

TESIS DOCTORAL



2017

RESPUESTA EMOCIONAL A ESTÍMULOS MUSICALES: EFECTO DE LA
ESTRUCTURA DEL ESTÍMULO SOBRE LA RESPUESTA COGNITIVA Y
FISIOLÓGICA

Elena Saiz Clar

Máster en Metodología en Ciencias del Comportamiento y de la Salud

DEPARTAMENTO DE METODOLOGÍA EN CC. DEL COMPORTAMIENTO Y DE LA SALUD

FACULTAD DE PSICOLOGÍA

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA (UNED)

Director:

Dr. José Manuel Reales Avilés

Profesor Titular, área de Metodología en CC. del Comportamiento

Facultad de Psicología (UNED)

Co-director:

Dr. Miguel Ángel Serrano Rosa

Profesor Titular, área de Psicobiología

Facultad de Psicología (Universidad de Valencia)

RESPUESTA EMOCIONAL A ESTÍMULOS MUSICALES: EFECTO DE LA
ESTRUCTURA DEL ESTÍMULO SOBRE LA RESPUESTA COGNITIVA Y
FISIOLÓGICA

DEPARTAMENTO DE METODOLOGÍA EN CC. DEL COMPORTAMIENTO Y DE LA SALUD

FACULTAD DE PSICOLOGÍA

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA (UNED)

Autora:

Elena Saiz Clar

Máster en Metodología en Ciencias del Comportamiento y de la Salud

Director:

Dr. José Manuel Reales Avilés

Profesor Titular, área de Metodología en CC. del Comportamiento

Facultad de Psicología (UNED)

Co-director:

Dr. Miguel Ángel Serrano Rosa

Profesor Titular, área de Psicobiología

Facultad de Psicología (Universidad de Valencia)

Agradecimientos

Dar el paso y embarcarse en un doctorado es una decisión que produce principalmente una sensación de vértigo. A esta sensación le acompañan a lo largo de este camino otras emociones y experiencias, algunas muy positivas y otras negativas. A parte de todo el aprendizaje práctico que implica un doctorado, lo más impresionante para mí ha sido descubrir tanta gente maravillosa que te ayuda y te enseña a ser mejor investigadora, pero también mejor persona. A lo largo de este tiempo debes enfrentarte a innumerables dificultades, barreras y obstáculos que hacen el avance sumamente lento. Es imprescindible aprender a resolver estas dificultades, con ayuda y con constancia.

Porque una de las cosas que aprendes en este camino es que no llegas hasta este punto si no es con el apoyo y la ayuda de los demás. Sin el apoyo en primer lugar de las personas más implicadas en este proyecto, José Manuel y Miguel Ángel, directores de esta tesis. Quiero agradecer especialmente el esfuerzo diario y el empeño de José Manuel en esta tesis, ha sido el mejor director que podía haber tenido. He adquirido muchos conocimientos a su lado, pero las lecciones más importantes que me llevo son su propio ejemplo de integridad y honestidad en el trabajo, lecciones que me han hecho mejor científica. Aún nos queda camino por recorrer y cosas por aprender.

A Miguel Ángel tengo que agradecerle su trabajo, su buen criterio y sus ánimos constantes. Pero también le tengo que agradecer todo lo que he aprendido de él, porque es el ejemplo de que una pequeña idea puede cambiarlo todo. Porque con él inicié este camino. Todo empezó para un congreso de estudiantes cuando aún no había alcanzado la licenciatura. Gracias a los dos porque pienso sinceramente que no hubiera llegado hasta aquí sin vosotros.

Soy consciente de que llegar hasta escribir estas líneas no es sencillo y muchos de los que empiezan quedan atrás. Además de un gran conocimiento de tu área de estudio, debes tener multitud de cualidades que te permitan afrontar las situaciones difíciles, manejar la frustración, ser constante en todo momento, luchar, defenderte, caer y levantarte, pero también mantener la humildad, encajar las malas críticas y aceptar las buenas. Por fortuna ya tenía estas habilidades cuando empecé este camino porque las he aprendido de Hilario y Ana, mis padres, gracias a ellos porque me han hecho fuerte y constante. Sin sus lecciones tampoco habría llegado hasta aquí.

Esta travesía no puede hacerse solo, es mejor estar acompañado y yo he tenido el mejor compañero posible a mi lado, Fernando. Los días buenos y los días malos siempre ha estado ahí, mi templanza y mi palo en la tormenta. Pero también a ratos he disfrutado de la compañía de “mis niñas”, Sonia que siempre ha estado ahí cuando necesitaba levantarme, y Natalia que es un ejemplo en el que fijarme. Gracias, porque sin vosotros tampoco hubiera llegado aquí. Quiero agradecer también a Diana su trabajo expresar lo importante que ha sido para mí su ayuda. Espero algún día poder devolverte este favor. Sin ti hubiera sido también muy difícil llegar a este punto.

Lanzo un agradecimiento general a todos aquellos que en algún momento creyeron en mí, profesores que me enseñaron en la escuela o en el instituto, gente que aparece puntualmente en algún momento en el camino pero que llega a marcar un instante. Aunque también quiero acordarme de los que se preguntaron por qué la música y no algo más práctico, algo con más posibilidades de financiación, o algo “más importante” o más elevado, de los que dijeron que no podía, de los que dijeron que no valía, de los que apostaron que no lo conseguiría. Gracias porque con vuestras piedras levanté este muro.

Lista de figuras

CAPITULO 1

Figura 1: Modelo Neuropsicológico de Peretz y Coltheart (2003)	15
Figura 2: Modelo Circunflejo de la Emoción de Russell (1980).....	20
Figura 3: Clasificación de las piezas musicales en el trabajo de Bigand y cols. (2005).	24
Figura 4: Posición de los estímulos musicales en el experimento de Viellard y cols. (2008).....	26
Figura 5: Ejemplo de láminas para el estudio del reconocimiento facial de emociones de Ekman (1976).....	30
Figura 7: Resultados del trabajo de Khalfa y cols. (2003).	36
Figura 8: Variación de la frecuencia cardiaca entre las diferentes categorías emocionales a lo largo del tiempo.	37
Figura 9: Variación del ciclo respiratorio entre las diferentes categorías emocionales a lo largo del tiempo.....	37
Figura 10: Respuesta de músculo cigomático ante piezas tristes vs. alegres.	39
Figura 11: Respuesta de músculo corrugador de las cejas ante piezas tristes vs. alegres.	39
Figura 12: Cambios en la respuesta de conductancia eléctrica de la piel a lo largo del tiempo como respuesta a los estímulos alegres vs. tristes.	40
Figura 13: Cambios en la presión sanguínea diastólica a lo largo del tiempo como respuesta a los estímulos alegres vs. tristes.	40
Figura 14: Resultados en el trabajo de Roy y cols. (2009) en las diferentes medidas fisiológicas consideradas en función del grado de placer que genera el estímulo en el oyente.....	43
Figura 15: Modelo de redundancia de la señal de Balkwill y Thompson (1999).....	46
Figura 16: Resultados de Webster y Weir (2005).....	45.
Figura 17: Salidas en el modelo de Korhonen (2004).....	51
Figura 18: Modelo de Red Neuronal de Coutinho y Cangelosi (2009 y 2011).....	53
CAPÍTULO 2.....	59
Figure 1: Neural Network model used in our study	67

Figure 2: A schematic view of Peretz et al.'s (2003) neuropsychological model showing the structural similarity intended with our ANN model.....	68
Figure 3: SOM map after training.	76

CAPÍTULO 3

Figure 1: Integrative model of the emotional perception of music proposed by de double path hypothesis (Balkwill & Thompson, 1999; Saiz-Clar & Reales, 2017)	89
Figure 2: Graphical example of measures derived from the <i>Onsets</i> Curve for a single <i>onset</i> (t_1). ...	95
Figure 3: Eigenvalues as a function of factor in the rotated space.....	104

CAPÍTULO 4

Figura 1: Pantalla de registro continuo de las dimensiones emocionales Activación y Valencia..	124
Figura 2: Ejemplo de distribución media para la EDA en un ítem alegre (a), un ítem tenso (b), un ítem triste (c) y un ítem tranquilo (d). El eje X representa el tiempo en milisegundos y el eje y la intensidad en μS	129
Figura 3: Gráfico de dispersión del pronóstico de la activación predicha y el valor empírico de activación.	136
Figura 4: Gráfico de dispersión del pronóstico para la valencia emocional y el valor empírico de la valencia emocional.....	137

CAPITULO 5

Figura 1: Diagrama sobre el porcentaje de varianza que explicaría un modelo en el que las variables fisiológicas actúan como mediadoras.....	153
Figura 2: Pantalla de respuesta en el registro continuo de las dimensiones emocionales en el estudio de Saiz-Clar et al., (2017b).....	156
Figura 3: Arquitectura típica de una red recurrente tipo Elman.....	159
Figura 4: Gráficos de dispersión para los valores de a) <i>output</i> y <i>target</i> en Activación en la fase de entrenamiento, b) <i>output</i> y <i>target</i> en Activación la fase de generalización, c) <i>output</i> y <i>target</i> en Valencia emocional para la fase de entrenamiento y d) <i>output</i> y <i>target</i> en Valencia emocional para la fase de generalización.....	165

CAPÍTULO 6

Figura 1: Modelo Neuropsicológico de Peretz y Coltheart (2003)	172
Figura 2: Estructura RNA obtenida en el capítulo 2 compuesta por dos ADALINE para el procesamiento de la activación y valencia emocional y una SOM para el juicio categórico.....	181

Material Suplementario

Gráfico 3: Grado de activación autoinformado por el oyente segundo a segundo de las piezas tranquilas	225
Gráfico 4: Grado de activación autoinformado por el oyente segundo a segundo de las piezas alegres.....	225
Gráfico 5: Grado de activación autoinformado por el oyente segundo a segundo de las piezas tensas	226
Gráfico 6: Grado de activación autoinformado por el oyente segundo a segundo de las piezas tristes	226
Gráfico 7: Grado de VE autoinformado por el oyente segundo a segundo de las piezas tranquilas	227
Gráfico 8: Grado de VE autoinformado por el oyente segundo a segundo de las piezas alegres....	227
Gráfico 9: Grado de VE autoinformado por el oyente segundo a segundo de las piezas tensas	228
Gráfico 10: Grado de VE autoinformado por el oyente segundo a segundo de las piezas tristes ...	228

Lista de tablas

CAPITULO 1

Tabla 1: Porcentajes de acuerdo en la categorización emocional entre los oyentes del primer grupo de Vieillard y cols. (2008) y la intención del compositor de estas piezas musicales 25

Tabla 2: Medidas del estímulo musical usadas como entradas en el modelo de Korhonen (2004).. 50

CAPÍTULO 2

Table 1. Descriptive and inferential musical properties of Vieillard et al. (2008) stimuli as a function of emotions elicited..... 69

Table 2. Percent correct, means and standard deviations for activation and emotional valence as a function of emotion, correlations between musical properties and emotional dimensions and Adaline's weight for emotional dimension. 70

Table 3. Comparative analysis between output and target distributions for emotional dimensions. 73

Table 4. SOM's correct classification in our network and his comparison with Vieillard et al.'s experimental rates..... 75

Table 5. SOM's output layer weight by emotional dimensions. 77

Table 6. Means and Standard deviations for activation and emotional valence by emotional category on experimental phase and statistical indices for the comparison between output and target distributions..... 79

CAPÍTULO 3

Table 1: Description of the measurements obtained from the *Onset* Curve..... 97

Table 2: Rescaled factorial scores from the *Onsets* Curve measurements and their correlations with the emotional dimensions in the first study..... 101

Table 3: Rescaled factorial scores from the *Onsets* Curve measurements and their correlations with the emotional dimensions in the second study. 105

Table 4: Correlation matrix for factorial weights and Congruence Index of Wrigley y Neuhaus (1955) in first and second studies..... 107

CAPÍTULO 4

Tabla 1: Comparación entre las valoraciones al estímulo en nuestro estudio y en el de Vieillard et al. (2008).	127
Tabla 2: Matriz de correlaciones entre medidas fisiológicas y las dimensiones emocionales de Activación y Valencia Emocional.....	133
Tabla 3: Matriz de correlaciones entre medidas fisiológicas y parámetros estáticos del estímulo.	135

CAPITULO 5

Tabla 1: Magnitud de la asociación entre los valores <i>output</i> y <i>target</i> y ECM de las seis simulaciones.....	163
Tabla 2: Estudio del ajuste de la distribución de los valores del <i>output</i> , el <i>target</i> y del error en la RNA de Elman.	164
Tabla 3: Pesos sinápticos en las redes de Activación y Valencia Emocional tipo Elman.....	166

Material Suplementario..... 214

Tabla 1. Descripción de las piezas musicales empleadas en el segundo experimento del capítulo 2.	217
Tabla 2. Descripción de las piezas musicales empleadas en el segundo experimento del capítulo 3	222

Índice de Contenidos

Agradecimientos	3
Lista de figuras	5
Lista de tablas	8
Índice de Contenidos	10
CAPÍTULO 1. Emoción, percepción y procesamiento del estímulo musical	14
Modelos emocionales y música.....	19
El estudio psicofisiológico de la respuesta emocional	27
Enfoques en el estudio de la respuesta psicofisiológica de la emoción.....	28
Naturaleza de la respuesta emocional a la música: Congruencia entre la categorización emocional y la respuesta fisiológica del oyente	32
La dimensión de activación.....	32
La dimensión de valencia emocional	40
¿Qué desencadena la respuesta emocional del oyente?.....	44
Objetivos	54
CAPÍTULO 2. Predicting emotional response to music through a compound neural network ..	59
Abstract	60
Introduction	61
Objectives.....	66
Method	68
Stimulus variables	68
Stimuli	68
Architecture of the Artificial Neural Network	70
Results	71
Results for the Adaline networks	71
Results of the SOM network	74
Generalization of results to another set of stimuli.....	77
Discussion	81

CAPÍTULO 3. Prediction of emotional response of music using principal component factors extracted from the <i>onset</i> curves	85
Abstract.....	86
Introduction	87
Onset Curves (OC) as a Source of Physical Predictors	92
Method.....	96
Procedure and Stimuli	96
Plan of Analysis	98
Results	101
The First Study	101
The Second Study.....	105
Study of the Relationship between the OC and the Emotion Dimensions	107
General Discussion	109

CAPÍTULO 4. Predicción de la respuesta emocional continua del oyente a la escucha musical: Efecto de la estructura del estímulo y de la respuesta fisiológica.	115
Método.....	122
Descripción de la muestra	122
Descripción de la tarea y medidas relativas a la respuesta emocional	123
Variables relativas a la respuesta psicofisiológica	124
Variables relativas a la estructura del estímulo	126
Estímulos musicales empleados en el análisis.....	126
Plan de Análisis.....	128
Resultados	132
Evaluación continua de la emoción.....	132
Relación entre la CO, la respuesta fisiológica y las dimensiones emocionales.....	134
Estructura estática del estímulo como predictor de la Rs. Fisiológica	135
Predictores de la activación y la valencia emocional	136
Conclusiones	138
Sobre el juicio emocional del oyente.....	138
Sobre la respuesta fisiológica y el juicio emocional.....	141

Sobre la estructura del estímulo	141
CAPÍTULO 5. Modelado de la respuesta emocional y fisiológica al estímulo musical.....	146
Una aproximación desde las redes neuronales recurrentes.....	146
Introducción.....	147
Propiedades físicas y teóricas del estímulo musical.....	148
Medición continua del estímulo musical.....	149
Objetivo.....	154
Método.....	155
Descripción de las variables	156
Medidas Fisiológicas	156
Variables del estímulo musical.....	157
Medida de la emoción percibida.....	157
Descripción de la red.....	158
Plan de análisis	161
Resultados	163
Conclusiones	167
CAPÍTULO 6. Conclusiones	171
Sobre la estructura del estímulo	177
Sobre la respuesta fisiológica	182
Sobre la aplicación de los modelos de redes neuronales en la investigación	183
Otros factores internos del sujeto	186
Aplicaciones de nuestro trabajo.....	189
Aplicaciones en el ámbito clínico.....	189
Aplicaciones en ingeniería y computación.....	192
Bibliografía	197
Material Suplementario	214
Capítulo 2	214

ANN Training phase	214
Stimulus Variables	215
Capítulo 3.....	219
Capítulo 4.....	225

CAPITULO 1

Emoción, percepción y procesamiento del estímulo musical

Antes de comenzar a tratar la relación entre la emoción y la música, es interesante introducir un modelo neuropsicológico de la percepción musical que sirva de guía en este proceso. Los modelos son simplificaciones de procesos y sistemas usualmente complejos, y como tal, el modelo que proponemos a continuación es una representación simplificada de cómo pueden ser las interacciones entre la música y los diferentes procesos cognitivos señalados. Para su construcción, Peretz y Coltheart (2003) se basó no sólo en los resultados de sus propios trabajos, sino que este modelo es fruto de todo el *corpus* investigador acumulado durante décadas de trabajo aportado por multitud de investigadores en todo el mundo. En la Figura 1 podemos observar el modelo Cognitivo-Neuropsicológico de la Percepción Musical descrito por Peretz y que vamos a ir explicando paso a paso.

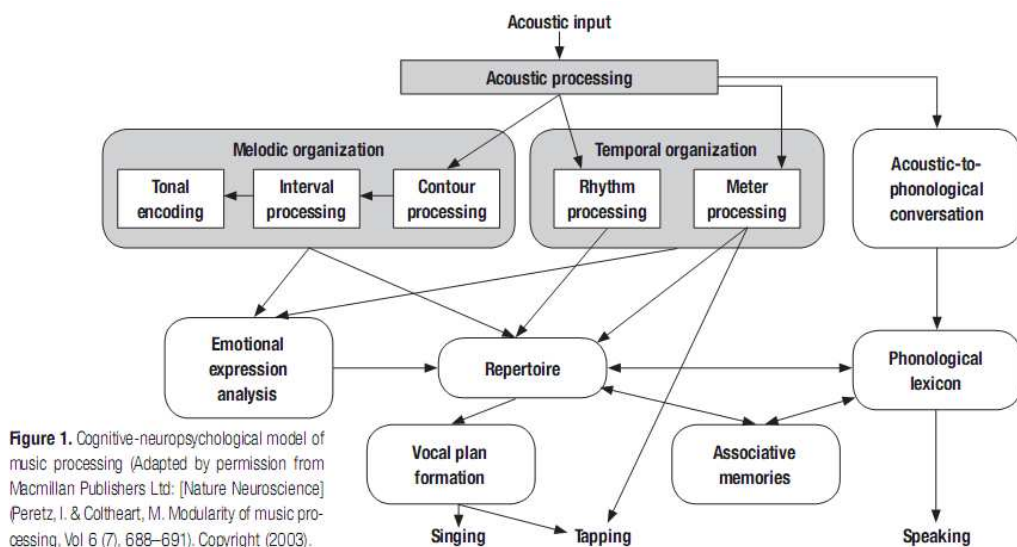


Figura 1. Modelo Neuropsicológico de Peretz y Coltheart (2003)

En la parte superior de la figura encontramos la entrada al procesamiento acústico. Esta entrada no es otra cosa que la pieza musical presentada al oyente. Cualquier pieza musical expuesta a un oyente está formada por diferentes elementos. Del mismo modo que el discurso está formado por párrafos y cada párrafo por palabras y cada palabra por letras, la

pieza musical (en adelante también lo denominaremos estímulo musical), está formada por elementos más pequeños cuya unión da lugar a un todo complejo. En el modelo observamos que tras la entrada del *input* musical llega la fase de procesamiento del mismo. El procesamiento del estímulo musical se refiere a la etapa donde la información acústica (en forma de onda sonora) es transformada en una señal bioeléctrica comprensible para nuestro cerebro.

El proceso de transducción de la señal acústica en señal eléctrica es un proceso complejo que excede los objetivos de esta introducción pero remitimos al lector interesado a los manuales de fisiología de la audición. Nosotros comenzamos nuestro camino una vez que ya la información, en forma de señales eléctricas, pasa a ser procesada por nuestro cerebro. En este paso, lo que sabemos hasta ahora es que esos pequeños elementos que forman el estímulo (lo que denominaremos en adelante **propiedades** del estímulo musical) son procesados por nuestro cerebro presumiblemente por, al menos, dos vías diferentes: por un lado, es procesada lo que denominamos **información temporal** del estímulo, y por otro lado, la **información tonal** del estímulo musical. Estas dos rutas de procesamiento se refieren en el modelo como “Organización Melódica” vs. “Organización Temporal” (*Melodic Organization* y *Temporal Organization* respectivamente en la Figura 1). Si nos fijamos en estas propiedades podemos llegar a la conclusión de que, por un lado, el cerebro procesaría toda la información relativa al ritmo separadamente de la información relativa a la tonalidad de la pieza. Sin embargo, veremos que en algunos momentos del procesamiento parece que estos dos sistemas no actúan independientemente sino que se apoyan el uno en el otro.

Viendo más de cerca esta parte del proceso de la percepción musical observamos que las dos rutas de procesamiento se separan a su vez en otras vías de procesamiento que trabajan en paralelo. La vía de la Organización Temporal (OT) distingue el procesamiento de la información métrica en conjunto y el procesamiento de lo que denominan “ritmo”. Sabemos

que la definición del ritmo en investigación es compleja. Podemos definirla desde un plano cualitativo como la sucesión de sonidos y silencios dentro de la partitura musical. Sin embargo, esta definición es poco operativa cuando tratamos de llevarla al laboratorio pues la posible combinación de esos elementos se hace inconmensurable. No obstante, podemos llevarlo a un terreno más cuantitativo centrándonos en las diferentes propiedades que puede tener el ritmo, por ejemplo, la regularidad. Bajo la etiqueta de “procesamiento métrico” se engloban otras propiedades musicales como son el tempo, la articulación o el compás musical. La vía de la Organización Melódica (OM) según el modelo de Peretz y Coltheart (2003) está dividida en tres sub-procesos relacionados entre sí: procesamiento del contorno, procesamiento del intervalo y codificación del tono

Esta distinción entre propiedades melódicas y propiedades temporales tiene su base en el modelo de procesamiento de Balkwill y Thompson, (1999) y Balkwill, Thompson, y Matsunaga (2004). Este modelo propone que cuando el oyente se enfrenta a la escucha musical de piezas no escritas bajo las reglas tonales propias de la cultura musical del oyente, éste se apoya, no obstante, en información que sí es compartida por todas las culturas cuando emiten un juicio emocional sobre la pieza musical: la información métrica.

Por último, vemos a la derecha del modelo (Figura 1) un apartado separado con respecto al procesamiento melódico y temporal: la conversación acústica-fonológica que conecta directamente con el aspecto fonológico del lenguaje y, a su vez, con el habla. Esta distinción separa por tanto el estímulo musical nuevamente en dos aspectos muy importantes: por un lado, estudiamos la pieza musical (con sus rutas temporales y melódicas) y, por otro lado, la información que proviene de la letra que puede contener la pieza musical. Es comprensible que hagamos esta distinción entre “letra y música” pues en principio, las rutas fonológicas son *a priori* diferentes a las rutas melódicas y temporales

Los siguientes niveles del modelo nos muestran cómo el procesamiento del estímulo musical está relacionado con diferentes procesos (considerados como “cajas negras” en el sentido de que se propone una función pero no se indica cómo la ejecutan) que corresponden principalmente a los tres procesos implicados en el procesamiento de la pieza musical: emoción, memoria y lenguaje, y como entre ellos hay un elemento más interrelacionado con todos ellos que es el repertorio. El **repertorio** englobaría todo nuestro conocimiento previo sobre música, sobre nuestra experiencia como oyentes o intérpretes. Como podemos observar, para Peretz todos los caminos del modelo terminan confluyendo en el “almacén” permanente del oyente que denomina repertorio: todo confluye ahí y todo producto está relacionado con este nivel.

Cuando introducimos el concepto de “caja negra” en el modelo de Peretz y Coltheart (2003) lo hacemos porque este modelo no nos explica qué sucede en estos procesos cognitivos cuando el oyente es musicalmente estimulado. Ese es precisamente el objetivo de este capítulo, introducimos en el conocimiento de lo que sucede en una de esa “cajas negras”: la emoción. Por último, el modelo neuropsicológico nos muestra los tres posibles productos de la estimulación musical: el “*tapping*” se refiere a la capacidad del oyente para reproducir los ritmos presentados musicalmente, bien de forma simultánea (por ejemplo, un bajista que sigue el ritmo de la batería) o bien tras un proceso de escucha (por ejemplo, cuando pedimos a una persona que reproduzca el ritmo de palmas de “*We will rock you*” de *Queen*). Podemos observar que según este modelo, la reproducción de ritmos vendría directamente relacionada con el procesamiento primario del ritmo del estímulo y de forma indirecta, estaría asociado al repertorio del oyente.

Otro de los productos sería directamente el canto que depende a su vez de un plan de formación vocal que consiste en la integración de un léxico musical con un léxico verbal para originar el plan vocal que se llevará a cabo. Planteado ya un marco de referencia en el que

guiarnos en el proceso de percepción musical, procedemos a centrar nuestra atención en la “caja negra” de la emoción.

Una de las principales cuestiones en el estudio de la respuesta a estímulos musicales es determinar si nos encontramos ante una respuesta en la que se encuentra presente un componente realmente emocional (postura emotivista) o si consiste en un proceso de **reconocimiento** de emociones y categorización del estímulo (postura cognitivista). Es decir, un aspecto básico en la investigación psicológica sobre la relación entre la música y las emociones es si las piezas musicales nos pueden hacer sentir realmente una emoción o si nos encontramos ante un proceso de reconocimiento de la emoción sin que necesariamente ésta deba sentirse. Para tratar de responder a esta pregunta debemos primeramente revisar los modelos que han relacionado música y emoción.

Modelos emocionales y música

Uno de los primeros modelos de la emoción fue planteado por Wundt (1874) con una teoría tridimensional de los sentimientos. Según Wundt, los sentimientos varían en función de tres dimensiones: placer-displacer, excitación-apaciguamiento y tensión-alivio. Wundt afirmaba que “... *cada sentimiento concreto puede pertenecer a todas estas dimensiones, o únicamente a dos, o incluso tan sólo a una de ellas*” (Wundt, 1874, págs. 91-92). No obstante, el desarrollo actual de estas ideas mediante técnicas estadísticas actuales (entre ellas, los modelos factoriales y de escalamiento multidimensional), ha conllevado la eliminación de la tercera dimensión propuesta por Wundt (tensión-alivio) manteniendo como variables primarias de la emoción la excitación-apaciguamiento (o activación) y el placer-displacer (o valencia emocional) dando lugar a los actuales modelos bidimensionales de la emoción.

Uno de estos modelos es el llamado “Modelo Circunflejo de la Emoción” de Russell (1980). Este modelo plantea que todos los estímulos afectivos pueden ser caracterizados como una combinación de la activación y la valencia emocional (Feldman-Barret, 1999, 2004). La *valencia emocional* se concibe en el modelo circunflejo de la emoción como una dimensión hedónica donde un estímulo es evaluado como placentero o no placentero, mientras que la *activación* es definida como la excitación que produce dicho estímulo en el organismo (Feldman-Barret, 2004). La combinación de estas evaluaciones primarias del estímulo (valencia y activación) da lugar a una posición ordenada de las emociones dentro de un sistema de ejes cartesianos 2-D (bidimensional) formando una representación circular de las emociones. Según este modelo, altos niveles en activación y placer producirán emociones de alegría y excitación, mientras que bajos niveles de activación y displacer conducirían a las emociones de tristeza y aburrimiento (véase la Figura 2).

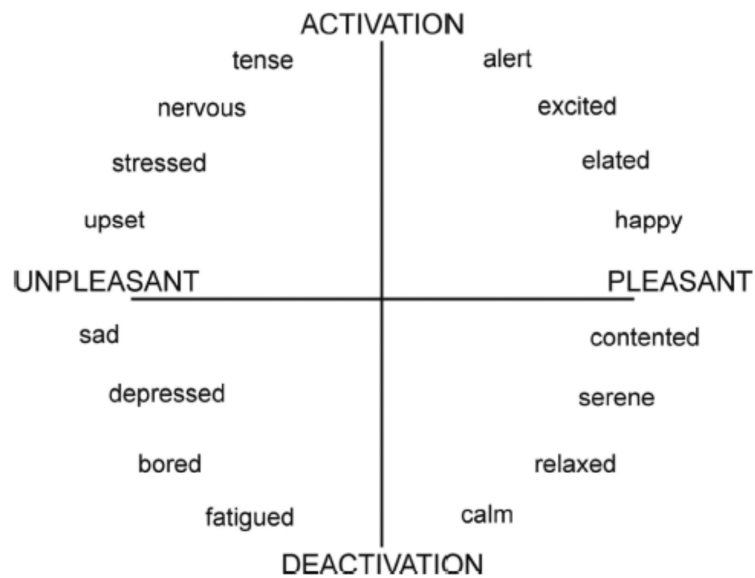


Figura 2. Modelo Circunflejo de la Emoción de Russell (1980)

El modelo emocional que propone Russell se encuadra dentro de los llamados modelos dimensionales de la emoción que plantean que cualquier emoción puede ser definida a partir de unas pocas dimensiones más básicas. Russell emplea en su modelo dos dimensiones semejantes a las dimensiones de “placer-displacer” y “excitación-apaciguamiento” de Wundt, excluyendo la tercera dimensión. Esta desaparición de la tercera dimensión del modelo no responde a una falta de evidencia empírica sobre la misma sino a un criterio económico: Russell observó en su modelo que emergían tres dimensiones, sin embargo, bastaban dos para poder situar todas las emociones discretas en el plano (Russell, 1980).

La mayoría de los estudios actuales que investigan la reacción afectiva a la música se basan en el modelo dimensional de Russell (v.g., Bigand, Vieillard, Madurell, Marozeau y Dacquet, 2005b; Gosselin, Peretz, Johnsen y Adolphs, 2007; Gosselin, Peretz, Noulhiane, Hasbound, Beckett, Baulac y Samsom, 2005; Peretz, Gagnon y Bouchard, 1998; Peretz y Gagnon, 1999; Vieillard, Peretz, Gosselin, Khalfa, Gagnon y Bouchard, 2008). En estos estudios es frecuente centrarse únicamente en cuatro emociones básicas y contrapuestas, fácilmente identificables por las personas, como son la alegría o la tristeza (v.g., Gagnon y Peretz 2003; Khalfa, Roy, Rainville, Dalla Bella y Peretz, 2008; Schellenberg, Peretz y Vieillard, 2008). Así por ejemplo, observando la Figura 2, vemos que la alegría y la tristeza distan entre sí 180° considerándose emociones contrapuestas: la alegría se definiría por una valencia emocional positiva y altos niveles de activación mientras que la tristeza se caracteriza por presentar una evaluación emocional negativa y bajos niveles de activación. También es frecuente en estas investigaciones trabajar con las emociones ortogonales (distancia angular de 90°) de las emociones básicas de alegría y tristeza. En este sentido, las emociones que distan 90° tanto de la alegría como de la tristeza son justamente las emociones de miedo/tensión y calma/tranquilidad. Las emociones que distan entre sí 90 grados son complementarias en una dimensión y contrapuestas en otra (Feldman-Barret, 2004). Así por

ejemplo, la alegría y la tranquilidad son emociones contrapuestas en la dimensión de activación, pero ambas son emociones placenteras; sin embargo, las emociones de tristeza y tranquilidad son emociones contrapuestas en su valencia emocional pero ambas emociones son poco activadoras.

Hay que señalar que el modelo reflejado en la Figura 2 se correspondería con una ordenación perfecta y armónica de las emociones donde las emociones contrapuestas como la alegría y la tristeza distan entre sí exactamente 180°. Sin embargo, esta disposición no es exacta a nivel práctico en todos los individuos, pudiéndose observar estructuras más cercanas a la elipse o con irregularidades. El origen de estas desigualdades en la forma del circunflejo vendría dado según Feldman-Barret (2004) por la focalización diferencial en la valencia emocional o en la activación. La focalización en la valencia emocional está relacionada con una mayor sensibilidad (o peso) otorgada a la evaluación hedónica del estímulo en el continuo de placer-displacer; mientras que una focalización en la activación estaría más relacionada con una sensibilidad interoceptiva (Feldman-Barret, Quigley, Bliss-Moureau y Aronson, 2004). Es decir, según este postulado, podemos encontrar diferencias individuales en la evaluación emocional de los estímulos musicales dependiendo de si el oyente focaliza más su evaluación en el plano hedónico o en el plano activador.

