the tonic ($r(20) = .862$, $p < .0001$) and the leading tone ($r(20) =.554$, $p < .011$). Means and standard deviations of the image widths are shown in Table 1. An ANOVA showed that image widths were significantly larger in the imagery than the cued-attention conditions ($F(1,18) = 65.68$, $p< .0001$), and for the leading tone than for the tonic ($F(1,18) = 12.39$, $p < .002$). By definition, there was a significant difference between the groups ($F(1,18) = 39.97$, $p < .0001$). The interactions of Tonal Context x Group ($F(1,18) = 1.93$), Task x Group ($F(1,18) = 1.23$), Tonal Context x Task ($F(1,18) = 4.22$) and Tonal Context x Task x Group ($F(1,18) = 3.28$) were not significant.

Table 3. Mean differences in image widths measured in cents for the cued-attention and imagery conditions for good and bad performers.

	TARGET	TASK	Mean (cents)	Std. Error
Bad performers	Tonic	Cued-attention	108.5	8.19
		Imagery	164	6.2
	Leading tone	Cued-attention	117.5	13.46
		Imagery	175.5	6.93
Good performers	Tonic	Cued-attention	59.25	4.96
		Imagery	82.75	8.21
	Leading tone	Cued-attention	62.75	8.22
		Imagery	126.5	14.25

Averages for the reaction times of the responses were analyzed in order to identify further possible differences between groups. When the tonic was the last note in the scale, the averages for the response latencies were 2.53 and 1.57 seconds after the onset of the last note, for the bad and the good performers respectively. In the leading tone condition, the averages for the response latencies in the leading tone condition were 2.86 and 2.41 seconds after the onset of the target, for the bad and the good performers respectively. For the response latencies, there were no significant differences between the tonic and leading tone (F(1,18) = .004) or between good and bad performers (F(1,18) = 4.01). The interaction Tonal Context X Group (F(1,18) = 1.28) was not significant.

To assist in maintaining a steady tempo in the imagery task, participants were asked to tap along by pressing the space bar with every heard and imagined note in both the cued-attention and imagery tasks. To analyze how accurately participants tapped along with the notes and images, the onset of the last tap was compared for the good and bad performers.

When the tonic was the last note in the scale, there were no significant differences between groups for the onset of the last tap (t(1,18) = .993). The averages for the onset of the last tap were 81 ms before and 7 ms after the onset of the last note, for the bad and the good performers, respectively. In the leading tone condition, there were no significant differences between groups for the onset of the last tap (t(1,18) = 1.73). The averages for the onset of the last tap were 3 ms and 11ms after the onset of the target, for the bad and the good performers respectively.

To have a better idea of the imagery strategies used by each group we examine the responses to the questions that followed the experiment. For the question, "Overall, how difficult was it to imagine the melody? (1 is 'very easy' and 7 is 'very difficult')", the average for the good performers was 3.3, and the average for the bad performers was 5.3. A *t-test* reveled significant differences between these two averages (*t*(1,18) = 3.34; p = .004) showing that good performers had less difficulty imagining the melody.

To the question, "Overall, how often did you find yourself forgetting to imagine the melody? (1 is 'very often' and 7 is 'not often')", poor performers appeared to forget to imagine the melody (3.1) more often than good performers (4.1) though this difference was not significant *t*(1,14.063) = 1.54).

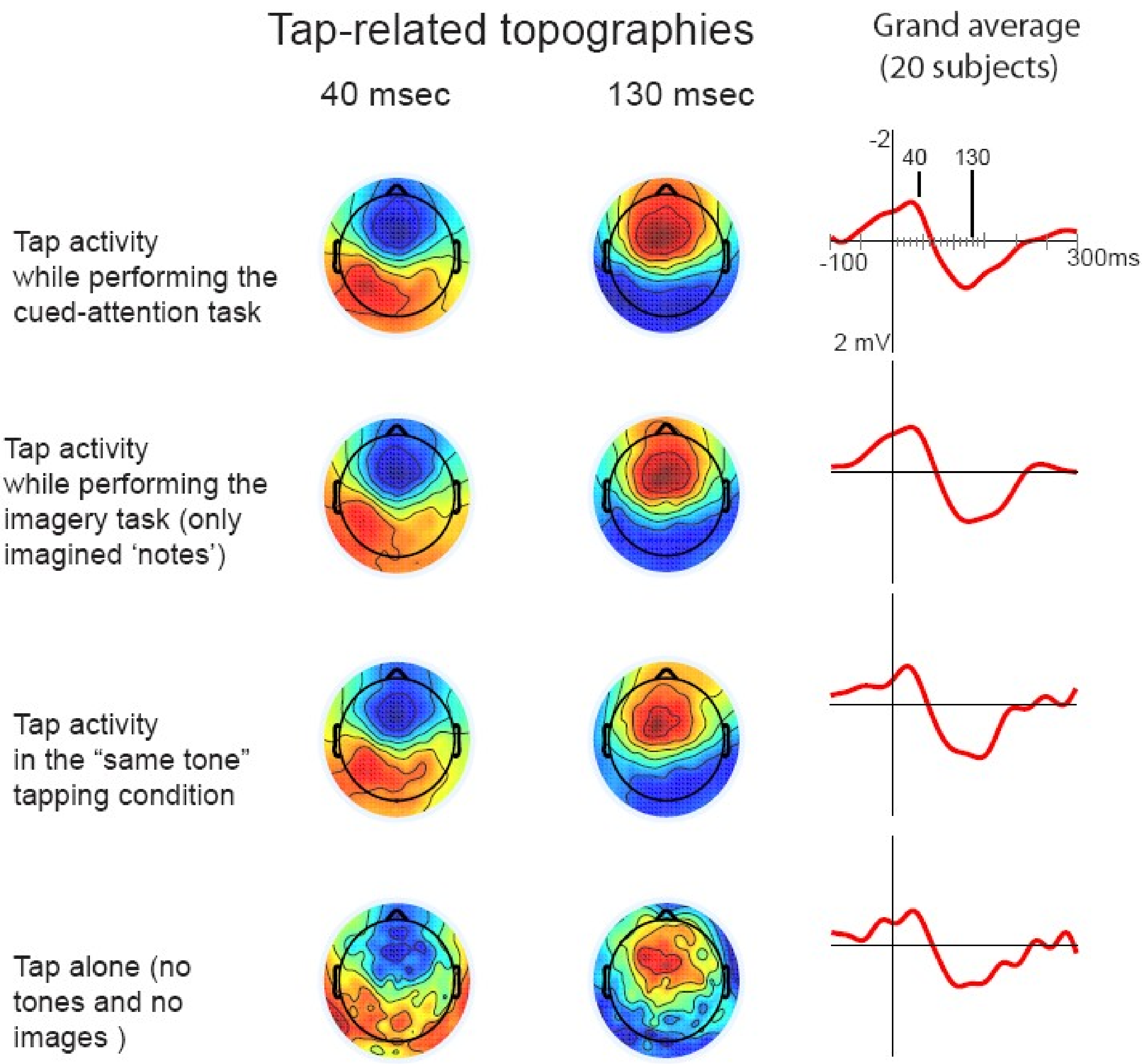
When they were asked about the vividness of their images (1 is "very vivid" and 7 is "had no image"), the good performers tended to have more vivid images (3.5) than the bad performers (4.5), though this difference was also not significant (*t*(1,18) = 1.3).

Finally, they were also asked about the necessity to imagine the notes in the imagery condition, ("How necessary did you find it to imagine the notes in the IMAGINE condition in order to make the pitch judgment? (1 is 'completely unnecessary,' 4 is 'somewhat necessary,' and 7 is 'very necessary')". Both good (3.6) and bad performers (4.2) found it somewhat necessary to imagine the notes, again with no significant differences between groups ($t(1,18) = 0.76$).

ERP results

Figure 7 shows scalp distributions of the ERPs at 100 (N100), 350 (P3a) and 500 ms (P3b) after the deviant onset for the cued-attention (upper portion) and the imagery (bottom portion) conditions for both good and bad performers. Figure 10 shows the scalp maps associated with the taps that participants made in the cued attention, imagery, and synchronization/continuation tapping tasks. The ERPs are locked to the time of the tap, and the scalp maps indicate the topography at the peaks (40 and 130 ms post-tap onset).

Overall the N100 was significantly larger in response to targets (both the standard and the deviant) in the imagery than in the cued-attention condition (Task main effect: $F(1,18) = 80.42$; $p < .0001$). The interaction Task x Group ($F(1,18) = 5.56$; $p = .03$) was significant. None of the other interactions were significant. There were no significant differences between standard and deviant scales ($F(1,18) = .155$), or between leading tone and tonic targets ($F(1,18) = .255$).



<u>Figure 10.</u>* Scalp voltage distributions at 40 ms, and 130 ms after tap onset for the cued-attention, the imagery (only imagined "notes"), the tap condition (tapping along with frequency-constant tones) and the continuation of the tap condition (in which they continued tapping while tones were not presented). The ERP average for the 20 participants is also illustrated for each of these conditions, from 100 ms before the onset to 300 ms after tap onset.

The groups of good and bad performers did not differ significantly (F(1,18) = 2.51). The interaction Task x Group was explored further with pairwise comparisons which showed that there were significant differences between the groups of good and bad performers in the imagery condition (p = .048) but not in the cued-attention condition (p = .88), such that the N100 was smaller for good performers in the imagery condition. Figure 11 illustrates a significant negative correlation between the N100 amplitude and the behavioral thresholds in the image condition for the deviant target (r(20) = -.404; p < .039). As behavioral thresholds increased, the amplitude of the N100 response became larger (more negative). The correlation between the N100 amplitude and standard targets was also significant (r(20) = -.48: p = .017).

For the P3a there were significant differences between the imagery and cued-attention tasks (F(1,18) = 10.64; p < .004) with a stronger P3a effect in the cued-attention task. The P3a was larger in response to the tonic than the leading tone (F=6.39, p < .021). The groups of good and bad performers differed significantly (F(1,18)= 4.68; p < .044), with the good performers showing a larger P3a response. The interaction of Tonal Context X Group was significant (F(1,18) = 6.39; p < .021). A pairwise comparison showed that there were significant differences between leading tone and tonic for the group of good performers (Figure 12, B; Cued-attention condition: p = .015; Imagery condition: p = .037) but there are not differences for the group of bad performers.