En el estudio del efecto emocional de la música sobre el oyente, el modelo circunflejo de la emoción de Russell (1980) ha sido usado en multitud de trabajos (Bigan y cols. 2005; Gosselin y cols., 2007; Gosselin y cols. 2005; Khalfa, Peretz, Blondin y Manon 2002; Vieillard y cols., 2008). Así por ejemplo, Bigand, y cols. (2005) evaluaron mediante la técnica de análisis multidimensional¹ la categorización emocional de un conjunto de piezas musicales a partir de la valencia emocional y la activación que suscitaban las piezas musicales en los

¹El escalamiento multidimensional es una técnica de análisis de datos que permite estudiar a partir de una matriz de similitud (o disimilitud) entre variables observables el número de dimensiones de la misma.

oyentes (véase la Figura 3). En una serie de tres experimentos, los oyentes debían escuchar un conjunto de piezas musicales facilitadas por los experimentadores. Una vez escuchadas todas las piezas musicales se pedía a los oyentes que focalizaran su atención en las sensaciones que les producían cada pieza y que las clasificaran en función de la similitud con la emoción que les despertaba. De este modo, los oyentes tenían que agrupar en una misma categoría aquellas piezas que les resultaran estimulantes y placenteras (alegría), o estimulantes pero no placenteras (miedo-tensión). Este experimento se repitió con músicos expertos y con voluntarios sin experiencia musical observando que la categorización de los estímulos era semejante en ambos grupos; en otras palabras, comprobaron que el juicio emocional de la pieza musical no se relacionaba con la experiencia musical del oyente o con su formación sino que, aparentemente, era una respuesta genérica, que no depende de la experiencia musical formal (posiblemente sea una respuesta universal pero los autores no variaron la cultura para poder realizar esta afirmación). También estudiaron la estabilidad en el juicio emocional evaluando la respuesta de los oyentes una semana más tarde observando que ésta mantenía una elevada consistencia temporal²: los oyentes clasificaban los estímulos de forma similar en el segundo ensayo en relación a como los habían situado en el primer ensayo.

En la línea de este trabajo, Vieillard y cols. (2008) estudiaron la respuesta emocional a estímulos musicales de un grupo de 45 participantes también desde la perspectiva teórica del modelo bidimensional o circunflejo. Estos autores presentaron un total de 56 piezas musicales que inducían estados de ánimo de alegría, miedo/tensión, tranquilidad y tristeza. En un primer experimento, pidieron a un conjunto de oyentes que categorizaran el conjunto de estímulos dentro de una de las cuatro emociones propuestas: Alegría, Tristeza, Miedo y Tranquilidad.

²La consistencia temporal de la emoción que genera el estímulo es un tema controvertido. que viene determinado también por factores como la habituación, la personalidad o la sobrexposición.

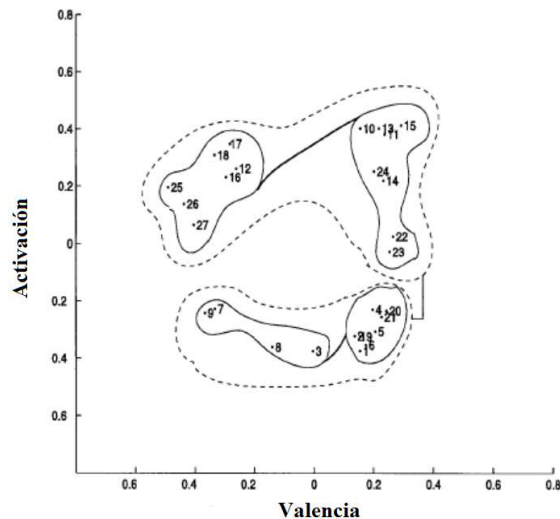


Figura 3: Clasificación de las piezas musicales en el trabajo de Bigand y cols. (2005). En esta figura vemos un plano de dos dimensiones (activación y valencia emocional) sobre el cual podemos ver numerados la posición de cada uno de los estímulos presentados. En el mismo se aprecia como hay un conjunto de estímulos que los oyentes evalúan como muy activadores pero poco placenteros (grupo superior izquierdo) mientras que otro conjunto presentaba también una alta activación pero resultaban placenteros a los oyentes (grupo superior derecho). Otros dos grupos de estímulos se situaban por debajo del valor cero de activación (resultando poco activadores para los oyentes) y diferían principalmente en su valencia emocional. Así, podemos ver que un conjunto se situaba por debajo del valor cero en la valencia emocional (displacer) mientras que el otro grupo se situaba por encima de dicho valor (placer).

Observaron que el 99% de los estímulos que habían sido compuestos por el autor³ con la intención de transmitir alegría eran categorizados efectivamente por los oyentes como alegres. El 84% de las piezas compuestas para transmitir tristeza y el 82% de las piezas compuestas para transmitir miedo eran correctamente clasificadas por los oyentes en estas categorías emocionales.

³El conjunto de estímulos era totalmente novedoso para el oyente. Todas las piezas musicales fueron escritas e interpretadas para piano y tenían una duración aproximada de 16 segundos.

Tabla 1: Porcentajes de acuerdo en la categorización emocional entre los oyentes del primer grupo de Vieillard y cols. (2008) y la intención del compositor de estas piezas musicales

	Respuesta				
	Felicidad	Tristeza	Amenaza	Tranquilidad	Ambivalente
Felicidad	99	0	0	0	1
Tristeza	0	84	1	7	8
Amenaza	5	8	82	0	5
Tranquilidad	8	12	1	67	12

Solo se produjo una ambivalencia en la categorización entre los estímulos con intención de transmitir tranquilidad: únicamente el 67% de los estímulos fue categorizado como “tranquilo” por los oyentes. Si nos fijamos en la Tabla 1 observamos que algunos de los estímulos tranquilos eran confundidos como alegres (8%) o tristes (12%). Es decir, se confundían principalmente con estímulos que distan 90° de la emoción de tranquilidad según el modelo dimensional de Russell, o lo que es lo mismo, los estímulos que debían transmitir tranquilidad variaban en el juicio emocional sobre la activación que producían (alegría) o en su valencia emocional (tristeza). Exceptuando esta discrepancia, es importante señalar la consistencia en la evaluación entre los oyentes pero también en la consistencia entre la intención comunicativa del autor y la recepción del mensaje por el oyente.

En un segundo experimento, estos mismos autores pidieron a otro conjunto de oyentes que evaluaran los mismos estímulos pero ahora en sus dimensiones emocionales. En esta tarea los oyentes no debían clasificar los estímulos dentro de una categoría emocional sino que debían puntuar cada uno en una escala de 0 a 9 los activadores que le parecían por un lado y la valencia emocional de los mismo por otro.

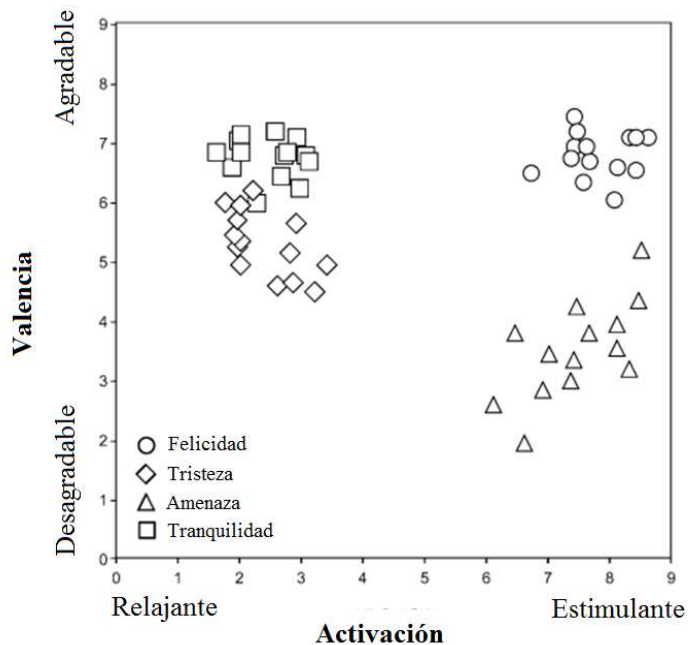


Figura 4: Posición de los estímulos musicales en el experimento de Viellard y cols. (2008). En esta figura podemos ver la valoración de cada estímulo musical en las dimensiones de valencia emocional y activación del segundo grupo de los oyentes, y que previamente había categorizado el primer grupo de oyentes de su experimento. Podemos apreciar por ejemplo, que los estímulos que habían sido categorizados como alegres por el primer grupo eran valorados por los oyentes como activadores y placenteros, mientras que los estímulos categorizados como “Scary” por los oyentes del primer grupo eran valorados como no placenteros pero activadores.

En la Figura 4 podemos observar el resultado de este experimento. Los estímulos que habían sido clasificados por los oyentes del primer grupo como alegres fueron valorados por los oyentes del segundo experimento, como activadores y placenteros, mientras que los estímulos valorados como activadores y poco placenteros en el segundo grupo pertenecían al grupo de estímulos categorizados en la emoción de miedo por el primer grupo. Este experimento demostró dos aspectos muy importantes en el estudio de la emoción elicitada por la música; por un lado, la consistencia en la evaluación entre oyentes, y por otro, la validez del modelo dimensional de Russell para el estudio de la emoción y la música. Sin embargo

también desvelaron una cuestión interesante y poco considerada en estos estudios y es el grado de placer que generan en el oyente los estímulos clasificados como tristes. Si nos fijamos nuevamente en la Figura 4, aunque el primer grupo categorizó correctamente el 84% de estos estímulos, la valencia emocional asignada por el segundo grupo a estos estímulos osciló entre 4 y 6 por término medio lo cual podría considerarse “relativamente placentero”. Es decir, en la investigación aplicada se observa que los estímulos que son categorizados como tristes por los oyentes no suelen ser considerados por los mismos como especialmente placenteros.

El estudio psicofisiológico de la respuesta emocional

En el apartado anterior hemos visto dos ejemplos de investigaciones en los que se exploraba la categorización emocional de estímulos musicales por los oyentes. La característica principal de estos estudios es que en ellos se pide al oyente un juicio cognitivo sobre el mismo; es decir, queremos conocer el producto de su interpretación de la pieza musical. Sin embargo, tal y como hemos visto previamente, este enfoque ha generado controversia porque no demuestra que el oyente esté sintiendo realmente la emoción sobre la que categoriza el estímulo. Por este motivo, durante la última década del s. XX, uno de los focos de atención en el estudio de la música fue demostrar que la respuesta en que se basan los oyentes para formular sus juicios es realmente de naturaleza emocional. Para ello, algunos investigadores se propusieron demostrar que existía una respuesta objetiva de naturaleza fisiológica y semejante a la encontrada en otras modalidades sensoriales como respuesta a la música y que esta respuesta era congruente con la respuesta cognitiva del oyente.

En este apartado estudiaremos en primer lugar los enfoques que se han usado para evaluar la naturaleza de la respuesta emocional con el objetivo de conocer que parámetros definen el concepto de emoción. Seguidamente analizaremos diferentes investigaciones en las

que se han empleado el estudio de la respuesta fisiológica para demostrar que en el contexto de la escucha musical, ésta es de naturaleza emocional. Finalizaremos el apartado estudiando cuales son los determinantes del estímulo que desencadenan una respuesta emocional ante la escucha musical.

Enfoques en el estudio de la respuesta psicofisiológica de la emoción.

El estudio de la emoción humana desde la Psicofisiología se centra, tal y como señala Carretié (2009), en describir los patrones de actividad autonómica, muscular y cerebral que se producen tanto en la experiencia emocional como en el procesamiento de la información emocional. Tradicionalmente podemos dividir los resultados obtenidos en esta área atendiendo a los dos enfoques principales desde los que se ha estudiado la emoción: el enfoque discreto y el enfoque dimensional. En este apartado, analizaremos no las características de estos enfoques sino los hallazgos encontrados en la psicofisiología de la respuesta emocional.

En los inicios del estudio de la psicofisiología de la respuesta emocional, el enfoque desde el que se estudiaba la emoción era desde el **enfoque discreto**. Se entendía cada emoción como una entidad propia y el objetivo desde la psicofisiología era encontrar el patrón biológico que desencadenaba cada una de las emociones o, por lo menos, de las emociones básicas. Es decir, se entendía que para cada emoción el organismo desencadenaría una respuesta diferenciada. Desde el enfoque discreto se atomiza el mundo emocional listando las emociones en grupos más o menos numerosos. Muchos de los autores se centran en estudiar las denominadas “emociones básicas” o “emociones primarias” como son el miedo, la alegría o la tristeza; estas emociones son universales en la especie humana, por tanto, no son dependientes de la cultura de individuo. En otro nivel, estarían las emociones secundarias cuyo desarrollo tiene un fuerte componente educacional y cultural, y suelen ser fruto de la

combinación de las emociones primarias. La primera teoría discreta de la emoción fue descrita por James–Lange. Estos autores propusieron que la experiencia psicológica de las emociones estaba determinada por cambios periféricos desencadenados por un estímulo emocional.

El estudio del patrón individual característico de cada emoción se aborda desde dos tipos de medidas objetivas: el estudio de la musculatura (especialmente la expresión facial de los músculos corrugador y cigomático) y el estudio de la respuesta autónoma. El objetivo del estudio de la **expresión facial** de las emociones es identificar el patrón muscular asociado y propio de cada emoción. Para comprender este enfoque se parte de la hipótesis de la universalidad de las emociones: si las emociones básicas son innatas, entonces la forma de expresarlas deberá ser la misma en todos los individuos independientemente de los factores culturales propios para cada uno. Por ejemplo, la emoción de alegría presentaría bajo este enfoque una expresión facial característica en la que entre otros grupos musculares se produce la contracción del músculo cigomático (que permite el movimiento de determinados grupos musculares alrededor de las comisuras de los labios reflejando la sonrisa), mientras que en la emoción de tristeza observaríamos, por ejemplo, la contracción del músculo corrugador de las cejas que permite fruncir el entrecejo. La sonrisa y el fruncimiento del ceño serían rasgos característicos de estas emociones y según la hipótesis de la universalidad de las emociones básicas, **no inherentes** a la cultura del individuo (véase la Figura 5).

La investigación sobre la universalidad de las emociones ha mostrado que el reconocimiento y la expresión facial de las mismas (al menos de las básicas), es universal. Esto supone que, independientemente de la cultura del individuo, no solo la expresión de la alegría o la tristeza serán semejantes a la nuestra sino que podremos reconocerla. También mediante el estudio evolutivo se ha demostrado que la respuesta facial a una emoción

determinada es consistente durante todo el ciclo vital, es decir, la expresión facial de las emociones básicas es semejante en la infancia y durante la vejez

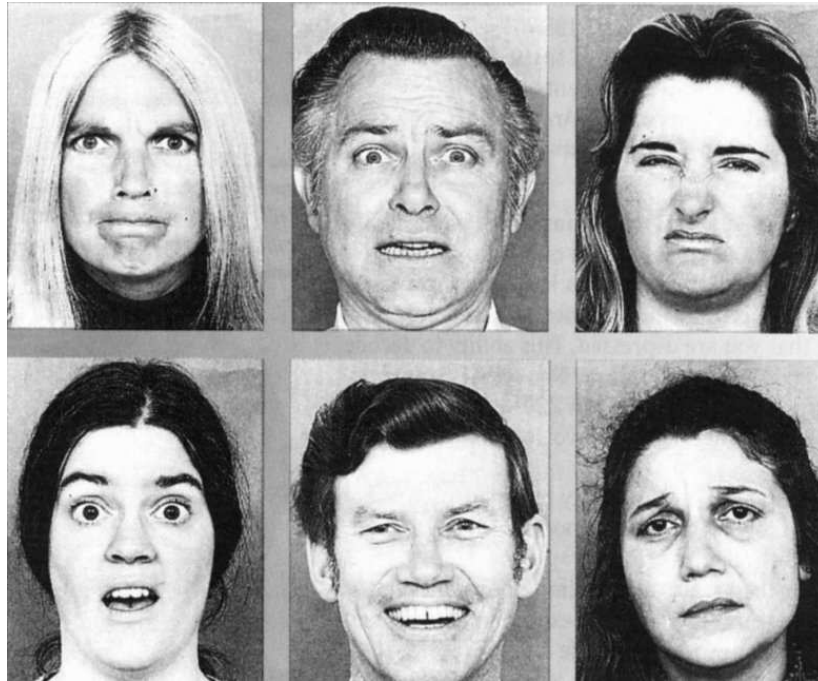


Figura 5: Ejemplo de láminas para el estudio del reconocimiento facial de emociones de Ekman (1976). Ekman elaboró un método para estudiar la universalidad en el reconocimiento de las emociones mediante la expresión facial que consistía en la presentación de una serie de rostros en las que se expresaban diferentes emociones. En estas imágenes de muestra se representan las expresiones faciales de enfado, miedo, asco, sorpresa, alegría y tristeza. Los trabajos de Ekman demostraron que la expresión de algunas emociones básicas como las señaladas trasciende el factor cultural. En el ámbito de la música, Fritz y cols. (2009) usaron este mismo método para que los oyentes de un pueblo indígena (los Mafas) con poco contacto con la cultura occidental indicaran la emoción que les suscitaba una selección de piezas musicales. Los Mafas escuchaban cada pieza musical y debían señalar la imagen que se ajustaba con la emoción que sentían.

Tal y como hemos señalado, otra área de estudio de la respuesta emocional es mediante el registro de la **respuesta autónoma** (sistema nervioso autónomo). Las medidas autónomas más frecuentemente usadas han sido tradicionalmente la actividad electrodermal de la piel y la respuesta cardiovascular, siendo usadas en menor medida las medidas

respiratorias. En el terreno de la música también observamos que la mayoría de los estudios que evalúan la respuesta autónoma se centran en las citadas medidas aunque también encontramos otras como la temperatura corporal y las medidas hormonales.

Desde el enfoque discreto de la emoción el objetivo es describir el patrón específico de respuesta autónoma a cada emoción. Si bien es cierto que determinadas reacciones corporales como las náuseas o el rubor se dan ante emociones concretas como el asco o la vergüenza, los resultados distan mucho de haber identificado un patrón de respuesta autónoma unívoco a cada emoción (Carretié, 2009). Aunque autores como Ekman y cols. (1983) o Wagner (1989) justifican el fracaso a la hora de establecer patrones claros de respuesta unívocos a cada emoción al enfoque metodológico utilizado⁴, lo cierto es que solo podemos esperar una especificidad fisiológica única a cada emoción si y solo si cada emoción se asocia a un patrón motor específico siendo esto contrario al pragmatismo biológico de nuestro organismo. El principio de economía sería contrario al enfoque discreto de la emoción ya que un patrón único para cada emoción sería algo poco práctico y poco económico en términos de adaptabilidad (Norris, Gollan, Bernston y Cacioppo, 2010), por el contrario el enfoque dimensional de la emoción ofrece una visión más pragmática

El **enfoque dimensional** da una vuelta al enfoque discreto planteando la existencia de determinadas dimensiones (como la activación y la valencia emocional que propone Russell) que, en combinación, dan lugar a las distintas emociones discretas. Este enfoque supone una revolución con respecto a sus predecesores ya que permite captar matices entre las emociones. Desde esta perspectiva el objetivo es estudiar la respuesta autónoma a las dimensiones y de ahí inferir la emoción que suscita el estímulo. Si en el enfoque discreto los pilares eran el estudio de la musculatura facial y la respuesta autónoma, con el auge de las técnicas de

⁴ Estos autores plantean que las diferencias en la respuesta fisiológica a la emoción se originan porque en los diferentes trabajos que han abordado este tema se usan diferentes paradigmas y diferentes instrumentos y métodos para registrar la respuesta fisiológica.

neuroimagen en la investigación psicológica, el enfoque dimensional se va a caracterizar por dos líneas: seguir estudiando la actividad periférica (respuesta autónoma) pero, a la vez, introducir el estudio de la actividad cerebral (áreas que se activan y conexión entre las mismas).

Sin embargo, independientemente del auge de los enfoques dimensionales, el estudio de la respuesta autónoma a emociones discretas ha seguido desarrollándose paralelamente a los planteamientos dimensionales y a sus críticas. Por ejemplo, Kreibig (2010) cuestiona las posturas dimensionalistas de Feldman-Barrett (2006) sobre el modelo de Russell (1980) y sus objeciones al enfoque discreto siendo estas la ausencia de patrones específicos asociados a cada emoción (como hemos señalado previamente). Kreibig (2010), además sitúa en un punto medio las posturas que plantean la existencia de un patrón de respuesta diferencial pero sólo para las emociones negativas y otro para las emociones positivas, tal y como plantea el modelo del espacio evaluativo de Norris y cols. (2010) que es precisamente la principal crítica al modelo bidimensional de Russell (1980) y que se refleja en el estudio de Vieillard y cols. (2008) en la valoración en el plano hedónico de los estímulos tristes.

Naturaleza de la respuesta emocional a la música: Congruencia entre la categorización emocional y la respuesta fisiológica del oyente

La dimensión de activación

Los trabajos sobre el efecto de la música en el oyente se han centrado tradicionalmente en su potencial para activar o desactivar el organismo. En general, la revisión de estos trabajos pone de manifiesto que tanto a nivel de respuesta cognitiva (evaluada mediante autoinformes)

como en el registro psicofisiológico, aquellos estímulos que suscitan en el oyente emociones de alegría o miedo/tensión generan una mayor activación que los estímulos que inducen emociones de tristeza o tranquilidad. El potencial activador/desactivador del estímulo musical se ha mostrado (como veremos a continuación) como una respuesta consistente inter e intra sujetos. La investigación plantea, por tanto, que la música puede generar en el oyente una respuesta de excitación fisiológica acompañada de una evaluación cognitiva del estímulo como activador y de la sensación subjetiva del oyente como de activación. Del mismo modo se ha comprobado que otros estímulos generan en el oyente una respuesta desactivadora de la respuesta autónoma acompañada de una sensación de relajación. Un aspecto interesante de estas investigaciones es la consistencia entre oyentes en la respuesta fisiológica y cognitiva ante la escucha de estímulos concretos. Salvando las diferencias individuales, en general hay cierto consenso entre oyentes en los juicios emocionales sobre piezas musicales y su categorización, pero además, se han encontrado patrones similares de respuesta entre oyentes a los mismos estímulos.

Uno de los trabajos pioneros en el estudio de la congruencia entre la respuesta fisiológica del oyente ante la música y su juicio cognitivo fue llevado a cabo por Krumhansl (1997). Esta autora evaluó en una muestra de 40 participantes sus respuestas psicofisiológicas de activación (medidas mediante la frecuencia cardiaca, la presión arterial en las extremidades superiores y la oreja, la variación en la conductancia eléctrica, la temperatura periférica y el ritmo respiratorio) ante la escucha de un *set* de piezas musicales. La primera conclusión relevante en este estudio fue que la presentación de los estímulos generaba un cambio en las medidas de activación con respecto a la línea base, es decir, la música generaba un efecto activador en los participantes. Además, se detectaron patrones de activación diferentes en función de la emoción asociada a cada estímulo. De esta forma, los estímulos tristes se caracterizaban por mayores intervalos temporales entre ciclos cardiacos, es decir, por una

frecuencia cardiaca más lenta, mientras que los estímulos de miedo y alegría se caracterizaban por intervalos temporales más rápidos entre los citados ciclos cardiacos. Krumhansl halló que existía una congruencia entre el juicio cognitivo y la respuesta fisiológica desde un enfoque dimensional de la emoción: ante emociones que según el modelo de Russell son más activadoras (alegría y tensión), la respuesta fisiológica del organismo es una respuesta de excitación, mientras que ante emociones que bajo este de la emoción presentan como característica una baja activación (*v.g.*, tristeza) la respuesta del organismo es la desactivación del sistema autónomo.

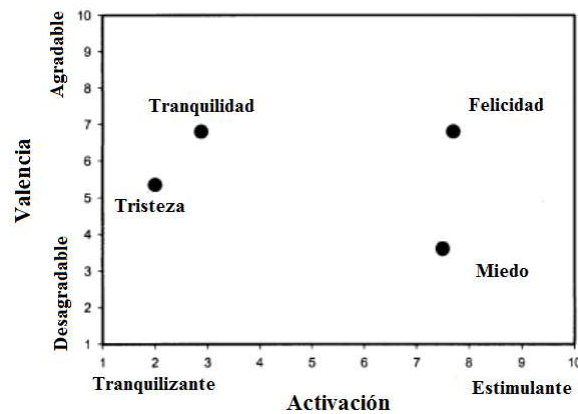


Figura 6: Valores promedio de activación y valencia emocional de los extractos musicales usados en el trabajo de Khalfa y cols. (2002). Observe que el juicio hedónico (valencia emocional) promedio de los estímulos categorizados como tristes por los oyentes vuelve a situarse en torno al valor 5 como posteriormente (y con otros estímulos) observarían Vieillard y cols. (2008) y que comentamos en apartados anteriores.

Siguiendo la línea de Krumhansl (1997), Khalfa y cols. (2002) evaluaron la respuesta de conductancia eléctrica de la piel ante otro *set* de estímulos musicales. Estos estímulos musicales estaban asociados a una emoción que podía ser: alegría, tranquilidad, tristeza o miedo/tensión. Se registraba la respuesta de conductancia eléctrica de la piel de cada oyente

y ante cada una de las piezas musicales seleccionadas por los investigadores. Los oyentes debían valorar posteriormente la activación que les producía cada uno de estos estímulo y en qué categoría emocional situarían cada una de las piezas (véase la Figura 6).

En la Figura 6 podemos comprobar que los estímulos más activadores fueron los categorizados como alegres o inductores de miedo/tensión, mientras que los estímulos categorizados como tristes o tranquilos eran valorados por los oyentes como menos activadores. A nivel fisiológico observaron un aumento significativamente mayor de la conductancia eléctrica de la piel ante estímulos alegres y de miedo/tensión en comparación con los estímulos asociados a las emociones de tranquilidad y tristeza. Es decir, a medida que los estímulos variaban en el continuo de activación su respuesta de conductancia era más o menos intensa. Sin embargo, no se encontraron diferencias cuando los estímulos variaban en su dimensión hedónica: los estímulos alegres presentaban una intensidad similar en la respuesta de conductancia eléctrica que los estímulos categorizados como de miedo o tensos, mientras que los estímulos tristes presentaban una intensidad de respuesta de conductancia similar la de los estímulos calificados como tranquilos.

Sin embargo este trabajo no mostraba que la música pudiera modificar el estado de activación fisiológica y cognitiva del oyente. Habían comprobado que efectivamente hay una respuesta de excitación ante estímulos activadores y una respuesta significativamente menos intensa cuando al oyente se le presentan estímulos menos activadores (v.g., tristes o tranquilos) pero no que un estado de excitación previo pudiera ser modificado por la exposición del oyente a piezas musicales desactivadoras. Esta hipótesis fue explorada también por Khalfa, Bella, Roy, Peretz y Lupien (2003).

En su trabajo expusieron a los participantes a una tarea de estrés social⁵; tras la misma, algunos participantes se recuperaron en una habitación en silencio mientras que otros eran expuestos a música relajante. Estos autores hallaron que la exposición a música relajante tras una tarea inductora de estrés psicológico reducía los niveles de cortisol⁶ en saliva más rápidamente que una condición de silencio.

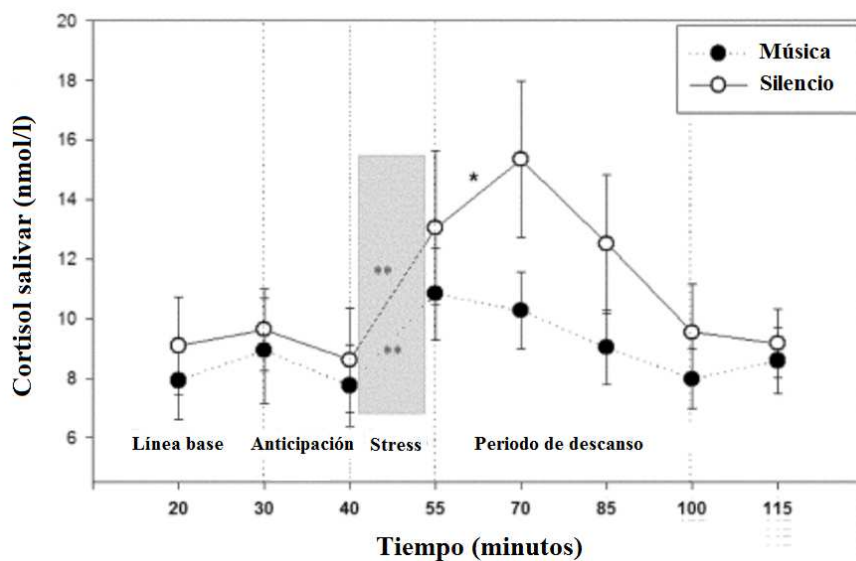


Figura 7: Resultados del trabajo de Khalfa y cols. (2003).

En la figura 7 podemos observar incrementos en los niveles de cortisol en saliva hasta 15 minutos después de la finalización de la tarea estresora entre los participantes cuya recuperación se hizo en la condición de silencio. Por el contrario, los niveles de cortisol en saliva comienzan a reducirse desde el inicio de la exposición a música relajante en el segundo

⁵ La tarea de estrés social consistía en una breve exposición de los participantes ante una audiencia y, a continuación, la verbalización en público de la solución a una serie de tareas aritméticas. El estrés social se genera como consecuencia de la realización de una tarea en la que la persona se siente evaluada y juzgada por un público. Este tipo de tareas generan una respuesta de estrés en el participante donde uno de los parámetros medibles son los niveles hormonales como el cortisol.

⁶ El cortisol es una hormona secretada por el organismo como respuesta a situaciones de estrés prolongado.

grupo. Es decir, tras una tarea generadora de estrés se observó que una exposición a música relajante permite al organismo recuperarse más rápidamente que en una condición de silencio.

Lamentablemente, la respuesta fisiológica registrada en los diferentes trabajos dista aún de ser consistente encontrando numerosos estudios en los que o no se halla el mismo efecto o se detectan patrones de respuesta diferentes. Por ejemplo, Etzel, Jhonsen, Dickerson, Tranel y Adolphs (2006) quisieron comprobar si existía un patrón diferente de respuesta cardiovascular y respiratoria como respuesta a la presentación de piezas musicales con diferente categorización emocional (alegría, tristeza y miedo).

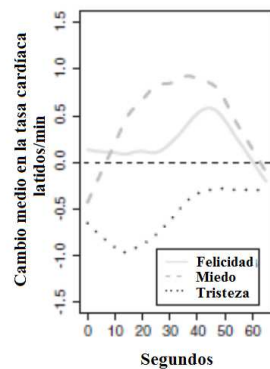


Figura 8: Variación de la frecuencia cardiaca entre las diferentes categorías emocionales a lo largo del tiempo.

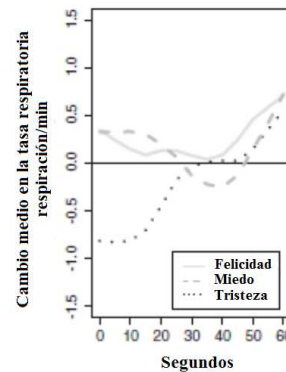


Figura 9: Variación del ciclo respiratorio entre las diferentes categorías emocionales a lo largo del tiempo.

La Figura 8 representa los cambios en la frecuencia cardiaca en tres diferentes categorías emocionales. Los autores señalan que a simple vista parece que los estímulos alegres generan ciclos cardiacos de menor duración en comparación con los ciclos (más lentos y de mayor duración) que se observa ante las piezas musicales tristes, sin embargo esta diferencia no resultó estadísticamente significativa. La Figura 9 representa las variaciones en el ciclo respiratorio durante la escucha musical la cual según los autores tampoco se mostró diferente entre las categorías emocionales; a pesar de que parece que hasta el segundo 30 la frecuencia cardiaca es inferior para los estímulos tristes que para los estímulos alegres e inductores de miedo, esta diferencia no mencionan que se sea significativa

La hipótesis que se plantearon estos autores es que si, como concluyeron Khalifa y cols. (2002), los estímulos más activadores (alegría y miedo) generan una respuesta de

activación autónoma más intensa que ante estímulos poco activadores (tristeza y tranquilidad) entonces esta activación/desactivación debería reflejarse en otros parámetros de la respuesta autónoma como la respuesta cardiovascular o respiratoria. Por tanto, esperaban que los estímulos musicales más activadores produjeran ciclos respiratorios y cardiovasculares más cortos y sucesivamente más rápidos que los estímulos menos activadores. Sin embargo los resultados apuntaron hacia conclusiones opuestas. Tanto la frecuencia cardiaca (Figura 8) como la duración de los ciclos respiratorios (Figura 9) fueron semejantes tanto entre los estímulos más activadores como entre los menos activadores. Únicamente encontraron medidas estadísticamente diferentes en las medidas subjetivas de autoinforme.

Khalifa y cols. (2008), a partir de las conclusiones de Etzel y cols. (2006) y de su estudio previo (Khalifa y cols. 2003), se propusieron analizar el patrón completo de respuesta autónoma en estímulos musicales tristes y alegres. En esa ocasión incluyeron como medida de la activación fisiológica la respuesta de conductancia eléctrica de la piel, la presión sanguínea, la frecuencia cardiaca y respiratoria y la respuesta del músculo zigomático facial. Una vez más, la medida subjetiva mostró un alto porcentaje de clasificaciones correctas del estímulo musical: los participantes identificaban correctamente los estímulos alegres como pertenecientes a la emoción alegre y los estímulos tristes a la emoción de tristeza. Los resultados mostraron valores estadísticamente superiores en la respuesta de conductancia eléctrica de la piel ante los estímulos alegres frente a los estímulos tristes tras 30 segundos de exposición.

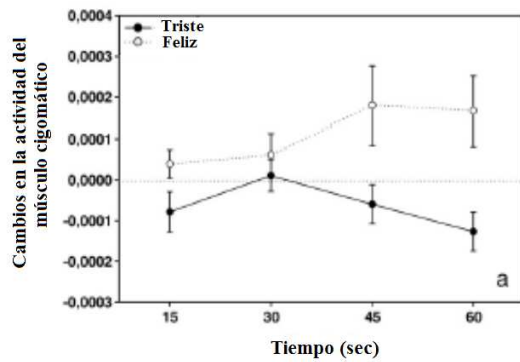


Figura 10: Respuesta de músculo cigomático ante piezas tristes vs. alegres.