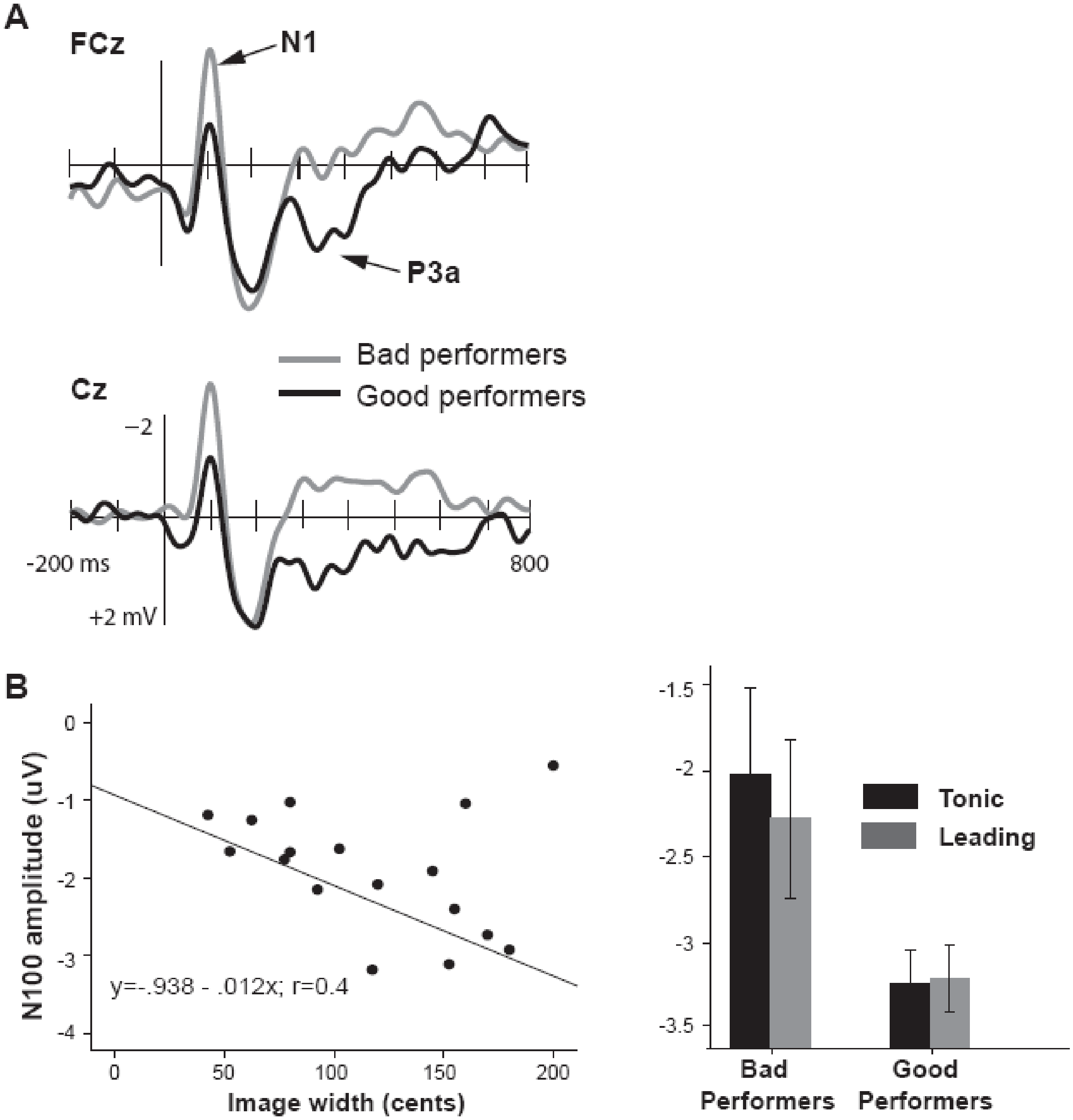
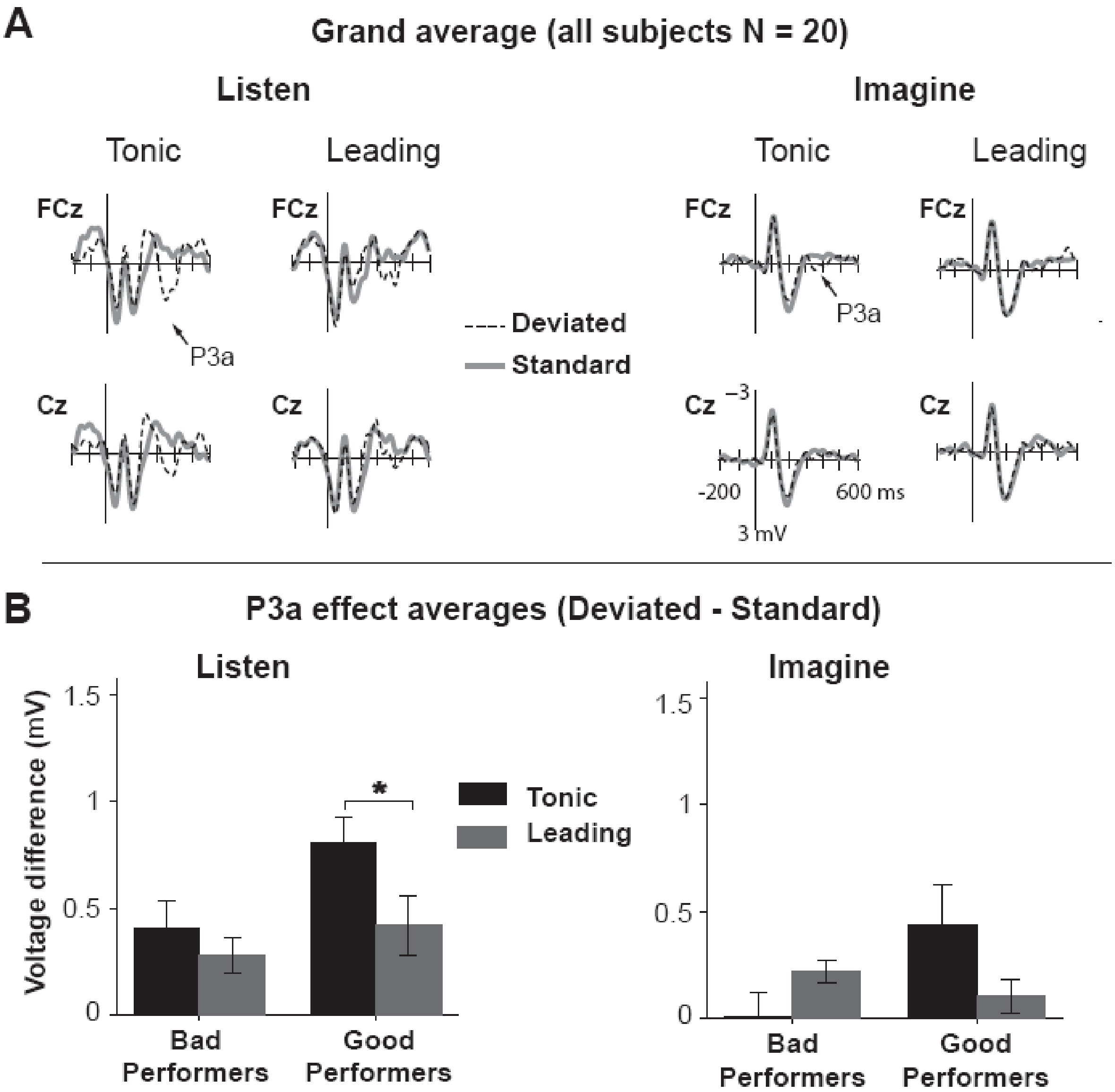


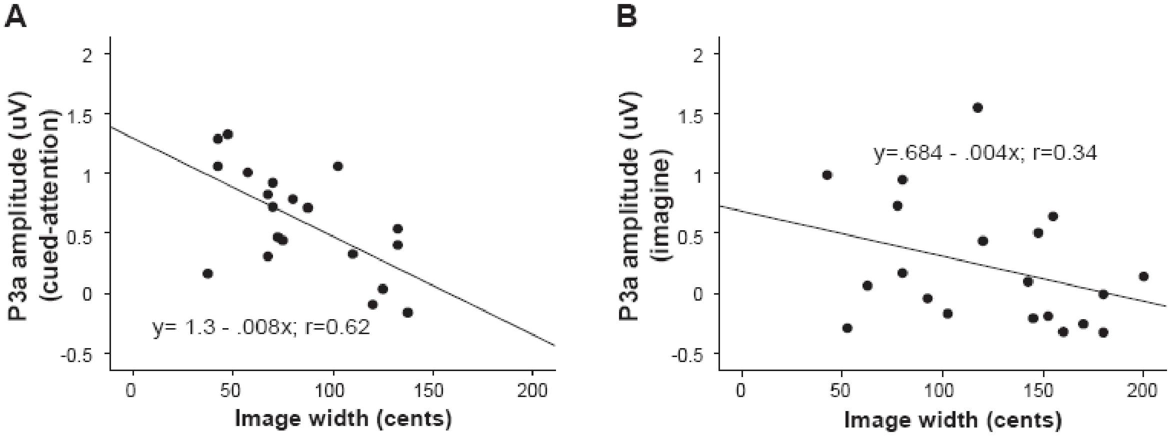
***Figure 11.* N100 amplitude differences as a function of imagery performance. A) Average ERP waveforms in response to the targets in the imagery condition plotted separately for the groups of good and the bad performers. B) Scatter plot showing the relationship between N100 amplitude and mental image widths. C) N100 averages differences between groups of subjects for the Tonic and the Leading tone in the imagery conditions**



<u>Figure 12.</u> ERP differences for the Tonic and the Leading tone conditions in electrodes of interest (A). P3a averages (deviated - standard) differences between groups of subjects for the Tonic and the Leading tone in the cued-attention and the imagery conditions (B). The statistical significant comparisons are marked (*p < .05).

The interactions of Task x Group (F(1,18) = .517), Tonal context x Task (F(1,18) = .252) and Tonal context x Task x Group (F(1,18) = .401) were not significant. Figure 13 shows a significant negative correlation (r(20) = -.615; p < .002) between P3a amplitude and the behavioral thresholds in the cued-attention condition. Narrower thresholds resulted in larger P3a responses. The correlation between P3a amplitude and behavioral thresholds in the imagery condition was weaker and did not reach significance (r(20) = .336; p < .073).

Finally, in the time window corresponding to the P3b, the factor Task (F(1,18) = 5.35, p < .033) was significant, with larger P3b effects for the cued-attention task. There was no significant difference between the groups of listeners (F(1,18) = .566) or between tonic and leading tone (F(1,18) = 1.85). None of the interactions were significant.



<u>Figure 13</u>. Scatter plot showing the relationship between P3a amplitude and mental image widths for the cued-attention (A) and the imagery (B) conditions.

General discussion

The present experiment used ERPs to study the brain mechanisms underlying the formation of pitch representations and their comparison with incoming sensory input. To better identify brain processes that are associated with the formation of accurate pitch representations, two types of tasks were used (Janata & Paroo, 2006). In a cued-attention task, auditory mental images were formed under conditions of strong bottom-up sensory support in which all of the notes in an ascending scale leading up to a target note were heard. By contrast, the imagery task relied strongly on top-down memory retrieval mechanisms in that only the initial notes of the scale were presented and the remaining notes leading up to the target had to be imagined. In both tasks, subjects had to judge the intonation of a target tone that occurred either at the time at which the tonic, the perceptually most stable note in the key primed by the ascending scale, would have occurred, or at the time that the leading tone, one of the perceptually least stable notes in a key, would have occurred.

Pitch discrimination performance in the imagery and the cued-attention tasks was highly correlated. Those participants who made good use of the sensory information to judge the pitch of the target also formed more accurate images of the target based only on the contextual cues given by the first four notes of the scale. These results in performance might be explained partly by the differences in musical experience. Both sensory (Shahin, Bosnyak, Trainor, & Roberts, 2003; Shahin, Roberts, Pantev, Aziz, & Picton, 2007) and cognitive (Tervaniemi, Just, Koelsch, Widmann, & Schröger, 2005) processes seem to be enhanced in people with more musical training when they are presented with a pitch discrimination task. In those studies, enhanced N1, P2 (Shahin et al., 2003) and P3a (Tervaniemi et al., 2005) responses were found in response to musical tones and pitch changes. Constraints on the acuity of pitch representations may already be present subcortically, as evidenced by more sharply tuned frequency-following responses to complex tones in musically trained listeners (Lee, Skoe, Kraus, & Ashley, 2009). In our study, participants were not professional musicians but significant differences were found between groups, with more years of musical training for the group of good performers. Although this difference might partly explain a benefit for the good performers due to musical experience, we also found several people in this group that had less training than the average of bad performers, and in contrast, some of the participants of this last group had more experience than the average of good performers. Therefore, innate abilities seem to be a condition necessary to form accurate pitch representations, while more years of musical training may help increase acuity.

Overall, both good and bad performers formed more accurate images for the tonic than the leading tone, illustrating an effect of tonality on intonation judgments. Figure 9, for example, clearly shows the effects of the implicit memory for tonal context on imagery. Even inexperienced participants possess an implicit knowledge of the structure of music which influences tonal expectations (Bigand, Madurell, Tillmann, & Pineau, 1999). In the present experiment, this knowledge likely influenced the image of the tonic, since this note is primed more strongly by the tonal context. However, the creation of an accurate image for the leading tone will probably require more practice. In this sense, only those subjects with more musical experience or innate abilities will be able to perform equally well in the imagery and cued-attention tasks when dealing with the leading tone (Figure 9).

The ERP data paralleled the behavioral results, showing stronger effects evoked by a mistuned tonic tone than by a mistuned leading tone (Figure 12). However, this effect was only present in the group of good performers. The results shown in this experiment corroborate those of other studies (Besson & Faïta, 1995; Brattico, Näätänen, & Tervaniemi, 2001) that have shown that accuracy in music perception is facilitated by the tonal context (for example, Western vs. non Western scales). This improvement has been found in both musicians and non-musicians for familiar melodies, but only people with musical training seem to be helped by the tonal context when they are presented with unfamiliar melodies (Besson & Faïta, 1995). A precise representation built in long-term memory through experience is necessary in order to recognize when its pitch is not in tune. Not only accurate representations built in long-term memory, but also the connections between tones in a tonal context are reinforced through experience. It has been found that stimuli more closely related to the tonal context are processed more quickly and accurately (Bharucha & Stoeckig, 1986; Janata & Reisberg, 1988).

In this experiment, we found evidence the tonal context contributing to the creation of accurate images and stronger effects on those brain components that are related to pitch discrimination. It should be noted here that the tonic is the only tone that is repeated in the sequence, since it is presented in a lower octave at the beginning of each scale. In contrast to words in language processing, repetition priming effects have not been found in music processing (Bigand, Tillmann, Poulin-Charronnat, & Manderlier, 2005), with the relative importance of events within tonal contexts having a greater influence than repetition. A recent study (Hutchins & Palmer, 2008) found that priming effects in musical melodies depended on prime-target distances, with a weakened repetition benefit in distances of 4 notes (~ 2000 ms.). Moreover, the same authors also found priming benefits explained by the tonal context. They reported stronger priming effects for the repetition of the tonic than for the repetition of other tones. These findings support the assertion that the effects found in our study are due to the tonal knowledge rather than the presentation of the tonic in a lower octave at the beginning of the scale. Repetition priming effects should also be consistent across groups. The fact that the ERP components showed differences between groups in the comparison of the leading tone vs. tonic further support this idea.

Of particular interest in this study was a comparison of auditory ERP components elicited under task conditions that weight perceptual and cognitive processes associated with the formation of pitch representations differently. In both the cued-attention and imagery tasks the ERP response to a target note reflects the interaction of an internally maintained representation of the expected input with a representation of the momentary sensory input. The interaction is readily studied because mismatches between the internal representation and the sensory representation give rise to a number of measurable neural processes that shape the auditory evoked potential (e.g., MMN, P300). Variation in the presence and magnitude of these different components as a function of task and/or listener characteristics can help determine what component processes underlie accurate mental image formation.

The neural activity associated with the detection of pitch change has already been studied with ERPs (Sams, Paavilainen, Alho, & Näätänen, 1985; Tiitinen, May, Reinikainen, & Näätänen, 1994). The main difference between the present and other pitch discrimination studies was the use of mental images by the participants to make a decision about the intonation of the last note. Previous experiments reporting expectancy violations in musical material presented a complete melody or sequences of tones or chords so that subjects made their decisions based on the sensory information preceding the target (Janata, 1995; Koelsch, Gunter, Friederici, & Schröger, 2000). In the current study, participants were presented with incomplete scales in half of the trials so they could make use only of imagery and memory retrieval to make their decisions. Some studies have reported an MMN and later neural responses (P300, P600) to auditory rule violations when expectancies are based on visual information (i.e. reading musical notation), without any auditory sensory traces to help participants make a decision (Schön & Besson, 2005; Yumoto et al., 2005).

N100

The N100 component of the auditory evoked potential is of particular interest because it serves as a marker of neural processing at what might be regarded as a sensory/cognitive interface. The auditory cortex may serve as the neural substrate for this interface. N100 generators have been localized to the planum temporale region of the auditory cortex (Verkindt, Bertrand, Perrin, Echallier, & Pernier, 1995), and the amplitude of the N100 is influenced both by exogenous (sensory) and endogenous (cognitive) factors (Näätänen & Picton, 1987). For example, the N100 is larger in response to a sound emanating from an attended compared to an unattended location (Woldorff et al., 1993). It is also larger in musicians than non-musicians (Pantev et al., 1998). A hallmark of the sensory aspects of the N100 is the rapid adaptation of its amplitude in response to a sequence of auditory stimuli (e.g. Figure 14), with a marked reduction in amplitude from the first to second stimulus in a sequence (Nelson & Lassman, 1973; Woods & Elmasian, 1986). As the inter-stimulus-interval increases, so does the size of the N100 (Näätänen & Picton, 1987). Although the N100 amplitude reduction is more pronounced when the repeated tones are identical (Butler, 1973), the habituation of the N100 has also been observed with tones of different frequencies with loudness maintained at a constant level (Nelson & Lassman, 1968). Sequences of changing complex auditory stimuli, e.g. chord progressions, also elicit adaptation of the N100 (see Fig. 1 in Janata, 1995).

Further support for the idea that the auditory cortex serves as a substrate at which top-down and bottom-up representations interact, comes from studies that observed N100-like responses in response to the first of several imagined notes (Janata, 2001), or expected but omitted sounds (Besson & Faïta, 1995; Besson, Faita, Czternasty, & Kutas, 1997; Janata, 2001; Simson, Vaughan, & Ritter, 1976). A more recent MEG study using an imagery paradigm reported an MMN-like response (iMMN) to incorrect tones at the end of partly imagined, familiar melodies (Herholz et al., 2008), though the MMN component was found only for a group of musicians with a considerable amount of musical experience (*M=18.6 SD=5.7*).

By examining the differential effects of task and psychophysical performance factors on N100 amplitude, the present experiment provides complementary support for the idea that mental images formed via top-down processes activate the auditory cortex. Specifically, the smaller amplitude N100 to the target notes (standard and deviant notes) in the Imagery condition in those listeners who formed more accurate images than those who did not (Figures 11, 14) suggests that the auditory cortex was engaged in a way during the imagery task that made it seem to the auditory cortex as though a note had recently been heard. In other words, among poor performers, the amplitude of the N100 to the target note returned to full size in the imagery task, as would be expected given the temporal separation of the last heard event and the target note. However, among the good performers the amplitude of the N100 was comparable to the size of the N100 elicited by the target in the cued-attention condition, suggesting that similar mechanisms that account for normal

N100 adaptation during the processing of a note sequence were active during imagery. The N100 amplitude differences between good and bad performers for targets in the imagery condition can only be due to the imagery processes occurring before the target since no other sounds were presented before the target in the imagery condition for either of the groups. While we cannot be certain that the participants were imagining each of the notes preceding the target, the behavioral results indicate that the good performers were able to activate a stronger and more precise representation of the target than were the bad performers.

P3a

Evidence that the neural processes engaged during the imagery period that led to N100 amplitude suppression were associated with the formation of reasonably accurate pitch images comes from an examination of how a subsequent ERP component, the P3a, behaved. The P3a is a positive peak with a fronto-central distribution that is evoked in response to infrequent or novel stimuli, even when those stimuli are not task relevant (Comerchero & Polich, 1998; Goldstein, Spencer, & Donchin, 2002; Schroger & Wolff, 1998; Squires, Squires, & Hillyard, 1975), and it has been interpreted as an index of attentional orienting to distracting stimuli (Schröger & Wolff, 1998; for a review see (Polich, 2007). Although primarily studied in response to infrequent stimuli in various auditory oddball paradigms, and analogous paradigms using other stimulus modalities, the

P3a is also observed in richer musical contexts. For example, chords that are harmonically improbable given the preceding sequence of chords elicit a P3a (Janata, 1995; Koelsch et al., 2000). Interestingly, even when harmonically improbable chords are equiprobable with other chords within the experimental context, a P3a is elicited (Janata, 1995), indicating that the P3a is sensitive to the context rather than to the actual target probabilities across the trials of the experiment.

In the present experiment, we found, as expected, that in the cued-attention condition, deviant target notes elicited a P3a component (Figures 7, 14). In the imagery condition, however, the P3a was present for the good performers but virtually absent in the bad performers, suggesting that poor performers had not formed a sufficiently accurate pitch image prior to the onset of the target to allow a deviant stimulus to be perceived as a deviant stimulus. As indicated in the Methods, the deviations presented in the experiment varied depending on the individual pitch thresholds. We would expect a benefit in the group of poor performers since they were presented with easier (larger) pitch deviations in general. However, the ERP data did not show an effect of the larger pitch deviations for the group of poor performers. Also, it could be argue that since both groups were listening to deviations near their thresholds, their performance should be equivalent. However, the behavioral data show that the thresholds for the bad performers were in many cases worse than the starting deviation (75 cents). In some cases they even reached the deviation limit (one semitone). Therefore, their discrimination thresholds may be underestimated in some cases. We believe participants in the group of poor performers, even in those trials with deviations close to their thresholds, had difficulties to develop an image of the target and therefore to detect pitch changes.

The presence of the P3a in situations in which mental images of pitch are being formed successfully raises the question of how the neural generators that give rise to the P3a fit into the broader set of processes that are involved in the formation and evaluation of auditory mental images and expectations. The focal fronto-central distribution of the P3a is suggestive of midline neural generators in the vicinity of the medial premotor areas (the supplementary motor area (SMA), and pre-SMA) and anterior cingulate, with the latter region implicated in the attentional reorienting signal interpretation of the P3a (Polich, 2007). In neuroimaging, these midline areas have been shown to respond to targets in auditory oddball tasks, though the regions involved in auditory target detection are fairly widespread (Kiehl, Laurens, Duty, Forster, & Liddle, 2001; Kiehl, Stevens, Laurens, Pearlson, & Liddle, 2005). In musical contexts, the pre-SMA and adjacent cingulate respond to deviant target chords at the end of harmonic sequences (Tillmann, Janata, Birk, & Bharucha, 2003), which is consistent with P3a observations in similar contexts (Janata, 1995; Koelsch et al., 2000).

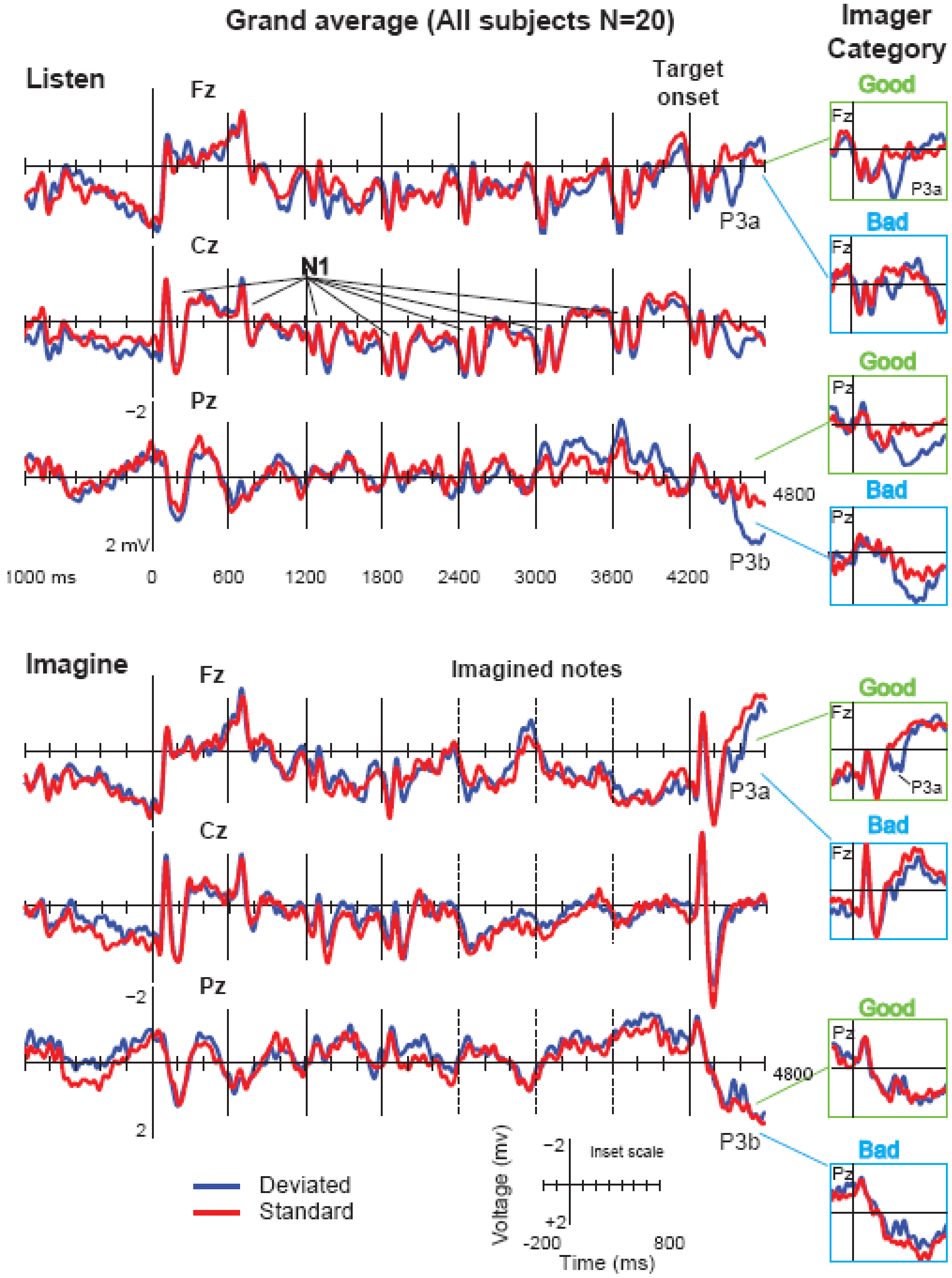
Midline premotor areas are also engaged by auditory imagery and listening tasks, such as imagining notes in familiar melodies (Halpern & Zatorre, 1999; Leaver, Van Lare, Zielinski, Halpern, & Rauschecker, 2009; Zatorre et al., 1996), listening to familiar music

(Janata, 2009), or repeated attentive listening to the same piece of music (Janata, Tillmann, & Bharucha, 2002). The premotor areas are activated in tandem with secondary auditory areas in the temporal lobe, indicating that sensorimotor loops are engaged during a variety of musical tasks in which auditory mental images are actively formed. The P3a that we observe and that has been previously observed in musical and other auditory contexts may reflect the activity of a sensorimotor monitoring process that signals the mismatch between the covert action that is associated with the formation of a mental image of a specific pitch and the pitch that is actually perceived. This interpretation is supported by a study of vocal regulation during singing, in which functional coupling was observed between the auditory cortex, the SMA, and the ACC, and activity in the pre-SMA and ACC was greater when singers had to adjust the pitch they produced to compensate for mistuning of their vocal production (Zarate & Zatorre, 2008).

Although the auditory P3a has been examined primarily within the context of the three-stimulus oddball paradigm (Polich, 2007), where it arises in response to a low-probability task-irrelevant distractor, the present findings, together with P3a results in earlier studies that examined deviant processing in musical contexts, suggest that the P3a arises when there is a mismatch between a sufficiently accurate mental image of what sound could or should occur and the sound that actually occurs, irrespective of the overall probability of the eliciting event within the experiment. (Besson & Faïta, 1995; Janata, 1995; Paller, McCarthy, & Wood, 1992). In this study, the probability of a deviant occurring was the same as the probability of a standard occurring, though the deviation level for the deviant varied from very salient to non-noticeable. In this sense, it was probably only in those trials in which the deviant was salient enough (depending on the individual threshold) that subjects were able to notice the last note as the deviated one. We interpret this result as evidence for the P3a component as dependent on the salience, and not the novelty, of the stimulus. Further support for this finding comes from studies in which the salience of the non-target (and not the novelty) and the difficulty of the task are sufficient to elicit a P3a (Comerchero & Polich, 1998, 1999; Goldstein et al., 2002).

P3b

This component is generally regarded as a task-related P300 component (Squires et al., 1975), with a later latency than the P3a and a positive voltage distribution across temporo-parietal scalp sites. In the present study, the P3b showed significant differences between standard and target in the cued-attention task (although there was also a P3b increment in the standard for the bad performers, -see figure 14-), and no differences in the imagery condition, since there was an increment of the P3b in both the target and standard scales amplitude. This is in agreement with the P3b as a task-related component. However, the fact that the P3b amplitude increased in some cases for standard targets is likely due to the uncertainty of whether the last note was mistuned or in-tune.



***<u>Figure 14.</u>* ERP time series of three selected electrodes showing the P3a and P3b effects for the good and bad performers conditions when the tonic was the target. Left figure shows ERP responses for the 'tonic' scale (8 notes). Inset figures (right) show ERP effects for the target note.**

The influence of uncertainty on P3b generation has been already studied using a paradigm in which a first stimulus is informative to differing degrees about a second stimulus (Ruchkin, Johnson, Canoune, Ritter, & Hammer, 1990). In that study, participants were presented with two numbers and had to predict whether the second number would be higher or lower than the first one. The presentation of the first number was informative, and it could resolve none, some or all of the uncertainty about their predictions. Some of the results showed that when the first number predicted the second one (all uncertainty was resolved), the P3b following the presentation of the second number was almost absent. In contrast, when the outcome was not predicted by the first number, a larger P3b was elicited. In agreement with those results, in the present experiment, the different levels of uncertainty with respect to the category of the target presented (standard vs. deviant) may have caused a P3b in response to the standards in cases, such as the imagery condition, in which participants were less certain of the intonation category of the target pitch.

Conclusion

A direct comparison of the electrophysiological components evoked by perception vs. imagery had not been made previously. In the present study, we demonstrated for the first time how later components such as the P3a and P3b can rely on the comparison of sensory traces with pure mental images. We also showed the differences in the activation of such ERPs as a function of the ability to evoke mental images/representations of pitch.

More importantly, the neuronal relationship between perception and imagery was further studied, showing how a phenomenon characteristic of sensory processing can also be observed in relation to imagery. Specifically, we observed habituation of the N100 component of the auditory evoked potential when participants imagined a sequence of notes in a scale. Using ERPs, we showed how perception and imagery interact so that the perception of a target will be modulated by the strength of the imagery processes preceding that target.

References

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain Topography, 125*, 238-251.

Besson, M., & Faïta, F. (1995). An Event-Related Potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance, 21*(6), 1278-1296.