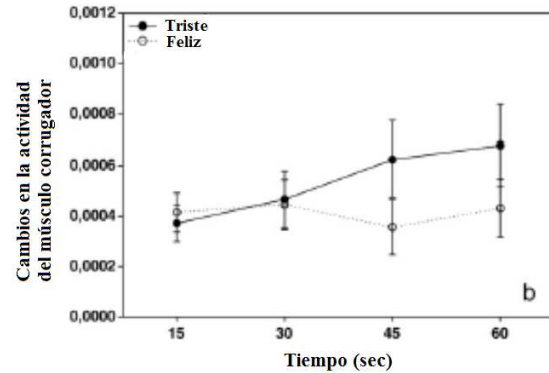


Figura 11: Respuesta de músculo corrugador de las cejas ante piezas tristes vs. alegres.

Estas diferencias se encontraron también a los 30 segundos en la respuesta del músculo cigomático observando una mayor actividad ante los estímulos alegres que ante los tristes (véase la Figura 10). Sin embargo en esta ocasión no se encontraron diferencias en la actividad del músculo corrugador (veáse figura 11).

En cuanto a las medidas cardiovasculares, Khalfa y cols. (2008) hallaron que los estímulos alegres incrementaban la frecuencia cardiaca en mayor medida que los estímulos tristes pero esta diferencia tampoco fue estadísticamente significativa en la línea de los resultados de Etzel y cols. (2006). Sin embargo la presión sanguínea diastólica entre los estímulos alegres y tristes a partir de los 60 segundos de exposición sí que resultó mayor entre los estímulos más activadores (véase la Figura 12).

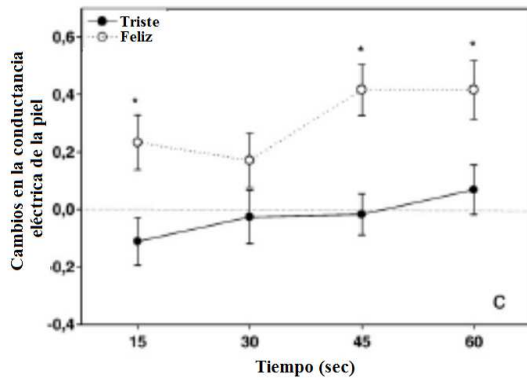


Figura 12: Cambios en la respuesta de conductancia eléctrica de la piel a lo largo del tiempo como respuesta a los estímulos alegres vs. tristes.

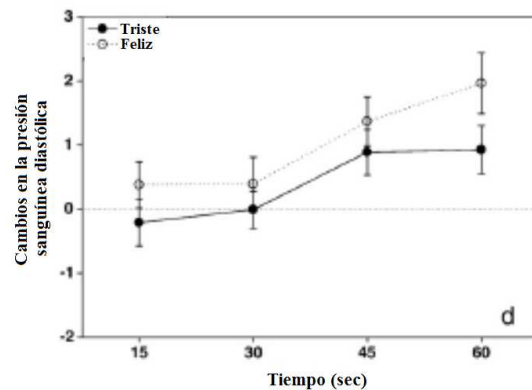


Figura 13: Cambios en la presión sanguínea diastólica a lo largo del tiempo como respuesta a los estímulos alegres vs. tristes.

Existen numerosos trabajos que siguen la línea de los aquí citados y que demuestran que la categoría emocional en la que el sujeto clasifica una pieza musical a la que es expuesto, es congruente con una respuesta de activación/desactivación del sistema nervioso autónomo. Asimismo, estos trabajos muestran un patrón de respuesta fisiológica congruente con la evaluación del estímulo musical en las dimensiones de la emoción. De forma que los oyentes que evalúan un estímulo como activador (independientemente de su valencia emocional) también generan una respuesta fisiológica de activación ante ese estímulo. Esto nos permite discriminar estímulos más o menos activadores en función de la respuesta fisiológica del oyente. Sin embargo, esto no es tan evidente cuando tratamos el patrón fisiológico de la valencia emocional del estímulo.

La dimensión de valencia emocional

En cuanto a la valencia emocional del estímulo musical, la investigación ha mostrado una congruencia en el juicio cognitivo: en términos generales los estímulos alegres y tranquilos son descritos como más placenteros que los tristes. Algunos de estos trabajos se

han centrado en la evaluación fisiológica de la respuesta de placer ante estímulos musicales clasificados en diferentes categorías emocionales. Uno de estos trabajos es el de Nater, Abbruzzese, Krebs y Ehlert (2006) quienes compararon las medidas de conductancia eléctrica de la piel, la frecuencia cardíaca y la temperatura corporal ante estímulos placenteros y no placenteros. Estos autores observaron un incremento de la frecuencia cardíaca y de la respuesta de conductancia eléctrica significativamente superiores en los estímulos no placenteros frente a los placenteros. En cuanto a la temperatura corporal, se observaron valores significativamente inferiores en la música no placentera que en la placentera. Por ello, según estos autores los estímulos con valencia emocional negativa generan en el oyente una respuesta de activación del sistema autónomo de mayor intensidad que los estímulos calificados por los oyentes como más placenteros.

En otro estudio, Roy, Mailhot, Gosselin, Paquette y Peretz (2009) analizaron la diferencia en el reflejo de sobresalto ante estímulos musicales que variaban en su valencia emocional (estímulos placenteros frente a estímulos no placenteros). Dado que en estudios previos habían podido describir un patrón de activación autónoma coherente con la activación auto-informada por los oyentes, en este nuevo trabajo se propusieron evaluar la existencia de un patrón semejante con la dimensión de valencia emocional pudiendo distinguir un patrón de activación autónoma para la valencia emocional negativa y otro para la valencia emocional positiva. A diferencia de los trabajos anteriores, Roy y cols. (2009) controlaron la activación suscitada por cada estímulo en el oyente de forma que tanto el grupo de estímulos musicales con valencia emocional positiva como el grupo con valencia emocional negativa tenían el mismo promedio en la dimensión ortogonal de activación. Esto permitió poder atribuir las diferencias en la respuesta autónoma a la valencia emocional del estímulo, independientemente de la activación.

A parte de las medidas autónomas ya citadas (respuesta de conductancia eléctrica de la piel y respuesta cardiovascular) y de la electromiografía del músculo corrugador y cigomático, estudiaron también la respuesta de sobresalto (*acoustic startle blink reflex*). Esta respuesta consiste en un movimiento involuntario y defensivo del oyente ante un estímulo no placentero (incluso desagradable) que desencadena un parpadeo; cuanto más frecuente y rápido es el parpadeo, más intensa es por tanto esta respuesta. Por tanto, la amplitud de la *acoustic startle blink reflex* (es decir, lo intensa que es la respuesta) debería ser mayor ante estímulos calificados por el oyente como menos placenteros, mientras que la latencia de la misma (o tiempo en el que tarda la persona en completar el movimiento de parpadeo) deberá ser mayor ante estímulos placenteros para el oyente. Como vemos en la Figura 14, los resultados confirmaron la hipótesis de los investigadores: la amplitud de la respuesta era mayor ante los estímulos calificados como no placenteros frente a los estímulos placenteros. Por el contrario, la latencia de la respuesta se mostró mayor entre los estímulos evaluados por el oyente como placenteros frente a los no placenteros.

En cuanto al estudio de la musculatura facial, el electromiograma mostró una activación superior en el músculo corrugador para los estímulos no placenteros. Sin embargo, aunque la actividad del músculo cigomático era mayor ante estímulos placenteros que ante estímulos no placenteros, esta diferencia no fue significativa (véase la Figura 13).

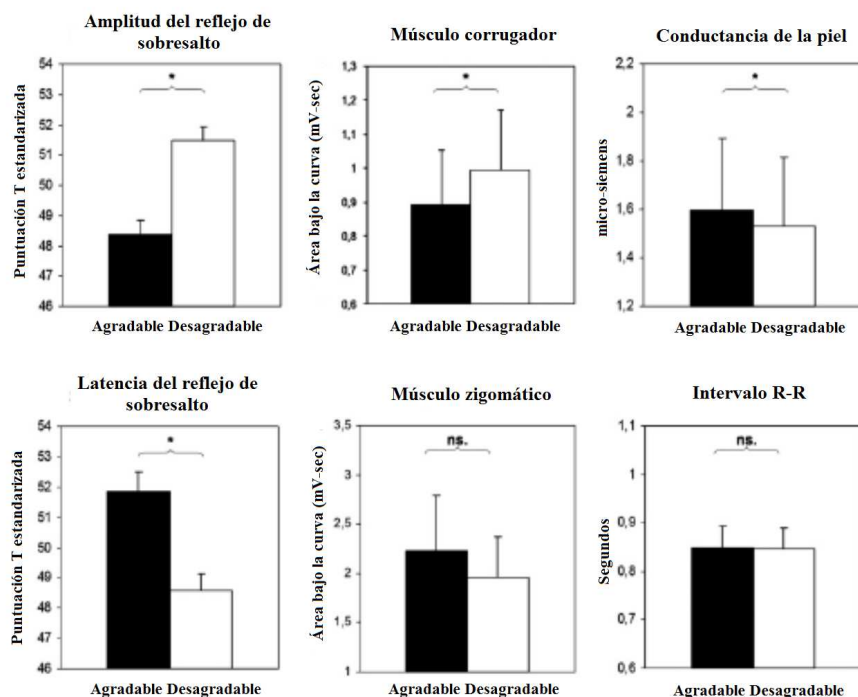


Figura 14: Resultados en el trabajo de Roy y cols. (2009) en las diferentes medidas fisiológicas consideradas en función del grado de placer que genera el estímulo en el oyente.

Al igual que en el trabajo de Nater y cols. (2006), la respuesta cardiovascular no se mostró discriminativa de la valencia emocional de los estímulos, sin embargo sí se observó un aumento significativo en el nivel de conductancia eléctrica de la piel para los estímulos placenteros, contradiciendo los resultados de Roy y cols. (2006). La posible razón de esta diferencia es que en el trabajo de Roy y cols. (2009) se controlaba la activación subjetiva de los estímulos mientras que en el trabajo de Nater y cols. (2006) no tenemos información sobre los niveles de activación subjetivos de los estímulos. A la vista de estos resultados podemos comprobar que las medidas que mejor rendimiento han producido a la hora de discriminar el juicio hedónico del oyente son la musculatura facial y el *acoustic startle blink reflex*.

Del conjunto de estos trabajos podemos concluir que la respuesta a la escucha musical es una respuesta de naturaleza emocional que puede ser medida bien desde un nivel

fisiológico o bien desde un nivel cognitivo. Hemos comprobado que esta respuesta es consistente entre personas y es consistente en el tiempo, y que, además, es similar a la encontrada en otras modalidades sensoriales. Sin embargo, aún no podemos comprender por qué existen resultados contradictorios entre algunos estudios: si en estos la metodología es similar (presentar la pieza musical y registrar la respuesta fisiológica y/o cognitiva) ¿por qué se encuentran resultados ligeramente diferentes ante estímulos encuadrados por el oyente en la misma categoría emocional? Para resolver esta cuestión algunos investigadores comenzaron a mirar hacia la pieza musical buscando elementos intrínsecos a la misma que influenciaran la respuesta emocional del oyente.

¿Qué desencadena la respuesta emocional del oyente?

En apartados anteriores hemos podido observar que la respuesta emocional a una pieza musical puede evaluarse desde un plano cognitivo pidiendo a los oyentes que autoinformen de la sensación que le produce la pieza musical o bien desde un plano fisiológico. A pesar de que la respuesta parece consistente en ambas áreas entre los oyentes de un mismo estudio, se observan diferencias importantes entre estudios. La investigación de esta discrepancia condujo la atención de los investigadores hacia un factor pobremente considerado en las sucesivas investigaciones y extrínseco al oyente; es decir, si el procedimiento era semejante y los oyentes categorizaban del mismo modo los estímulos ¿qué podía estar en la raíz de las diferencias? Precisamente lo que era diferente en cada uno de los estudios que hemos visto hasta el momento era el propio estímulo musical utilizado en cada estudio: cada uno de ellos había seleccionado su propio conjunto de piezas musicales sin atender a un concepto fundamental en psicología y es la importancia de la estructura y el control del estímulo.

A pesar de que el estudio del efecto psicológico de la estructura del estudio se remonta a principios del s. XX, uno de los primeros modelos que pretende explicar la relación entre la

estructura del estímulo y la respuesta emocional del oyente es el propuesto por Balkwill y Thompson (1999) tras una serie de estudios transculturales. Estos autores observaron que existía una congruencia en el juicio emocional a diferentes piezas musicales cuando los oyentes pertenecían a una misma cultura; es decir, oyentes occidentales categorizaban de forma similar piezas musicales occidentales. Ahora bien, cuando a estos mismos oyentes se les presentaban piezas musicales escritas en otras reglas tonales distintas de las occidentales, sus juicios emocionales eran también consistentes dentro del grupo pero ligeramente diferentes a los que realizaban los oyentes de la cultura particular del estudio. Concretamente en el trabajo de Fritz y cols. (2009) con la tribu de los Mafas, observaron que cuando se les pedía que juzgasen emocionalmente piezas musicales occidentales, éstos las categorizaban de forma similar a como lo hacían los oyentes occidentales pero con mayor grado de ambivalencia y menor intensidad en la respuesta. Según el modelo de redundancia de la señal de Balkwill y Thompson (1999), este fenómeno se explicaba porque a la hora de emitir un juicio emocional sobre el estímulo los oyentes reparan en elementos concretos de la estructura del estímulo. De este modo, cuando el oyente no está familiarizado con las reglas tonales del estímulo recurre a la métrica del estímulo para establecer un juicio emocional sobre el mismo (véase la Figura 15).

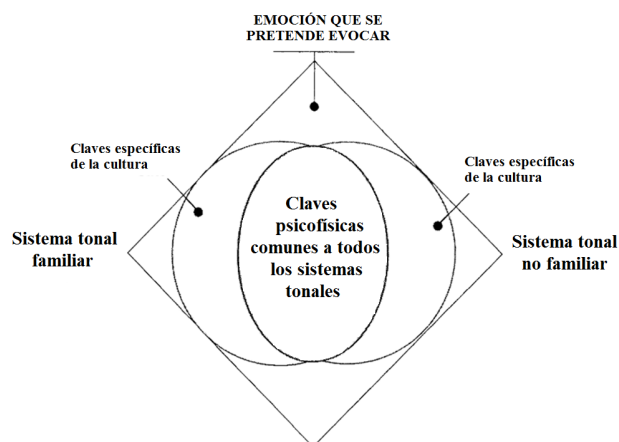
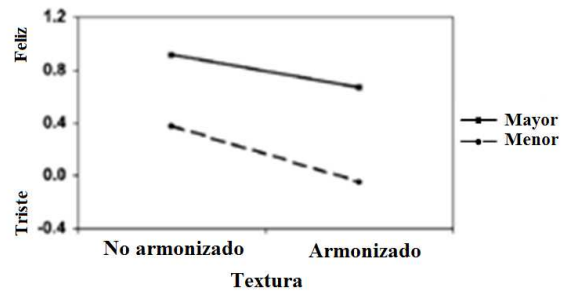
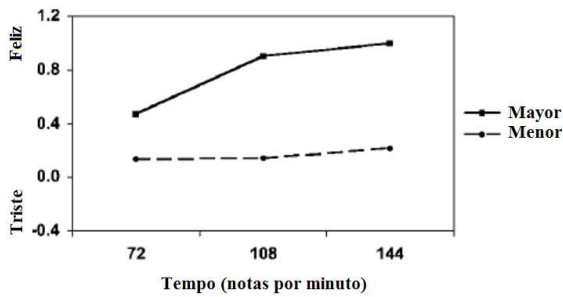


Figura 15: Modelo de redundancia de la señal de Balkwill y Thompson (1999).

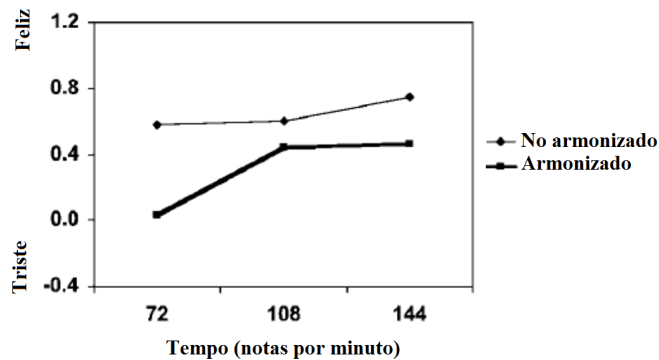
Posteriormente otros trabajos han ido intentando desvelar la relación entre estas rutas de procesamiento y la respuesta emocional. Por ejemplo, el trabajo de Husain, Thompson y Schellenberg (2002) encontró un efecto significativo de la manipulación del tempo y el modo de la sonata K.448 de Mozart sobre la activación y la valencia emocional. Estos autores observaron que cuando incrementaban el tempo de la pieza, aumentaba la valoración subjetiva de la activación entre los participantes mientras que las variaciones en el modo (mayor vs. menor) producían cambios en el afecto. Se observó que el modo mayor producía cambios hacia la valencia emocional positiva mientras que el modo menor producía cambios hacia la valencia emocional negativa. Es decir, la propiedad métrica del tempo musical afectaba de forma exclusiva a los juicios cognitivos sobre la activación suscitada por el estímulo mientras que la propiedad tonal del modo afectaba al juicio hedónico del oyente.

En otro trabajo, Webster y Weir (2005) analizaron la interacción entre el tempo, el modo y la textura sobre las medidas subjetivas de alegría y tristeza. A la hora de analizar los efectos, encontraron que a medida que se incrementaba el tempo de la pieza, se incrementaba también la valoración subjetiva del estímulo como alegre; asimismo, el modo mayor presentaba puntuaciones más altas en alegría que el modo menor (véase la Figura 16). También fue significativa la interacción estadística entre el tempo y el modo de forma que a medida que se incrementaba el tempo en los extractos escritos en modo mayor se incrementaban también las valoraciones subjetivas de alegría del extracto. Sin embargo, cuando el modo era menor, el incremento del tempo no afectaba a la valoración subjetiva del estímulo como alegre.



(a) Efecto del modo y el tempo sobre el juicio emocional

(b) Efecto del modo y la textura sobre el juicio emocional



(c) Interacción entre tempo y textura

Figura 16: Resultados de Webster y Weir (2005).

Estos autores observaron que cuando se incrementaba el tempo de una pieza musical, los oyentes calificaban la misma como más activadora; es decir, describen una relación lineal entre el tempo musical de la pieza y la intensidad emocional de la misma. Sin embargo observaron que este efecto se producía únicamente si la pieza estaba escrita en un modo mayor (a); por el contrario, la manipulación del tempo en piezas escritas en modo menor no modificaba lo alegre o triste que resultaba la pieza También hallaron otras interacciones relevantes entre la textura y el modo (b) o entre el tempo y la textura (c).

Otros estudios como por ejemplo en Dalla-Bella, Peretz, Rousseau y Gosselin (2001) hallaron el mismo efecto de interacción entre el tempo y el modo en la valoración del estímulo como triste o alegre. En el análisis de los efectos simples de esta interacción

encontraron también que los extractos en modo mayor y los tempos rápidos eran considerados como más alegres que los extractos escritos en modo menor.

Los trabajos sobre la influencia en la activación y la valencia emocional de otras propiedades métricas o tonales, sin embargo, son escasos. De ellos, cabe destacar el trabajo de Gomez y Danuser (2007) en el que evaluaron la influencia de diferentes propiedades del estímulo musical sobre la categorización emocional de dichos estímulos. Para este estudio, tomaron medidas subjetivas y objetivas de activación y valencia emocional. Estos autores evaluaron concretamente la influencia de propiedades rítmicas del estímulo como el tempo y el ritmo, propiedades armónicas como el modo y la amplitud tonal así como propiedades físicas tales como la intensidad del estímulo musical y su efecto sobre medidas fisiológicas clásicas como la respuesta de conductancia eléctrica de la piel, respuesta cardiaca y respiratoria, así como sobre la valoración subjetiva del estímulo.

Gomez y Danuser (2007) situaron las diferentes propiedades del estímulo musical sobre el cuadrante emocional de Russell en función de la activación y la valencia emocional que suscitaban los extractos. Observaron que los tempos rápidos y el ritmo se situaban especialmente en los cuadrantes de alta activación y valencia positiva y alta activación y valencia negativa; mientras que el modo menor y la amplitud tonal extensa se situaban preferentemente en el cuadrante de baja activación y valencia emocional negativa. En cuanto a las variables psicofisiológicas asociadas a las diferentes propiedades del estímulo musical, estos autores observaron que la frecuencia cardiaca estaba positivamente relacionada con el tempo, siendo mayor la frecuencia cardiaca a tempos rápidos, así como con el modo, encontrando una mayor frecuencia cardiaca ante estímulos en modo menor. Las variables positivamente relacionadas con los incrementos en la respuesta de conductancia eléctrica de la piel fueron por ejemplo el tempo, la amplitud tonal y la intensidad mientras que las

propiedades negativamente relacionadas con esta medida fueron el modo, el ritmo y el nivel tonal.

Una vez constatada la relación entre las propiedades estructurales del estímulo musical y la respuesta fisiológica y cognitiva del oyente, el siguiente paso en la investigación se centró en la construcción de modelos que permitieran simular estos procesos. Si bien los primeros modelos de clasificación de piezas musicales a partir de su estructura se sitúan en los últimos años del s. XX procedentes de la ingeniería informática, desde el campo de la psicología el objetivo de los mismos se centraba no en crear clasificadores óptimos sino en simular el proceso de categorización humano en sistemas de inteligencia artificial. Uno de estos primeros modelos es el propuesto por Korhonen (2004) desde la ingeniería que mediante un tipo de modelado llamado *Autoregressive Models with Exogenous Variables* o ARX. Este autor estudió el efecto que tenían diferentes propiedades físicas del estímulo musical sobre la valoración subjetiva y continua del estímulo en el tiempo. Como peculiaridad en el modelo, Korhonen (2004) no emplea medidas del estímulo extraídas de la teoría musical como el tempo o el modo. Por el contrario emplea propiedades físicas de la onda acústica (véase la Tabla 2).

Los resultados de su modelo mostraron que este conjunto de medidas físicas del estímulo musical permitía predecir de forma muy precisa las valoraciones subjetivas del oyente en la dimensión de activación. Sin embargo, el rendimiento a la hora de predecir la valencia emocional del oyente fue menos preciso.

Tabla 2: Medidas del estímulo musical usadas como entradas en el modelo de Korhonen (2004).

No.	Propiedad musical	Rasgo musical	Método de extracción
1	Dinámica	Volumen	PsySound
2		Volumen máximo a corto plazo	PsySound
3	Tono medio	Centroide de la potencia espectral	PsySound
4		Centroide medio STFT	MARSYAS
5	Variables del tono	Flujo medio STFT	MARSYAS
6		Desviación típica del flujo STFT	MARSYAS
7		Desviación típica del centroide STFT	MARSYAS
8	Timbre	Anchura del timbre	PsySound
9		Media de la caída STFT	MARSYAS
10		Desviación típica de la caída STFT	MARSYAS
11		Nitidez	PlaySound
12	Armonía	Disonancia espectral	PsySound
13		Disonancia espectral	PsySound
14		Disonancia tonal	PsySound
15		Disonancia tonal	PsySound
16		Tonalidad compleja	PsySound
17	Tempo	Notas por minuto	Método de Schuber
18	Textura	Multiplidad	PsySound

En la Figura 17 podemos encontrar a la derecha las valoraciones a lo largo del tiempo de cada uno de los oyentes del estudio (líneas gris claro) a un extracto del concierto de Aranjuez tanto en la dimensión de activación (gráfico superior) como en la dimensión de valencia emocional (grafico inferior). El conjunto de estas valoraciones se promedia en cada unidad de tiempo (por ejemplo cada segundo) obteniendo una valoración “global” de la pieza en cada dimensión (línea negra) a la cual se le añaden unos límites (líneas rojas) de

desviación. Desde la propuesta de Korhonen (2004), podemos considerar que un modelo se encuentra ajustado cuando la respuesta o salida que ofrece la simulación por ordenador devuelve una línea semejante al promedio encontrado entre los oyentes en el contexto experimental y dentro de los límites de error establecidos. En la imagen izquierda de la figura 17 podemos comprobar el ajuste que presentaba la simulación de Korhonen (2004) a los datos experimentales (líneas en verde).

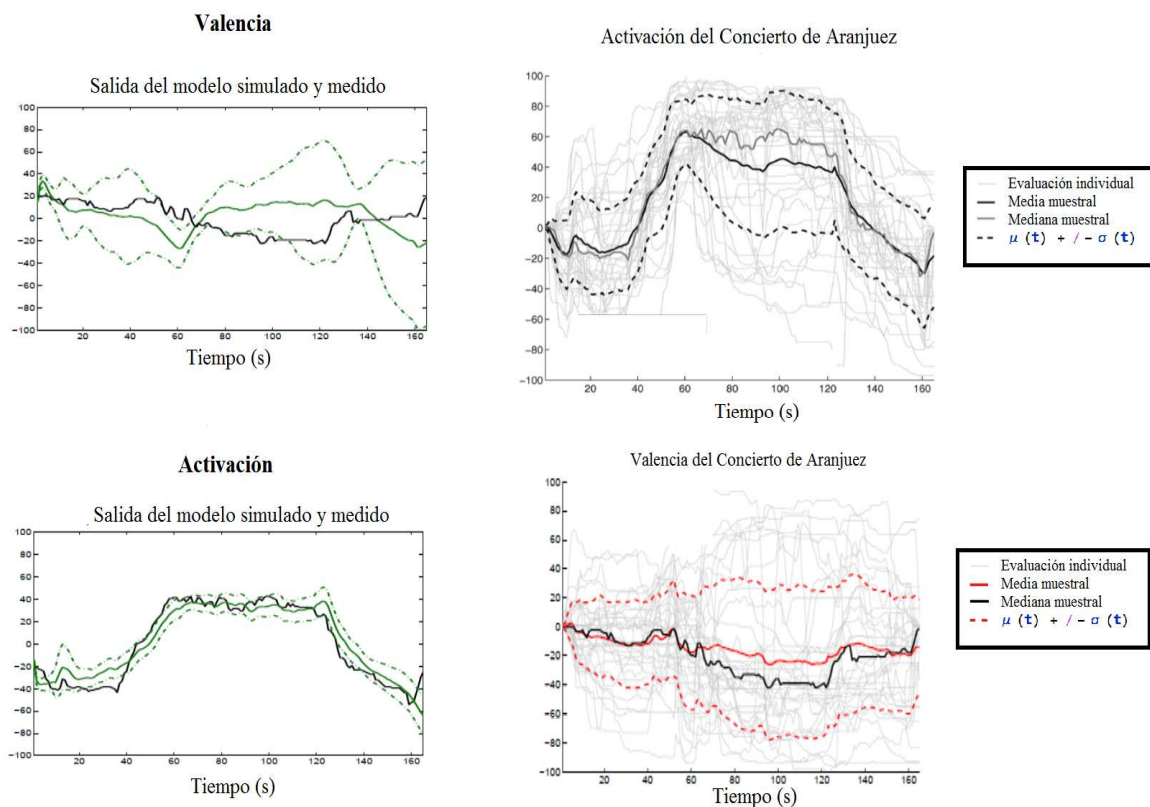


Figura 17: Salidas en el modelo de Korhonen (2004).

Podemos comprobar que en esta pieza musical tanto la simulación de la activación (la línea negra de la figura pertenece a los resultados de la simulación) como la valencia emocional se mantienen la mayor parte del tiempo dentro de los límites establecidos (líneas punteadas verdes). Sin embargo, pensamos que su trabajo presentaba dos aspectos cuestionables. Por un lado, este tipo de modelos se encuadran dentro de los modelos lineales

lo cual implicaba considerar *a priori* que la relación entre las propiedades tonales y métricas del estímulo musical y las valoraciones subjetivas de activación y valencia emocional eran también lineales. Por otro lado, la introducción de múltiples medidas físicas del estímulo en el modelo que definían diferentes propiedades del estímulo musical genera “redundancia” en las variables predictivas.

Siguiendo las recomendaciones de mejora de estos estudios y basándose en los datos experimentales de Korhonen (2004 y 2006), Coutinho y Cangelosi (2009) estudiaron las propiedades dinámicas (intensidad del sonido), armónicas (variación tonal, promedio tonal y timbre), de textura⁷ y métricas (tempo) sobre la valoración emocional de la pieza musical (véase la Figura 18). Estos autores plantearon un modelo de red neuronal artificial⁸ (RNA) basado en la arquitectura de los perceptrones multicapa con una unidad oculta. El objetivo de esta red era simular los valores de activación y valencia emocional obtenidos por el modelo matemático de Korhonen (2004) pero a partir de un tipo de modelado que no se limitase a la relación lineal entre los elementos y utilizando un menor número de medidas físicas para evitar la redundancia a que hemos aludido anteriormente.

⁷La textura musical es una propiedad del estímulo musical que evalúa la relación entre diferentes voces dentro de la pieza. Una melodía es simple o monódica si es ejecutada utilizando un único instrumento o voz o si varios instrumentos suenan como una única línea melódica. La melodía será compleja si la melodía está acompañada por múltiples instrumentos o voces en diferentes líneas melódicas.

⁸Los modelos RNA son modelos conexionistas que se basan en el sistema de procesamiento de las redes neuronales biológicas. Por tanto la estructura general consiste en una capa de entrada donde cada variable de entrada en el modelo (en este caso, cada propiedad del estímulo musical) funciona como una neurona que se encuentra conectada a una o varias neuronas de la siguiente capa, que en las arquitecturas más sencillas suele ser la capa de salida donde la red toma una decisión sobre la salida que va a emitir. En las arquitecturas más complejas existe una (o varias) capas ocultas entre las capas de entrada y salida, estas capas ocultas tienen la misión de transformar la señal que reciben de la capa inmediatamente anterior de la cual reciben información. Cada neurona está conectada con la capa suprayacente con una intensidad variable; si el efecto de la neurona input sobre otra neurona es pequeño, la fuerza de conexión entre ambas será pequeña y su aportación a la hora de determinar la salida en la red será poca. Esta cuantificación de la fuerza de conexión entre una neurona y otra en el siguiente nivel se denomina peso.

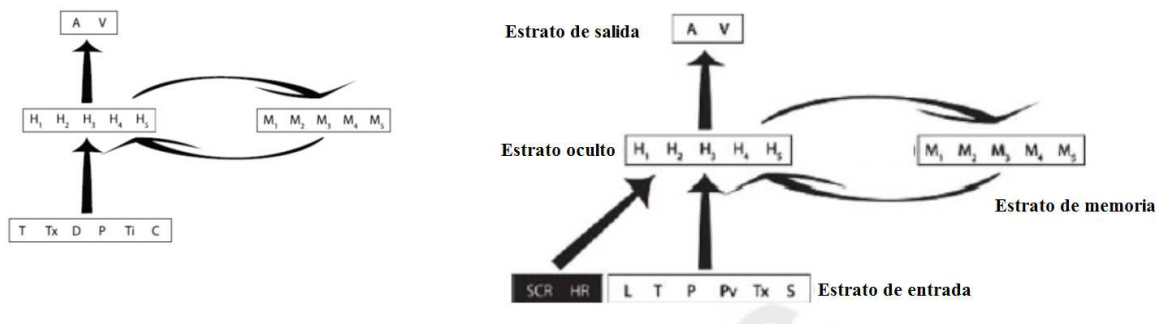


Figura 18: Modelo de Red Neuronal de Coutinho y Cangelosi (2009 y 2011).

Nota: Unidades de entrada: Rasgos del sonido (T, Tx, L, P, S y C) y variables fisiológicas (SC y HR); Unidades ocultas – H₁ a H₅; Unidades de memoria – M₁ a M₅; Unidades de salida: A (Activacion) y V (Valencia).

El primer modelo de Coutinho y Cangelosi (imagen izquierda en la Figura 18) permitía predecir la activación y la valencia emocional de los estímulos musicales con mayor precisión que el modelo de Korhonen, utilizando un menor número de parámetros y sin la restricción de los modelos lineales. En un segundo modelo quisieron introducir el efecto de las variables fisiológicas en la predicción de las dimensiones emocionales, sin embargo comprobaron que su aportación a la predicción apenas mejoraba el rendimiento del primer modelo que únicamente incluía como entradas las propiedades físicas del estímulo.

A pesar de que los sucesivos modelos de Coutinho y Cangelosi (2009 y 2011) permitían predecir con gran precisión las salidas de la “caja negra” de la emoción en el modelo neuropsicológico de Peretz y Coltheart, no explicaban que era lo que podía estar sucediendo en dicha caja negra. Si nos fijamos en la arquitectura de sus modelos (Figura 18), observamos que en el primer nivel de las ARN, todas las medidas físicas del estímulo se introducen de una sola vez dentro del modelo para predecir tanto la valencia emocional como la activación, sin saber que ocurre en ese nivel intermedio que corresponde a la capa oculta. Es decir, no nos permite conocer cuál es la relación particular entre cada una de las medidas

consideradas y las dimensiones emocionales. A nivel práctico es muy relevante este aspecto ya que solo conociendo que propiedades afectan a qué dimensión podemos manipularlas prediciendo los posibles efectos que pueden ocasionar en el oyente. Por ejemplo, si queremos reducir el nivel de activación de un oyente, con los modelos de Coutinho y Cangelosi no sabríamos que propiedad deberíamos manipular para reducir la activación manteniendo constante la valencia emocional en un punto concreto.

Objetivos

A la vista de los planteamientos expuestos en la introducción, entendemos que existen aún muchas cuestiones por resolver sobre la estructura y funcionamiento de la “caja negra” de la emoción descrita en el modelo neuropsicológico de Peretz y Coltheart (2003). Por ejemplo, sabemos que la información relativa al estímulo musical entra en esta caja a través de dos vías de procesamiento diferentes: una vía procesa la información temporal del estímulo, mientras que la otra procesa la información relativa al tono. Sin embargo, el modelo de Peretz y Coltheart no entra a explicar cómo es el procesamiento del estímulo a nivel emocional, ni si éste procesamiento se realiza también en paralelo. Es decir, si tal y como plantea el modelo de redundancia de la señal de Balkwill y Thompson (1999), las propiedades rítmicas afectan principalmente al juicio emocional sobre la activación que genera el estímulo en el oyente, y las propiedades relativas al tono, afectan principalmente al juicio sobre la valencia emocional.

Al respecto hemos señalado que modelos recientes como el de Coutinho y Cangelosi (2009 y 2011), asumen que el procesamiento emocional del estímulo es un procesamiento en bloque donde no se distingue estas rutas de procesamiento. Para estos autores la información entra en bloque en el sistema y se computa sin distinción de la ruta de procesamiento de la

que proceden. Este planteamiento aunque en cierta medida resulta parsimonioso, no se ajusta a los modelos teóricos expuestos. Por este motivo, uno de los objetivos de esta tesis se centrará en comprobar si la doble ruta de procesamiento es eficiente en términos de mejora de la predicción con respecto a los modelos anteriores.

Si nos basamos en el modelo de redundancia de la señal de Balkwill y Thompson (1999), entonces asumimos una concepción bidimensional del espacio emocional de acuerdo al modelo de la emoción de Russell (1980). Esta premisa tiene importantes implicaciones a nivel metodológico dado que asumimos por un lado que la emoción como categoría discreta puede ser definida a partir de dimensiones continuas, y por otro lado, que dos dimensiones, Activación y Valencia Emocional son suficientes para explicar dicho espacio emocional. Por tanto, otro de los objetivos de esta tesis es encontrar evidencias que sustenten que el modelo bidimensional de la emoción de Russell (1980) es un marco teórico práctico y parsimonioso para definir el concepto de emoción sin descartar la existencia de otras dimensiones para definir este espacio. Entendemos que el modelo bidimensional es eficiente respecto a otras alternativas, cuando nos permite poder predecir el juicio emocional discreto del oyente, y cuando estos juicios son consistentes y estables para un mismo estímulo en diferentes oyentes.

Otra cuestión que no viene reflejada en el modelo neuropsicológico de Peretz y Coltheart (2003) es el papel que juega el plano fisiológico en el procesamiento emocional del estímulo. A la hora de valorar cual es la respuesta fisiológica del oyente al estímulo musical, no se ha podido determinar un patrón concreto de respuesta a cada categoría emocional o a cada dimensión del modelo de Russell (1980). Es posible que este problema venga del intento de encontrar un patrón asociado a la emoción y no un patrón asociado al tipo de estimulación. Es decir, si tal y como plantean los trabajos citados previamente la estructura del estímulo

afecta al juicio emocional, es coherente pensar que también dicha estructura del estímulo pueda afectar a la respuesta fisiológica del oyente.

El modelo de Coutinho y Cangelosi (2011), concebía el papel de la respuesta fisiológica del oyente al estímulo como un input más su modelo. Sin embargo, el poder predictivo de esta respuesta era muy pequeño con respecto al papel que juegan las propiedades estructurales del estímulo en dicha predicción. Este pobre rendimiento de la respuesta fisiológica puede ser debido a que su papel es “eclipsado” por la colinealidad entre esta respuesta y la estructura del estímulo. Es decir, nos planteamos conocer si el papel que juega la respuesta fisiológica en la predicción de la respuesta emocional del oyente se trata de un papel mediador o de una respuesta paralela que también influye en el juicio emocional del oyente.

Tal y como hemos reflejado en la introducción, los modelos de Schubert (1999), Korhonen (2004 y 2006) y Coutinho y Cangelosi (2009 y 2011), introducen la necesidad de concebir el estímulo musical como un tipo de estímulo dinámico cuyas propiedades estructurales se modifican en el tiempo. Es decir, una pieza musical puede mantener constantes ciertas propiedades a lo largo de su interpretación, como por ejemplo el modo o el compás musical (aunque no necesariamente). Pero otras propiedades es más probable que no se mantengan constantes a lo largo de dicha interpretación, como los acentos, las alteraciones en las notas musicales, la propia tonalidad de las notas musicales, los acordes o la duración de cada nota musical, entre otras propiedades. Si tal y como plantean los trabajos expuestos, la estructura del estímulo juega un papel muy influyente en el juicio emocional del oyente y esta estructura es dinámica en el tiempo, es coherente pensar, que las respuesta del oyente tanto en un plano subjetivo como en un plano fisiológico pueda ser también variable en el dominio tiempo. Y no solo variable en este dominio, sino también que el estado previo del sistema

influya en el estado del mismo en el momento de la respuesta, tal y como proponen los modelos señalados. Por tanto es necesario tener en consideración el dominio tiempo en los modelos propuestos y de este modo, estudiar la respuesta continua del oyente a la estimulación musical.

El objetivo último de esta tesis es la comprensión de la “caja negra” de la emoción que citábamos al comienzo de esta sección. Para comprender el proceso que subyace a la respuesta emocional del oyente al estímulo musical debemos ser capaces de simular dicho proceso. Por tanto el último objetivo que nos proponemos es integrar todo este conocimiento en un modelo que sea capaz de simular la respuesta del oyente recogida de forma experimental. Con esto no nos proponemos ser más eficientes que el sistema de procesamiento humano, sino realmente simularlo con sus tasas de acierto y sus tasas de error. Para asegurarnos de que el modelo resultante sea realmente predictivo, tiene que ser capaz de dar respuesta a situaciones distintas de las que ha usado para aprender. Por tanto, nuestro modelo debe ser capaz de generalizar su respuesta a diferentes sets de estímulos.

En resumen, podemos enumerar los objetivos de esta tesis en:

Objetivo 1. Conocer si es más eficiente y tiene mayor capacidad predictiva una estructura de entrada de la información en paralelo que en serie.

Objetivo 2. Conocer si de acuerdo al modelo redundancia de la señal de Balkwill y Thompson (1999) las propiedades temporales afectan en mayor medida al juicio sobre la activación y las propiedades tonales al juicio sobre la valencia emocional del estímulo.

Objetivo 3. Conocer si el papel de la respuesta fisiológica en el juicio emocional al estímulo musical es un papel mediador entre el input y el juicio emocional.

Objetivo 4. Conocer si la concepción dinámica en el dominio tiempo del estímulo musical y la respuesta del oyente al mismo incrementa la precisión del modelo en la simulación de la respuesta experimental.

Objetivo 5. Construcción de un modelo que permita simular el sistema de procesamiento emocional del estímulo musical humano. Lo cual implica demostrar que nuestro modelo no es más o menos eficiente que los humanos en las tareas de categorización sino igual: con sus tasas de acierto y error.

Objetivo 6. Estudiar la capacidad de generalización de nuestros modelos sobre otros estímulos.

CAPÍTULO 2

Predicting emotional response to music through a compound neural network

Elena Saiz-Clar & José M. Reales

Universidad Nacional de Educación a Distancia (UNED)

Department of Science Behavior and Health Methodology

Saiz-Clar, E. & Reales, J.M. (2017). *Psychology of Music*, 1-17.

doi: 10.1177/0305735617705452

Abstract

The emotional effects of music have a cross-cultural component that can be explained through the tonal and non-tonal properties of musical pieces. To investigate the relationship between music and the emotions it arouses, we have built a composite neural network with the aim of predicting both the emotional categorization and the emotional valence and activation of Vieillard et al.'s (2008) musical stimuli. Our neural network uses two Adalines in the first level of the structure to predict activation and emotional valence from a minimal set of temporal and tonal properties of the stimuli (rhythm, tempo, time signature, mode, absolute tonal range and the frequency of the lowest note). In the second level, the network uses a Self Organizing Map (SOM) network to classify the stimuli into four emotional categories (calm, happiness, fear and sadness). The results have allowed us to replicate the features of the Circumplex Model of Emotion. The percentage of explained variance obtained for activation is satisfactory and higher than in previous research for emotional valence. The percentage of music pieces correctly classified by the SOM was also very high (87%). We discuss the results in relation to competing models of music and emotion.

Keywords

emotion, artificial intelligence, music, activation, valence

Introduction

From the beginning of psychology as a science, theorists have explored the factors that relate music with emotions. Hevner's experiments (1935, 1936, 1937) were a landmark on this subject. She concluded that variables such as tempo, rhythm and mode are related to the emotional judgement or categorization of music. From these early studies, the number of variables or factors that explain the emotional categorization of music has increased considerably. As an example, Scherer & Zener (2001) included structural elements of music as well as elements related to the motivation, expectations and mood of listeners and situational and performance variables. In a similar vein as linguistic models, these authors proposed a multiplicative model of music with a transmitter, a receiver, the context and the message. In this model, to provide emotional content to any musical fragment, all of these elements must be present simultaneously. More recently, other authors (e.g., Koelsch, 2015) have broadened the number of factors in this model including the social function of music.

The predominant model in studies of emotional response to music is the Circumplex Model of Emotion (Russell, 1980). This model proposes that all emotions can be represented according to two orthogonal dimensions, activation and emotional valence (Russell, 1978, 1979). Both of them are defined as basic yet independent components of emotion (Barrett & Russell, 1999). Activation refers to a psychological dimension that reflects the arousal brought forth by the music. Valence is the hedonic judgement people make about that music. While activation is a unipolar dimension that goes from null activation to high activation, valence is a bipolar dimension in that there are two poles, a negative and a positive.

However, there are some models that include a third dimension (tension), already postulated by Wundt in 1874. Although the goal of including this third emotion is to improve the accuracy of the modelling of emotion (Bigand, Vieillard, Madurell, Marozeau, &

Dacquet, 2005; Eerola, Lartillot, & Toiviainen, 2009; Eerola & Vuoskoski, 2010; Vieillard et al. 2008), the general conclusion reached by these authors is that the improvement in the accuracy achieved by including “tension” in the model is real but nonetheless very low. In other words, the improvement in the explained variance is not worth the introduction of a third dimension in the model. Therefore, as our first criterion was to keep the neural model as simple as possible, we decided to use a two-dimensional space composed only by activation and valence.

Numerous studies have used the circumplex model to inquire into the relationship between music and the emotions it evokes. For example, Vieillard et al. (2008) separately analysed the categorization of emotional stimuli from subjective ratings of emotional valence and activation. In this study, a group of listeners had to recognize the emotion that each of the chosen pieces intended to convey. A second group assessed how the stimuli made them feel the emotion into which each musical fragment had previously been categorized. Finally, a third group evaluated the activation and emotional valence of each stimulus in order to check whether these assessments corresponded to the values expected for each emotion according to Russell’s emotional model (1980). The results indicated that listeners were able to associate each of the stimuli with a specific emotion category with high inter-rater reliability. They noted that the emotional valence and activation of each stimulus judged by the third group of participants was consistent with the emotional judgements made by the first group. These results are consistent with those obtained in other studies (Bigan, Filipic, & Lalitte, 2005; Gosselin, Peretz, Johnsen, & Adolphs, 2007).

Also working within the circumplex model, other authors have explored the properties of musical stimuli that affect their emotional categorization. For example, Husain, Thompson and Schellenberg (2002) noted that an increase in the tempo of Mozart’s Sonata for Two

Pianos (K. 448) also increased the activation judgements of participants. However, variations in mode (major and minor) produced changes in valence, that is, major mode produced a trend towards a positive affect while minor mode produced a trend towards a negative affect. Husain et al. (2002) concluded that tempo has an effect on the subjective evaluation of activation but not on emotional valence. Meanwhile, mode has an effect on emotional valence but not on emotional evaluation.

Webster and Weir (2005) also analysed the effect of tempo and mode variations on emotional categorization. They found that increasing tempo also increased the classification of stimuli as “joyful” and that scores in major mode were evaluated as happier than those written in minor mode. In addition, they found a significant interaction between tempo and mode. Increasing tempo in scores written in major mode produced an increase in subjective evaluations of happiness of these scores, but it did not modify the emotional judgement of scores written in minor mode.

From this revised set of studies, we conclude that there appear to be at least two major dimensions in music perception that are reflected in models such as those put forward by Balkwill and Thompson (1999), Peretz and Coltheart (2005) and Lerdahl and Jackendoff (2003). On the one hand, these studies distinguish a temporal dimension that includes the position of musical notes in the musical sequence. These properties are related to the temporal location of events and their transitions and influence perceptual aspects such as the speed at which the piece is performed (tempo), the grouping of notes into measures and the generation of temporal expectations (Schulze, 1978, 1989; Quim & Watt, 2006; Vos, van Assen, & Franek, 1997). On the other hand, these studies also distinguish a dimension that refers to the tonal quality of the notes, that is, the frequency at which they are emitted and how they are organized within the musical piece (aspects that are usually grouped into melody and

harmony). In relation to Webster and Weir's (2005) and Husain et al.'s (2002) results, properties related to the position of events (temporal properties) appear to primarily influence activation, whereas properties related to tone (tonal or pitch-related properties) relate to the emotional valence of the stimulus. Therefore, we adhere to the idea prevalent in this field that there are two processing channels (tonal and temporal), each one related to activation and valence in the emotional model of Russell (1980), that can explain in a simple and direct way the results obtained by Hussain et al. (2002) and Webster and Weir (2005).

However, research on the emotional response to music has not been the exclusive domain of psychology. There has also been much interest on this topic in software engineering. Computer engineers have developed sophisticated algorithms to classify musical stimuli, for example, by genre. Usually, these algorithms include acoustic parameters as their main input (Shao, Wang, Li, & Ogihara, 2009; Tzanetakis & Cook, 2002). Other categories of algorithms try to categorize soundtracks by emotion from a variety of acoustic parameters of the stimuli (e.g., Yang, Lin, Cheng, & Chen, 2008; Yang, Liu, & Chen, 2006). Although these engineering models are very significant to psychology, they usually use many predictors, and this generates overfitting and redundancy (Korhonen, 2014). As their goal is to achieve the best classification possible, they do not try to remain simple. For this reason, we will confine our discussion to psychological models.

Among the previous studies that have attempted to model the relationship between stimulus parameters and emotional judgement, Korhonen's (2004) study must be mentioned. This author used statistical models to predict subjective ratings of emotional valence and activation of the musical stimulus based on its acoustic properties. To this end, Korhonen compared autoregressive models with exogenous variables (ARX) and state space models. The results indicated that the ARX technique obtained better fit indices. However, his work

had two conflicting aspects. On one hand, the ARX models establish linear relationships between the independent and dependent variables, which involves consideration of the relationship between the tonal and temporal properties of the musical stimulus and subjective assessments of emotional valence and activation as linear relationships. Furthermore, multiple variables were included in Korhonen's model that defined different properties of the musical stimulus, which generated high redundancy in the model. From his results, Korhonen's conclusions recommended using modelling techniques that do not require assuming *a priori* a linear relationship between the variables, citing the use of Artificial Neural Networks (ANNs) as an example.

Following these recommendations and based on Korhonen's (2004) experimental data, Coutinho and Cangelosi (2009) developed an ANN model based on the architecture of multilayer perceptrons that reproduced the activation and emotional valence values obtained experimentally by Korhonen (2004) based on dynamic (acoustic intensity), tonal (tonal variation, average tone and pitch) and temporal (tempo) properties. In Coutinho and Cangelosi's (2009, 2011) model, the properties of the musical stimulus were considered as inputs. The output layer was designed with two output nodes, one to represent activation and another to represent emotional valence. Between these layers, a single hidden layer was included. The number of units of this hidden layer was determined after the training phase. As was the case with Korhonen's (2004) model, good fit indices were obtained for activation values but a worse performance in predicting the values of emotional valence was also observed. Notwithstanding, Coutinho and Cangelosi's (2009 and 2011) model showed several shortcomings. First, separately identifying the effect of the dynamic, tonal and temporal properties on both emotional dimensions was not possible due to the structure of the ANN. On the other hand, both models (Coutinho & Cangelosi, 2009 and 2011; Korhonen, 2004) were used to obtain values of valence and activation, but they did not pursue the use of

processing to also obtain the emotional category linked to each stimulus. We have put to test a simple neural network that mainly focuses on solving these problems, while trying, at the same time, to attain a good fit to the empirical data. Our model also fits neatly with a subset of the neuropsychological model of Peretz and Coltheart (2005).

Objectives

First, we tried to build a simple but plausible ANN model that allows us to predict the subjective judgement of activation and valence of musical stimuli in people (see Figure 1). The structure of our ANN model was devised to mirror the coarse structure of the neuropsychological model of Peretz et al. (2003), as seen by comparing Figure 1 and Figure 2. Following Korhonen (2004), we selected an adaptative linear element (Adaline) architecture because, unlike the perceptron used by Coutinho and Cangelosi (2009, 2011), it allowed us to obtain continuous outputs required for predicting activation and emotional valence, thereby making them comparable to scores obtained in experimental studies. Following the results of Webster and Weir (2005) and Husain et al. (2002), this kind of output also allowed us to test the hypothesis that temporal properties of music mainly affect the activation dimension, while tonal properties primarily affect the valence dimension.

Second, we used output values on the activation and emotional valence in the ANN to build a selector switch able to reproduce the emotional categorization of the same musical stimuli in humans and compared the performance with that obtained by Vieillard et al. (2008). We employed a self-organizing map (SOM) network for this task because this type of ANN is trained using a learning algorithm that reproduces the human categorization process (Kallinen, 2005). In this way, we attempted to verify that the output of our network would allow us to rate both the emotional dimensions of Russell's model and the discrete category in which the participants assign a standard stimulus set.

In the first part of our simulations with the ANN model, we used the stimuli, scores and valence/emotional ratings used in the study by Vieillard et al. (2008). In the last part, we also studied the generalization skills of our model using another set of stimuli proposed by Schubert (1999). To this end, we contrasted the performance of our model with the results obtained in a behavioural experiment with this new set of stimuli.

Figure 1. Neural Network model used in our study

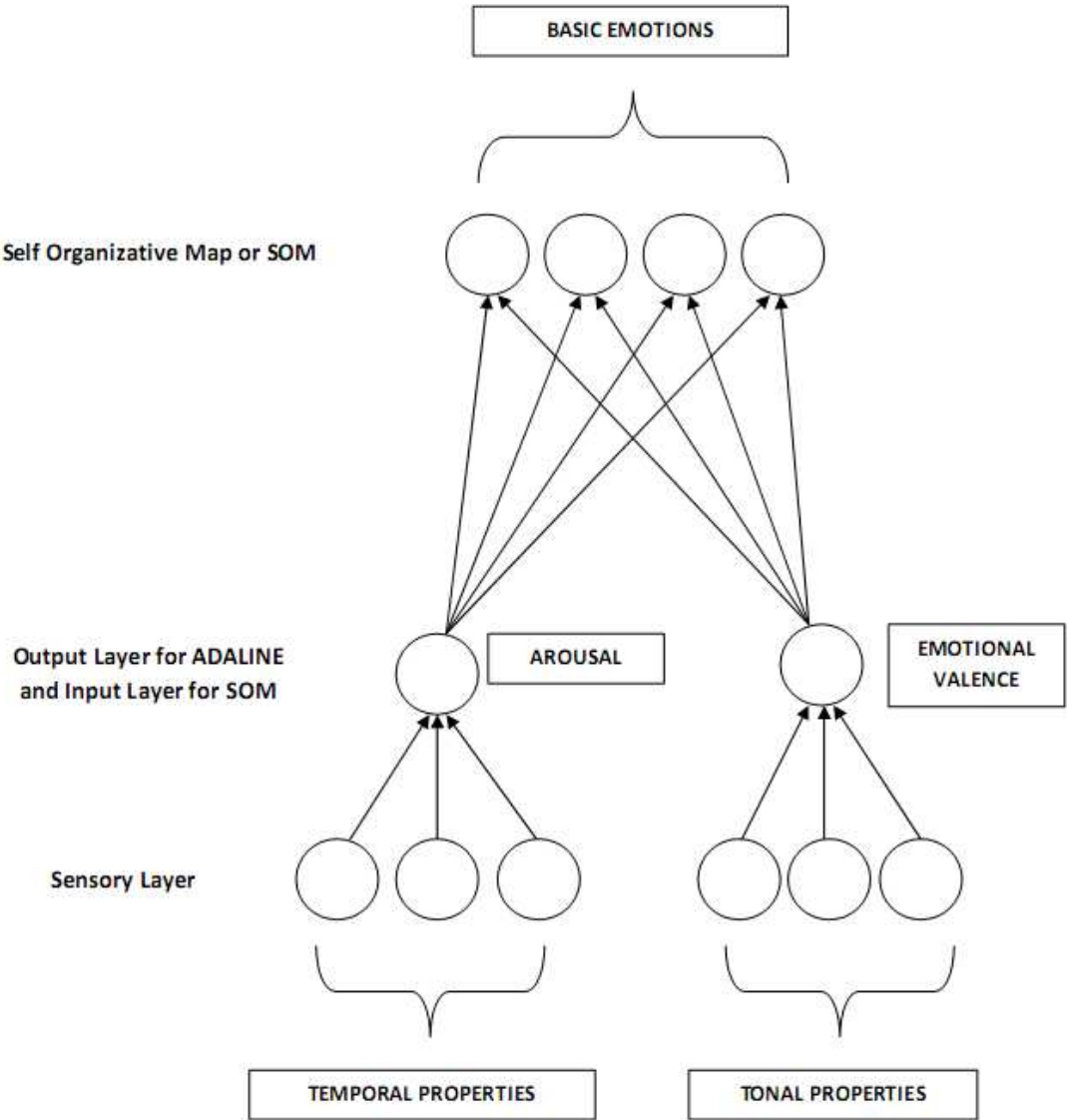
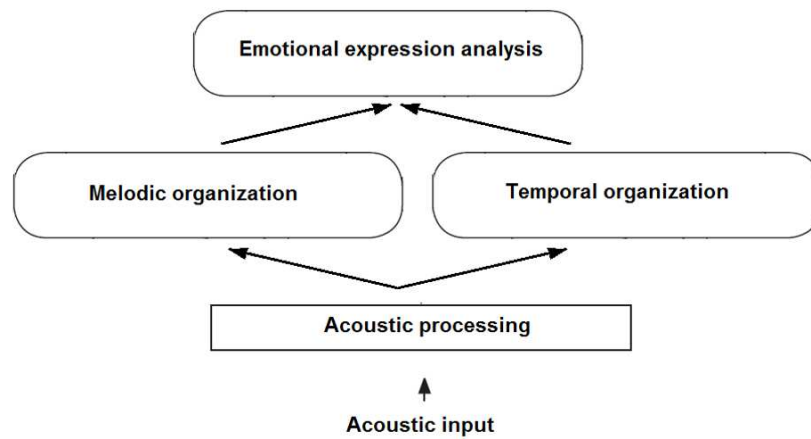


Figure 2. A schematic view of Peretz et al.'s (2003) neuropsychological model showing the structural similarity intended with our ANN model.



Method

Stimulus variables

We selected the tempo, the average number of notes per bar (ANNB) and the time signature as input variables for the activation Adaline and the mode, the absolute tonal range (ATR) and the frequency of the lowest note (FLN) as input variables for the valence Adaline. A more thorough explanation of these variables can be located in the online supplementary material.

Stimuli

The set of musical stimuli used in this study were the 40 musical stimuli provided in Vieillard et al.'s (2008) paper. Out of these, one-quarter were classified into emotional categories of calm, joy, fear or sadness (10 in each category). We computed the musical properties to use as the input in our network from the music scores provided by Vieillard et al.

(2008). Preliminary descriptive and inferential analyses of these properties were carried out in order to check whether they were valid for inclusion in our neural network. Table 1 shows the results in terms of emotion and the way they had been categorized. All of the pieces associated with the emotions of joy and calm were written in major mode, while the pieces associated with the emotions of fear and sadness were composed in minor key. Thus, splitting the valence (positive vs. negative) from the mode (major vs. minor, respectively) of the stimuli was not possible, which meant that the analysis of contingency could not be performed for this specific property.

Table 1. Descriptive and inferential musical properties of Vieillard et al. (2008) stimuli as a function of emotions elicited.

Property		Major mode		Minor mode		$t_{dof=38}$	d
		Calm	Joy	Fear	Sadness		
Time	3/4	6	1	3	8		
signature	4/4	4	9	7	2		
Tempo	M (SE)	71.40 (3.02)	126.50 (8.97)	99.50 (12.51)	49.20 (2.44)	2.13*	.674
	Min-Max	54 – 88	91 - 180	44 - 172	40 - 60		
ANNB	M (SE)	7.13 (.62)	10.81 (1.37)	4.85 (.70)	5.34 (.47)	4.12 ***	1.303
	Min-Max	5.20 – 10.67	5.63 - 16	1.88 – 8.50	3.25 – 8.67		
FLN	M (SE)	72.14 (2.55)	119.55 (12.12)	50.34 (2.41)	83.62 (7.66)	2.95 **	.933
	Min-Max	61.74 – 87.31	61.74 – 174.61	43.65 – 65.41	51.91 – 130.8		
ATR	M (SE)	911.8 (159.8)	869.9 (54.62)	673.2 (129.6)	540.9 (51.7)	2.63 *	.832
	Min-Max	185.2 – 2,027.5	637.1 – 1,222.3	23.0 – 1,269.5	284.5 – 806.6		

Note. M = Mean; SE = Standard Error; dof: degree of freedom; ANNB = Average Number of Notes per Bar;

FLN = Frequency of the Lowest Note; ATR = Absolute Tonal Range.

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$.

Table 1 also includes the overall rate of correct classifications obtained by Vieillard et al. (2008) in humans. The overall rates of correct classifications were obtained by averaging the percentage of correct classifications obtained in each stimulus for each emotion across participants, and they are presented for comparison with later results obtained in our network (Table 2).

Table 2. Percent correct, means and standard deviations for activation and emotional valence as a function of emotion, correlations between musical properties and emotional dimensions and Adaline's weight for emotional dimension.

	Emotional dimension		Correct classification
	Activation	Valence	
Joy	7.95 (.56)	6.79 (.35)	99%
Calm	2.38 (.53)	6.72 (.27)	67%
Fear	7.37 (.82)	5.37 (.96)	82%
Sadness	2.34 (.56)	3.53 (.50)	84%

	<i>r</i>	<i>w_{ij}</i>	<i>r</i>	<i>w_{ij}</i>
ANNB	.29*	.076	.52***	
Tempo	.76***	.832	.14	
Time signature	.42**	.183	.00	
ATR	.09		.35*	.235
Mode	.09		.88***	.682
FLN	.18		.49***	.044
Bias unit		.008		.031

Note. ANNB = Average Number of Notes per Bar; FLN = Frequency of the Lowest Note; ATR = Absolute Tonal Range.

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$.

Architecture of the Artificial Neural Network

The Adaline networks were designed so that, in the first phase of training, the values of activation and emotional valence associated with each stimulus could be predicted from the set of basic properties extracted. Subsequently, in a second phase, the SOM network was designed to correctly classify each musical stimulus into the emotional category to which it belonged, starting from the values of activation and emotional valence obtained from the output nodes of the Adaline network. As previously noted, two different types of ANN

architectures were coupled in succession: two parallel Adaline-type networks and an SOM-type network were used to achieve our objectives. The combination of both networks was carried out using the general structure shown in Figure 1 (the bias neurons are not shown). The first Adaline (R_1) calculates the activation values from the properties of rhythm, tempo and time signature, while the second Adaline (R_2) estimates the emotional valence values from the properties of mode, FLN and absolute tonal range. The SOM network uses the output values of the R_1 and R_2 networks as the input. The aim of the SOM network was to categorize the stimuli according to the four basic emotions of the circumplex model (Kallinen, 2005). To know more about the training phase, the reader can consult the online supplementary material.

Results

Results for the Adaline networks

As a first step, we studied relationship between activation and valence, and the musical properties of tempo, time signature, rhythm, mode, tonal amplitude and FLN. A statistically significant relationship between the dimensions of activation-valence and musical properties would ensure that the selected properties were appropriate for the classification that the complete neural network will have to perform.

We can observe (see Table 2) that the correlation between activation and tempo was highly significant, while the rhythm (measured by the ANNB) and the time signature showed a lower although significant associative strength. Although rhythm was shown to better predict emotional valence than activation, we decided to keep ANNB within the model in the activation Adaline to test the theoretical hypothesis presented in the introduction. Emotional

valence was moderately and significantly correlated with the three tonal variables (absolute tonal range, FLN and mode). In conclusion, the results of the descriptive and inferential analysis of the correlations between the properties used and the emotional dimensions showed that the extracted musical properties were related to activation and emotional valence with small variations from the expected outcome. Motivated by the theoretical arguments raised, the construction of the network continued with the structure initially proposed.

After that, we ran both Adaline networks and compared the output values with the activation and valence values obtained by Vieillard et al. (2008). In the Adaline network for activation, a mean square error (MSE) of .371 was obtained in the training phase and an MSE of .394 in the generalization phase. In the Adaline network for emotional valence, the MSE reached a value of .260 in the training phase and .466 in the generalization phase. The values of the weights for both networks (see Table 2) demonstrated that the properties with the greatest connection strength were tempo and mode. A low weight for the FLN was observed for the emotional valence outputs, although the association between the minimum tonal frequency and emotional valence was high (see Table 2). The ANNB also achieved a lower connection weight with activation, although the association of this property with activation reached a moderate value in this case (see Table 2).

The ANNB and ATR properties showed a moderate contribution to the output calculation in their respective networks. Although the connection weights were lower than those obtained for tempo and mode, their moderate contribution justifies the inclusion of ANNB and ATR in a model that seeks to predict the emotional judgement of a musical piece. Another relevant aspect to note is that the bias unit reached a very low weight in both Adaline networks. Since the bias unit sets the activation threshold of the output neuron, a small value indicates that a minimal stimulation of the output neuron will produce a response. On a

practical level, this result means that even in situations of ambiguity or lack of information in the network, a response from the output neuron will still be possible.

Table 3. Comparative analysis between output and target distributions for emotional dimensions.

	Activation		Valence	
	Training	Generalization	Training	Generalization
t-test	$t(19) = -.774$	$t(19) = -1.568$	$t(19) = .008$	$t(19) = -.403$
	$p = .448$	$p = .133$	$p = .993$	$p = .692$
	$ d = 0$	$ d = .324$	$ d = 0$	$ d = 0$
Correlation	$r = .799^{***}$	$r = .817^{***}$	$r = .866^{***}$	$r = .715^{***}$
	$ d = 2.657$	$ d = 2.833$	$ d = 3.463$	$ d = 2.045$
Run test	$Z = .368$	$Z = -.203$	$Z = 1.134$	$Z = -1.608$
	$p = .999$	$p = .818$	$p = .153$	$p = .108$
Error	$ d = .058$	$ d = .032$	$ d = .179$	$ d = .254$
K-S	$Z = .620$	$Z = .802$	$Z = .839$	$Z = .642$
	$p = .863$	$p = .541$	$p = .483$	$p = .805$
	$ d = 0.098$	$ d = 0.126$	$ d = 0.133$	$ d = 0.101$

Note. K-S: Kolmogorov-Smirnov's test. * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

Table 3 shows the inferential statistics of the adjustment in the training of both Adaline networks. First, we checked whether the median of the obtained *output* values differed from the expected output (target). The Wilcoxon's W test for dependent populations was used for this evaluation. The null hypothesis of equal medians could not be rejected in any case. As a complement to the previous statistic, the correlation coefficient between the values of the output and the target was also evaluated. For activation, a determination

coefficient of 62.25% was obtained, while the variability explained for the emotional valence was 62.88%. From these results, we can conclude that the values predicted by our Adaline (output) networks were statistically similar to their expected values (target).

However, these analyses did not provide much information about the nature of the prediction error of the outputs. We observed that the distribution of error does not follow a systematic pattern. We found that the difference between the target and output was -0.031 ($SD = .623$) on average for activation and -0.053 ($SD = .623$) for emotional valence. To check whether the pattern of errors had some systematic component, the independence assumption was analysed using the runs test, and the assumption of normality was tested using the Kolmogorov-Smirnov test for a single sample using a normal distribution as the comparative distribution. For both activation and emotional valence, the hypothesis that the error is distributed randomly and normally could not be rejected since none of the statistics reached significance (see Table 3). From the above results, we can conclude that the Adaline networks adequately reproduced the activation and emotional valence values found by Vieillard et al. (2008) experimentally. After that, these values were applied to the SOM network in order to extract the basic emotions from the activation and valence values provided by the Adaline networks.