Besson, M., Faita, F., Czternasty, c., & Kutas, M. (1997). What's in a pause: Event-related potential analysis of temporal disruptions in written and spoken sentences. *Biological Psychology, 46*, 3-23.

Besson, M., Faïta, F., & Requin, J. (1994). Brain waves associated with musical incongruities differ for musicians and non-musicians. *Neuroscience Letters, 168*, 101-105.

Besson, M., & Macar, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology, 24*, 14-25.

Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 12*(4), 403-410.

Bigand, E., Madurell, F., Tillmann, B., & Pineau, M. (1999). Effect of global structure and temporal organization on chord processing. *Journal of Experimental Psychology: Human Perception and Performance, 25*(1), 184-197.

Bigand, E., Tillmann, B., Poulin-Charronnat, B., & Manderlier, D. (2005). Repetition priming: Is music special? *The Quarterly Journal of Experimental Psychology, 58A*, 1347-1375.

Brattico, E., Näätänen, R., & Tervaniemi, M. (2001). Context effects on pitch perception in musicians and non-musicians: Evidence from ERP recordins. *Music Perception, 19*, 199-222.

Brattico, E., Winkler, I., Näätänen, R., Paavilainen, P., & Tervaniemi, M. (2002). Simultaneous storage of two complex temporal sound patterns in auditory sensory memory. *NeuroReport, 13*, 1747-1751.

Comerchero, M. D., & Polich, J. (1998). P3a, perceptual distinctiveness, and stimulus modality. *Cognitive Brain Research, 7*, 41-48.

Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology, 110*, 24-30.

Demany, L. (1985). Perceptual learning in frequency discrimination. *Journal of the Acoustical Society of America, 78*, 1118-1120.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2005). Automatic Encoding of Polyphonic Melodies in Musicians and Nonmusicians. *Journal of Cognitive Neuroscience, 17*, 1578-1592.

Goldstein, A., Spencer, K. M., & Donchin, E. (2002). The influence of stimulus deviance and novelty on the P300 and Novelty P3. *Psychophysiology, 39*, 781-790.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex, 9*(7), 697-704.

Herholz, S. C., Lappe, C., Knief, A., & Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. *European Journal of Neuroscience, 28*, 2352-2360.

Hutchins, S., & Palmer, C. (2008). Repetition priming in music. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 693-707.

Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *journal of Cognitive Neuroscience, 7*, 153-164.

Janata, P. (2001). Brain electrical activity evoked by mental formation of auditory expectantions and images. *Brain Topography, 13*(3), 169-193.

Janata, P. (2009). The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex, 19*, 2579-2594.

Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. *Perception and Psychophysics, 68*, 829-844.

Janata, P., & Reisberg, D. (1988). Response-time measures as a means of exploring tonal hierarchies. *Music & Perception, 6*, 161-172.

Janata, P., Tillmann, B., & Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, Affective, and*

Behavioral Neuroscience, 2, 121-140.

Jung, T. P., Humphries, C., Lee, T. W., Makeig, S., McKeown, M. J., Iragui, V., et al. (1998). Extended ICA removes artifacts from electroencefalografics recordings. *Advances in Neural Information Processing Systems, 10*, 894-900.

Kiehl, K. A., Laurens, K. R., Duty, T. L., Forster, B. B., & Liddle, P. F. (2001). Neural sources involved in auditory target detection and novelty processing: An event-related fMRI study. *Psychophysiology, 38*(133-142).

Kiehl, K. A., Stevens, M. C., Laurens, K. R., Pearlson, G. P., & Liddle, P. F. (2005). An adaptive reflexive processing model of neurocognitive function: Supporting evidence from a large scale (n = 100) fMRI study of an auditory oddball task. *Neuroimage, 25*, 899-915.

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: "Nonmusicians" are musical. *Journal of Cognitive Neuroscience, 12*, 520-541.

Koelsch, S., Schröger, E., & Tervaniemi, M. (1999). Superior attentive and pre-attentive auditory processing in musicians. *NeuroReport, 10*, 1309-1313.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R., & Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *Journal of Neuroscience, 29*, 2477-2485.

Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *Journal of Neuroscience, 29*, 5832-5848.

Marvit, P., Florentine, M., & Buus, S. (2003). A comparison of psychophysical procedures for level-discrimination thresholds. *Journal of the Acoustical Society of America, 113*(6), 3348-3361.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*, 36-47.

Miranda, R. A., & Ullman, M. T. (2007). Double dissociation between rules and memory in music: An event-related potential study. *NeuroImage, 38*, 331-345.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology, 24*(4), 375-425.

Navarro Cebrian, A., & Janata, P. (in press). Influences of multiple memory systems on auditory mental image acuity. *Journal of the Acoustical Society of America*.

Nelson, D., & Lassman, F. (1968). Effects of intersignal interval on the human auditory evoked response. *Journal of the Acoustical Society of America, 44*, 1529-1532.

Nelson, D., & Lassman, F. (1973). Combined effects of recovery period and stimulus intensity on the human auditory evoked vertex response. *Journal of Speech and*

Hearing Research, 16, 297-308.

Paller, K. A., McCarthy, G., & Wood, C. C. (1992). Event-Related Potencials elicited by deviant endings to melodies. *Psychophysiology, 29*(2), 202-206.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392*, 811-814.

Peretz, I., Brattico, E., Jarvenpaa, M., & Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain Topography, 132*, 1277-1286.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *PClinical Neurophysiology, 118*, 2128-2148.

Ruchkin, D. S., Johnson, R., Canoune, H. L., Ritter, W., & Hammer, M. (1990). Multiple sources of P3b associated with different types of information. *The Society for Psychophysiological Research, 27*, 157-176.

Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related protentials. *Electroencephalography and Clinical Neurophysiology, 62*, 437-448.

Schön, D., & Besson, M. (2005). Visually induced auditory expectancy in music reading: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 17*(4), 694-705.

Schroger, E., & Wolff, C. (1998). Behavioral and electrophysiological effects of task-irrelevant sound change: A new distraction paradigm. *Cognitive Brain Research,* *7*(71-87).

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *The Journal of Neuroscience, 17*, 5545-5552.

Shahin, A., Roberts, L. E., Pantev, C., Aziz, M., & Picton, T. (2007). Enhaced anterior-temporal processing for complex tones in musicians. *Clinical Neurophysiology, 118*, 209-220.

Simson, R., Vaughan, H. G., & Ritter, W. (1976). The scalp topography of potentials associated with missing visual or auditory stimuli. *Electroencephalography and Clinical Neurophysiology, 40*, 33-42.

Spiegel, M. F. (1984). Performance on Frequency-Discrimination Tasks by Musicians And. *Journal of the Acoustical Society of America, 76*(6), 1690-1695.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology, 38*, 378-401.

Tervaniemi, M., Just, V., Koelsch, s., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Experimental Brain Research, 161*, 1-10.

Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R. J., & Näätänen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians.

Learning and Memory, 8, 295-300.

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature, 372*, 90-92.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychological Review, 107*(4), 885-913.

Tillmann, B., Janata, P., Birk, J., & Bharucha, J. J. (2003). The costs and benefits of tonal centers for chord processing. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2), 470-482.

Verkindt, C., Bertrand, O., Perrin, F., Echallier, J. F., & Pernier, J. (1995). Tonotopic organization of thehumanauditory cortex: N100 topography and multiple dipole model analysis. *Electroencephalography and clinical neurophysiology, 96*, 143-156.

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception and Psychophysics, 64*, 198-207.

Wier, C. C. (1977). Frequency Discrimination as a Function of Frequency and Sensation Level. *Journal of the Acoustical Society of America, 61*(1), 178-184.

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., et al. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences, 90*, 8722-8726.

Woods, D. L., & Elmasian, R. (1986). The habituation of event-related potentials to speech sounds and tones. *Electroencephalography and Clinical Neurophysiology, 65*, 447-459.

Yumoto, M., Matsuda, M., Itoh, K., Karino, S., Saitoh, O., Kaneko, Y., et al. (2005). Auditory imagery mismatch negativity elicited in musicians. *NeuroReport, 16*, 1175-1178.

Zarate, J. M., & Zatorre, R. J. (2008). Experience dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage, 40*, 1871-1887.

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience, 8*, 29-46.

CHAPTER IV,

EFFECTS OF PERCEPTUAL LEARNING ON THE FORMATION OF MENTAL IMAGES

Abstract

The following experiment was conducted to investigate the effects of perceptual training on the formation of mental images. In one pretest and two post-tests (one immediately following a training session and one 48 hours after the pretest) participants were required to listen (bottom-up) or imagine (top-down) a musical scale and decide whether the last note presented was in tune. During the training session, the participants were trained only in a bottom-up (perception) task. Participants exhibited an improvement in image acuity during the imagery task in both post-tests relative to the pretest. There was not only an improvement in behavioral performance but also a stronger P3 effect in response to mistuned tones. These effects were stronger after 48 hours showing the fast and slow neural changes taking place after perceptual learning. Finally, the behavioral and ERP data showed that the training effects found in this experiment are also generalized to a different tonal key.

Introduction

Mental images can be studied in both a bottom-up task and a top-down task (Janata & Paroo, 2006). While in the first task an image is formed based on the previous sensory memory trace, in the second, images are retrieved from long-term memory. Despite this difference, similarities between these two mechanisms are well documented and there is some consensus that defends that both share similar neuronal systems (Halpern & Zatorre, 1999).

Accuracy in mental image formation is influenced to different degrees by short and long-term memory and attentional processes (Navarro-Cebrian & Janata, in press-b). Previous experience with the stimuli is beyond doubt one of the most influential factors in the formation of accurate images, and training has already been shown to improve the quality of representations in auditory perception tasks (Demany & Semal, 2002; Kraus et al., 1995; Micheyl, Delhommeau, Perrot, & Oxenham, 2006). Because of the neural overlapping of perception and imagery, it would be expected that perceptual training would also benefit the imagery processing of the same stimuli. In the following experiment we want to prove that supervised training in a perception task affects acuity in an imagery task. A recent behavioral experiment has proven this hypothesis in the opposite direction (Tartaglia, Bamert, Mast, & Herzog, 2009). In that study, the authors found that training in a visual imagery task led to an improvement in the perception task with the corresponding visual stimuli.

The event-related potential (ERP) technique has been used to study the neuronal changes underlying perceptual learning. Early potentials (N1, P2 and MMN) have been modulated by training in the auditory modality (Atienza, Cantero, & Dominguez-Marin, 2002; Tremblay, Kraus, McGee, Ponton, & Otis, 2001). In this experiment ERPs are used to investigate the effects of perceptual learning in the acuity of mental images. A pretest and two post-tests are presented in which subjects either listen to or imagine the preceding notes to the target (last) note of a musical scale. After the pretest, the participants are trained only in the 'listen' condition.

The tasks used in this experiment are significantly different than the tasks typically found in the perceptual learning literature. Learning to detect a mistuned tone in a sequence requires experience with both the individual stimuli and the task. On one hand, people need to discriminate small changes in the pitch of the individual stimulus. The improvement in perceptual acuity and in general the change in performance due to experience is known as perceptual learning (or 'stimulus learning'; (Robinson & Summerfield, 1996)). On the other hand, the relationships between the individual tones in the sequence as much as some attentional adaptation to the procedure also need of some practice before obtaining an optimal performance. In other words, people might need of some training before they optimally understand the characteristics of the procedure or even direct attention to the specific features of the stimuli required by the task.

Due to the complexity of the tasks presented here, we expect that after a training session, people will learn not only to discriminate the difference in pitch, which would imply a modification of the sensory representations, but also to better understand the relation between notes in the scale and other characteristics of the procedure, which would imply changes in other areas in the brain. In this sense,later components such as the P3 potential would be expected to be influenced by the training in the complex task that we present here (Song et al., 2005).

The P3 component has been related to pitch expectancy violations in the perception of melodies (Besson & Faïta, 1995; Besson, Faïta, & Requin, 1994). This component has also been studied in the context of imagery and a P3 effect to auditory rule violations has been reported when only visual information (musical notation) was presented (Schön & Besson, 2005) and when participants were required to imagine the stimuli preceding the violated tone (Navarro-Cebrian & Janata, in press-a). We hypothesize here that a stronger P3 effect in the imagery condition will follow the training session demonstrating the influences of perceptual learning on image acuity.

Finally, the effects of training are also studied in two different tonal keys. In pitch discrimination tasks, one debated issue has been the frequency specificity of perceptual learning. One hypothesis indicates that perceptual learning occurs because of the narrowing of frequency-sensitive cell receptors (Recanzone, Schreiner, & Merzenich, 1993). Training with a specific stimulus will modify characteristics of the neurons responding to that stimulus and that experience should not be transferred to other stimuli even when the same kind of task is used. The specificity vs. generalization in perceptual learning seems to be partly resolved by the complexity of the experimental tasks (Fahle, 2005; Wright & Zhang, 2009). Perceptual learning implies two different processes, one is specific for the stimulus characteristics and the other is task dependent (Doane, Alderton, Sohn, & Pellegrino, 1996; Sagi & Tanne, 1994). However other studies using less complex tasks have found generalization of pitch interval learning to other frequencies (Amitay, Hawkey, & Moore, 2005; Delhommeau, Micheyl, & Jouvent, 2005; Demany & Semal, 2002; Irvine, Martin, Klimkeit, & Smith, 2000) which leaves the question of generalization in frequency discrimination learning unresolved.

Methods

Participants Ten UC Davis students participated in exchange for course credit. Two participants were excluded from the analyses due to excessive EEG artifacts and one did not complete the experiment. Six of the subjects included in the analyses had at least one year of musical training (*M*=5.25 *SD*=4.61). None of them had hearing problems.

Stimuli Pairs of ascending major scales were used in the experiment. The scales consisted of eight notes and were in 2 different keys. The frequencies of starting notes were: D_3 (146.83 Hz) and E_3 (164.81 Hz). The duration of each note was 250 ms and the stimulus onset asynchrony was 600 ms. Each note was synthesized in MATLAB from its fundamental frequency and the next seven higher harmonics as previously described

(Janata & Paroo, 2006). A 5-ms linear ramp was applied to the envelope at the beginning and end of each note. The last note in the scale was either in-tune or mistuned. Pitch deviations for mistuned notes were either 20 or 45 cents higher (sharp) than the standard note (one cent being equivalent to one hundredth of a semitone).