Results of the SOM network

The SOM network was trained with the outputs generated in the training phase of the Adaline networks. The rate of correct classifications of emotions in the SOM network reached 95% in the training phase and decreased to 85% in the generalization phase, achieving an overall classification rate of 90% (see Table 4). The emotion with the best rate of correct classifications was calm, with 100% accuracy, followed by the emotions of joy and sadness

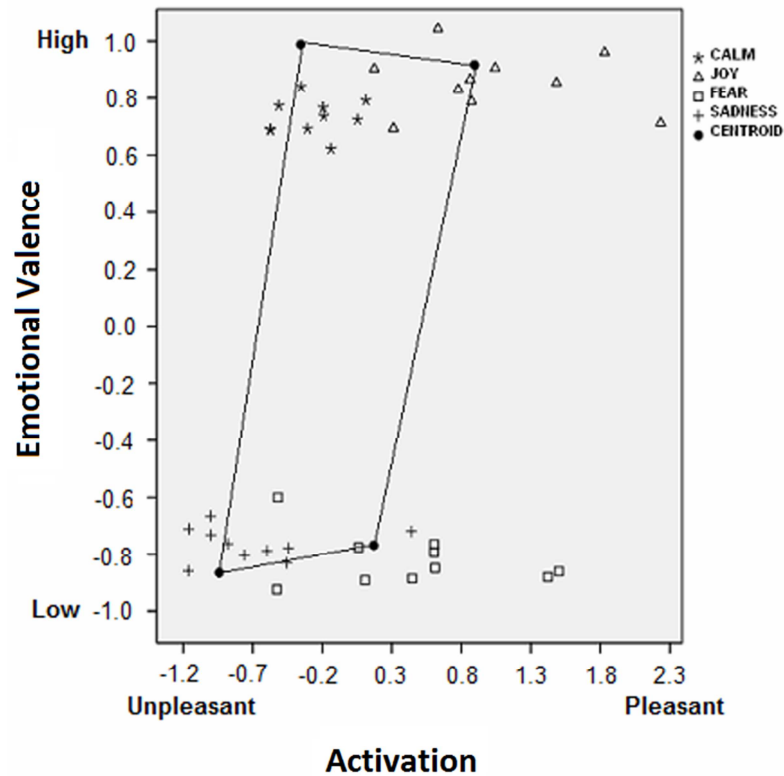
with a correct classification rate of 90%. The lowest ranked emotion was fear, with identical performance (80% correct classifications) in the training and generalization phase.

Table 4. SOM's correct classification in our network and his comparison with Vieillard et al.'s experimental rates.

	Phase		Total	Experimental
	Training	Generalization		
Calm	5 (100%)	5 (100%)	10 (100%)	67%
Joy	5 (100%)	4 (80%)	9 (90%)	99%
Fear	4 (80%)	4 (80%)	8 (80%)	82%
Sadness	5 (100%)	4 (90%)	9 (90%)	84%
Total	19 (95%)	17 (85%)	36 (90%)	83%

By comparing success rates obtained in our neural network with the rate of correct classifications obtained by Vieillard et al. (2008) in humans for each stimulus ("Experimental" column in Table 5), we found that the percentages of correct answers were similar for the emotions of joy ($Z = -0.790, p > .05$), fear ($Z = -0.073, p > .05$) and sadness ($Z = 0.479, p > .05$). However, our network classified stimuli belonging to the emotion of calm ($Z = -1.96, p < .05$) more efficiently than that in Vieillard et al.'s (2008) study. We checked that the output of the activation and valence neurons in our network formed a quadrangular structure consistent with the approach outlined in the introduction regarding the distance between emotions in the circumplex model of emotion (see Figure 3). The points located in the structure correspond to the coordinates of the input of the SOM network. These points are located around the four emotional categories identified by the neurons of the SOM network.

Figure 3. SOM map after training.



The activation and valence weights for the output unit of the SOM (see Table 5) representing the emotion of joy were both positive, whereas for the opposite emotion (sadness), the weights changed sign, which coincides with Russell’s (1980) circumplex model. A similar pattern occurred in the output unit of the SOM representing the emotion of tranquillity or calm. While its weight was positive for emotional valence, its weight was negative for activation due to its relaxing nature. The opposite emotion, fear, obtained a positive weight for activation due to its activating character and a negative weight for emotional valence. Moreover, Figure 3 shows that, as indicated by Russell’s (1980) circumplex model, the distance was greater between neurons that represent opposite emotions (fear vs. calm and joy vs. sadness), and neighbouring neurons formed an angle close to 90 degrees. As can be observed in Figure 3, the value representing the neuron that processes the emotion of fear has 4 points widely scattered around it. Its neighbouring neuron (lower left

neuron) has 6 coordinate points around itself, which shows that one of its stimuli that should have been classified as fear was incorrectly classified by its neighbouring neuron.

Table 5. SOM's output layer weight by emotional dimensions.

	Activation	Valence
Joy	.896	.917
Fear	.171	-.767
Calm	-.355	.988
Sadness	-.941	-.865

Generalization of results to another set of stimuli

The analysis of the results for the simulation carried out with a simple ANN, structurally similar to that of the Peretz & Coltheart (2005) model, has shown that a relationship can be established between certain elementary properties of the musical stimuli and the dimensions of emotion previously described by Russell through the bi-dimensional space of valence and activation. The results replicated the behavioural data in humans with a high degree of accuracy. Our results reinforce the hypothesis that properties of the musical stimulus are fundamental determinants that define and help explain subjective emotional ratings of the stimulus, as posed by Balkwill and Thompson (1999), Korhonen (2004) and Coutinho and Cangelosi (2009). However, unlike the models proposed by Coutinho and Cangelosi (2009, 2011), our network allows for the specification of the individual relationship between specific stimulus properties and specific dimensions of emotion. This is one of the highlights of our model. Moreover, the specific structure of our network disentangles two main processing steps, the extraction of activation and emotional valence and the categorization of these values into specific emotions. To highlight this, we must underline the high values of fit of our results obtained by our ANN to the human values. Notwithstanding,

we also need to ascertain the generalizability of this neuronal model to other sets of stimuli. This was our third goal.

To evaluate the generalizability of our previous results, we decided to test the performance of the network with a new set of stimuli. We tried to equate the new set of stimuli with the characteristics of the standardized set used by Vieillard et al. (2008) in parameters such as length, simplicity, etc., but we did not build the stimuli ourselves. Instead, we extracted them from Western classical music. This new set of 40 musical stimuli was presented randomly to 16 participants (10 women; $\chi^2 = 1, p > .05$) with a mean age of 33 years ($SD = 4.9$) by means of software built in Delphi (Object Pascal). Participants were instructed to evaluate every stimulus on activation and valence using two sliders that were labelled from 1 to 100 as well as to categorize them in one of the basics emotions: joy, peacefulness, tension and sadness.

Although 70% of stimuli were classified consistently in the same emotional category by the participants, 30% of the stimuli were ambiguous, with an agreement rate lower than 60%. Following the criteria used by Vieillard et al. (2008) with regards to the stimulus selection, we selected only the 20 stimuli with the higher rate of consistency (see online supplementary material), five for each emotional category.

The mean rate of agreement of the selected stimuli was 75%. After this selection phase, we demonstrated that the valence differed between the emotional categories (joyful, peaceful, sad and tense stimuli) by an ANOVA with emotional categorization as the between-subjects factor with four levels, $F(3, 16) = 19.912, p < .001, \omega^2 = .789$. The effect size for the difference between joyful-peaceful stimuli vs. sad-tense stimuli was high ($|d| = 3.458$). The results can be seen in Table 6.

Table 6. Means and Standard deviations for activation and emotional valence by emotional category on experimental phase and statistical indices for the comparison between output and target distributions.

		Activation	Emotional valence	
Emotional categorization	Joy	68.687 (8.297)	66.583 (4.31)	
	Calm	48.08 (20.992)	58.553 (9.327)	
	Fear	61.478 (20.736)	41.187 (9.781)	
	Sadness	40.687 (13.444)	39.406 (7.18)	
MSE		0.446	2.643	
t Student		$t(19) = .173$	$t(19) = .084$	
		$p = .864$	$p = .934$	
Correlation		$ d = .081$	$ d = .039$	
		$r = .732^{***}$	$r = .803^{***}$	
		$ d = 2.657$	$ d = 3.463$	
	Run test	$Z = -2.068$	$Z = -.689$	
Error		$p = .039$	$p = .491$	
		$ d = 1.043$	$ d = 0.$	
	K-S		$Z = .477$	$Z = .762$
			$p = .977$	$p = .607$
		$ d = .098$	$ d = .133$	

Note. K-S: Kolmogorov-Smirnov's test. * $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$.

We also compared results in the activation dimension by way of an ANOVA with emotional categorization as the between-subjects factor. The results showed that emotional categorization as the main factor was significant, $F(3, 16) = 23.662, p < .001, \omega^2 = .816$. Moreover, the effect size for activation was very high ($|d| = 1.185$).

Next, we computed the same six properties (ANNB, b.p.m., time signature, mode, ATR and FLN) considered previously for the new set of stimuli. We used the weights previously computed with the stimuli from Vieillard et al. (2008) in our ANN network and applied those properties as inputs for the new set to obtain the activation, valence and emotional category for these new stimuli. The fit between the judgement of people on the activation and valence dimensions and the one provided by the network was very similar to our previous results with the stimuli provided by Vieillard et al. (2008). The correlations between the behavioural results and the results from the simulation were 0.734 ($p < .001$) for activation and 0.803 ($p < .001$) for valence. These results are quantitatively very similar to that obtained for the stimulus set from Vieillard et al. (2008). Moreover, we found that the prediction error in both simulations was distributed normally and was random. The mean square error (MSE) was very close to the one obtained in the experimental phase for activation but was higher for valence. We can conclude from this pattern of results that some outputs from this dimension in the network were not well adjusted to match the judgement from the people. We demonstrated that both distributions are similar even with this high MSE value for valence (see Table 6, Student's t test for related samples).

We also studied the classification rate of the SOM. The results showed a satisfactory hit rate (75%), although it was lower than that obtained in the first study. The emotional categories with the lower rate of classification were tension and joy (60%, 3 items from 5), but a 100% and 80% correct classification rate was observed for sadness and peacefulness, respectively. Perhaps the achievement of the network was lower than that obtained with the stimulus set from Vieillard et al. (2008) due to the higher variability in activation already discussed. Notwithstanding, we have shown the generalization ability of our ANN network as it worked properly with both sets of stimuli.

Discussion

We have implemented an ANN for predicting valence, activation and emotional categorization in musical fragments. Our ANN was intended to be simple but congruent with the structure specified in the model by Peretz and Coltheart (2005). Simplicity was achieved by using few acoustic inputs and in the structure of the network. Moreover, we tried to simulate the values of activation and valence simultaneously with the emotional categorization both in already reported stimuli and behavioural data (Vieillard et al., 2008) as well as in data collected by ourselves with a new stimulus set. Finally, we used an ANN to try to avoid linear statistical models that would not capture nonlinearities.

We have observed that tempo and mode seem to be the properties strongly connected with activation and emotional valence, respectively (DallaBella et al., 2001; Gagnon & Peretz, 2003; Husain et al., 2002). Notwithstanding, we must acknowledge that the relevance of these two properties could be overestimated because the stimuli designed by Vieillard et al. (2008) were specifically composed so that the tempo was higher in joy and fear stimuli than that in calm and sadness stimuli. Moreover, the musical stimuli were designed in such a way that all scores linked to positive emotion were written in major mode, while the scores linked to negative emotion were written in minor mode.

The worst performing properties in the Adaline networks were rhythm and the minimum frequency (FLN). By defining the average number of notes per bar as a measure of the rhythm of the score, we thought that those tracks with a higher note average per bar would be perceived as faster pieces and, therefore, would result in greater activation. However, rhythm showed a greater effect on emotional valence prediction in our model, which can challenge the previous models that separate the temporal code from the tonal code (Peretz et al. 2003). Perhaps both processing mechanisms could interact somehow, that is, it could be

that both pathways are not completely isolated from each other. For this reason, in future research, incorporating the interaction effect between temporal and tonal properties on the emotional process will be interesting. In reference to the low-pitch frequency variable, we observed differences in the minimum frequency of pieces depending on the emotion to which they were associated. Thus, pieces associated with joy presented frequencies higher than other pieces, while those associated with fear/tension emotion were written one octave below the other emotions. The future improvement of our network would need to consider all ranges of tonal frequencies, not only the lower ones.

We have worked with a few properties of musical stimuli in our ANN. This fact allowed us to get around a problem already mentioned with previous networks: the introduction of redundancy in models by the use of several measures of sound waves as inputs. Using only three variables as predictors for activation and valence, we have achieved fit indices in the same range as previous models for activation and higher for valence. Therefore, parsimony in our model is a relevant feature in comparison with that of previous neural network models.

Moreover, our results showed that the output of both Adaline networks was distributed in the same way as values obtained in the experimental data (Vieillard et al., 2008), and the prediction error was normally and independently distributed. The error showed no systematic pattern. Therefore, we plan to decrease the error in future works by better exploring predictors of activation and valence in the Adaline networks. As the input of the SOM network is dependent on the output of the Adaline networks, we expect that the improved predictors would also improve the categorization of emotions in the SOM network. Another noteworthy feature of our model is that it achieved a very high rate of correct categorization in the SOM

(87%). That is, our proposed ANN model succeeded in both phases of processing: the simulation of valence and activation as well as the emotional categorization of these values.

Furthermore, we observed that the weights of the neurons of the SOM were positive for joy and negative for sadness (see Table 7). In addition, the neuron that represents the emotion of tranquillity obtained a negative weight for activation and positive weight for emotional valence, whereas the neuron that represents the emotion of fear obtained a negative weight for emotional valence and a positive weight for activation. These results are consistent with the approach of Russell's (1980) circumplex model of emotion and with the results obtained experimentally by other authors (Balkwill & Thompson, 1999; Bigan et al., 2005; Dalla Bella et al., 2001; Fritz et al., 2009; Gagnon & Peretz, 2003; Korhonen, 2004; Roy et al., 2008; Sloboda, 1991; Vieillard et al., 2008; Webster & Weir, 2005; Yang et al., 2008; Zenter, Grandjean, & Scherer, 2008).

Using the proposed composite ANN, we have observed that a few properties of the musical stimulus can predict the assessments made by listeners on the activation generated in them by the musical stimuli, as well as its emotional valence. We have built a neural model that has been able to reproduce this evaluation process of the musical stimulus by building two separate Adaline networks that have provided some output values similar to those target values obtained experimentally by Vieillard et al. (2008). Moreover, these output values allowed for the classification of such musical scores into discrete emotional categories using an SOM network, yielding a rate of emotional classification of musical stimuli similar to that found in humans (Fritz et al., 2009; Gosselin et al., 2007; Vieillard et al., 2008).

These results were extended in a new experiment with humans and a different but similar stimulus set extracted from Western popular music. The results were very similar to the previous simulation: the networks achieved a good fit of activation and valence using new

stimuli as well as predicted the emotional category of them. Overall, we can contend that there was not an overfitting problem as an explanation of the first simulations.

We consider that the main unsolved problem in our network is the one posed by the stimulus subset discarded by Vieillard et al. (2008) and those in our stimulus set that produced ambiguity in participants and, as a consequence, in the network prediction capabilities. We can only guess that this could be due to the restriction to only four emotional categories in our spatial model. Perhaps by using more than four categories for the emotion of music, the stimuli could be placed with less variability in the Russell model. Another explanation for our results is that by trying to maintain the simplicity of the model, we used very few properties of music, and this fact could diminish the predictor capabilities of our model. Moreover, we have used a narrow set of stimuli in terms of genre. That is, our stimuli were extracted from Western classical piano music without orchestral accompaniment. We have not used musical stimuli from other genres (e.g., jazz, popular music, etc.). Therefore, we are planning to verify the effect that broadening the number of emotional categories and the type and number of musical properties included as inputs and using other bodies of music could have on the results that we have obtained.

CAPÍTULO 3

Prediction of emotional response of music using principal component factors extracted from the *onset* curves

Elena Saiz-Clar*, José Manuel Reales* & Miguel Ángel Serrano

*Universidad Nacional de Educación a Distancia

Department of Science Behavior and Health Methodology

Universidad de Valencia

Department of Psychobiology

(Submitted to Music Perception)

Abstract

The relationship between parameters extracted from the acoustic wave of musical stimuli and the emotional response of people has been traditionally approached using several physical measures extracted from time or frequency domains. From the measures used in the time domain, the musical *onset* is defined as the definite moment in that a musical note is issued by any musical instrument or human voice. The sequence of *onsets* in the performance of a specific musical score creates what it is known as the *onset* curve (OC). The influence of the structure of OC on emotional judgment of people is not known. To this end, we have applied Principal Component Analysis on one full set of variables extracted from the OC to capture their statistical structure. We have found a trifactorial structure related to Activation and Valence dimensions of emotional judgment. The structure has been cross-validated in a second study using different participants and stimuli. In this way, we add the factorial scores of the OC as a stable and relevant piece of information to be considered in the prediction of emotional judgment of music.

Key Words: *emotion, arousal, psychoacoustics, mood and modelling*

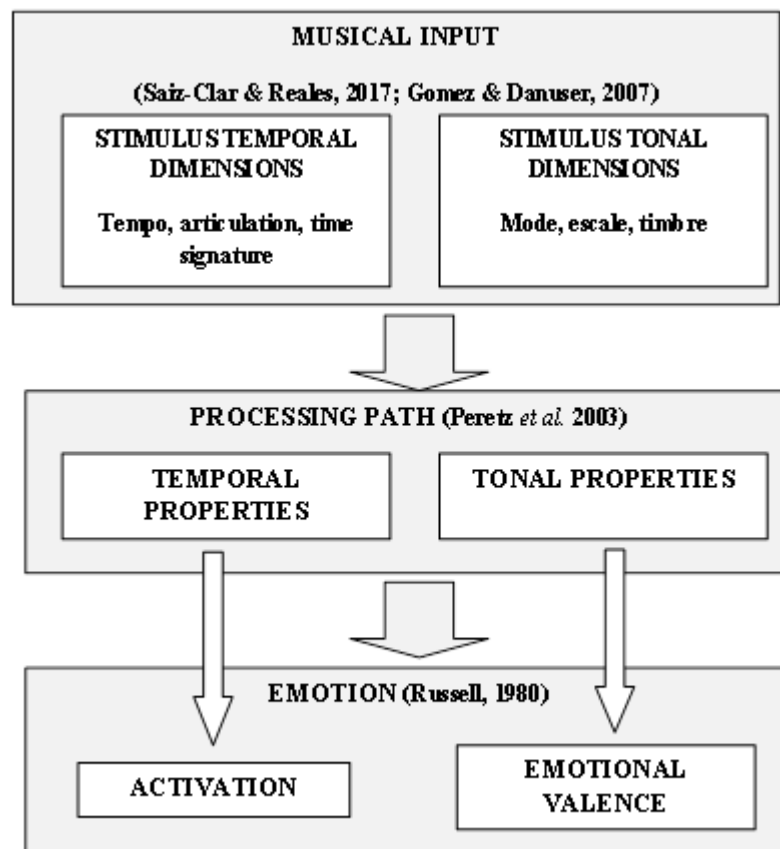
Introduction

The physical structure of musical stimuli plays a determining role in its emotional evaluation (Coutinho & Cangelosi, 2009 and 2011; Gomez & Danuser, 2007; Husain, Thompson & Schellenberg, 2002; Korhonen, 2003; Peretz, 2001; Peretz & Coltheart, 2003). Some recent theoretical models have demonstrated the predictive capacity of certain dimensions of musical stimulus (e.g., tempo, beat or mode) both in terms of the listener's emotional response (Saiz-Clar & Reales, 2017) and his physiological response (Gomez & Danuser, 2007). Some models have utilized dimensions or variables taken from music theory as predictive variables of subjective emotion or psycho-physiological response. On the other hand, other models have utilized physical measures of stimuli as predictive variables, normally (although not exclusively) taken from the frequency domain. For example, in successive connectionist models, Coutinho & Cangelosi (2009 and 2011) have demonstrated that the emotional dimensions of activation and emotional valence (Russell, 1980) in response to musical stimuli can be predicted based on certain descriptive parameters of the spectrum (e.g., the power-spectrum centroid) or measurements of the Short-Time Fourier Transform (STFT). The model by Coutinho & Cangelosi (2009) can be seen as a continuation of the model by Korhonen (2004), who proposed that the emotional evaluation of a piece of music is not static and singular but rather dynamic, as the structure of the stimulus changes over time. This initial proposition by Korhonen (2004), however, is not reflected in models that utilize parameters taken from the frequency domain to make predictions about emotion, as temporal information is lost in this domain. As a result, we believe it important to identify what descriptive parameters of musical stimuli can exist in the temporal domain with regard to the prediction of emotional dimensions. With this objective, we attempt to supplement descriptive statistics in the frequency domain with those found in the temporal domain in an attempt to combine both domains and improve prediction rates for emotional response.

Models that utilize variables proposed by music theory to predict emotion as well as those that rely on physical variables both propose the existence of two separate neural pathways for processing activation and emotional valence (Coutinho & Cangelosi, 2009 and 2011; Gomez & Danuser, 2007; Korhonen, 2004). This idea is congruent with the formal structure of the emotional model by Russell (1980), who conceives of activation and emotional valence as orthogonal and hence independent dimensions. Another important question, this time from the neuropsychological focus of musical perception (Peretz & Coltheart, 2003), is the distinction between two different processing pathways: one that processes the temporal elements of the stimulus (that is, properties such as tempo, articulation or rhythm) and another that processes tonal properties, such as mode or timbre. The idea of a double processing pathway is also put forth by Balkwill & Thompson (1999), whose model includes the idea that the physical properties of a stimulus could affect emotional judgment about it in a differential manner, such that those properties related to time (e.g., rhythm) would affect judgment regarding activation, while properties related to tone would affect emotional valence (see Figure 1).

This hypothesis was subsequently taken up by Husain, Thompson & Schellenberg (2002), who observe that the manipulation of the tempo of a piece of music affects the listener's judgment of the activation of the stimulus, while manipulation of the mode primarily affects their judgment at a hedonic level. Meanwhile, in a study on the transculturality of emotional categorization in music, Fritz et al. (2009) attribute the differences found in the categorization of pieces between Western listeners and the Mafa to the tonal rules in composition, while attributing universality to temporal and rhythmic properties.

Figure 1: Integrative model of the emotional perception of music proposed by the double path hypothesis (Balkwill & Thompson, 1999; Saiz-Clar & Reales, 2017)



Therefore, based on the model by Balkwill & Thompson (1999) and that by Peretz & Coltheart (2003), we could argue that the temporal properties of stimuli are processed independently from tonal properties and that the former would affect judgment regarding activation, while the properties processed via the tonal pathway would affect hedonic judgment (Figure 1).

However, this hypothesis has been further nuanced in other studies that have demonstrated that some temporal properties emerge as more efficient predictors of emotional valence than of activation. For example, in Saiz-Clar & Reales (2017), a temporal property such as rhythmic regularity presented a greater partial correlation coefficient for emotional valence than for activation. Yang, Lin, Su & Chen (2008) also observed that of the 114

predictors analyzed in their study, some temporal properties could explain emotional valence, while tonal properties would be good predictors of activation. A precursor to this reconceptualization of the independence of temporal and tonal processing can be found in Samson, Zatorre & Ramsay (1997), who, after applying the technique of Multidimensional Scaling (MDS) to study the dimensions that enable the definition of musical timbre, observed that listeners utilized both the rise-time duration of the note as well as the number of harmonics to judge whether two consecutive tones belonged to the same instrument. In this case, if temporal dimensions affect judgment regarding activation and tonal dimensions affect hedonic judgment, then musical timbre would be a property that affects activation and emotional valence simultaneously.

The interaction between the different properties of musical stimuli (defined in the domain of space, frequency or musical theory) can modulate the listener's response in both emotional dimensions. Thus, for example, Boltz (1998) found that perceptions of the speed of execution of a piece of music are influenced by the temporal envelope of that music. In other words, two pieces of music with the same musical tempo would seem more or less fast depending on the tonal variability of the piece. Hence, it might be hasty to conclude that tonal and temporal properties appear to be clearly theoretically separable. Despite empirical research supporting their separate effect on emotional dimensions, it is not consistent with all of the empirical data.

It is important to note that interest in modeling the emotional response to music is not limited to the realm of the behavioral sciences. We find interesting contributions in engineering, in which numerous algorithms have been developed that also employ physical measures of musical stimulus to detect patterns that allow for the classification of musical pieces by genre (Shao, Wang, Li, & Ogihara, 2009; Tzanetakis & Cook, 2002). Some of these

algorithms aim for an emotional classification of musical compositions based on the extraction of the physical properties of the acoustic wave, both temporal as well as tonal (e.g., Yang, Lin, Cheng, & Chen, 2008; Yang, Liu, & Chen, 2006). However, some of them prove problematic for direct use in the empirical sciences, although they can be useful for increasing the precision of prediction. We refer in particular to the use of multiple predictors, which leads to over-adjusted and redundant models, as noted by Korhonen (2004) in his conclusions, or also the tendency to introduce all of the predictors as a group within the model. Without denying the good intentions of these models, we should note that the inclusion of independent variables without prior study of the best predictors of the dependent variable or the collinearity between independent variables leads to models that are redundant and over-adjusted, with little power of generalization. This problem has been noted by Eerola, Lartillot & Toiviainen (2009), who studied the optimal construction of linear models for introducing predictors while avoiding overfitting.

We can conclude that both the number of predictors and the manner in which they are included affect the prediction of emotion and that the relationship between the temporal and tonal elements, on one hand, with the dimensions of activation and emotional valence appears to be more complex than what was initially considered; hence, it is relevant to propose an in-depth analysis of the predictors utilized as well as their nature, using a mixed or eclectic theoretical focus. For example, although Saiz-Clar & Reales (2017) obtain adequate results regarding the predictors of emotional dimensions based on direct extraction from the musical score, some important questions arise. On one hand, although the score indicates the necessary aspects for playing the piece, it is difficult to adhere to perfectly, as there can be discrepancies between the musical notation and its concrete interpretation. The creation of models that are based on both sources of predictors, the score and the sound wave considered physically, has as an advantage the fact that those properties of the piece of music extracted

from the score are invariable properties of the model for this score, while the physical measures vary based on the interpretation of the piece. On the other hand, if our aim as clinicians is the applicability of the music to the mood of the listener, we must know clearly which variables we can manipulate and what effect we hope they will produce. This is another key reason to study the particular effect of each measure of the stimulus on emotional judgment. Here, the comparison between measures of the physical properties of sound as well as the measures and concepts such as the properties extracted from the musical wave itself in the temporal domain, become important.

Onset Curves (OC) as a Source of Physical Predictors

Previous models have utilized very diverse variables extracted from the frequency domain (e.g., the successive connectionist models by Coutinho & Cangelosi, 2009 and 2011) or the temporal domain (e.g., the study of *onset* curves by Scheirer, 1998 or Kapluri, 1999). A musical *onset* is defined as the precise moment at which a musical note is emitted by a musical instrument (Gainza, Lawlor, Coyle, & Keller, 2004). The point at which we situate the musical *onset* in the temporal continuum has a subsequent evolution, with a rising phase and a falling phase, whose temporal evolution should also be considered relevant. It is important to observe that while our auditory system allows us to capture the moment at which a sound is initiated (its *onset*), this moment is not well defined in the sound wave itself, and complex algorithms are needed for its detection. For Kapluri (1999), the detection of *onsets* is a process that makes it possible to situate the beginning of an event (for example, a musical note or chord) in the temporal domain, as said identification is determined by a perceptible change in the intensity, tonality and/or timbre of the sound. That is, the *onset* detection algorithms yield temporal points that represent the musical notes of a score and their temporal succession giving rise to *onset* curves (OCs).

The study of OCs has been previously used in the analysis of the temporal properties of musical compositions such as rhythm or tempo. Thus, for example, in a review article, Yang & Chen (2012) propose different OC indices to define different dimensions of rhythm. For their part, Lu, Liu & Zang (2006) propose the calculation of the average across the various *onsets* as a measure of rhythm strength and the function of auto-correlation of OCs as an index of clarity of rhythm and its regularity. The frequency of *onsets*, calculated as the number of *onsets* per second, can be related to the density of events; as suggested by Lartillot & Toiviainen (2007), this index allows determining the periodicity of the OC for the measure of musical tempo. The diversity of indices extracted from the OC demonstrates the lack of a common frame of reference in the research, though it also points to the relevance of the systematic study of the existing variability in OCs between musical fragments as an important aspect of the study of music and its relationship to the psychological functions that depend on rhythm. That is why one of the objectives of this study, apart from those indicated above, is to identify the dimensions that underlie all the indices that can be rationally extracted from the OC. In this way, subsequent work can be performed using appropriate and exhaustive OC indices, thus avoiding the plurality of alternatives in the ideation and identification of these indices.

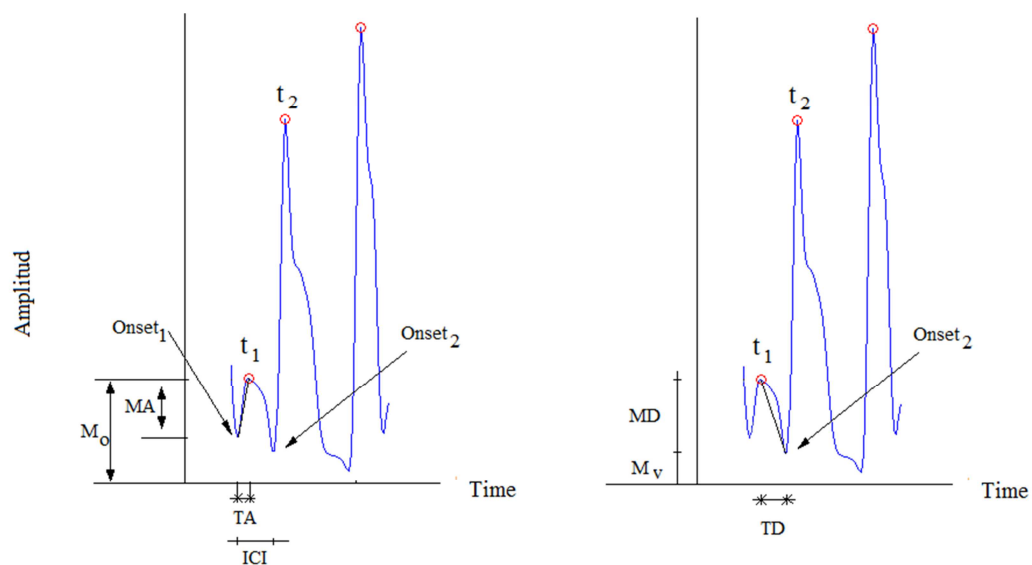
As we have indicated, the extraction of the OC based on the acoustic signal of the piece of music to order its rhythmic pulse is a complex task that has given rise to various algorithms. Undoubtedly, though, the greatest problem faced by these algorithms is the complexity of the harmonic grouping of tone and timbre, concretely, grouping them together in a single pulse defined in the temporal domain of the fundamental frequency and its harmonics (Klapuri, 1998 and Scheirer, 1999). The reason for this is that algorithms for detecting *onsets* operate based on the analysis of frequencies in the envelope (which includes both the fundamental note interpreted and its harmonics). This requires that the detection of

onsets include a prior phase of signal filtering that allows for the integration of the basic frequency and its harmonies in a single pulse. The algorithm developed by Scheirer (1998) aimed to develop a method capable of extracting a symbolic representation that would correspond with the perceptive experience of rhythm in humans. For Scheirer (1998), the rhythmic pulse can be described in terms of two components that define any sine wave: frequency and period. From this perspective, the frequency of the rhythmic pulse (or number of times the wave is repeated per unit of time) would be a concept equivalent to that of musical tempo (the number of notes or beats per unit of time), while the period (which includes the beginning, point of maximum amplitude, and decay of the *onset*) would be related to other rhythmic concepts such as articulation and the distribution of silences in the score, indicating the point at which each pulse falls.

While the elements described offer us a temporal position of the *onsets* represented by musical notes and are thus appropriate for representing rhythm, there are *onset* measurements that are also related to tonal properties. Hence, for example, the time of attack of the musical note (or time between when the note begins and when it reaches its maximum amplitude) has been frequently utilized as a measure of timbre (Lartillot & Toiviainen, 2007; Lartillot, Toiviainen, & Eerola, 2008) because different instruments present different forms of note attack. For example, percussion instruments have an attack (distance between the valley and the *onset*) that is short in duration and high in amplitude and hence has a greater slope, while other instruments such as wind instruments have a gentler slope of attack. This is precisely one of the great advantages of the study of sonic form from the perspective of OC analysis, that is, the possibility of considering both aspects belonging to the temporal route as well as aspects of the tonal route simultaneously.

As a result of the application of the algorithm by Scheirer (1998), we have a graphic representation of the rhythm of the piece (see Figure 2) in which we find two basic elements: the *onset* and the valley. The *onset* comprises the portion of the wave that lasts from the beginning of the attack of the note until it achieves its maximum amplitude, the point at which the intensity of the *onset* falls until reaching a minimum that would correspond to silence or the attack of the following note. This portion of the wave in which the magnitude of the *onset* declines we will call the valley of the *onset*. Both elements, the *onset* and the valley, are defined by specifying their amplitude and temporal position.

Figure 2. Graphical example of measures derived from the *Onsets Curve* for a single *onset* (t_1).



Note. M_o = *Onset* magnitude; MA = *Attack* magnitude; TA = *Attack* time; M_v = *Valley* magnitude; MD = *Decay* magnitude; TD = *Decay* distance; ICI = *Inter click onsets* interval.

Based on the magnitude and location of the *onsets* and valleys, we can obtain other measures that describe the OC more specifically; for example, the distance between two successive *onsets* or the magnitude of the valley (Table 1). Given that each piece of music is composed of a series of *onsets* and valleys that can be more or less prolonged, we must use

scores that can summarize all of the OC measurements we have obtained in relation to the series of *onsets* and valleys. In this case, the average as a central tendency parameter and standard deviation as a dispersion parameter are two parameters that represent the distribution of the OC.

In view of the above, the main objective of this article is to extract those physical properties of the sound wave that allow us to reliably predict, with an acceptable level of certainty, judgment regarding the emotional dimensions of musical stimulus. We carried out two studies. The first served to extract, through an exploratory factor analysis (FA), the basic indices that define the OC. Meanwhile, in the second study, we attempted to evaluate the level of generality of the predictive model utilizing a new stimulus group. We subsequently used both sets to determine the relationship between the resulting dimensions of the OC and Activation and Emotional Valence through a predictive model.

Method

Procedure and Stimuli

In the first study, we utilized the 40 stimuli described by Vieillard et al. (2008), and by means of the MIRToolbox (Lartillot, Tovianen & Eerola 2008b) for Matlab, we extracted basic OC measurements (see Table 1 and Annex 1), which were later expanded in SPSS (Table 1) for a better description of OCs using the code from Annex 2.

Table 1: Description of the measurements obtained from the *Onset Curve*

ONSET	Magnitude	Maximum amplitude reached by an <i>onset</i> .
	Attack magnitude	Absolute amplitude from the beginning of the attack to the maximum amplitude reached.
	Attack time	Time elapsed from the beginning <i>onset</i> time to the maximum amplitude moment.
	Attack slope	Slope of the line between <i>onset</i> beginning and the maximum amplitude reached.
	Attack distance	Euclidean distance between <i>onset</i> beginning and the maximum <i>onset</i> amplitude coordinates.
VALLEY	Magnitude	Minimum amplitude reached by an <i>onset</i> on his decline phase.
	Magnitude decay	Absolute amplitude from <i>onset</i> maximum amplitude to <i>onset</i> minimum amplitude.
	Decline time	Time elapsed from <i>onset</i> maximum amplitude moment to the next <i>onset</i> beginning time.
	Decline slope	Slope of the line between the maximum <i>onset</i> amplitude coordinate and next <i>onset</i> beginning.
	Decline distance	Euclidean distance between maximum <i>onset</i> amplitude and the next <i>onset</i> beginning.
Inter Click Interval (ICI) mean		Average time elapsed between consecutive <i>onsets</i> .
Measures obtained from MIRToolbox		
Onset position and magnitude		Horizontal and vertical coordinates of each of the points of maximum amplitude from the <i>onsets</i> .
Attack slope		Slope of the line between <i>onset</i> beginning and the maximum amplitude reached.
Attack time		Time elapsed from the beginning of <i>onset</i> time to the tiem of maximum amplitude.
Valley position and magnitude		Horizontal and vertical coordinates of each of the points of minimum amplitude from the <i>onsets</i>

This process allowed us to obtain 11 temporary series of variables of the OC for each musical fragment, for which we obtained their descriptive statistics (average and standard deviation) as defining moments of the central tendency and dispersion. Finally, the values for activation and emotional valence for each stimulus were gathered from the annexes of the work by Vieillard et al. (2008). The authors obtained said values by averaging the subjective valuations for activation and valence provided by their listeners upon hearing each piece of music. They utilized Likert scales of 1-9, with 9 being maximum activation (or positive emotional valence) and 1 being minimum activation (or negative emotional valence).

For the second experiment, we employed a set of 40 pieces of music from the popular classical music repertoire (see supplementary material). This set was presented to an incidental sample of 16 participants with no formal musical training (10 women and 6 men; $\chi^2 = 1$; $p = 0.317$) and with an average age of 33 years ($SD=4.9$). The task (entirely computerized) consisted of a first judgment in which the participants provided a general evaluation of each of the nine pieces of music according to the dimensions of activation and emotional valence on a scale from 1 to 100 and then categorize them as happy, sad, calm or tense. The OC measurements for this second set of stimuli were found by following the same procedure utilized in the prior experiment.

Plan of Analysis

With the objective of the study in mind, our first step was to confirm the magnitude of the Pearson's correlations between the 22 OC measurements and measurements of activation and emotional valence in the first set of stimuli. Because the results (see results section) reflected high collinearity between the OC measurements in the correlation matrix, we chose to reduce its dimensionality by applying the Factor Analysis (FA) technique through the

method of Principal Component Analysis based on the covariance structure and utilizing an orthogonal rotation to maximize uncorrelation between the resulting factors to use them subsequently as multiple predictive variables of the emotional scores. All of the OC measurements included in the FA were typified. Prior to the analysis, we verified the compliance of the parametric assumptions necessary to apply the FA.

To determine the stability of the factorial solution obtained, we carried out a Parallel Factor Analysis (PA) according to the method described by Horn (1965) and subsequently adapted by O'Connor (2000). The parallel analysis makes it possible to find the magnitude of each eigenvalue obtained through the average in a simulation of 1,000 random matrices with the same number of observations and variables. That is, it allows us to determine the random values of the eigenvalues and, based on that, to judge, in the original matrix, which of them are superior to what was expected according to the simulation (the code used to carry out an FA can be found in O'Connor, 2000).

To test the consistency of the factorial matrix of the first set of stimuli, we carried out a new FA with the same conditions and restrictions carried out in the prior experiment but with a different set of stimuli. The next step was to evaluate the similarity between the factorial matrices obtained for the two previous sets of stimuli, utilizing two complementary methods. The first of them consists of a correlation analysis between the factorial weights of each factor in each matrix. As described by Tabachnick & Fidell (1989), the idea underlying this procedure is that the factorial weights of a factor should be similar, independent of the sample from which they are obtained. This leads to two hypotheses: 1) The correlation between two similar factors in two different samples should have a greater magnitude; and 2) Two dissimilar factors from different populations should have different factorial weights, and therefore, their correlation should not be significantly different from zero. However, applying

the correlation method, we find that the relationship between the weights of the two dissimilar factors would be negative; this would indicate not just that the factors are not similar but also that when the weights of a factor increase for an item, the weight of that same item for the other factor decreases.

To resolve the problems that can stem from the exclusive use of this first method for evaluating the similarity between the factorial matrices, we employed a second method based on the congruence coefficient for each pair of factors. The calculation was based on the formula proposed by Wrigley & Neuhaus (1955) and described in Barrett (1986). As recommended by Tabachnick & Fidell (1989), this index can be used as a proven alternative to Cattell's S index (1978) when the percentage of variables in the hyperplane for each compared pair is less than 60%, as in this case the usual statistical tables cannot be used to determine the level of significance.

Finally, and given that the main objective of this article was based on studying the existing relationship between OC measurements and the dimensions of activation and emotional valence, we built a multiple linear regression model that would enable us to determine the predictive capacity of these emotional dimensions based on the factor scores extracted from the OC. For this, we combined the data from the two samples utilized, thus obtaining a total sample of 80 stimuli, which included the activation and emotional valence averages for each item previously classified as well as the classified score in each factor extracted in their respective FAs.

Results

The First Study

As we have indicated, the main objective of this study was to investigate the type of relationship that exists between the physical properties extracted from the OC and the dimensions of activation and emotional valence. To do this, we calculated the linear correlation coefficients between the direct physical dimensions of the OC and the emotional dimensions (Table 2). The results demonstrated that those measures that were significantly correlated with activation did not demonstrate said relationship with emotional valence and vice versa, with the exception of three variables that were significantly correlated with both emotional dimensions, though in the opposite directions: the standard deviation of the declining slope, the standard deviation of the magnitude of the valleys and their average magnitude. Only four of the 22 OC measurements did not demonstrate a significant relationship with any of the emotional dimensions: average duration of attack, average slope of attack, magnitude of attack, and average distance of attack.

Table 2: Rescaled factorial scores from the *Onsets* Curve measurements and their correlations with the emotional dimensions in the first study.

	Rescaled factorial scores			Correlations	
	Factor 1	Factor 2	Factor 3	Activation	Emotional Valence
Standard deviation mean of the attack distance	.924			-.605**	-.010
Standard deviation mean of the attack magnitude	.922			-.602**	-.013
Standard deviation mean of decline magnitude	.897			-.649**	.271
Standard deviation mean of slope attack	.878			-.465**	-.026

Average euclidean distance of decline	.841			-.511**	-.050
Average magnitude of decline	.799			-.174	.325*
Standard deviation mean for the magnitude of onsets	.799			-.558**	.035
Average duration of decline	.707	.511		-.609**	-.192
Average distance between onsets	.705	.518		-.594**	-.189
Standard deviation mean of the distance between onsets		.957		.119	-.665**
Standard deviation mean of the decline duration		.955		.112	-.668**
Standard deviation mean of the decline distance		.919		.073	-.573**
Standard deviation mean of the attack duration		.840		.261	-.580**
Average magnitude of the onsets			.921	.472**	-.233
Average slope of decline slopes			.843	.843**	-.171
Average magnitude of the valley			.793	.694**	-.524**
Standard deviation mean for the valley magnitude			.740	.484**	-.616**
Standard deviation mean of the declining slope			.671	.592**	-.415**
Average duration of attack	-	-	-	-.178	-.123
Average slope of attack	-	-	-	.141	.265
Average magnitude of attack	-	-	-	-.124	.258
Average distance of attack	-	-	-	-.136	.218
FACTOR 1				-.534**	.113
FACTOR 2				.068	-.669**
FACTOR 3				.635**	-.272

*** $p < .001$; ** $p < .01$; * $p < .05$

To carry out the FA, we included in the analysis those variables that had demonstrated a significant relationship with activation, emotional valence or both, and hence, we had a total of 18 OC measurements. The results demonstrated that all of the variables considered met the assumption of normality, with the exception of the standard deviation of the distance of decline ($Z = 1.772$; $p < .05$), average deviation of the time of decline ($Z = 1.398$; $p < .05$), average deviation of the time of attack ($Z = 1.469$; $p < .05$), and standard deviation of the distance between *onsets* ($Z = 1.375$; $p < .05$). The calculation of the determinant of the covariance matrix ($\det|A| = 0.322 * 10^{-25}$) demonstrated that the collinearity was too high, and hence, the objective of carrying out the FA to reduce the number of variables as a prior step before the multiple regression was shown to be justified. The sphericity assumption was verified, demonstrating that the correlation matrix is significantly different from the identity matrix ($\chi^2 = 1719.847$; $p < .001$). We obtained a modest value for the Kaiser-Meyer-Olkin sampling adequacy coefficient ($KMO = .690$), slightly lower than the minimum recommended (.70), although higher than the minimum that the FA would warn against (.50). This value suggests that the number of variables per factor is sufficient.

After carrying out the FA, we confirmed that all of the OC measurements obtained a commonality of more than 0.7, with the exception of the standard deviation of the slope of decline, standard deviation of the magnitude of the *onsets*, and standard deviation of the magnitude of the valleys that obtained commonalities above 0.6. From this first approximation, we obtained a tetrafactorial structure with a fourth factor formed exclusively by two OC measurements. We forced a trifactorial structure, obtaining a solution that explained 83.93% of the total variance: the first factor accounted for 38.2% of the variance, the second factor accounted for 23.92%, and the third factor accounted for 21.8%.

Figure 3. Eigenvalues as a function of factor in the rotated space

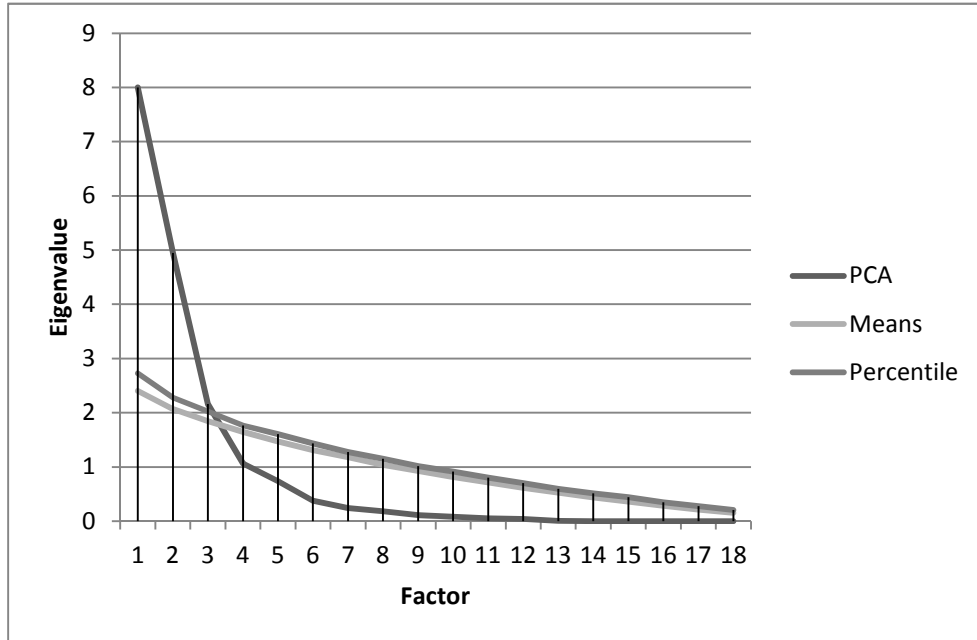


Figure 3 indicates the magnitude of the eigenvalue for each one of the factors obtained through the FA with the original matrix of data and those obtained through the Parallel Analysis (PA). We were able to observe that just three factors in the exploratory FA obtained an eigenvalue higher than $\lambda = 1$ and that they are higher than the average value obtained through the PA method. A trifactorial structure can be concluded from the covariance matrix of the OC.

As stated by Scheirer (1998), a model that attempts to predict the listener's emotional response to music should be capable of generalizing its predictions to different situations than those used for its construction. Given that the stimuli by Vieillard et al. (2008) present a structure designed *ex post facto* for concrete emotional induction, we decided to test our model on a different set of stimuli in a second study.

The Second Study

After selecting the new pieces of music that would serve as a sample for the validation of the initial AF, we confirmed that those that had been classified as happy or calm had on average a more positive emotional valence than those pieces that subjects had classified as sad or tense ($F_{[4,36]} = 16.123, p < .001, \omega^2 = .575$). In the same way, the pieces classified by listeners as happy or tense obtained a significantly higher average level of activation than the pieces categorized as sad or calm ($F_{[4,36]} = 57.352, p < .001, \omega^2 = .827$).

We carried out a second FA for the OC measurements that was identical to the one previously applied for the set of calibration stimuli. The results of this FA returned a determinant near zero ($\det|A| = 6.5 * 10^{-13}$) and a KMO value of .718, exceeding the value obtained in the prior study. Bartlett's test of sphericity was significant ($\chi^2 = 902.667; p < .001$). The three factors obtained explain 73.75% of the total variance, which following rotation, is distributed in terms of 27.25% for the first factor, 26.98% for the second factor, and 19.52% for the third factor. According to the sedimentation graphic, we would have four factors with an eigenvalue higher than that of the unit. However, the fourth factor would increase by only 7% of the total explained variation. The factorial matrix after orthogonal rotation is shown in Table 3.

Table 3: Rescaled factorial scores from the *Onsets* Curve measurements and their correlations with the emotional dimensions in the second study.

	Rescaled Factorial Scores			Correlations	
	Factor 1	Factor 2	Factor 3	Activation	Emotional Valence
Standard deviation mean of the attack distance	,871			-.414**	-.287
Standard deviation mean of the attack	,938			-.459**	-.279

magnitude				
Standard deviation mean of decline magnitude	,821		-.525**	-.150
Standard deviation mean of slope attack	,885		-.161	-.251
Average euclidean distance of decline	,550		-.741***	-.255
Average magnitude of decline	,066		-.024	-.134
Standard deviation mean for the magnitude of onsets	,883		-.270	-.336*
Average duration of decline	,459	,667	-.564***	-.366*
Average distance between onsets	,444	,723	-.795***	-.295
Standard deviation mean of the distance between onsets		,795	-.370*	-.288
Standard deviation mean of the duration of decline		,911	-.533***	-.456**
Standard deviation mean of decline distance		,883	-.479**	-.424**
Standard deviation mean of attack duration		,686	-.287	-.124
Average magnitude of the onsets		,793	.112	-.062
Average magnitude of the decline slopes		,853	.642***	.062
Average magnitude of the valleys		,764	.487**	-.030
Standard deviation mean for the valley magnitude		,628	.177	-.019
Standard deviation mean of the decline slopes		,765	.498**	.062
FACTOR 1			-.540***	-.240
FACTOR 2			-.330*	-.279
FACTOR 3			.485**	-.062

*** $p < .001$; ** $p < .01$; * $p < .05$

We compared the factorial matrices from studies one and two through the correlation of the factorial weights of both structures. As Table 4 shows, there is a high magnitude in the correlation coefficients of the main diagonal. This same table also shows the congruence

index for each pair of factors and the level of significance achieved by this index. We can see a higher congruence between the factorial weights for the factors of the main diagonal and a lower congruence in the rest of the comparisons, except that between Factor 1 in the first experiment and Factor 2 in the second experiment.

Table 4: Correlation matrix for factorial weights and Congruence Index of Wrigley y Neuhaus (1955) in first and second studies

		Second study					
		Factor 1		Factor 2		Factor 3	
		r	IQ	r	IQ	r	IQ
First study	Factor 1	.886***	.929***	.321	.596**	-.709**	-.340
	Factor 2	-.325	0.294	.761***	.855***	-.367	-.041
	Factor 3	-.632**	-0.277	-.575*	-.118	.855***	.905***

*** $p < .001$; ** $p < .025$

Study of the Relationship between the OC and the Emotion Dimensions

After verifying the correspondence between the factorial structures of the OC between the two sets of stimuli used, we proposed studying the relationship between the factor scores obtained for each stimulus and the dimensions of emotion (activation and valence). First, we confirmed that the five variables for analysis (the three scores obtained for each fragment of music in their factor scores, plus activation and valence gathered for the two behavioral studies, the first by Vieillard et al. in 2008 and the second by the authors of this work) are distributed according to normal distribution, except for activation ($Z = 1.552$; $p < .05$). It was also verified that there is no difference in their averages or in the variability of the groups of participants as a result of the stimuli group utilized.

With the assumptions of the analysis confirmed, we proceeded to model a multiple linear regression analysis, employing the method of introducing the factor scores as predictive variables in one block to predict activation on the one hand and emotional valence on the other. In the case of activation, we obtained a model that explained 53.7% of the total variance. The goodness of fit of the model was highly significant ($F_{[3,76]} = 31.576, p < .001, \omega^2 = .555$). Upon analyzing the coefficients of the model, we observe that the three factors of the OC emerge as good predictors of activation. Factor 3 presents a standardized coefficient in the model of $\beta = .523$ ($t = 6.837; p < .001$), emerging as the predictor that explained the most variance, followed by Factor 1 with a standardized coefficient in the model of $\beta = -.518$ ($t = -6.774; p < .001$), and finally, Factor 2, with a standardized coefficient in the model of $\beta = -.111$ ($t = -1.450; p > .05$). Given that the factors were orthogonally rotated, there is no Variance Inflation Factor (VIF) between the factors, and their tolerance in the three cases is equal to that of the unit. Hence, we also observed that the semipartial correlation coefficients for each factor regarding activation are similar to the value of their standardized coefficient.

In the case of emotional valence, we followed the same procedure, obtaining a model that reaches a determination coefficient of $R = .422$, with a percentage of explained variance of 21.8%. In this case, the sum of squares of the model ($SS = 16.974$) is lower than the sum of squares of the residual ($SS = 61.026$). This aspect is reflected in the modest percentage of variance explained by the model, the ANOVA, notwithstanding, being significant, thus signaling the predictive capacity of the model ($F_{[3,76]} = 7.046, p < .001, \omega^2 = .218$). For the regression model for emotional valence, only Factor 2 obtained a standardized coefficient distinct from zero ($\beta = -.431, t = -4.249, p < .001$).

General Discussion

The main objective of this study was to investigate the underlying dimensionality of the various OC measurements and how they can be employed as predictors of the emotional judgment of musical stimuli in listeners. In this way, we attempted to complement the statistical indices useful for predicting activation and emotional valence that are usually extracted from the frequency domain with information taken from the temporal domain. Given that we did not possess a consistent prior theoretical basis with regard to the OC, we decided to work with the greatest possible number of OC measurements to select the best predictors. We observed that the OC measurements obtained tended to relate in a particular way to the emotional dimensions, and given the high collinearity between them, we submitted them to an FA that yielded a factorial structure very similar to the correlation matrix described previously. We see this dimensionality reduction technique as advantageous for three reasons. First, the linear relationship between the predictors made a factorial solution through PCA optimal, given that this method makes it possible to group the variables in factors that maximize the total variance of the elements, calculating the percentage of variance for each factor based on the variance that was not explained by the preceding factor. We avoided collinearity between the resulting factors through an orthogonal rotation of the factorial matrix by using the varimax method. Second, we sought to confirm that the grouping of the variables following the FA was similar to the structure described based on the correlation matrix. Third, with the objective of creating a regression model to predict activation and emotional valence, the FA allowed us to balance the number of cases and predictors to construct the model.

Factor 1 includes measures of the OC related primarily to the variability of the magnitude of the attack and decline of the *onset*, which are also negatively correlated with

activation. After analyzing these elements, we can consider Factor 1 as that which ultimately describes the rhythmic variability of the stimulus. The direction of the relationship with activation suggests that pieces of music with a more regular structure over time generate an increased sensation of activation. This is congruent if we take into account that the musical genres that are traditionally most arousing (hip-hop, trance, dance or pop) are characterized by simple and regular rhythms. This factor also encompasses the Inter Click Interval (ICI), which indicates the average time between the succession of two musical events. With a lower density of events, the slower the succession of *onsets*. Hence, a negative association with activation is expected. Also included in Factor 1 are the magnitude and duration of the decline, which can be explained if we take into account that when the duration of the decline is short and the succession between *onsets* more rapid, a greater sense of activation is expected in the listener.

Factor 3 includes measures that describe the magnitude and variability of the valleys, the magnitude of the *onsets*, and the slope of decline and is positively related to activation. These measures concretely describe how the fall in the musical notes in the piece is produced: the greater the slope of decline (that is, the more abrupt the fall in the musical note) or the greater the magnitude of *onsets* and valleys, the more arousing the stimulus is for the listener. It is noteworthy that measurements of this factor were negatively correlated with emotional valence. Consequently, we believe that this factor can be particularly relevant for determining the discrete emotion that the stimulus elicits in listeners.

The variables included in **Factor 2** are primarily related to emotional valence, including measures of the variability of the OC that indicate that the pieces of music with greater variability in the ICI and in the attack/decline times generate a less pleasing sensation in the listener. The origin of this variability may be owed to the existence of different figures

and punctuation marks in the musical score, and hence, it would not be strange to expect a relationship between the emotional valence awakened in the listener by the stimulus and the rhythmic regularity of the piece. In other words, taking into account that the fall in the time of musical notes also depends on the type of note (eighth notes, quarter notes, etc.), the rhythmic regularity could favor a positive hedonic evaluation of the stimulus. These findings would require us to reconsider the hypothesis that temporal properties primarily influence judgment regarding activation (e.g., Husain, Thompson, & Schellenberg, 2002; Gagnon & Peretz, 2003; Saiz-Clar & Reales, 2017 and Webster & Weir, 2005) but are congruent with the relationship between rhythmic regularity and emotional valence found in Saiz-Clar & Reales (2017).

Based on the interpretation of the resulting factors, we can say in broad strokes that those stimuli with a greater regularity or rhythmic predictability (Factor 1) with a constant transition of notes (Factor 3) and with a lesser variability in the musical notes utilized generate more activation and pleasure in the listener, which, according to the emotional model by Russel (1978, 1979 and 1980), corresponds with the emotion of happiness. This would be one of the most important contributions of this study, although it is a hypothesis that should be replicated in subsequent research.

These results led to the proposal of a second study with a different set of stimuli (although similar in style) on which to repeat the FA from the first study. The objective of the second study was to verify the factorial structure obtained for the first set using a different set of stimuli. Based on the results obtained in the second study and omitting the variables of magnitude of decline, duration of decline and ICI, we can conclude that the factorial structure of the first experiment is reproduced with the second set of stimuli. That is, we were able to prove that the properties of the OC are grouped in a more or less similar way in the two

samples with different stimuli (calibration and validation), which were also tested on different samples of participants.

The high correlation in the main diagonal in the correlation matrix between the factorial weights in the first and second study indicates that there is a high covariation between the factorial weights for common factors in the two samples used. On the other hand, we also confirmed that the magnitude of correlation between factors is null or inversely correlated, as we expected. Regarding the similarity found between Factor 1 in the first experiment and Factor 2 in the second experiment, this could be explained by the observation by Barrett (1986), who found that one disadvantage of the congruence index is the overestimation of the association between factors when they have the same sign in their weights or similar average factorial magnitudes. Taking into account the magnitude of correlation for this pair of factors (Table 4), we can consider that congruence for this pair has resulted in an overestimation due to the method. In view of the results, we can conclude that in the second FA, obtained using a different sample of stimuli and different participants than the first study and with similar factorial weights, we have obtained a factorial structure similar to that obtained in the first study and with similar factorial weights. This fact allows us to affirm that the underlying dimensional structure of the OC of any fragment of music is composed of three dimensions and that the factor scores obtained in this analysis are the basic indices that should be utilized to characterize stimuli that utilize OCs as their independent variables or primary dependent variables.

However, not all OC measurements have demonstrated an effect on emotional judgment. It is interesting that the OC measures that correspond to the attack of the note have not emerged as good indicators of emotional dimensions in view of their correlation. As we stated in the introduction, some prior studies have employed magnitude, slope and attack

times to discriminate between different musical timbres. It is possible that the poor yield of these measures is due to the use of musical stimuli with a single timbre. It would be advisable to explore the effect of different timbres on OC measurements and how these affect the factorial structure obtained.

Finally, we employed the factor scores of the AFs in each of the regression models, which enabled an estimation of the percentage of variance explained by this group of OC measurements for the emotional dimensions. The use of different OC measurements has been shown to be a good indicator of the emotional response of the listener in the dimensions of activation and emotional valence. Working with these curves has been contextualized in recent years, primarily in the field of engineering (Bello et al., 2005; Lee & Kuo, 2006; Marchi et al., 2014; Klapuri, 1999), with few references to the effect of these variables as predictors of the perceptive response in the listener. A recent study (Cameron & Grahn, 2014) finds that musical *onsets* act as acoustic markers in the listener that provide auditory cues influencing the perception of more complex phenomena such as rhythm, meter or beats. The work by Sturm et al. (2015) describes an interesting relationship between the position of the *onset* and brain reactions in the listener through EEG analysis, observing that this is more salient for music interpreted by keyboards. In fact, new algorithms for detecting *onset* position are increasingly based on perceptive and neuroscientific models such as that by Coath et al. (2009). However, to date, we do not know of any study that has examined the capacity to predict emotional responses in the listener through OC measures.

Based on the results obtained in this preliminary study, we can conclude that the physical structure of musical stimuli in the temporal domain explains a large percentage of the variance regarding the emotional response of the listener. That is, it is not only the score that influences emotional response (Saiz-Clar & Reales, 2017); its interpretation could also play

an important role in the model. We do not refer exclusively to the fact that the properties of the wave affect emotional judgment, given that this has been addressed in previous articles (Coutinho & Cangelosi, 2009 and 2011; Gomez & Danuser, 2007; Korhonen, 2004), but are instead proposing that to construct a reliable model of emotional response in the listener, we should include aspects of the interpretation of the piece as a relevant factor. To better predict the listener's emotional judgment of the musical stimulus, these aspects appear to require a mixed approximation that includes measurements of both the notation and interpretation. It would also be important to determine the comparative weight of interpretation on that judgment and how the variation of musical elements in a single piece can modify the listener's judgment. This final point would have important clinical implications for areas such as emotional disorders or certain neurocognitive disorders in which emotion is affected.

CAPÍTULO 4

Predicción de la respuesta emocional continua del oyente a la escucha musical: Efecto de la estructura del estímulo y de la respuesta fisiológica.

Saiz-Clar, E. *; Abad-Tortosa, D.; Serrano, M.A. & Reales, J.M. *

*Departamento de Metodología en CC. Del comportamiento y de la salud -
Universidad Nacional de Educación a Distancia (UNED)

Departamento de Psicobiología – Universitat de València

(manuscrito en preparación)

Uno de los debates latentes en el estudio de los efectos de la música sobre el oyente hace referencia a la naturaleza de la respuesta del oyente a la categorización emocional del estímulo. Este debate gira en torno a si esta respuesta es realmente una reacción emocional (postura emotivista) o una reacción de reconocimiento de la intención comunicativa del estímulo (postura cognitivista). La primera posición plantea que la estimulación musical genera en el oyente una respuesta realmente emocional, mientras que la postura cognitivista no niega la naturaleza emocional de la reacción del oyente pero si cuestiona que siempre que haya una exposición musical el oyente sienta una emoción congruente con la intención comunicativa de la pieza, defendiendo que en la mayor parte de las ocasiones no hay emoción implicada sino un proceso de reconocimiento.

En la dilucidación de esta cuestión y a la vista de los resultados empíricos obtenidos hasta la fecha, puede afirmarse que parece existir una respuesta del sistema nervioso autónomo ante la presentación de estímulos musicales y esta es semejante a la encontrada ante estímulos emocionales presentados en otras modalidades sensoriales como la modalidad visual (Baumgartner, Esslen y Jäncke., 2006; Gomez y Danuser, 2004; Gomez, Stahel y Danuser, 2004; Gomez, Zimmermann, Guttimsen-Schar y Danuser, 2005). Estos autores observan como a medida que se incrementan los valores de activación subjetivos disminuyen los parámetros respiratorios y se incrementa la respuesta de SCL (*Skin Conductance Level*) ante la exposición a estímulos visuales clasificados por sus valores de activación y valencia emocional.

Sin embargo, en el estudio de la respuesta autónoma a las dimensiones de la emoción se encuentran muchos trabajos contradictorios. Por ejemplo, Blood y Zatorre (2001) estudian la respuesta a estímulos clasificados como placenteros, neutros o no placenteros. Estos autores no observan diferencias en la *Skin Conductance Response* (SCR) ni en la temperatura

corporal ante las condiciones descritas aunque en su trabajo no indican el potencial activador de los estímulos usados, por lo que no podemos dilucidar si esta falta de efecto se debe a que en las tres categorías existe el mismo nivel de activación o a que la relación entre la SCR y la activación es inconsistente. Siguiendo esta línea de investigación, Nater, Abbruzzese, Krebs y Ehlert (2006) describen una intensidad menor en el SCL ante estímulos calificados como placenteros y relajantes (Música Clásica) frente a estímulos calificados como no placenteros y activadores (*Heavy Metal*). Sin embargo en este trabajo también es difícil atribuir este fenómeno unívocamente a la valencia emocional ya que los mismos autores reconocen que las evaluaciones subjetivas del potencial de activación de los estímulos utilizados difieren y a que existen diferencias en el tempo de la pieza y en la sonoridad.

Resultados similares sobre la valencia emocional negativa y la respuesta de conductancia eléctrica de la piel fueron descritos por Dellacherie, Roy, Hugueville, Peretz y Samson (2010). Estos autores también encuentran incrementos en la SCR asociados a estímulos musicales con valencia emocional negativa por estructura disonante, pero solo localizaron esta respuesta en un grupo de expertos en música. Es de destacar que el grupo de participantes no expertos no valoró como especialmente negativos los estímulos disonantes, sino que fueron valorados de forma semejante en valencia emocional que los estímulos consonantes. Estos resultados apoyarían, en principio, la hipótesis de la relación entre valencia emocional negativa y los incrementos en SCR. Sin embargo, si la disonancia es una característica del estímulo que genera una respuesta de desagrado en el oyente, esta respuesta tendría que repetirse tanto en sujetos expertos como no expertos en música.

A nivel de respuesta cardiovascular también Blood y Zatorre (2001) describen incrementos en la frecuencia cardíaca ante estímulos clasificados como placenteros (autoseleccionados por los participantes). Los valores promedio obtenidos fueron

estadísticamente superiores con respecto a las condiciones de control (ruido, silencio y música no seleccionada por el participante). Sin embargo, Gomez y cols. (2004), Gomez y Danuser, (2004) y Roy y cols. (2008) no encuentran dicha asociación entre las valoraciones subjetivas de valencia emocional del estímulo y la frecuencia cardíaca. Por el contrario, Gomez y Danuser (2007) encuentran una relación lineal significativa entre los decrementos en la frecuencia cardíaca y las valoraciones subjetivas de valencia emocional negativa, pero no observan una relación positiva entre incrementos en la valoración positiva del estímulo y la frecuencia cardíaca. Este mismo resultado fue obtenido por Sammler, Grigutsch, Fritz y Koelsch (2007) ante estímulos disonantes que habían obtenido una valoración emocional negativa.