Procedure Participants were presented with three tests (one pretest and two post-tests) and a training period. They were required to come for two different sessions separated by at least 48 hours. In the first session, a pre-test, a training period and the first post-test were presented. The session took approximately 3 hours and 30 minutes, including the time to prepare the electrode cap. They had several breaks in between the tests and during the training and there was a longer (10 minutes) break at the end of the pretest. During the second session (after 48 hours) they were presented with the second post-test. This session took approximately 1 hour and 30 minutes.

Tests

The participants were seated in a comfortable chair in a dimly illuminated, sound-attenuated room. The stimuli were presented using headphones and the volume was adjusted to ~70 dB. The task was presented using Presentation software, running under Windows XP. Each trial consisted of two scales separated by an interval of 1 sec. One of the scales in the trial was the standard scale in which all notes were in tune. The other was a probe scale in which the last note was out of tune.

Trials were drawn from one of two task conditions (see Figure 2 -Chapter 2-). On each trial, a cue word appeared on the middle of the screen one second before the first scale started to indicate the trial type. The word was either “Listen” or “Imagine” and remained on the screen throughout the trial. Participants were asked to keep their eyes fixed on the words and not to blink until the end of the trial. In the Listen trials (cued-attention), participants heard every note of the scale. In the Imagine trials, participants heard only the first four notes and the last note of the scale. They were instructed to imagine the three notes that were not played. The target note was played at the time that one of the last notes should have occurred if all the notes had been played (4200 ms).

The task of the participants was to either listen to or imagine the scales and judge, pressing “F” or “J” on the keyboard, which of the two scales contained the mistuned note. The order of the target scale was interleaved randomly across trials. Participants were instructed not to move or make any vocalizations along with the scales. Imagine and Listen trials were interleaved randomly across trials.

Participants were presented with a total of 120 scales, 60 standard and 60 target (half of the targets deviated by 20 cents and the other half by 45 cents) for every condition (imagine and cued-attention) and each of the two keys. Every test took approximately 40 minutes. Participants were provided with a break every 40 trials and a longer break at the end of the pretest.

Training

During the training period participants listened to trials similar to the trials presented in the tests. However only the cued-attention condition was used to provide training. Participants were not presented with scales in which they had to imagine several notes. Moreover, only scales starting with the note D_3 (146.83 Hz) were used in the training.

During the training, pitch thresholds were estimated 10 times using the ZEST procedure (Marvit, Florentine, & Buus, 2003) with the starting deviation set at ± 60 cents. In the ZEST procedure, a probability density function represents the probable distribution of the listener's threshold, and it is modified after every trial using Bayes' theorem given the participant's response. Pitch deviations for the target note stepped from 0 to +100 cents (one cent being equivalent to one hundredth of a semitone) in 2.5 cent intervals. Every time their threshold was determined, participants were told whether their threshold was better or worst than the time before. Also, the first two times that the pitch threshold was determined they were given feedback after every trial on the correctness of their responses.

EEG recording and analyses

EEG data were collected from 128 channels at a sampling rate of 512 Hz using Biosemi System. Vertical and horizontal eye movements were recorded using EOG electrodes. EEG data analysis was performed using Matlab scripts based on functions from the EEGLAB toolbox. Data were filtered offline to remove frequencies above 50 Hz and below 0.3 Hz.

Independent components (ICs) were estimated for individual subjects using the filtered but unsegmented EEG data. ICA decomposition of the EEG provides spatially fixed and temporally independent components of the underlying processes. Analyses of the data relied on the independent component analysis functions of EEGLAB. 128 ICA components were identified for each subject. As a preprocessing stage for each subject, IC scalp topographies, time courses, and spectral characteristics were inspected visually to identify eye-movement and muscular artifacts (Jung et al., 1998) as well as anomalous deviations restricted to single electrodes. Artifactual ICs were marked and their estimated contribution to the original EEG waveform was excluded from subsequent analyses. An average of 33 components were rejected for each subject. These typically reflected aberrant activity within individual channels. On average, 95 components (range: 80 - 123) were retained in each dataset. One second epochs were selected, including a 200 ms pre-target baseline, for the pretest, post-test 1, and post-test 2

No effects were found in the time windows corresponding to the P50 and N1 components. A P3a component was defined as the largest positive-going peak occurring within latencies from 250 to 500 ms after the onset of the target tone and it was analyzed in Fz and Cz electrodes.

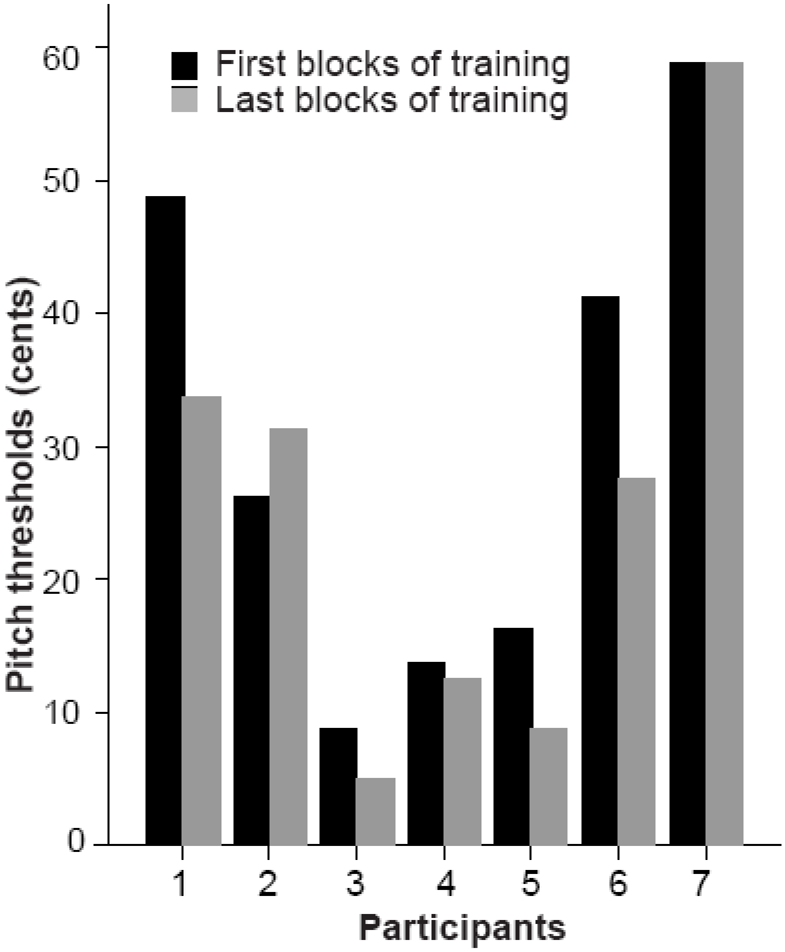


Figure 15. Learning degree for each subject in the training session. The black bar shows the average of the pitch thresholds (in cents) for the first two blocks of the training session. The gray bar shows the average of the last two blocks of the training session.

Results

Behavioral results

The improvement of pitch thresholds during the training session was analyzed. Figure 15 shows the average of the pitch threshold of the first two blocks in which the threshold was determined compared with the last two blocks for every subject.

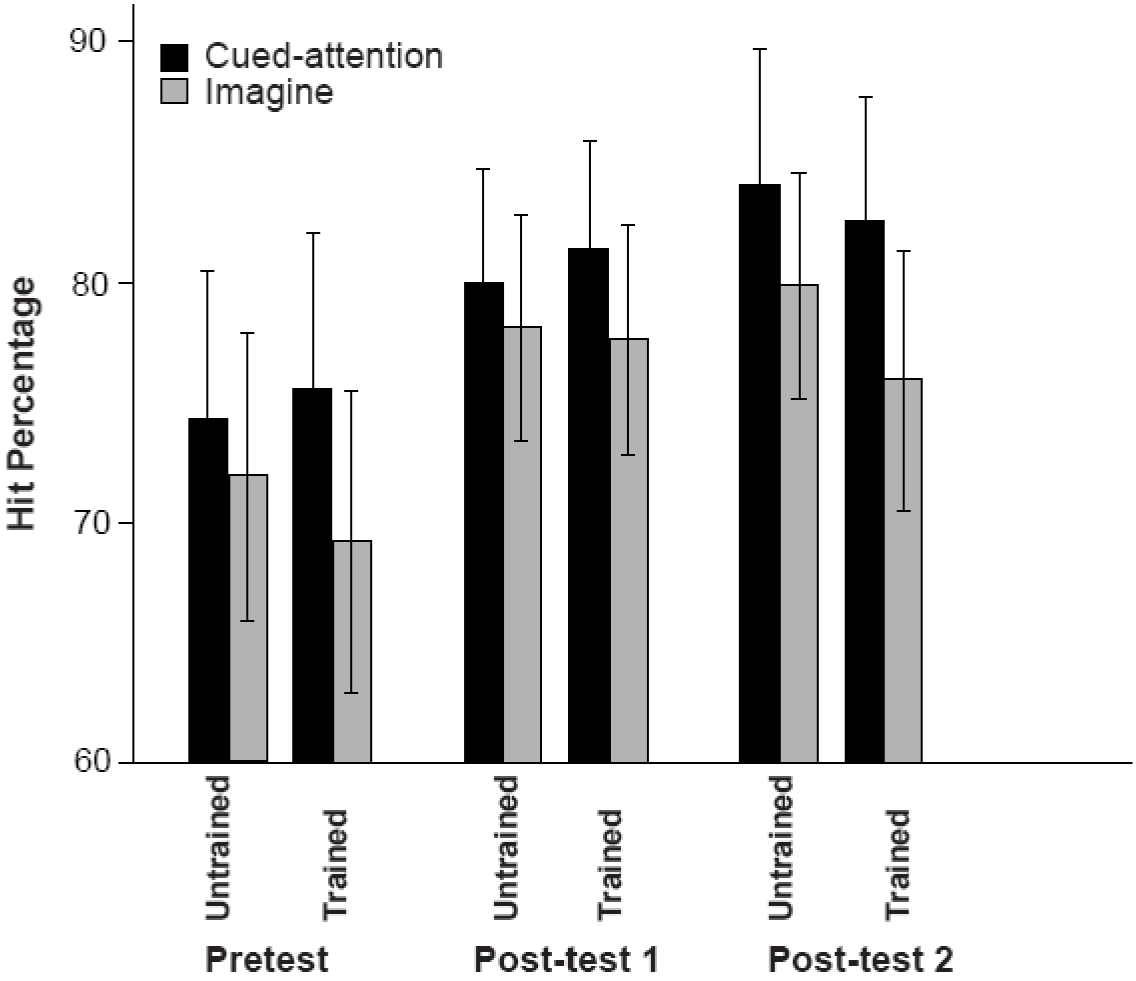


Figure 16. Percentage of correct responses for the cued-attention and the imagine conditions in each of the sessions (Pretest, Post-test 1 and Post-test 2). The "trained scale" condition is shown in gray and the "untrained scale" condition is shown in black.

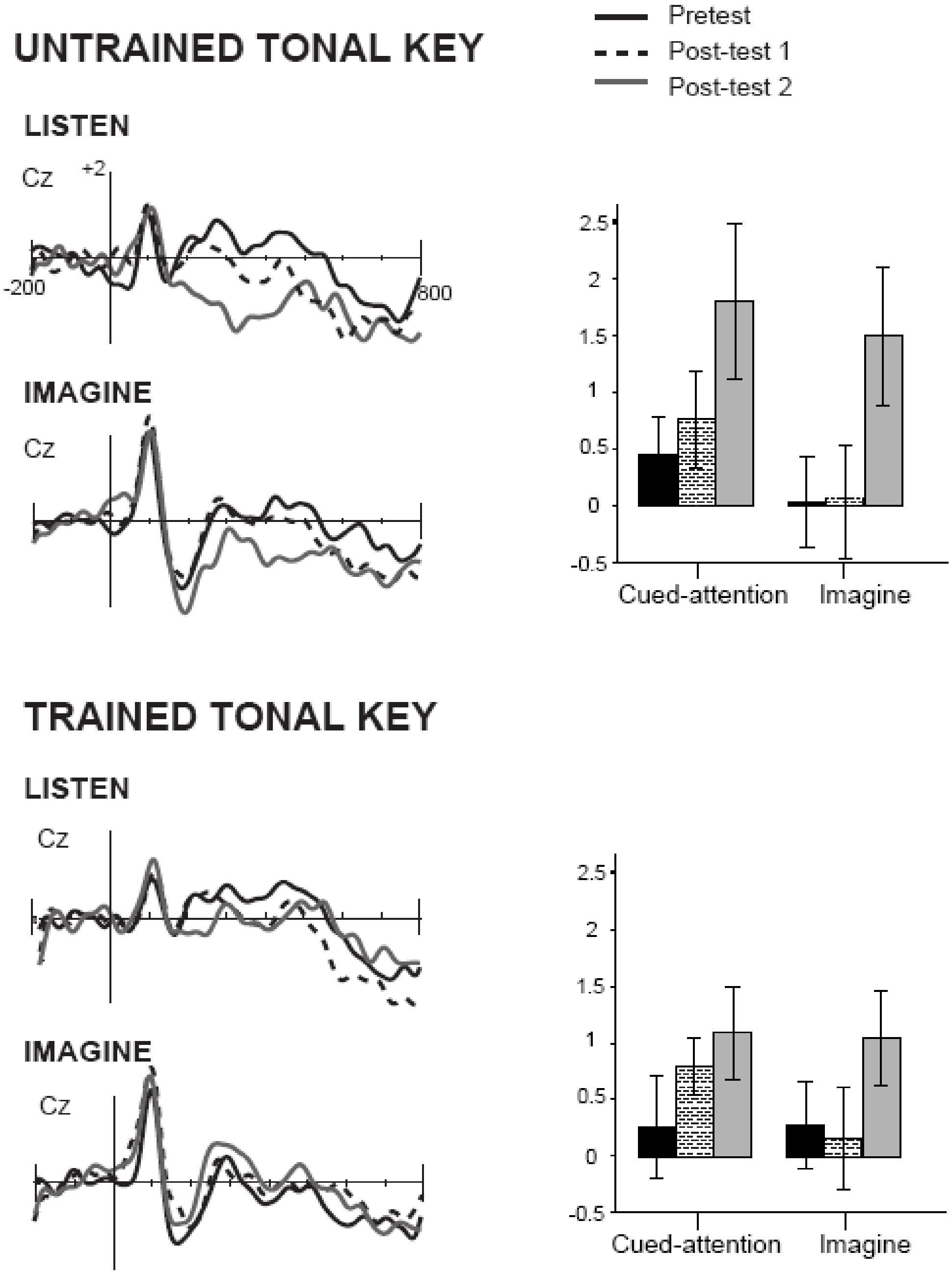
Although overall thresholds were more accurate after performing several training blocks (Average of the first two thresholds: 30.54 cents; Average of the last two thresholds:

25.36 cents) a *t*-test comparison shows that this improvement is not significant (*t*(1,6) = 1.87). A significant increment in the number of hits was found from the pretest to the post-test 2. Figure 16 shows the average of the percentage of hits in each of the three tests. On average, subjects seemed to be more accurate in the post-test 1 (after the training period) compared to the pretest, and they were still more accurate in the post-test 2 (after 48 hours) compared to the post-test 1. Only the imagine condition in the trained key was worse in the post-test 2 compared with the post-test 1. A repeated-measures ANOVA with Training (pretest, post-test1, and post-test 2), Key (trained vs untrained tonal key), Task (cued-attention vs imagine) as within-subjects factors shows significant differences in the Training factor (F(1,6) = 6.31; p < .011) with a greater number of hits for the post-test 2 compared with the pretest. *T*-tests show a significant increment from the pretest to the post-test 2 in the cued-attention condition (trained tonal key (p = .024); untrained tonal key (p = .035) and the imagine condition (trained tonal key (p = .033); untrained tonal key (p = .05). There are a significant greater number of hits in the post-test 1 compared with the pretest only in the untrained key for the cued-attention task(p = .048). Factors Key (F(1,6) = .263), Task (F(1,6) = 13.62) and the interactions were not significant.

ERP results

Both scalp map and latency indicate that such ERP effect corresponds to P3a component. Grand-average waveforms for ERPs to the target tone recorded from Cz are shown in Figure 17. ERPs elicited by the target showed a stronger P3a effect in the Post-test 2 (after 48 hours) compared with the Post-test 1 and Pretest. No differences in latency were found either between the conditions or between the different keys (trained vs. untrained).

An ANOVA was performed to study these changes in the P3a effect produced by training, with Task (cued-attention vs. imagine condition), Key (Untrained and Trained key scales) and Training (Pretest, Post-test 1, and Post-test 2) as within-subjects variables. The ANOVA shows a significant increment in the effect due to training ($F(1,6) = 6.56$; $p = .012$) with a greater P3a effect in the Post-test2 compared with the Post-test1 ($p = .31$) and compared with the Pretest ($p = .029$). The variables Task ($F(1,6) = 2.28$), Key ($F(1,6) = .719$) and the interactions were not significant.



<u>Figure 17.</u> Superimposed grand average ERPs (n = 7) to Pretest (black line), Post-test 1 (dashed line) and Post-test 2 (gray line). The graphs show mean of maximum peak of Fz and Cz electrodes.

General discussion

In the present experiment electrophysiological activity was recorded while subjects listened to and imagined musical scales and decided whether the last note of the scale was in tune. After a pretest they were given 10 blocks of training in which only the 'listen' condition was presented. The post-tests showed that the P3 effect related to the detection of the deviance was incremented with training in both the listen and the imagine condition. This increment was significantly more different after 48 hours.

Perceptual learning transference to imagery

Perceptual areas, associative auditory cortex, together with frontal and prefrontal areas seem to mediate the process of auditory image retrieval and generation (Halpern & Zatorre, 1999). ERPs have been applied to the study of imagery showing that mental images lead to similar neural responses than perceptual stimuli, and brain responses to deviated tones such as the MMN (Herholz, Lappe, Knief, & Pantev, 2008) and the P3 (Navarro-Cebrian & Janata, in press-a) are evoked when auditory images precede the target.

In this experiment we further studied this relation by looking at the transfer of perceptual learning. We show how the process of learning to recognize an out of tune tone when listening to a musical scale was transferred to the process of detecting an out of tune tone when imagining a musical scale. Because the P3a effects found in the imagine condition are transferred to the scale in which the participants were not trained, it can be ruled out that this transference to the imagine condition is due to a learning of the target. We believe that both the learning of the individual pitches and the learning of the intervals and the complex relation between the notes of the scale occurs during the training session which facilitates later the process of imagining the tones and relations between tones in the imagine condition.

Perceptual learning is defined as the experience-induced improvement in perception. When people are trained to discriminate pitch, for example, frequency-sensitive receptors narrow, and two close frequencies that were indistinguishable, start being perceived as different. In the auditory domain, perceptual learning studies using ERPs have investigated changes in early components related to the perception of the stimulus (Tremblay et al., 2001) and to the mismatch occurring when a deviated stimulus is presented (Atienza, Cantero, & Dominguez-Martin, 2002).

Although learning effects may be expected in earlier components, it seems reasonable that more training is needed to appreciate changes at these time windows. In fact, in perceptual learning studies using simple tasks that only imply learning of stimulus features, behavioral changes only start being appreciated after several days of training (Schoups, Vogels, & Orban, 1995). The kind of training provided in this experiment also seems to imply improvement in procedural aspects for this complex task. This training, at the same time, benefits the comparison of pitch between notes in the scale and thus, helps to identify the out of tune target. In this sense, a change in a later component, the P3a, was expected in the present experiment. Further support for this comes from a study in which subjects with a more accurate performance listening to and imagining scales have a stronger P3a when the last tone of the scale is deviated (Navarro-Cebrian & Janata, in press-a). The P3a potential may be explained here as reflecting an acquired ability to form accurate images so that the target is perceived as a deviant. The brain areas in charge of this ability may be shared by perceptual and imagery tasks. Only the fact that some brain systems are shared between imagery and perception can explain that the ability of accurately perceiving tones in a scale is transferred to the ability of accurately imagining tones in a scale.

Frequency and task generalization

Early and later neural processes seem to be both affected by the training in a complex task like the one presented here. While the changes of early processes produced by perceptual learning have been showed to be less generalizable (Sagi & Tanne, 1994; Sireteanu & Rettenbach, 1995), our data shows that the processes implicated in the elicitation of the P3 component can be generalized to other frequencies. We showed how the perceptual training in a musical scale can be generalized not only to another task (imagery) but also to an untrained key.

Some studies propose that the general learning that can be transferred to other similar stimuli often reflect learning in procedures needed to carry out the task but no changes in the stimulus processing are produced (Recanzone et al., 1993; Robinson & Summerfield, 1996). However, although the tasks used in this study present some complexity and procedural characteristics might be certainly learned and transferred in the post-tests, the learning of the correct interval between the penultimate and the last pitch is also needed to explained the improvement found from the pretest to the post-tests. Other recent studies have proven frequency learning generalization (Delhommeau et al., 2005; Irvine et al., 2000) and have shown that although pitch discrimination learning might be partly specific, a rapid generalization of the learning occurs however when subjects are presented other frequencies (Delhommeau et al., 2005). In this sense, once people have practice with one frequency, they will learn much faster to discriminate other frequencies.

Moreover, perceptual learning may take place due to supervised training, driven by feedback, or unsupervised training, without any instructions or feedback (Goldstone, 1998). Even when no instructions or feedback were given for the untrained tonal key or the imagery task, the training effects in post-test 2 might be partly due to some learning occurring in post-test 1. However, although not as pronounced as in the post-test 2, an improvement in both the electrophysiological records and the behavioral data was found in post-test 1 compared with the pretest (see Figure 15) which can only be explained by the generalization of the perceptual and procedural learning taken place during the training session.

Furthermore, training in pitch discrimination improves thresholds of all listeners (Amitay et al., 2005; Demany & Semal, 2002; Irvine et al., 2000) but a lot of individual variability is found in this type of tasks. Some studies have found that this variability between subjects could partly explain the nature of learning and the unresolved question of transference vs. frequency specificity in perceptual learning (Amitay et al., 2005). The present data shows that the ability to detect a mistuned note in a musical scale can be transferred to other tonal keys, and this generalization could be explained by the specific properties of the task, the unspecific practice during the tests or the individual characteristics of the listeners.

Fast vs. slow neural changes

Behavioral and physiological data show changes in the post-test 1 (right after the training session) but significantly stronger changes in the post-test 2 (after at least 48 hours) compared to the pretest. This finding is in agreement with previous physiological data that proves slower neural changes happening after sleeping. While the perceptual learning changes produced during or right after the training might be explained by fast field modulation of cortical neurons (Gilbert, 1994), the changes produced after hours, and while no sensory information is presented, might be responsible of long-term memory consolidation and explained by the reorganization of cortical representations (Karni & Sagi, 1993). Interestingly, these slow neural changes affected also the performance of the untrained imagery task. While some improvement was observed in this task right after the training session (post-test 1), significant differences in performance and the P3 effect were found after 48 hours showing that the slow neural reorganization after training in a perception task explains the perceptual consolidation in memory and therefore the subsequent ability to internally evoke these objects with greater accuracy.

References

Amitay, S., Hawkey, D. J., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Perception and Psychophysics, 67*, 691-698.

Atienza, M., Cantero, J. L., & Dominguez-Marin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learning and Memory, 9*, 138-150.

Atienza, M., Cantero, J. S., & Dominguez-Martin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learning & Memory, 9*, 138-150.

Besson, M., & Faïta, F. (1995). An Event-Related Potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance, 21*(6), 1278-1296.

Besson, M., Faïta, F., & Requin, J. (1994). Brain waves associated with musical incongruities differ for musicians and non-musicians. *Neuroscience Letters, 168*, 101-105.

Delhommeau, K., Micheyl, C., & Jouvent, R. (2005). Generalization of frequency discrimination learning across frequencies and ears: implications for underlying neural mechanisms in humans. *Journal of Assoc. Res. Otolaryngol., 6*, 171-179.

Demany, L., & Semal, C. (2002). Learning to perceive pitch differences. *Journal of Acoustic Society of America, 111*, 1377-1388.

Doane, S. M., Alderton, D. L., Sohn, Y. W., & Pellegrino, J. W. (1996). Acquisition and transfer of skilled performance: Are visual discrimination skills stimulus specific? *Journal of Experimental Psychology: Human Perception and Performance, 22*(5), 1218-1248.

Fahle, M. (2005). Perceptual learning: specificity versus generalization. *Current opinion in neurobiology, 15*(2), 154-160.

Gilbert, C. D. (1994). Early perceptual learning. *Proceedings of the National Academy of Sciences, 91*, 1195-1197.

Goldstone, R. L. (1998). Perceptual Learning. *Annual Review of Psychology, 49*, 585-612.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex, 9*(7), 697-704.

Herholz, S. C., Lappe, C., Knief, A., & Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. *European Journal of Neuroscience, 28*, 2352-2360.

Irvine, D. R., Martin, R. L., Klimkeit, E., & Smith, R. (2000). Specificity of perceptual learning in a frequency discrimination task. *Journal of Acoustic Society of America, 108*, 2964-2968.

Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. *Perception and Psychophysics, 68*, 829-844.

Jung, T. P., Humphries, C., Lee, T. W., Makeig, S., McKeown, M. J., Iragui, V., et al. (1998). Extended ICA removes artifacts from electroencefalografics recordings. *Advances in Neural Information Processing Systems, 10*, 894-900.

Kraus, N., McGee, T., Carrell, T. D., King, C., Tremblay, K., & Nicol, T. (1995). Central auditory system plasticity associated with speech discrimination training. *Journal of Cognitive Neuroscience, 7*(1), 25-32.

Marvit, P., Florentine, M., & Buus, S. (2003). A comparison of psychophysical procedures for level-discrimination thresholds. *Journal of the Acoustical Society of America, 113*(6), 3348-3361.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*, 36-47.

Navarro-Cebrian, A., & Janata, P. (in press-a). Electrophisiological Correlates of Accurate Mental Image Formation in Auditory Perception and Imagery Tasks. *Brain Research*.

Navarro-Cebrian, A., & Janata, P. (in press-b). Influences of multiple memory systems on auditory mental image acuity. *Journal of the Acoustical Society of America*.

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neurosciences, 13*, 87-103

Robinson, K., & Summerfield, Q. A. (1996). Adult auditory learning and training. *Ear and Hearing, 17*(3), 51-65.

Sagi, D., & Tanne, D. (1994). Perceptual learning: Learning to see. *Current opinion in neurobiology, 4*, 195-199.

Schön, D., & Besson, M. (2005). Visually induced auditory expectancy in music reading: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 17*(4), 694-705.

Schoups, A. A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: Retinotomy, orientation specifity and monocularity. *Journal of Physiology, 483*, 797-810.

Sireteanu, R., & Rettenbach, R. (1995). Perceptual learning in visual search: fast, enduring, but nonspecific. *Visual Research, 35*, 2037-2043.

Song, Y., King, Y., Fan, S., Qu, Z., Xu, L., Lu, C., et al. (2005). Neural substrates of visual perceptual learning of simple and complex stimuli *Clinical Neurophysiology, 116*(3), 632-639.

Tartaglia, E. M., Bamert, L., Mast, F. W., & Herzog, M. H. (2009). Human Perceptual Learning by Mental Imagery. *Current Biology, 19*, 2081-2085.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training

Ear and hear, 22, 79-90.

Wright, B. A., & Zhang, Y. A. (2009). A review of the generalization of auditory learning. *Phil. Trans. R. Soc. B., 364*, 301-311.

CHAPTER V,

GENERAL DISCUSSION

Summary and conclussions

The present doctoral research has focused on the investigation of the formation of auditory mental images of pitch. We studied the cognitive and physiological characteristics underlying the formation of accurate pitch representations in music. More specifically, in the first place, we tried to elucidate the relative influence of the different types of memory systems on perception and imagery. In chapter 2, in a series of behavioral experiments, we show that the acuity of pitch representations depended on the combination of several memory factors, such as the tonal relation of the images/representations with the musical context (long-term memory), the training or repetition of the stimulus (short-term memory) and the assistance of the preceding sensory information presented. More importantly, we show how both a bottom-up perception task (in which tones preceding the representation are presented) and a top-down imagery task (in which tones preceding the representation are imagined) are influenced by these different memory systems in a different grade. In this sense, the accuracy of representations formed in an imagery task is more influenced by the experience with the tonal context (long-term memory) while the accuracy of the representations in a bottom-up perception task seems to correlate more with a task in which the representation of the target is equal to a sensory memory trace.

In a second stage (chapter 3), we used the ERP technique to study the neural correlates of accurate mental representations. We make for the first time a direct comparison between imagery and perception using ERP. The results show that, to a lesser degree, a sequence of images preceding an auditory target elicit similar neural correlates than a sequence of auditory stimuli. In this sense, the P3a response to mistuned targets was also found in the imagery condition because of the comparison between a pure mental image and the mistuned target tone. Moreover, this neuronal response was correlated with performance and was larger for listeners who formed more accurate images. Subjects with more accurate images were able to better discriminate the pitch of the target. We also show physiological differences between good and bad performers in an earlier ERP component. In the imagine condition, the amplitude of the N100 component of the auditory evoked potential in response to the target tone was smaller for those subjects with more accurate mental images. We interpret this result as a short-term neuronal habituation of the audtiory cortex due to the formation of a sequence of auditory mental images. Due to a greater difficulty or an impossibility to form images, this habituation effect will not occur or will occur to a lesser degree in the group of bad performers, and, as a result, the N100 amplitude will increase after a few seconds of silence. Additionally, the influence of long-term implicit memory for the tonal structure of Western music on the acuity of mental images was examined by comparing responses to leading-tone (contextually unstable) and tonic (contextually stable) targets. Images were more accurate and led to a stronger P3a effect for targets that were related more closely to the established tonal context.

Finally, in a different ERP experiment, we studied whether the training effects found after perceptual learning could be generalized to an improvement of the acuity of images. We presented a pretest, a post-test (after a training period), and a second post-test (after 48 hours). The participants were trained only in a bottom-up (perception) task, and the data showed that there was also an improvement in the top-down (imagery) task in the post-tests with respect to the pretest. In the post-tests there was not only an improvement in performance but also a stronger P3a effect occurring after the last deviated tone in an 'imagined' sequence of tones. These effects were more significant after 48 hours than right after the training session, demonstrating the slow neural changes that develop in the absence of stimulation and that are responsible for consolidation in long-term memory.

Discusión general

Influencia de los distintos tipos de memoria en la formación de imágenes mentales

El hecho de que algunos sujetos tengan más habilidades que otros entonando melodías o discriminando una nota desafinada depende en gran medida de los años de experiencia musical previa. Sin embargo, dos personas con la misma experiencia podrán tener más o menos facilidades en estas tareas debido a factores innatos (Loui, Alsop, & Schlaug, 2009). Las diferencias individuales encontradas en relación a aptitudes musicales están relacionadas en gran medida con la memoria. El estudio de tales diferencias individuales ha sido uno de los principales objetivos de esta tesis.

Por otro lado, nuestro trabajo se ha centrado en el estudio de las representaciones e imágenes mentales formadas vía "top-down" vs. "bottom-up". Es importante entender que en este trabajo hemos identificado representaciones e imágenes mentales como un mismo fenómeno. Esta hipótesis está basada en investigación previa (Janata & Paroo, 2006) que indica un solapamiento neuronal de las predicciones o representaciones mentales (en una tarea de percepción) y las imágenes mentales (en imaginería). Cuando se escucha una melodía, por ejemplo, cada nota evoca predicciones de la siguiente nota en la secuencia, y estas predicciones equivalen a representaciones mentales de la memoria a largo plazo al igual que las imágenes mentales en imaginería. Lo que diferencia a estos dos fenómenos es que en el caso de la imaginería, las imágenes mentales son evocadas mediante un mecanismo top-down en vez de bottom-up.

Principalmente, en este trabajo se han analizado los factores que influencian la calidad de las imágenes o representaciones mentales y qué es lo que distingue que la calidad varíe de una tarea a otra (imaginería vs. percepción). Janata y Paroo (2006) realizaron una serie de experimentos conductuales en los que comparaban la calidad o precisión de las representaciones en dos tareas que históricamente han sido consideradas como diferentes (percepción e imaginería). El capítulo segundo describe una serie de experimentos conductuales en los que se compararon estas dos tareas y se estudió la influencia que las distintas formas de memoria tienen en la calidad de las representaciones / imágenes. A lo largo de tres experimentos, se hizo variar la relativa influencia de los distintos tipos de memoria, memoria sensorial, a corto plazo y a largo plazo.

Algunos autores han defendido que existe una similitud, pero no identidad, entre la activación neuronal provocada por un estímulo auditivo vs. aquella producida por procesos mentales (Cowan, 1995). Nosotros sugerimos que debe haber un almacén de memoria común capaz de mantener representaciones y/o imágenes. Este almacén sería más bien el sustrato donde se hace posible la comparación entre representaciones evocadas por estímulos externos e imágenes mentales internas. El modelo de memoria de trabajo de Baddeley y Hitch (Baddeley & Hitch, 1974; Baddeley & Logie, 1992) podría servir de base también para entender este almacén de memoria del que hablamos. Baddeley y Logie (1992) proponen el almacén de información fonólogica (parte del bucle fonológico de su modelo de memoria de trabajo) como base para la imaginería auditiva. Los autores describen que los objetos pueden llegar a este almacén por medio de inputs auditivos o mediante la activación de la memoria a largo plazo (imaginería).

Nosotros sugerimos que en este almacén se comparan los objetos internos y externos, y que la calidad de los objetos del almacén estará influenciada por las distintas fuentes de memoria dependiendo de las demandas de la tarea. A continuación se discuten los resultados encontrados en el capítulo 2 con relación a los distintos tipos de memoria.

En cuanto a la memoria a largo plazo, en las tareas presentadas en nuestros experimentos, la experiencia musical tiene una gran influencia en la ejecución de los sujetos. La experiencia musical modula la actividad neuronal implicada en la detección de los pequeños cambios de tonos tanto preatencional (Koelsch, Schröger, & Tervaniemi, 1999) como atencional (Tervaniemi, Just, Koelsch, Widmann, & Schröger, 2005). Además, se ha probado que estos dos grupos de sujetos, con y sin experiencia musical, obtienen mejores resultados en una tarea de imaginería, no sólo con estímulos musicales sino también con otro tipo de estímulos auditivos, como por ejemplo, comparando las imágenes de características acústicas de sonidos como el canto de un pájaro (Aleman, Nieuwenstein, Böcker, & Haan, 2000). Los resultados descritos en el capítulo 2 muestran que la calidad de las imágenes mentales está en gran parte influenciada por la memoria a largo plazo. Nuestros datos muestran, por ejemplo, que los sujetos con más experiencia musical pudieron evocar mejores representaciones mentales en ambas tareas (top-down -imaginería- y bottom-up -percepción-). Es más, mientras que se obtuvieron grandes diferencias entre imaginería y percepción para aquellos sujetos sin ninguna experiencia, los sujetos con experiencia musical mostraron un mayor beneficio en la tarea de imaginería con los que las diferencias entre estas dos tareas se vio reducida en este grupo de sujetos. A pesar de estos resultados, es importante señalar que tanto en la tarea de percepción como en la de imaginería algunos sujetos sin experiencia obtuvieron resultados similares que los sujetos músicos. Factores innatos que aportan un beneficio a la hora de discriminar un tono deben ser la causa de tales diferencias individuales (ej. Loui y cols., 2009).

Por otro lado, los efectos de la memoria a largo plazo aparecen reflejados en el experimento 1 del capítulo 2 como consecuencia del aprendizaje implícito de la estructura tonal de la música. Los efectos priming debidos al contexto que han sido previamente descritos en lenguaje (Stanovich & West, 1979) han sido también replicados en música de forma que el contexto tonal (o incluso una sóla nota o tono) puede facilitar el procesamiento de una nota *target* (Tillmann & Bigand, 2002). Los capítulos 2 y 3 describen cómo la tonalidad en cada escala crea un contexto musical que hace que las predicciones sean distintas para cada nota de la escala. De esta forma, como se sugirió en el capítulo 1, las representaciones mentales de la última nota de la escala, la tónica, fueron más esperadas y por tanto más precisas que las representaciones para la séptima nota, *leading tone*.

Por otro lado, los efectos de la memoria a corto plazo en tareas de discriminación tonal han sido observados en experimentos en los que se mantiene constante la frecuencia de los tonos o se repite el tono *target* dentro del mismo ensayo (Demany & Semal, 2002;

Micheyl, Delhommeau, Perrot, & Oxenham, 2006). En el capítulo 2, algunos sujetos obtuvieron iguales resultados en la tarea de imaginería que en la tarea de percepción cuando la misma escala se repetía de ensayo a ensayo. Los resultados fueron significativamente mejores cuando se presentó una única tonalidad que cuando varió la tonalidad de ensayo a ensayo. Aunque se ha probado (Deutsch, 1972, 1975) que en música la información relativa (relación entre notas) se retiene más fácilmente que la información absoluta (altura de una nota), sobretodo en el caso de individuos sin experiencia musical, aquí, debido a las características de la tarea y tras la repetición estimular, los sujetos parecen beneficiarse más de la información absoluta que de la relativa de las escalas.

En cuanto a los resultados obtenidos en este trabajo con relación a la memoria a corto plazo, es importante señalar también la influencia que la memoria a largo plazo tiene en la memoria a corto plazo. El capítulo 2 muestra cómo los beneficios debidos a la repetición en la memoria a corto plazo se ven incrementados en los casos en los que la nota *target* está más relacionada con el contexto musical que ha sido aprendido de forma implícita por la mera exposición a la música. La relación entre estos dos sistemas de memoria ha sido ya estudiada, por ejemplo, en lenguaje (Hulme, Roodenrys, Brown, & Mercer, 1995). Nuestro estudio muestra por primera vez la influencia que el aprendizaje implícito de la estructura tonal tiene en la consolidación tonal en la memoria a corto plazo.

Con el fin de estudiar la relación de estas dos tareas con la memoria sensorial, se presentó una tarea de discriminación tonal típica en la que las decisiones de los sujetos se basan en huellas de memoria sensorial. En esta se obtuvieron resultados esperados con respecto a los encontrados en la literatura previa (ej., Micheyl & cols., 2006). Como se mostró en el capítulo 2, la ejecución en la tarea bottom-up fue equiparable a la encontrada en la tarea de discriminación entre dos tonos. Por el contrario, en la tarea de imaginería, la calidad de las imágenes mentales empeoró con respecto a la tarea de discriminación, lo cual indica una mayor relación de la memoria sensoria con nuestra tarea de percepeción.

Datos psicofisiológicos en imaginería y percepción

Existen bastantes estudios psicofisiológicos que presentan tareas típicas de discriminación tonal (Sams, Paavilainen, Alho, & Näätänen, 1985; Tiitinen, May, Reinikainen, & Näätänen, 1994). Además, algunos estudios han investigado por ejemplo las diferencias entre músicos y no músicos encontrado un incremento en potenciales preatencionales (N1, P2) y atencionales (P3a) en sujetos músicos en respuesta a tonos o pequeños cambios de tono. La principal diferencia entre nuestro estudio y estudios previos ha sido el uso de imágenes mentales como ayuda a la hora de tomar una decisión sobre la afinación de la nota. Estudios previos han empleado tareas de imaginería visual (por ejemplo, leer notación musical) y se ha obtenido un incremento de potenciales como el MMN y la P300 al presentarse un estímulo auditivo desafinado en relación a la melodía leída. Un estudio más reciente con MEG encontró una MMN respuesta tras una terminación incorrecta de una melodía conocida que había sido en parte imaginada

(Herholz, Lappe, Knief, & Pantev, 2008). Nuestro estudio (capítulo 3) es el primer estudio de potenciales evocados que hace una comparación directa de imaginería y percepción musical.

En el capítulo 3 hablamos principalmente de un potencial temprano, la N100, que tiene lugar en la corteza auditiva y está influenciado por factores endógenos y exógenos (Näätänen & Picton, 1987). De esta forma, tanto características físicas del estímulo como factores atencionales podrán modular la latencia y amplitud del componente N100 (Woldorff et al., 1993). Por otro lado, en nuestro experimento se puede apreciar cómo la amplitud del N100 disminuye con cada nota de la escala. Debido a una habituación sensorial en la corteza auditiva, se produce una adaptación de la amplitud del componente N100 en respuesta a una secuencia de estímulos (Woods & Elmasian, 1986). Aunque ésta adaptación es más pronunciada cuando los estímulos de la secuencia tienen iguales características (mismo volúmen y frecuencia), el efecto ha sido también observado con una secuencia de tonos de frecuencias distintas (Nelson & Lassman, 1968).

Debido a la interacción de imágenes mentales y percepción en la corteza auditiva, nosotros habíamos sugerido que las imágenes mentales tendrían un efecto similar a las representaciones bottom-up en el córtex auditivo. De esta manera, observamos que tras una secuencia de imágenes mentales auditivas se produce un efecto de habituación en la corteza auditiva similar al producido cuando se presenta una secuencia de estímulos. En el caso de los sujetos con una pobre habilidad de imaginería, tras el intervalo de silencio anterior a la nota *target* en la condición de imaginería (intervalo en el que se supone que tienen que estar imaginando), el componente N100 tiene una amplitud similar a la amplitud de la primera nota de la escala. Sin embargo, para aquellos sujetos capaces de crear mejores imágenes mentales, se da un decremento de la amplitud de la N100 para la nota *target*, lo cual se ha interpretado como una adaptación del componente N100 debido a la sensibilización en la corteza auditiva. Aunque es difícil saber con certeza si los sujetos están imaginando o no las notas que preceden a la última nota, la ejecución de los sujetos, así como los componentes fisiológicos que siguen a la N100, nos permite concluir que aquellos sujetos con mejores resultados pudieron activar imágenes de más calidad y probablemente más vívidas que aquellos sujetos con peores resultados.