Otros estudios han relacionado patrones psicofisiológicos semejantes directamente con categorías emocionales del estímulo musical en lugar de con las dimensiones subyacentes. En estos estudios las respuestas y patrones encontrados no son consistentes, encontrando resultados contradictorios entre los mismos. Khalfa, Peretz, Blondin y Manon (2002) estudiaron la SCR ante estímulos musicales que habían sido categorizados como alegres, tristes, tranquilos o inductores de miedo/tensión. Los estímulos clasificados en las categorías de miedo y alegría obtuvieron mayores índices de valoración subjetiva de activación y mayores incrementos en la respuesta de SCR con respecto a los estímulos calificados como tranquilos o tristes. Sin embargo, Baumgartner y cols. (2006) encuentran estas diferencias en la SCR en los estímulos calificados como tristes e inductores de miedo, atribuyendo estos incrementos en SCR a la valencia emocional del estímulo. Es decir, estos autores observan que los estímulos que ofrecen una respuesta de SCR más intensa son aquellos que son valorados por los participantes como displacenteros en lugar de discriminar, como señalaban Khalfa y cols. (2002), entre los estímulos con mayor potencial activador de los estímulos con menor potencial activador. Así mismo, Baumgartner y cols. (2006), no encontraron diferencias

en la frecuencia cardiaca ante estímulos calificados como alegres, tristes o inductores de miedo. Mientras que Etzel y cols. (2006) sí observaron incrementos en la frecuencia cardiaca ante estímulos inductores de miedo y decrementos ante estímulos tristes. La relación entre estas dos categorías de estímulos es que tienen valencia emocional negativa pero presentan diferentes niveles de activación.

Resultados similares obtuvieron Khalifa, Roy, Della Bella y Peretz (2008). Estos autores observaron una respuesta SCR más intensa ante estímulos calificados como alegres que ante estímulos calificados como tristes. En este estudio los autores analizaron también el efecto del tempo y el ritmo sobre la activación del sistema autónomo pero no observaron respuestas diferenciales. Estos autores atribuyen los efectos descritos a la armonía y no a las propiedades métricas: sin embargo, estos mismos autores observan que las valoraciones subjetivas con las condiciones de solo tempo y solo ritmo resultan bastante indiferentes para los participantes. Es posible que ritmo y tempo, o cualquier otra propiedad aislada de su contexto no produzca el efecto que produce en combinación.

Tal y como señalan autores como Kenner-Mabiala, Gorges, Alpers, Lehmann y Pauli (2007), Gomez y Danuser (2007) o Iwanaga, Kobayashi y Kawasaki (2005), el origen de la variabilidad de resultados obtenidos en los diferentes trabajos señalados, puede deberse a la influencia de determinadas propiedades del estímulo musical sobre la respuesta autónoma. Sin embargo, el estudio de la influencia de dichas propiedades sobre la respuesta autónoma es escaso centrándose en mayor medida directamente en categorías emocionales discretas. Si, tal y como señalaban Balkwill y Thompson (1999) y Balkwill, Thompson y Matsunaga (2004), las propiedades musicales combinadas del estímulo proporcionan las pistas al oyente para reconocer la expresión emocional de la pieza, es decir, para emitir una respuesta cognitiva;

sería coherente pensar que estas propiedades también generan una respuesta en el organismo en el sistema de respuesta fisiológico.

Al respecto, Iwanaga y cols. (2005) plantean en sus conclusiones la existencia de una relación indirecta entre el tempo y la frecuencia cardíaca mediada por parámetros respiratorios. Estos autores señalan la relación entre la modulación de la frecuencia cardíaca como producto de la variación del ritmo respiratorio y cómo éste ha sido relacionado con variaciones en el tempo musical. Este planteamiento se observa también en el trabajo de Gomez y Danuser (2007) en el que encuentran un efecto del tempo sobre algunas medidas respiratorias como los tiempos de expiración e inspiración y sobre la respuesta cardiovascular medida a través de la frecuencia cardíaca. Se hace por tanto necesario estudiar la interacción entre diferentes propiedades del estímulo musical sobre la respuesta autónoma como, por ejemplo, la relación entre sonoridad y tonalidad. Es decir, si bien es importante estudiar el efecto de las propiedades individuales del estímulo musical, es asimismo importante tener en cuenta la relación entre las mismas.

Sin embargo ninguno de los estudios citados tiene en cuenta otra posible fuente de variación de los resultados excepto el trabajo de Coutinho y Cangelosi (2011). Estos autores se basan en los trabajos previos de Schubert (1999), Korhonen (2004) y Coutinho y Cangelosi (2009) quienes de forma general plantean que la estructura del estímulo musical juega un papel importante en la predicción de la respuesta emocional del oyente a las dimensiones de activación y valencia emocional. Esta hipótesis ha sido tratada en trabajos anteriormente citados pero la diferencia con estos autores es que ellos conciben el estímulo musical como un estímulo dinámico, cambiante en el dominio del tiempo y que evoluciona, y por tanto, la respuesta que obtengamos del oyente tampoco debe considerarse estática y global, sino continua, local y considerando el estado previo del sistema.

A la vista de lo expuesto, parece existir cierta estabilidad en los resultados a la hora de predecir qué efecto produce sobre el juicio emocional la manipulación de determinados elementos estructurales del estímulo musical. Sin embargo, no encontramos dicha consistencia cuando tratamos la relación entre diferentes medidas de la respuesta autónoma (como la SCL, la frecuencia cardíaca, etc.) y la respuesta emocional del oyente a la pieza musical. Por ejemplo, podemos esperar que la manipulación del tempo de una pieza musical afecte al juicio del oyente sobre el nivel de activación que le despierta dicha pieza musical, o que una pieza en modo menor genere una valencia emocional más negativa que una positiva (Husain, Thompson y Schellenberg, 2002). Pero resulta más complicado saber qué efecto podemos esperar a nivel fisiológico cuando empleamos estímulos musicales con una cualidad emocional concreta.

Esto resulta problemático por ejemplo, cuando queremos trasladar estos resultados a un ámbito terapéutico: si nuestra intención fuera reducir la respuesta de activación fisiológica de un paciente mediante este tipo de estimulación, sería difícil elegir un set de piezas musicales para este fin. Esta variabilidad de resultados no implica que la estimulación musical no sea una buena herramienta para la intervención terapéutica en este contexto fisiológico, pero si es complicado elegir qué estímulos emplear para ello.

La primera hipótesis que podría explicar la variabilidad entre la respuesta fisiológica al estímulo musical y el juicio emocional es la estructura de la pieza. Si las propiedades del estímulo afectan de forma diferente al juicio emocional, es coherente pensar como Gomez y Danuser (2007) que también la estructura del estímulo tiene un efecto sobre la respuesta. La segunda hipótesis que puede explicar esta variabilidad y que no resulta incompatible con la anterior, es que la respuesta fisiológica se trata de una respuesta que evoluciona con el tiempo y es posible que encontremos más especificidad en la relación entre estructura del estímulo y

respuesta fisiológica cuando introducimos el parámetro tiempo en el modelo. Por tanto, el objetivo general de este trabajo se centra en conocer la relación entre el juicio emocional del oyente a la música, la respuesta fisiológica del mismo y la estructura del estímulo, no solo de forma global como en algunos de los trabajos anteriormente citados, sino teniendo en cuenta el dominio temporal como en Schubert (1999), Korhonen (2004) y Coutinho y Cangelosi (2009 y 2011).

Método

Descripción de la muestra

En este estudio participaron 26 estudiantes universitarios (18 mujeres y 8 hombres) procedentes de diferentes facultades de la Universidad de Valencia, España. Todos los estudiantes que se ofrecieron voluntarios completaron un cuestionario de preselección. De entre los voluntarios, fueron excluidas las personas fumadoras (más de cinco cigarrillos al día), consumidores habituales de drogas, personas con trastornos cardiovasculares, neurológicos, psiquiátricos o con problemas de audición, y aquellos que tomaran alguna medicación que pudiese afectar a la respuesta fisiológica. Además, se seleccionaron solo a aquellas personas que informaron no tener conocimientos formales de música. Los participantes que cumplían los criterios indicados fueron convocados por teléfono dos días antes del experimento solicitando que mantuvieran su patrón normal de ingesta y sueño y evitaran tanto el consumo de alcohol como las actividades físicas extremas durante las 24 horas previas al experimento. También se les solicitó que evitaran fumar y tomar bebidas estimulantes dos horas antes del experimento.

El presente estudio fue aprobado por el *Ethics Research Committee* de la Universidad de Valencia de acuerdo con las normas éticas de la Declaración de Helsinki (1964). Todos los participantes recibieron información verbal y escrita acerca del estudio y firmaron un

consentimiento informado antes de empezar con el experimento. La muestra final fue de 23 participantes (15 mujeres y 8 hombres) de entre 18 y 27 años con una edad promedio de 21,5 años ($DT = 3,10$).

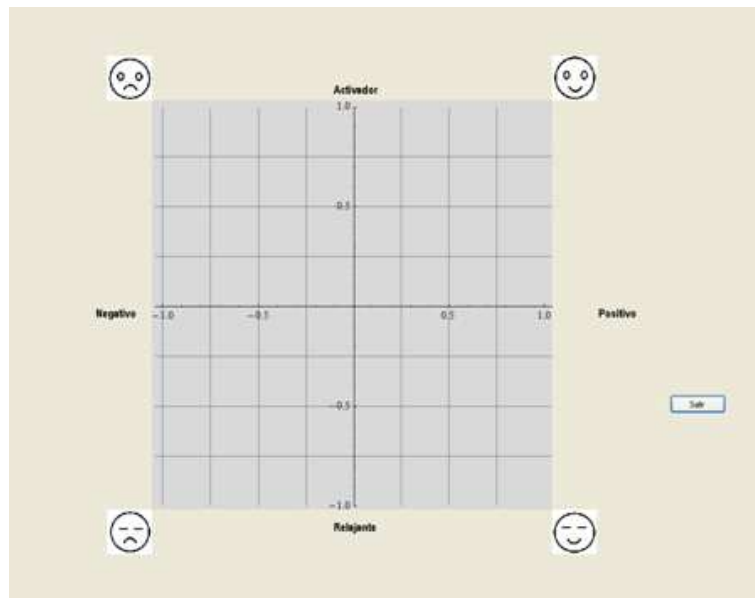
Descripción de la tarea y medidas relativas a la respuesta emocional

Los sujetos participaron de forma individual en sesiones de unos 30 minutos. Todas las sesiones experimentales se realizaron en horario de tarde (entre las 15:30 y las 20:00) con el fin de controlar los ciclos circadianos y sus efectos sobre la respuesta cardiovascular. Mientras los oyentes eran preparados para la adquisición de la respuesta fisiológica, iban leyendo las instrucciones de la tarea por escrito para que todos recibieran la misma instrucción. Si al final de la misma el oyente tenía alguna duda, ésta era resuelta por el experimentador. Tras esta fase, el participante pasaba 10 minutos en reposo para adquirir una línea base de los parámetros fisiológicos y, a continuación, comenzaba la tarea experimental. La tarea encomendada al oyente consistía en evaluar de forma continua en el tiempo el grado de activación y la valencia emocional que le producía la escucha de un set de piezas musicales escritas e interpretadas para piano. Los primeros cuatro estímulos se consideraban de prueba para comprobar que el sujeto había entendido el procedimiento de debía seguir durante el resto de exposiciones. El sujeto elegía el momento de iniciar cada presentación estimular pulsando un botón en el centro de la pantalla.

Durante la escucha musical, los oyentes debían situar el ratón en un plano bidimensional (véase la Figura 1) donde el eje horizontal representaba la valencia emocional que le generaba ese estímulo en ese momento: a la izquierda se situaba el polo de la valencia emocional negativa y a la derecha el de la valencia emocional positiva. El eje vertical representaba el nivel de activación en ese momento durante la escucha siendo el límite superior el nivel de activación más alto y el límite inferior el nivel de activación más bajo. De

esta forma, y de acuerdo al modelo emocional de Russell (1980), el cuadrante superior derecho correspondería a la emoción de alegría mientras que el cuadrante inferior derecho, con valencia emocional positiva y baja activación, cuadraría con la emoción de tranquilidad.

Figura 1: Pantalla de registro continuo de las dimensiones emocionales Activación y Valencia



Del mismo modo, el cuadrante superior izquierdo correspondería bajo este modelo emocional a la emoción de miedo/tensión y el cuadrante inferior izquierdo a la emoción de tristeza. Al final de cada presentación de cada estímulo, los oyentes debían categorizar de forma global el estímulo en una de las cuatro emociones básicas posibles: Alegría, Miedo/Tensión, Tristeza, Tranquilidad, así como emitir una valoración global del nivel de activación y valencia emocional que le había generado el estímulo. Realizada esta tarea comenzaba de nuevo la siguiente presentación estimular.

Variables relativas a la respuesta psicofisiológica

Para el registro del electrocardiograma (ECG) se utilizaron 3 electrodos adhesivos desechables con hidrogel conductor. A partir de la señal del ECG se calculó la **frecuencia**

cardíaca (*Heart Rate* o HR) como indicador de la influencia dual de la actividad simpática y parasimpática del sistema nervioso autónomo (ANS). Dos electrodos Ag/AgCl se utilizaron para registrar la **actividad electrodérmica** (*Electrodermal Activity* o EDA). Los electrodos se colocaron en las falanges de los dedos índice y corazón de la mano dominante. La EDA fue registrada empleando la corrección de línea base inicial (*subject zeroing*) para sustraer el valor absoluto de EDA de cada participante de todos los registros (Hein, Lamm, Brodbeck, y Singer, 2011). La EDA fue analizada a través del Software Matlab 7.12 utilizando el mismo método y procedimiento que en el caso de la HR. Por último, para el registro de la actividad respiratoria se utilizó un transductor que se ajustó en el pecho mediante un cinturón elástico. Esta medida es útil para determinar como de profunda es la respiración de los participantes y calcular la ratio de respiración. Las señales fisiológicas fueron adquiridas y digitalizadas a 1.000 Hz a través del Hardware Biopac MP 150 (BIOPAC Systems, Inc.) y el Software AcqKnowledge 4. Todos los datos fueron analizados usando el Software Matlab. Tanto la HR, la EDA y la actividad respiratoria registradas durante la visualización de cada uno de los estímulos musicales fueron analizadas de forma continua en ventanas de 1 segundo.

En el trabajo de Coutinho y Cangelosi (2011), las medidas fisiológicas consideradas fueron la actividad electrodérmica (EDA) y la frecuencia cardíaca (HR) por ser las medidas más frecuentemente empleadas y que mejor rendimiento han obtenido (Coutinho y Cangelosi, 2011). En nuestro caso, introducimos también un parámetro respiratorio como en los trabajos de Gomez y Danuser (2007) e Iwanaga y cols. (2005) quienes establecen la existencia de una relación entre diferentes propiedades estructurales del estímulo y algunos parámetros del ritmo respiratorio. Concretamente en este estudio hemos empleado como predictores de la actividad del sistema nervioso autónomo 1) la **intensidad de la actividad electrodérmica** (EDA) durante la escucha de cada pieza musical medida en microsiemens (μS), 2) la **frecuencia cardíaca** medida en función de los *beats per minute* (bpm), 3) la amplitud de la

onda R (RWave) en mV, 4) la duración **temporal promedio entre sucesivas ondas R** (RR) en segundos, 5) el **volumen pulmonar** (VP) durante la respiración medido en $V \cdot sec$, y por último 6) la **ratio del ritmo respiratorio** (RResp) medido también en bpm.

Variables relativas a la estructura del estímulo

En nuestro trabajo, hemos decidido usar las medidas empleadas en los trabajos de Saiz-Clar y Reales (2017) y Saiz-Clar y cols. (2017) debido, entre otras razones, a que se encuentran medidos en el dominio del tiempo al igual que las variables fisiológicas (los segundos) y a que caracterizan globalmente las partituras musicales (los primeros). Desde este marco, hemos incluido como variables estáticas en el modelo el tempo, el compás musical, el promedio de notas por compás (PNC), el modo, la frecuencia mínima y la amplitud tonal absoluta (ATA). Para una descripción más detallada de estas variables remitimos al lector a los trabajos citados. Las medidas de la Curva de *Onsets* (CO) fueron calculadas a partir de los parámetros descritos en Saiz-Clar y cols. (2017) y fueron ponderadas a partir de los pesos factoriales obtenidos en dicho trabajo. La obtención de las variables necesarias para calcular dichos parámetros siguió el procedimiento descrito en el trabajo de Saiz-Clar y cols. (2017), utilizando el MIRTtoolbox de Matlab (Lartillot y Toiviainen, 2007). Esta herramienta tan solo permite obtener la posición de los *onsets* y valles por lo que el número de registros por cada segundo era variable en cada pieza musical.

Estímulos musicales empleados en el análisis

Los estímulos empleados fueron recogidos de Vieillard, Peretz, Gosselin, Khalfa, Gagnon y Bouchard (2008). Este conjunto estimular se componía de 10 fragmentos musicales, uno para cada una de las cuatro emociones básicas. No obstante, para nuestro trabajo seleccionamos aquellos ítems que mayor índice de acuerdo en la categorización emocional habían tenido entre los oyentes. De este modo, incluimos 4 piezas musicales de

cada emoción (alegría, tranquilidad, tristeza y tensión) que hubieran sido categorizadas bajo la misma categoría emocional por al menos el 70% de los participantes ($n = 23$). Tan solo dos ítems no cumplían este requisito a partir de los resultados de Vieillard y cols. (2008): los ítems A07 y A10 para la emoción de tranquilidad.

Tabla 1: Comparación entre las valoraciones al estímulo en nuestro estudio y en el de Vieillard y cols. (2008).

	Vieillard y cols. 2008		Nuestro estudio	
	Activación	Valencia Emocional	Activación	Valencia Emocional
A01	1,9	6,6	-0,50	0,13
A02	2	7,1	-0,37	0,29
A07	2	6,9	-0,30	0,21
A10	3,1	6,7	-0,28	0,19
G01	8,3	7,1	0,23	0,29
G13	8,4	7,1	0,36	0,23
G04	8,6	7,1	0,31	0,23
G11	7	7,6	0,13	0,26
P02	8,5	4,4	0,20	-0,01
P04	8,5	3,5	-0,04	-0,17
P05	7,5	4,3	0,15	-0,26
P12	7,4	3	0,03	-0,31
T01	2,6	4,6	-0,31	-0,14
T03	1,8	6	-0,45	0,00
T09	2,9	5,7	-0,21	-0,29
T11	1,9	5,5	-0,28	-0,03

Debido a que las escalas utilizadas por Vieillard y cols. (2008) y las que nosotros hemos utilizado (véase la Figura 1) no se corresponden, hemos evaluado la relación entre los resultados en Activación y Valencia en la muestra original y la nuestra mediante el índice de correlación. Esta correlación la hemos realizado entre los promedios experimentales de activación y valencia emocional que obtuvieron Vieillard y cols. (2008) y los promedios experimentales obtenidos en nuestro estudio (Tabla 1). La correlación entre los promedios de activación ($r = .938$; $p < .001$; IC. 95% = .725/.980) y de valencia emocional ($r = .882$; $p < .001$; IC. 95% = .725/.980) obtenidos en nuestro estudio y en el estudio de Vieillard y cols. (2008) fueron muy elevadas (véase la Tabla 1) mostrando que los sujetos de nuestra muestra actuaron de manera similar a la muestra en donde se tipificaron estos estímulos.

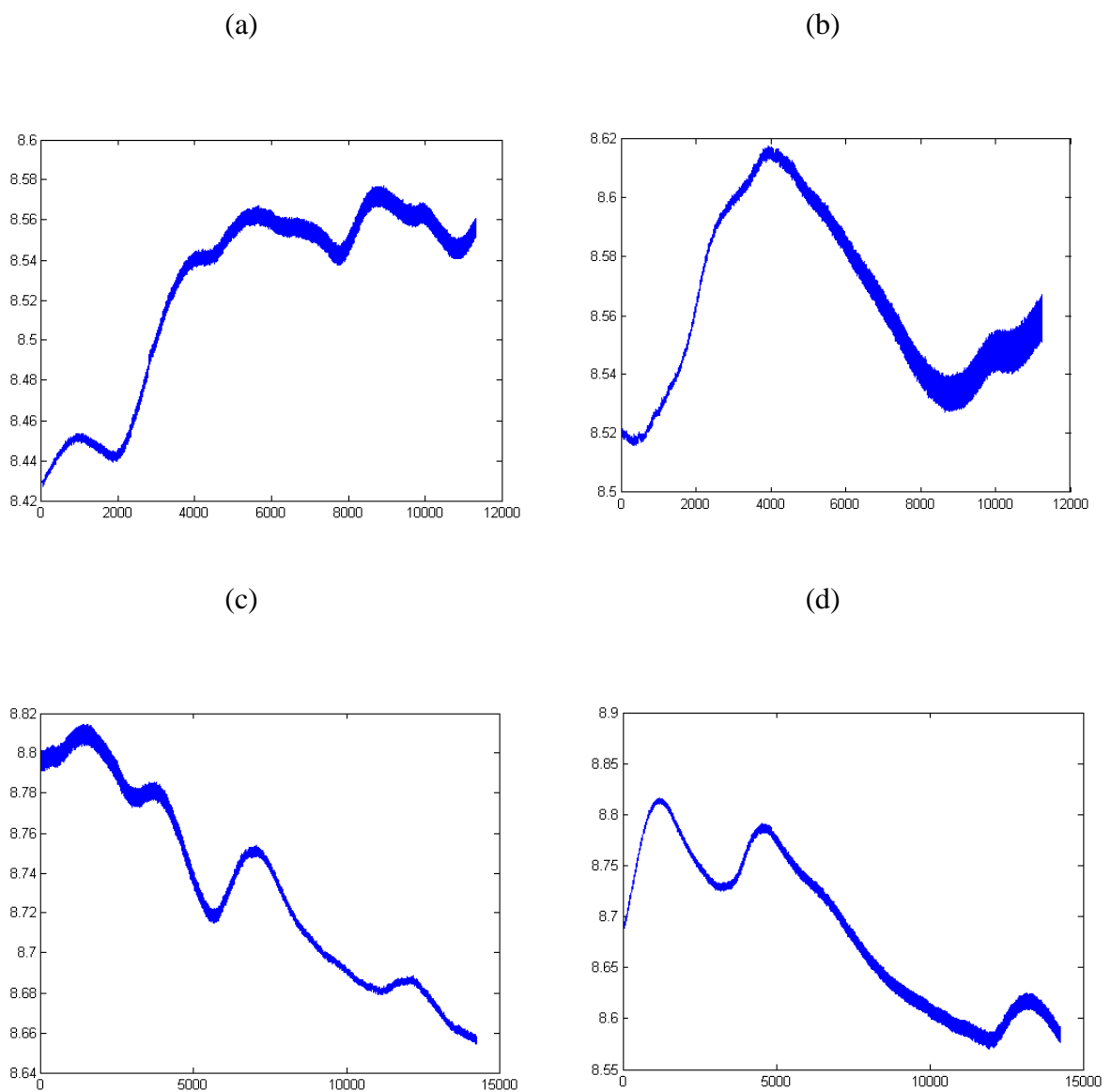
Plan de Análisis

La primera fase del análisis consistió en construir las matrices de datos para las propiedades del estímulo musical, para la respuesta fisiológica del oyente a dicho estímulo y para su evaluación emocional. Las variables fisiológicas fueron muestreadas a 1.000 Hz. Para la evaluación continua de la emoción, obtuvimos muestre de 100Hz y por tanto, unos 100 registros por segundo. Para las propiedades del estímulo musical derivadas de la curva de *onsets*, no podemos hablar de una frecuencia de muestreo ya que, en este caso, las propiedades se definían en relación a los *onsets* de cada pieza musical.

Las piezas musicales fueron segmentadas en su totalidad para cada sujeto. Una vez obtenidas las 23 respuestas para cada estímulo y en cada una de las variables señaladas, se obtenía el valor promedio en cada instante de registro. De esta forma obteníamos una distribución media del estímulo a partir de las respuestas emitidas por los 23 sujetos (véase un ejemplo en la Figura 2). Una vez obtenidas las distribuciones promedio para la respuesta fisiológica y el juicio emocional continuo, estas variables fueron segmentadas segundo a

segundo obteniendo un valor promedio que representara dicho intervalo temporal. De esta forma, tanto las medidas fisiológicas como la evaluación emocional quedaban reducidas a series temporales de entre 8 y 16 observaciones, en función de la duración temporal de los estímulos. De esta forma conseguimos simplificar el análisis notablemente.

Figura 2: Ejemplo de distribución media para la EDA en un ítem alegre (a), un ítem tenso (b), un ítem triste (c) y un ítem tranquilo (d). El eje X representa el tiempo en milisegundos y el eje Y la intensidad en μS .



Los factores de la curva de *onsets* descritos en Saiz-Clar y cols. (2017) fueron extraídos mediante la herramienta MIRtoolbox. Tal y como ha sido indicado previamente, los comandos específicos para estas medidas no permiten extraer de forma directa coordenadas continuas más allá de aquellas que nos permiten situar el inicio y final del *onset*, y el punto de máxima elongación. Por tanto para estas medidas teníamos un número de registros por segundo muy variable (entre 1 y 6 observaciones por segundo). No obstante, se obtuvo el valor promedio para cada una de las 18 medidas descritas en Saiz-Clar y cols. (2017) en cada segmento temporal considerado. A continuación se ponderó el peso de cada medida por su peso factorial en cada uno de los tres factores y se obtuvo el valor final en cada factor para cada pieza musical en cada segundo. Por último, el resto de medidas del estímulo globales consideradas y descritas en Saiz-Clar y Reales (2017), fueron consideradas como valores constantes en el modelo. Esto implicaba que no existía variabilidad dentro del estímulo para cada segmento en estas medidas pero si existía dicha variabilidad inter-estímulo.

En primer lugar, realizamos separadamente dos ANOVAs de un factor inter-sujetos para conocer la diferencia de promedios entre las dimensiones emocionales y la emoción final suscitada por el estímulo, por un lado, y entre las medidas fisiológicas y la emoción por otro. Con estos análisis pretendíamos por un lado verificar si las dimensiones emocionales podían predecir la emoción de acuerdo a la propuesta de Russell (1980). Por otro lado, queríamos saber si podíamos identificar un patrón en la respuesta fisiológica del oyente en función de la categoría emocional del estímulo. Para la realización de estos análisis fueron comprobados inicialmente los supuestos de normalidad, homocedasticidad e independencia. En aquellos casos en los que no se cumplió alguno de ellos se optó por el test no paramétrico de Kruskal-Wallis, equivalente al ANOVA cuando no se cumplen los supuestos de aquel. Todas las variables fueron estandarizadas para tener la misma unidad de medida y se empleó un nivel de confianza del 95% para todos los contrastes.

El siguiente paso fue analizar la relación entre las variables que definen la estructura del estímulo (curva de *onsets*) y la respuesta fisiológica por un lado, y entre estas variables y las dimensiones emocionales por otro. Tras la estandarización comprobamos que, de las 198 medidas de la CO para los 16 estímulos a estudio, dos medidas superaban las tres desviaciones típicas con respecto al valor promedio por lo que fueron considerados *outliers* y eliminados del análisis. Estos datos fueron analizados mediante pruebas paramétricas como el coeficiente de correlación Pearson o la prueba *t* para muestras independientes cuando se cumplían los supuestos paramétricos; en ausencia de ellos se optó por sus homólogos no paramétricos.

Debido a la alta colinealidad de que detectamos entre los factores de la CO en los ítems, consideramos apropiado combinarlos linealmente para poder introducirlos en los posteriores modelos lineales. Inicialmente comprobamos que el determinante de la matriz de correlación era próximo a cero ($\det|A| = .028$) y que la prueba de esfericidad de Bartlett resultaba significativa ($\chi^2[3] = 705.755; p < .000$) esto verificaba que la matriz de correlaciones entre las puntuaciones a los factores de la CO no es una matriz de identidad y el análisis de componentes principales resulta un procedimiento adecuado. Aunque la medida de adecuación muestral de Kaiser-Meyer-Olkin mostraba un valor pequeño ($KMO = .521$), consideramos que existían criterios suficientes para considerar la combinación lineal de estas variables y poder introducir las en el modelo. Obtuvimos un único factor basado en la matriz de correlaciones que explicaba un 86.9% de la varianza total. La puntuación final a la combinación lineal de los elementos se obtuvo mediante el método Anderson-Rubin.

Por último llevamos a cabo sendos modelos de regresión lineal por pasos sucesivos para comprobar cuáles eran los mejores predictores de cada dimensión emocional y que precisión podríamos esperar de la predicción de estos modelos. Nuestro primer modelo incluyó como variable dependiente la activación, segundo a segundo, para cada una de las

piezas musicales. Como predictores fisiológicos incluimos la intensidad de la EDA, la frecuencia cardíaca y el volumen pulmonar en cada intervalo considerado. Por otro lado, como predictores propios de la estructura del estímulo incluimos el tempo, el compás y el ritmo, y la combinación lineal de los factores de la CO. No fueron introducidos más parámetros fisiológicos en el modelo por un problema de colinealidad entre los predictores fisiológicos que no pudo ser eficientemente resuelto mediante la combinación lineal de los elementos. A continuación repetimos este mismo procedimiento pero para encontrar los mejores predictores de la valencia emocional. En este caso hemos introducido como predictores fisiológicos la frecuencia cardíaca, la intensidad de la EDA y el volumen pulmonar a pesar de que no mostraron una asociación lineal con el juicio hedónico pero con el objetivo de mantener el mismo número y tipo de variables introducidas en el modelo para la activación. El resto de variables introducidas fue la combinación lineal de los factores de la CO y como medidas estáticas la frecuencia mínima, el modo y la amplitud tonal absoluta.

Resultados

Evaluación continua de la emoción

En primer lugar, llevamos a cabo dos ANOVAs de un factor para comprobar si la distribución de las medidas continuas en las dimensiones de activación y valencia emocional resultaban consistentes con el modelo teórico de Russell. Empleamos como variable dependiente los valores segundo a segundo de la activación y de la valencia emocional, y como factor en ambos ANOVAS las cuatro categorías emocionales consideradas (alegría, tristeza, tranquilidad y tensión). Los resultados indicaron que el efecto del tipo de emoción fue altamente significativo tanto sobre el promedio de activación referido por los oyentes ($F[3, 194] = 118.797, p < .001, \omega^2 = .647$) y por el promedio de valencia emocional

($F[3, 194] = 91.609, p < .001, \omega^2 = .586$). De forma específica hemos podido observar en los subconjuntos homogéneos de Tukey, las emociones de tensión y alegría obtienen una valoración significativamente más alta en la dimensión de activación que la emociones de tristeza y tranquilidad ($t[196] = 17.620, p < .001, d = 2.509$), siendo incluso significativamente mayor para la emoción de alegría respecto la emoción de tensión ($t[80.99] = 4.120, p < .001, d = .833$).

Tabla 2: Matriz de correlaciones entre medidas fisiológicas y las dimensiones emocionales de Activación y Valencia Emocional

	Activación (Promedio)	Valencia Emocional (Promedio)
Actividad Electro dérmica (Promedio)	.166*	-.043
Frecuencia Cardíaca (Promedio)	.301***	-.006
Tiempo R-R (Promedio)	-.275***	-.030
Intensidad onda R (Promedio)	.165*	-.032
Volumen pulmonar	-.004	.031
Ratio Respiratorio	-.016	-.080
Activación (Promedio)	1	-.056
Valencia Emocional (Promedio)	-.056	1

* $p < .05$; ** $p < .01$; *** $p < .001$

Asimismo observamos un patrón semejante para la valencia emocional donde las emociones positivas de alegría y tranquilidad eran por término medio, valoradas con una valencia emocional más positiva que las piezas asociadas a emociones negativas como la tensión y la tristeza $t[182.62] = 15.450, p < .001, d = 1.846$. Los estímulos asociados a la emoción de tensión son valorados significativamente como menos placenteros que los estímulos tristes ($t[98] = 3.517, p < .01, d = .681$). En los gráficos 3 a 10 (remitimos al lector al capítulo 4 de la sección de material suplementario) podemos observar cómo se

distribuye la respuesta promedio del oyente a cada dimensión emocional en función de la categoría emocional de la pieza.

También pudimos observar un efecto del tipo de emoción sobre la intensidad promedio de la EDA ($\chi^2[3] = 9.384, p < .05$), el promedio de la Frecuencia Cardíaca ($F[3, 198] = 6.424, p < .001, \omega^2 = .089$), el promedio del intervalo R-R ($F[3, 198] = 4.671, p < .01, \omega^2 = .066$), la amplitud promedio de la onda R ($\chi^3[3] = 22.287, p < .001$) y el ritmo respiratorio promedio ($\chi^3[3] = 11.578, p < .01$) aunque en todos los caso este efecto es muy pequeño. En cuanto a la relación entre las medidas fisiológicas consideradas y las dimensiones emocionales (tabla 2), encontramos una asociación significativa entre la activación y la intensidad de la EDA ($r = .166, p < .000$), la frecuencia cardíaca ($r = .301, p < .001$), el intervalo R-R ($r = -.275, p < .001$) y la amplitud de la onda R ($r = -.165, < .001$). La valencia emocional no ha mostrado una asociación significativa con ninguna de las medidas fisiológicas consideradas

Relación entre la CO, la respuesta fisiológica y las dimensiones emocionales

Se ha encontrado una relación lineal entre la activación promedio y las puntuaciones en los factores 1 ($r = -.248, p < .001$) y 3 de la CO ($r = .538, p < .001$). No hayamos una relación significativa entre ninguno de los factores de la CO y la valencia emocional promedio. En cuanto a la relación entre los factores de la CO y las medidas fisiológicas, únicamente encontramos una asociación relevante entre la frecuencia cardíaca y el factor 1 de la CO ($r = .225, p < .001$).