En segundo lugar, el componente evocado más relevante en éste paradigma ha sido la P3a. Este potencial ha sido hallado principalmente dentro del paradigma *oddball* apareciendo normalmente tras estímulos infrecuentes y nuevos (Comerchero & Polich, 1998). Sin embargo, la P3a aparece también en otro tipo de contextos y la novedad del estímulo no tiene por qué ser una condición necesaria. Por ejemplo, se ha observado un incremento de la P3a en contextos musicales cuando se presenta una nota o un acorde desafinado (Janata, 1995; Koelsch, Gunter, Friederici, & Schröger, 2000). Los datos electrofisiológicos en el capítulo 3 siguen este patrón y como era de esperar se halló una P3a después una nota desafinada dentro de una secuencia de notas. En este experimento, cada escala del ensayo presentaba 50% de probabilidades de que la última nota estuviese desafinada. Al contrario que en paradigma *oddball*, la nota target (desafinada) era presentada con igual probabilidad que el resto de notas y que la nota estándar (escala con última nota afinada). Estos resultados indican que el componente P3a depende de cuánto destaca el estímulo *target* (y no la novedad o la probabilidad) aunque ésta "notabilidad" del *target* pueda ser equivalente a "novedad" en algunos casos. Otros estudios anteriores han llegado a conclusiones similares (Comerchero & Polich, 1998; Goldstein, Spencer, & Donchin, 2002).

En la condición de imaginería se halló también una P3a tras la última nota de la escala. Además, se encontraron diferencias significativas en la amplitud de la P3a para los dos grupos sujetos, sobretodo en la condición de imaginería. Encontramos que aquellos sujetos con unas imágenes mentales de mayor calidad tuvieron una mayor amplitud en el componente P3a en ambas condiciones. Este hallazgo muestra más evidencia sobre las interacciones entre imaginería y percepción. En un paradigma *oddball* la comparación de comparación de un estímulo nuevo con la huella de memoria sensorial del estímulo anterior puede evocar un componente P3a. En el capítulo 3 y 4 mostramos cómo la comparación de estímulos auditivos con imágenes mentales puras también evoca una P3a en los casos en los que el estímulo y la imagen no son equivalentes.

Este estudio electrofisiológico muestra que las imágenes mentales llevan a respuestas neuronales similares que los estímulos en percepción. Sin embargo, todos nuestros datos muestran que son pocos los casos en los que las imágenes mentales en imaginería sean de igual calidad que las representaciones de los objetos en percepción. Y es que, la habilidad para escuchar un estímulo con el "oído de la mente" requiere que se haya percibido y practicado tal estímulo primero. Aparte de las investigaciones que estudian las diferencias entre músicos y no músicos, no existe hasta la fecha investigación que se halla encargado de ver los efectos de la práctica en las imágenes mentales y resulta por tanto interesante ver cómo el aprendizaje perceptivo o entrenamiento en una tarea de percepción produce mejorías en una tarea equivalente de imaginería.

Aprendizaje perceptivo

El aprendizaje perceptivo se define como un cambio (normalmente mejora) en la percepción. En una tarea de discriminación tonal, por ejemplo, dos tonos que son parecidos y no se distinguían, pasan a diferenciarse tras un periodo de entrenamiento en el que los receptores sensitivos a las distintas frecuencias se hacen más específicos (se estrechan). De esta forma, en estudios electrofisiológicos, tras un periodo de aprendizaje perceptivo, se han observado cambios (aumento en amplitud) de componentes evocados tempranos como la N100 (Tremblay, Kraus, McGee, Ponton, & Otis, 2001) o la MMN. (Atienza, Cantero, & Dominguez-Marin, 2002).

Es importante distinguir entre los distintos procesos neuronales que pueden verse modificados tras un periodo de entrenamiento. Ha habido bastante debate, por ejemplo, en cuanto a la transferencia o generalización estimular del aprendizaje perceptivo. Así, mientras que algunos estudios no encontraron que los efectos del aprendizaje perceptivo fuesen generalizables a otros estímulos (Sagi & Tanne, 1994; Sireteanu & Rettenbach, 1995), otras investigaciones concluyeron que algunas características del aprendizaje perceptivo sí se pueden generalizar.

Algunos autores defienden que aquel aprendizaje perceptivo que puede ser transferible a estímulos nuevos se trata más bien de un aprendizaje procedural (Recanzone, Schreiner, & Merzenich, 1993; Robinson & Summerfield, 1996). Así, los sujetos estarían transfiriendo los conocimientos adquiridos durante el entrenamiento que tienen que ver con la tarea y no el aprendizaje relacionado con características físicas del estímulo. En este tipo de entrenamiento procedural son otras áreas (lóbulo prefrontal), y no la corteza auditiva, las que se ven implicadas, y las neuronas se especializan más bien en tareas atencionales (por ejemplo, aprender a focalizar la atención en características estimulares más importantes que permiten desarrollar la tarea más eficazmente).

La controversia encontrada en la literatura en cuanto a la generalización o no del aprendizaje perceptivo, parece estar debida al uso de tareas que requieren la ejecución de distintos procesos. Así, en tareas simples, en las que por ejemplo se aprende a discriminar el ángulo de una línea, se observará poca o ninguna transferencia del aprendizaje a otros estímulos. Mientras que en otro tipo de tarea más compleja, por ejemplo, discriminar líneas de distintos ángulos que además están en movimiento, se encontrará una generalización del aprendizaje. Otros autores han sugerido que en estas tareas complejas no existe tampoco una generalización propiamente dicha, y más bien lo que ocurre es que el aprendizaje del "segundo" estímulo ocurre más rápidamente debido a que el sujeto ha aprendido algunas estrategias de ejecución en el tipo particular de tarea (Delhommeau, Micheyl, & Jouvent, 2005). Este estudio de discriminación tonal, por ejemplo, muestra que aunque el aprendizaje para discriminar un tono pueda ser en parte específico para ese tono, cuando a los sujetos se les presenta con otras frecuencias se da algún tipo de generalización que les permite aprender más rápido a discriminar los nuevos tonos (Delhommeau et al., 2005).

En general, los estudios típicos de discriminación tonal de los que se ha hablado en este estudio y que han estudiado en parte el aprendizaje perceptivo, han encontrado algún tipo de generalización a otras frecuencias (ej., Irvine et al.., 2000). En nuestro estudio (capítulo 4) hemos mostrado como el entrenamiento perceptivo con un tipo de escalas musicales puede ser transferible a una escala en una tonalidad diferente. Es más y más importante, hemos probado que el entrenamiento perceptivo puede transferirse a una tarea de imaginería, un tipo de tarea totalmente distinta pero a la vez (como hemos mostrado a lo largo de este trabajo) íntimamente relacionada. De esta forma, mostramos que los procesos relacionados con aprender a discriminar una nota desafinada dentro de una escala musical pueden ser transferidos a los procesos relacionados con la detección de una nota desafinada cuando se imagina una escala.

Se podría concluir aquí que el aprendizaje en esta tarea es simplemente debido a un mejor reconocimiento de la nota *target.* Si esto fuese así, no podríamos concluir que ha habido algún tipo de transferencia de una tarea a otra (de percepción a imaginería) ya que los sujetos habrían mejorado únicamente en la percepción de la última nota (que es la misma en ambas tareas). Sin embargo, existen dos datos que nos indican que esto no es lo que ha sucedido. Por un lado, el aprendizaje ocurrido durante el entrenamiento perceptivo ha sido transferido a otra escala en una tonalidad diferente. Este hecho afirma que, como hemos explicado más arriba, los sujetos han podido mejorar sus estrategias en cuanto a la tarea (aprendizaje procedural) y no sólo la percepción de la última nota, ya que de otra forma sería imposible encontrar esta trasferencia. Por otro lado, el hecho de que los mayores cambios fisiológicos encontrados se den en el componente P3a, nos indica también que el aprendizaje no ha sido tanto en la percepción del estímulo *target* sino más bien en la comparación de éste con la nota o "imagen" anterior. De haberse producido únicamente un aprendizaje perceptivo de la última nota, se habrían notado cambios en componentes más tempranos (P50, N100). El componente P3a, como se explica en el capítulo 3 de este trabajo, define una relación de desajuste o diferencia (*mismatch*) entre fenómenos internos y externos. Estos datos nos permiten concluir que la tarea de imaginería se ha visto beneficiada debido al entrenamiento en la percepción de las escalas.

Por otro lado, el hecho de que componentes evocados más tempranos no se hallan visto explícitamente afectados por el entrenamiento podría explicarse por el hecho de que más entrenamiento se necesite para poder apreciar tales cambios. Como se indicó en la introducción de este trabajo, los cambios neuronales en la corteza auditiva pueden requerir de mucho más entrenamiento como ocurre en la modalidad visual (Sagi & Tanne, 1994). Sería interesante prolongar por más sesiones el entrenamiento en la tarea de percepción que hemos presentado en este trabajo. Nosotros sugerimos que tras más sesiones de entrenamiento, la N100 de la última nota en la tarea de imaginería podría verse reducida debido al efecto de habituación que las imágenes mentales precedentes tienen en la corteza auditiva (ver efecto de la N100 encontrado en el capítulo 3).

Por último, nuestros resultados muestran un mayor beneficio (aunque no significativo) debido al aprendizaje en la escala de la tonalidad en la que los sujetos no fueron entrenados. Este hallazgo inesperado podría deberse a una mayor facilidad general para discriminación tonal en esta escala (escalas de tonalidades más altas), sin embargo, sólo con los datos de que disponemos no podemos llegar a ninguna conclusión a ese respecto. No existe hasta el momento investigación que haya cuestionado este hecho con lo que no podemos ofrecer una explicación definitiva y más investigación sería necesaria para desvelar que verdaderamente puedan existir diferencias en la percepción de escalas musicales de distintas tonalidades.

References

Aleman, A., Nieuwenstein, M. R., Böcker, K. B. E., & Haan, E. H. F. (2000). Music training and mental imagery ability. *Neuropsychologia, 38*, 1664-1668.

Atienza, M., Cantero, J. L., & Dominguez-Marin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learning and Memory, 9*, 138-150.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 47-89). New York: Academic Press.

Baddeley, A. D., & Logie, R. (1992). Auditory imagery and working memory. In D. Reisberg (Ed.), *Auditory Imagery.* Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

Comerchero, M. D., & Polich, J. (1998). P3a, perceptual distinctiveness, and stimulus modality. *Cognitive Brain Research, 7*, 41-48.

Cowan, N. (1995). *Attention and Memory.* New York: Oxford University Press.

Delhommeau, K., Micheyl, C., & Jouvent, R. (2005). Generalization of frequency discrimination learning across frequencies and ears: implications for underlying neural mechanisms in humans. *Journal of Assoc. Res. Otolaryngol., 6*, 171-179.

Demany, L., & Semal, C. (2002). Learning to perceive pitch differences. *Journal of the Acoustical Society of America, 111*, 1377-1388.

Deutsch, D. (1972). Effect of repetition of standard and comparison tones on recognition memory for pitch. *Journal of Experimental Psychology, 93*, 153-162.

Deutsch, D. (1975). The organization of short-term memory for a single acoustic attribute. In D. Deutsch & J. A. Deutsch (Eds.), *Short Term memory* (pp. 107-152). New York: Academic Press.

Goldstein, A., Spencer, K. M., & Donchin, E. (2002). The influence of stimulus deviance and novelty on the P300 and Novelty P3. *Psychophysiology, 39*, 781-790.

Herholz, S. C., Lappe, C., Knief, A., & Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. *European Journal of Neuroscience, 28*, 2352-2360.

Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology, 86*, 527-536.

Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *journal of Cognitive Neuroscience, 7*, 153-164.

Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. *Perception and Psychophysics, 68*, 829-844.

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: "Nonmusicians" are musical. *Journal of Cognitive Neuroscience, 12*, 520-541.

Koelsch, S., Schröger, E., & Tervaniemi, M. (1999). Superior attentive and pre-attentive auditory processing in musicians. *NeuroReport, 10*, 1309-1313.

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *Journal of Neuroscience, 29*, 10215-10220.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*, 36-47.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology, 24*(4), 375-425.

Nelson, D., & Lassman, F. (1968). Effects of intersignal interval on the human auditory evoked response. *Journal of the Acoustical Society of America, 44*, 1529-1532.

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neurosciences, 13*, 87-103

Robinson, K., & Summerfield, Q. A. (1996). Adult auditory learning and training. *Ear and Hearing, 17*(3), 51-65.

Sagi, D., & Tanne, D. (1994). Perceptual learning: Learning to see. *Current opinion in neurobiology, 4*, 195-199.

Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related protentials. *Electroencephalography and Clinical Neurophysiology, 62*, 437-448.

Sireteanu, R., & Rettenbach, R. (1995). Perceptual learning in visual search: fast, enduring, but nonspecific. *Visual Research, 35*, 2037-2043.

Stanovich, K. E., & West, R. F. (1979). Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory and Cognition, 7*, 77-85.

Tervaniemi, M., Just, V., Koelsch, s., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Experimental Brain Research, 161*, 1-10.

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature, 372*, 90-92.

Tillmann, B., & Bigand, E. (2002). A comparative review of priming effects in language and music. In P. McKevitt, S. O Nualláin & C. Mulvihill (Eds.), *Language, vision and music* (pp. 231-240). Amsterdam: John Benjamins.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training

Ear and hear, 22, 79-90.

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., et al. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences, 90*, 8722-8726.

Woods, D. L., & Elmasian, R. (1986). The habituation of event-related potentials to speech sounds and tones. *Electroencephalography and Clinical Neurophysiology, 65*, 447-459.