Estructura estática del estímulo como predictor de la Rs. Fisiológica

En la tabla 3 se presentan las correlaciones entre propiedades estáticas del estímulo y las medidas fisiológicas consideradas. Se ha observado una relación significativa entre el tempo de la pieza musical y los parámetros de la respuesta cardiovascular estudiados, detectando un incremento en el ritmo cardiaco a medida que se incrementa el tempo musical de la pieza. Este patrón no se aprecia con la regularidad rítmica. La frecuencia mínima de la pieza musical se relaciona positivamente también con el ritmo respiratorio. Aunque sin duda la relación más notable es la encontrada entre la amplitud tonal de la pieza musical y la intensidad de la respuesta electrodérmica.

Tabla 3: Matriz de correlaciones entre medidas fisiológicas y parámetros estáticos del estímulo

	EDA	HR	Resp Vol	R-R	R-Wave	R Resp
Tempo	.016	.232***	.079	-.187**	.216**	.069
Ritmo	.096	.027	.019	-.073	.076	.048
FM	.072	.123	-.030	-.091	.082	.220**
ATA	-.646***	-.033	.016	.059	-.153*	-.161*

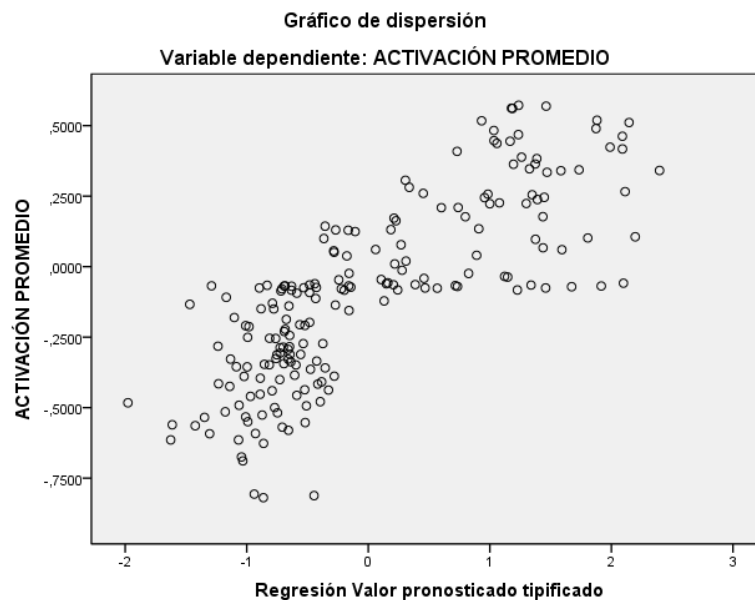
* $p < .05$; ** $p < .01$; *** $p < .001$

Por último, y dada la naturaleza categórica de las variables modo y compás, analizamos la relación entre estas variables y las medidas fisiológicas consideradas mediante una prueba t de grupos independientes en cada caso. El tipo de compás musical muestra un efecto sobre la intensidad promedio de la onda R ($t[167.771] = -2.854, p < .01$), la razón respiratoria ($t[186.687] = -3.639, p < .001$) y la ratio del pulso periférico ($t[195.291] = 5.283, p < .001$). Se ha observado un patrón semejante para el modo, con un efecto significativo sobre la ratio respiratoria ($t[132.788] = 2.947, p < .01$)

Predictores de la activación y la valencia emocional

La primera variable incluida en el modelo es el tempo musical ($\beta = .561, t = 11.087, p < .001$) que vuelve a emerger como el mejor predictor de la activación. Esta variable por sí sola explica el 52% de la variabilidad observada. La siguiente medida que cumple el criterio de entrada y que produce un cambio significativo en el coeficiente de determinación es la combinación lineal de los factores de la CO ($\beta = .212, t = 4.408, p < .001$), seguido por las medidas fisiológicas EDA ($\beta = .149, t = 3.339, p < .01$) y FC ($\beta = .137, t = 2.999, p < .01$), siendo la última variable incluida el ritmo ($\beta = .139, t = 2.919, p < .01$). El modelo final ($F[5, 193] = 64.271, p < .001; \omega^2 = .631$) incluye cinco predictores y un coeficiente de determinación de $R = .794$ que explica un 63% de la varianza de la activación.

Figura 3: Gráfico de dispersión del pronóstico de la activación predicha y el valor empírico de activación.

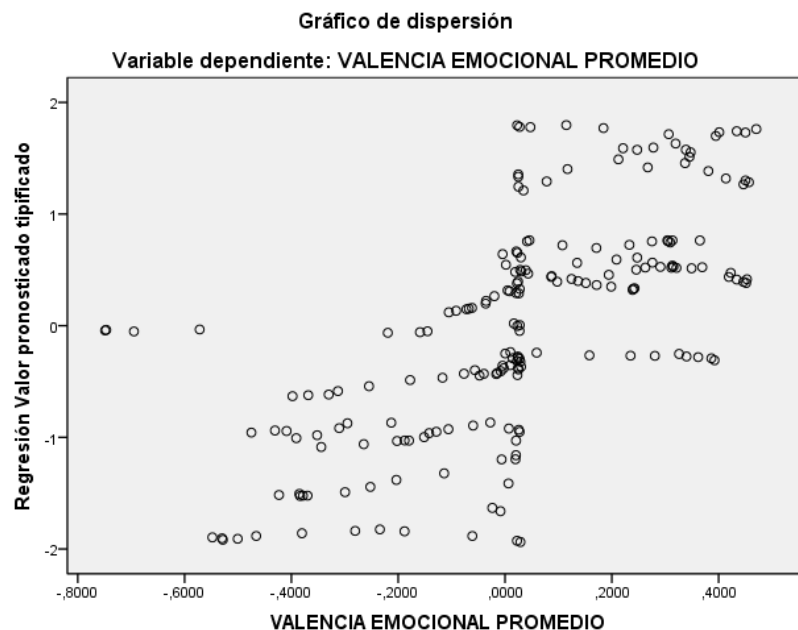


Como puede apreciarse en la figura 5, el ajuste entre los valores de activación obtenidos experimentalmente y los predichos a partir de estas variables es muy elevado. La

correlación entre outputs y targets es muy elevada ($r=.794$; $p<.001$) y el residuo se distribuye normalmente ($Z=.564$; $p>.05$) pero no es aleatorio según el test de rachas.

Los predictores incluidos en el modelo presentan factores de inflación de la varianza (FIV) y tolerancia próximos a la unidad, exceptuando la variable tiempo la cual presenta una ligera elevación. Esta elevación del FIV para el tiempo, puede considerarse normal a la vista de la capacidad predictiva que presenta por si misma esta variable sobre la activación. Por el contrario, las variables excluidas de compás musical y volumen pulmonar presentan niveles de tolerancia y FIV inaceptables para ser incluidos como predictores.

Figura 4: Gráfico de dispersión del pronóstico para la valencia emocional y el valor empírico de la valencia emocional.



A continuación, tratamos de realizar un modelo predictivo de la valencia emocional (véase la Figura 6). Observamos que obteníamos un modelo lineal simple ($F[1,192] = 7.375, p < .001, \omega^2 = .550$) compuesto exclusivamente por la variable modo ($\beta = -.742, t = -15.333, p < .001$) que explicaba un 55% de la varianza ($R = .742$). El resto de variables no cumplían el criterio estadístico para la entrada en el modelo o resultaban

altamente colineales con la variable modo. Por ello realizamos un segundo modelo eliminando esta variable.

En esta ocasión obtuvimos un modelo con dos predictores ($F[1, 192] = 72.085, p < .001, \omega^2 = .430$), donde únicamente entraban en él la Frecuencia Mínima ($\beta = .569, t = 10.331, p < .001$) y la amplitud tonal absoluta ($\beta = .259, t = 4.697, p < .001$). Este modelo explica un 43% de la varianza de la valencia emocional ($R = 0.655$). A la vista de los valores próximos a la unidad de Tolerancia y FIV, estos predictores no resultan colineales. La relación entre los valores predichos por el modelo y los valores experimentales de valencia emocional (véase la Figura 6) es muy elevada ($r = .744, p < .001$). El error de predicción se distribuye normalmente ($Z[198] = .608, p < .854$) de manera no aleatoria según el test de rachas. Ninguna variable fisiológica evaluada resultó ser predictora de la valencia emocional.

Conclusiones

El objetivo principal de este trabajo se centraba por un lado en conocer el posible papel mediador de la respuesta fisiológica del oyente al juicio emocional, intentando saber si existía una conexión entre la estructura del estímulo, la respuesta del SNA y el juicio emocional del oyente.

Sobre el juicio emocional del oyente

Hemos podido comprobar que existe una coherencia entre la evaluación emocional que hace el oyente del estímulo en el dominio tiempo y la valoración global que realiza del mismo. Las piezas que eran valoradas globalmente como más activadoras por nuestros oyentes fueron las piezas de alegría y tensión. Estas mismas piezas obtienen promedios superiores de activación en las series continuas frente a las piezas musicales tranquilas y

tristes. Estos datos no sólo confirman la adecuación y vigencia del modelo bifactorial de Russell (1978, 1979, 1980) para el estudio de la emoción, sino también su estabilidad intersujetos dado que los valores obtenidos en Vieillard y cols. (2008) son semejantes a los obtenidos en nuestra muestra. También Schubert (2013), en una revisión sistemática de diferentes trabajos en este ámbito, concluye que existe una consistencia entre la emoción percibida en el oyente y la intención comunicativa de la pieza musical expuesta.

La coherencia encontrada entre la respuesta global del oyente y su respuesta continua tiene interesantes aplicaciones prácticas y en investigación. Podemos considerar el juicio global del oyente, como una medida fiable del juicio emocional sobre el estímulo musical, siendo en determinados contextos aplicados y de investigación más práctico y sencillo obtener solamente la evaluación global del estímulo. Del mismo modo, esta coherencia demuestra que de la evaluación global que hayamos obtenido de un estímulo musical, podemos esperar que la evaluación en el dominio tiempo sea semejante, pudiendo ampliar, en futuras investigaciones, el corpus estimular abarcable sin necesidad de una experimentación previa continua con la pieza musical. Es decir, podemos aumentar el corpus estimular para el estudio de la respuesta continua explorando la respuesta global de un conjunto de oyentes.

Por otro lado, también se constata de nuevo en este trabajo que las piezas calificadas como alegres y tranquilas son evaluadas como más placenteras que las piezas calificadas como tensas por los oyentes. Una especial mención recibe la relación entre las piezas categorizadas como tristes y la valencia emocional del oyente. Como se ha venido repitiendo en otros trabajos (Saiz-Clar y Reales, 2017), las piezas tristes no parecen resultar especialmente displacenteras para el oyente obteniendo promedios cercanos al valor intermedio del rango posible de respuestas. Recientemente Kawakami, Furukawa, Katahira y Okanova (2013) explican que la evaluación positiva general y recurrente que se observa con

las piezas musicales categorizadas como tristes puede tener una explicación que implicaría aumentar el rango posible de emociones en las tareas de categorización. Algunas de estas piezas que para el oyente encajan categóricamente como tristes realmente son vistas por los oyentes como románticas o nostálgicas.

También hemos observado un patrón interesante en la respuesta continua al estímulo musical, detectando que los oyentes emplean al menos entre 0.5 y 1 segundo en emitir un juicio sobre el poder activador de la pieza musical, es decir, registramos movimiento en el cursor pasado por término medio este periodo de tiempo. En cuanto a la valencia emocional, hemos observado que este lapso puede incrementarse hasta los dos segundos de exposición (ver gráficos 3 a 10 en la sección de material suplementario). Es decir, parece que los oyentes realizan primero el juicio emocional sobre la activación y después sobre la valencia emocional. De este experimento desconocemos si el fenómeno se debe a que resulta más sencillo desplazar el curso horizontalmente vs. verticalmente, o se debe a que en las instrucciones siempre se comienza mencionando la activación y después la valencia emocional. En futuros trabajos tenemos que estudiar este efecto cambiando la distribución de los ejes entre sujetos.

Independientemente del origen de la discrepancia entre el lapso de respuesta a cada una de las dimensiones emocionales, sí que podemos concluir que a la hora de emitir un juicio emocional el sujeto necesita por término medio de 0.5 a 2 segundos para realizar el mismo. Durante este tiempo, también hemos **observado en la EDA una respuesta de orientación**, es decir, existe una respuesta fisiológica que antecede al juicio emocional de forma que, una vez que el oyente evalúa su estado (que necesita muy poco tiempo para hacer ese juicio), se encuentra preparado para evaluar o más bien, etiquetar dicho estado. Por último es también interesante la persistencia en el juicio emocional del estímulo, no solo inter-sujetos como

hemos podido comprobar previamente, sino también intra-sujetos a excepción de los estímulos tristes. En el resto de piezas musicales observamos una consistencia en las **curvas** que representan la posición de la evaluación emocional de la pieza promedio y el tipo de emoción que suscita.

Sobre la respuesta fisiológica y el juicio emocional

A la vista de trabajos como el de Khalifa, Roy, Rainville, Dalla-Bella y Peretz (2008), debemos analizar detenidamente la evolución temporal de la respuesta fisiológica del oyente y el ajuste de la misma a su respuesta emocional. Estos autores observaron una latencia de respuesta de alrededor de 30 segundos para detectar patrones de respuesta fisiológica discriminantes de la emoción suscitada por la pieza. En nuestro caso, creemos conveniente al menos eliminar del análisis los primeros instantes de la respuesta del oyente para eliminar la respuesta de orientación inicial (Bradeley, Codispoti, Cuthbert y Lang, 2001). Esto ayudaría también a controlar el desfase encontrado en la respuesta cognitiva. Sin embargo, también sería interesante dedicar otro estudio al modelado de esta respuesta inicial dado que es aquí donde el factor tiempo y el estado previo del sistema juegan el papel más importante debido a los *delays* de los diferentes planos de respuesta.

Sobre la estructura del estímulo

Nuevamente hemos observado una relación entre las medidas de la CO y el juicio emocional del oyente pero esta vez, se ha podido comprobar cómo esta relación se encuentra asociada también con el tiempo. Es decir, existe una asociación entre la estructura del estímulo y la evaluación global de la pieza musical, especialmente para la activación que suscita la pieza musical en el oyente. Sin embargo no hemos hallado una relación relevante entre la respuesta fisiológica del oyente en el dominio tiempo y los factores de la curva de *onsets* descritos en Saiz y cols. (2017). Aunque esto puede deberse a un pobre muestreo de las

propiedades de la CO en el dominio tiempo y que consideramos debería ser resuelto en futuros trabajos. Sin embargo, sí que hemos hallado en la línea de Gomez y Danuser (2007), algunas relaciones significativas entre parámetros de la respuesta fisiológica y la estructura estática del estímulo. Esto nos invita a pensar que si bien en el dominio tiempo puede no ser tan manifiesta la relación entre la estructura del estímulo y la medida fisiológica, a nivel de parámetros globales si encontramos un efecto de las propiedades del estímulo y este plano de la conducta.

Otra posible explicación a estos resultados implicaría considerar un orden diferente para los elementos del modelo que estamos considerando y pensar que la evaluación subjetiva del oyente sobre el juicio emocional comprende por un lado información sobre la estructura del estímulo y por otro lado, información sobre el estado fisiológico del oyente. Es decir, ante la tarea solicitada el oyente emite un juicio emocional que comprende por un lado las características del estímulo y su procesamiento (*¿qué estoy escuchando?*), y por otro lado, una evaluación del estado interno del oyente (*¿Qué estoy experimentando?*), para dar respuesta a la demanda de la tarea (*¿qué estoy sintiendo?*), esto sería congruente además con las demoras en el tiempo de respuesta observadas en la evaluación cognitiva del estímulo. De ser así, entonces la concepción de Coutinho y Cangelosi (2011) sería la más próxima a la experiencia real del oyente sin embargo, esta concepción no puede explicar la relación entre parámetros estáticos del estímulo como el tempo, modo, etc. y la respuesta fisiológica que también fue encontrada previamente por ejemplo en Gomez y Danuser (2007) o Husain y cols. (2002).

Sobre la relación entre la estructura del estímulo musical y la respuesta fisiológica del oyente, hemos hallado un patrón interesante de respuesta entre el **tempo musical** y la respuesta cardiovascular. Concretamente, observamos un incremento significativo de la

frecuencia cardiaca y de la intensidad de la onda R, así como una disminución del tiempo entre beats como consecuencia del incremento del tempo. Es decir, en este estudio constatamos la relación planteada en otros trabajos como (Krumhansl, 1997; Gomez y Danuser, 2007; Kallinen, Saari, Ravaja, y Salminen, 2006; Khalfa y cols. 2002; Vander Ark y Ely, 1992; Witvliet y Vrana, 2007) entre el ritmo cardiaco y el tempo musical. Esta relación deberá ser explorada con más profundidad a partir de otros parámetros del registro cardiaco. Sin embargo no hemos encontrado una asociación significativa entre el tempo musical y el ritmo respiratorio, es posible que debamos incluir otras medidas como el tiempo y volumen de inspiración y expiración como en Gomez y Danuser (2007). Por otro lado, la relación entre la **amplitud tonal** de la pieza musical y la intensidad de la respuesta electrodérmica indica que las piezas con una menor variabilidad tonal generarían una mayor activación en el oyente, esta respuesta se refleja también en un incremento de la intensidad de la onda R y del ritmo respiratorio en el oyente. La relación entre el **compás musical** y las medidas fisiológicas consideradas muestra que las piezas escritas en un compás cuaternario parecen generar en el oyente una respuesta menos activadora que la que generan las piezas ternarias. Un patrón semejante parece generar también el **modo**, observando que las piezas escritas en modo mayor generan en el oyente un menor ritmo periférico y una tasa respiratoria más elevada, es decir, las piezas en modo mayor generarían una respuesta de activación más intensa en el oyente.

En resumen, una pieza con un tempo musical elevado, con una amplitud tonal pequeña, escritas en un modo mayor y un compás ternario sería el tipo de pieza musical que mayor activación fisiológica generaría en el oyente. Sin embargo, la magnitud de algunas de estas diferencias es muy pequeña por lo que deben ser tomadas con cautela. Resulta llamativa la relación entre la amplitud tonal absoluta y la intensidad de la EDA y que pese a esto, la relación entre la EDA y la valencia emocional sea nula. En este sentido, la ATA resulta un

buen predictor de la activación del oyente pero la intensidad de la EDA que genera la manipulación de la ATA, no la convierte en un predictor efectivo de la valencia emocional.

Tal y como hemos ido señalando en esta sección, nos hemos encontrado con algunas limitaciones que pueden resumirse en unos pocos puntos. Por un lado, es necesario que podamos obtener una muestra mayor de valores en cada segmento temporal para las medidas de la CO. Para ello es necesario mejorar la programación de la función empleada para la obtención de la curva de *onsets*. Incrementando el número de registros podremos obtener valores más precisos de la estructura del estímulo en el instante t. Por otro lado, hemos usado solo cinco variables para describir la respuesta fisiológica oyente, consideramos necesario explorar otras variables y si fuera necesario, combinarlas linealmente. También debemos introducir otras características del estímulo musical importantes muy importantes que no hemos tenido en consideración como variaciones en la sonoridad, en el timbre o en la orquestación. Por último, consideramos que será necesario incrementar el número de estímulos en los análisis para replicar estos resultados, y realizar un nuevo experimento en las mismas condiciones con otro corpus estimular para seguir estudiando la generabilidad de nuestro modelo en otros contextos estimulares.

Entender el mecanismo fisiológico que subyace a la escucha musical nos ayudaría a comprender y poder controlar el efecto que podemos esperar cuando empleamos este tipo de estimulación a nivel terapéutico. Si analizamos la bibliografía actual observamos múltiples contextos en los que se puede emplear la música como instrumento para reducir la activación fisiológica, por ejemplo ante pruebas médicas invasivas, cirugías, pequeñas intervenciones o tratamientos médicos prolongados como quimioterapias (Gómez-Romero, Jiménez-Palomares, Rodríguez-Mansilla, Flores-Nieto, Garrido-Ardila y López-Arza, 2017; Johnson, Raymond y Goss, 2012; Lee, Chao, Yiin, Chiang y Chao, 2011; Mejía-Rubalcava; Alanís-Tavira, Mendieta-Zerón y Sánchez-Pérez, 2015; Miller y Bertnatzky, 2005; Najafi, Mohades

Rafii y Haghani, 2017; Nieto-Romero, 2017; Voss, Good, Yates, Baun, Thompson y Hertzog, 2004; Wang, Kulkarni, Dolev y Kain, 2002, entre otros). Pero para ello necesitamos poder establecer una relación más clara entre la estructura del estímulo, el afecto y la respuesta fisiológica del oyente. Este trabajo avanza en ese aspecto teniendo en cuenta el efecto que tiene además el factor tiempo en el modelo.

CAPITULO 5.

Modelado de la respuesta emocional y fisiológica al estímulo musical.

Una aproximación desde las redes neuronales recurrentes.

Saiz-Clar, E.* ; Abad-Tortosa, D.; Serrano, M.A. & Reales, J.M.*

*Departamento de Metodología en CC. Del comportamiento y de la salud -
Universidad Nacional de Educación a Distancia (UNED)

Departamento de Psicobiología – Universitat de València

(manuscrito en preparación)

Introducción

Uno de los grandes debates en el estudio de la emoción y de la percepción musical gira en torno a clarificar cuál es la naturaleza del juicio emocional emitido por el oyente a nivel experimental (Coutinho y Cangelosi, 2011). En este sentido se plantea si la respuesta que obtenemos del oyente sobre la cualidad emocional del estímulo es una categorización del mismo sin aspectos emotivos (naturaleza cognitiva del juicio) o se trata realmente de una emoción realmente sentida durante la escucha (naturaleza emocional del juicio). Bigand, Vieillard, Madurell, Merozeau y Dacquet (2005) ya trataron esta cuestión en un intento de determinar definitivamente si la respuesta del oyente era realmente una sensación o se trataba de un juicio cognitivo. Para ello, estudiaron el tiempo necesario para que un oyente pudiera determinar la cualidad emocional de una pieza musical. Observaron que bastaba un segundo de exposición a una pieza musical para emitir dicho juicio, lo cual implicaba, según los autores, que el procesamiento de ese estímulo necesariamente debía haberse llevado a cabo a nivel de sistema nervioso autónomo no siendo posible, por una cuestión temporal, que este hubiera sido procesado cognitivamente.

En un estudio posterior, Gomez y Danuser (2004 y 2007) pusieron a prueba la hipótesis de que si la respuesta del oyente al estímulo musical era de naturaleza emocional, necesariamente esta respuesta debía tener un reflejo en el sistema de respuesta fisiológico comparable al observado en otras modalidades sensoriales como la visión. En sucesivos trabajos comprobaron que, efectivamente, existían diferencias en la respuesta fisiológica de los oyentes en función del juicio emocional que manifestaban, y que esta respuesta era similar a la encontrada en la modalidad visual con un test de asociación implícita (Gomez y Danuser, 2004). Otro hallazgo importante de este trabajo es que comprobaron que la estructura del estímulo podía ser un buen predictor de dicha respuesta fisiológica.

El efecto de la estructura del estímulo sobre la respuesta fisiológica del oyente es una hipótesis abordada ya por Krumhansl (1997) quien observó que diferentes propiedades del estímulo generaban diferencias en la respuesta de conductancia eléctrica de la piel o en la respuesta cardiaca (v.g., Gomez y Danuser, 2004 y 2007; Kallinen, Saari, Ravaja, y Salminen, 2006; Khalfa, Peretz, Blondin, y Manon, 2002; VanderArk y Ely, 1992; Witvliet y Vrana, 2007).

Propiedades físicas y teóricas del estímulo musical

A la hora de determinar las propiedades del estímulo que pueden jugar un papel modulador de la respuesta emocional fisiológica y cognitiva del oyente, debemos distinguir entre dos tipos de propiedades empleadas en los diferentes trabajos previamente citados. Mientras que en Krumhansl (1997), Gomez y Danuser (2007) o Saiz-Clar y Reales (2017) se emplean propiedades propias de la teoría musical como son el tempo, el modo, la acentuación, el compás musical, etc., en los trabajos de Coutinho y Cangelosi (2009 y 2011), Eerola, Lartillot y Toiviainen (2009), Korhonen (2004) y Saiz-Clar, Reales y Serrano (2017), entre otros, emplean medidas propias de la estructura física del estímulo, bien en el dominio de la frecuencia o en el dominio del tiempo.

En el primer grupo de trabajos, además, podemos clasificar las propiedades del estímulo en, al menos, dos grandes grupos. Por un lado tendríamos las propiedades relacionadas con la métrica del estímulo (Gomez y Danuser, 2007), es decir, con la sucesión temporal de los eventos sonoros o notas musicales (v.g., el tempo, el ritmo, la acentuación o el compás musical entre otras). Por otro lado, tendríamos aquellos que han utilizado propiedades relacionadas con la tonalidad del estímulo (es decir, con la cualidad sonora) donde encontraríamos propiedades como el modo, el timbre musical, la altura del tono, la escala musical, etc. Balkwill y Thompson (1999), en su teoría de la respuesta dual, plantean que las

propiedades métricas y tonales afectan de forma diferencial al juicio emocional, de forma que las primeras afectarían en mayor medida a la sensación de activación en el oyente, mientras que las propiedades tonales lo harían con la valencia emocional del estímulo. Esta estructura ha sido retomada en trabajos recientes como el de Fritz et al. (2009), y más recientemente en Droit-Volet, Ramos, Bueno y Bigand, (2013), Mathur, Vijayakumar, Chakrabarti y Singer (2015), Saiz-Clar y Reales (2017) o Wong, Clocca, Chan, Ha, Tan y Peretz (2012).

Sin embargo, en contraste con el elevado pero limitado número de propiedades que proporciona la teoría musical, existe una gran variedad de parámetros físicos del estímulo musical que pueden ser empleados para modelar la respuesta emocional del oyente y muy poca información sobre su verdadera contribución a la hora de predecir dicha respuesta. A pesar de que resultan predictores eficientes (Eerola et al. 2009; Coutinho y Cangelosi, 2009 y 2011; Saiz-Clar et al. 2017), la heterogeneidad de estas medidas y la dificultad de su manipulación real en contextos prácticos suponen una limitación cuando queremos aplicarlas, por ejemplo, en un contexto clínico.

Tal y como proponen por ejemplo Coutinho y Cangelosi (2009), Eerola et al., (2009) o Saiz-Clar y cols. (2017), a la hora de proponer un modelo de predicción de la respuesta emocional a la escucha musical, es importante que estudiemos los predictores que vamos a introducir en el mismo e incluyamos aquellos que resulten relevantes para predecir la respuesta, con el objetivo de evitar en la medida de lo posible la redundancia, la colinealidad y el sobreajuste en el input del modelo.

Medición continua del estímulo musical

Uno de los primeros trabajos que plantean la necesidad de estudiar la respuesta del oyente a la escucha musical en el dominio del tiempo es el realizado por Schubert (1999). Según su propuesta, la pieza musical es un estímulo dinámico y cambiante en el tiempo

debido a que evoluciona conforme se desarrolla. Por tanto, plantea la necesidad de estudiar la respuesta del oyente no solo de forma global como se venía realizando hasta la fecha, sino atendiendo a la respuesta del oyente en el dominio del tiempo. Para ello plantea el uso del análisis de series temporales pero observa que este método presenta limitaciones para representar la relación entre los diferentes componentes. Otro trabajo precursor de este enfoque temporal de la respuesta fue desarrollado por Korhonen (2004) quien aplicando los modelos ARX (*autoregressive with exogenous terms*) intentó, por un lado, establecer la consistencia temporal del juicio emocional del oyente al estímulo musical y, por otro lado, determinar cómo la estructura del estímulo podía predecir esta evolución temporal del juicio emocional. Lamentablemente, los modelos autorregresivos presentan una serie de problemas, de los cuales, el más relevante es que están basados en la relación lineal entre los predictores siendo éste un supuesto restrictivo que deja fuera del mismo otro tipo de asociaciones (cuadráticas, logísticas, etc.), resultando más conveniente otro tipo de modelos menos restrictivos como podrían ser las Redes Neuronales Artificiales o RNA.

Sin embargo, los modelos ARX presentan una ventaja que no tienen otros enfoques dadas sus características. La función de autocorrelación establece un nivel $t-1$ a partir del cual se establece el nivel en el momento t , es decir, los modelos autorregresivos tienen en cuenta el estado del sistema en un momento previo para estimar el momento actual. Esto supuso un avance con respecto a la propuesta de Schubert (1999). De esta forma, trasladando este concepto a la respuesta del oyente al estímulo musical, es coherente pensar que el nivel de activación o valencia emocional del oyente en el instante previo a la escucha actual influirá en su estado, o más bien, en el efecto del estímulo sobre el estado del oyente durante la escucha, y que este estado irá evolucionando conforme se desarrolle el estímulo en el tiempo.

Para comprender más precisamente este enfoque debemos tener en consideración la pieza musical como un estímulo dinámico que cambia en el dominio del tiempo. Un estímulo musical tiene una estructura que se modifica o desarrolla en el tiempo y este hecho conduce a pensar que la reacción del oyente también debe considerarse dinámicamente en dos sentidos: que se vea afectada dinámicamente por la variación temporal del estímulo y que sea afectada por el propio estado del oyente en momentos anteriores. Esta concepción del estímulo musical es también adoptada por Schubert (1999) en su pionero trabajo y retomado posteriormente en posteriores como Korhonen (2004) o Coutinho y Cangelosi (2009 y 2011) o en la revisión llevada a cabo por Yang y Chen (2012).

Este es uno de los motivos por los que necesitamos técnicas que permitan evaluar el patrón de respuesta al estímulo musical considerando la dinámica temporal intrínseca al estímulo musical así como la propia respuesta del sistema cognitivo y emocional de los sujetos. En el modelo de Korhonen (2004), los modelos ARX soportaban este tipo de características temporales del estímulo pero se encontraban con las limitaciones anteriormente señaladas. Una alternativa interesante fue aportada por Coutinho y Cangelosi (2009 y 2011) quienes, a partir de los resultados de Korhonen (2004), emplean una arquitectura de RNA para la predicción de la respuesta emocional del oyente al estímulo musical.

Estos autores proponen el uso de la arquitectura RNA recurrente tipo Elman cuyo elemento más interesante es la inclusión de una capa oculta de memoria que permite introducir, como factor en la predicción, el estado del sistema en el instante anterior a la entrada de información. Coutinho y Cangelosi (2009, 2011) conciben así una estructura en tres capas: la capa sensitiva que incluye paralelamente la información relativa a la estructura del estímulo y la respuesta fisiológica del oyente. La segunda capa se trata de una capa oculta de procesamiento de la información compuesta por cinco unidades o neuronas de

computación. Estas neuronas reciben la señal de las neuronas sensitivas y en función de las conexiones entre las neuronas de computación y sensitivas emiten una respuesta si se supera el umbral de activación proporcionado por las neuronas de sesgo de respuesta. Relacionada con esta capa se encuentra la capa de memoria que almacena información sobre la activación de las 5 neuronas de computación en el momento temporal previo a la entrada de información actual. La salida de la red en el modelo de Coutinho y Cangelosi (2011) será por tanto una combinación lineal entre la información sensorial y el estado del sistema en el momento anterior.

Sin embargo esta arquitectura presenta un problema estructural al emplear como predictor del juicio cognitivo la respuesta fisiológica del oyente en una combinación indiferenciada con la estructura del estímulo. Esta aproximación minimiza el efecto de la respuesta fisiológica del oyente en contra de la lógica inicial propuesta, muy posiblemente debido a que en su modelo se introducen como predictores en serie los parámetros físicos del estímulo musical y la respuesta fisiológica. Esto sería un error si, tal y como se observa en el citado trabajo de Gomez y Danuser (2007), la estructura física del estímulo predice la respuesta fisiológica del oyente. Es decir, en el modelo propuesto por Coutinho y Cangelosi (2011) podría existir un problema de inflación de varianzas o de colinealidad. El efecto de las variables fisiológicas queda reducido en el modelo porque la respuesta fisiológica ya está predicha por la estructura del estímulo y, por tanto, su aporte a la predicción de la respuesta cognitiva es muy pequeño, tal y como concluyen Coutinho y Cangelosi (2011).

De los resultados obtenidos en Saiz-Clar et al. (2017b), se puede observar que sin tener en cuenta el estado previo del oyente en el instante anterior como en Korhonen (2004) y Coutinho y Cangelosi (2009 y 2011), existe una relación entre el desarrollo en el dominio temporal de la estructura del estímulo y la variabilidad de la respuesta fisiológica del oyente,

y entre la respuesta fisiológica del oyente y el juicio emocional (véase la Figura 1). Sin embargo, con la estructura reflejada en la figura 1 cometemos un elevado error de predicción al dejar un gran porcentaje de varianza sin explicar. Desde esta perspectiva analítica la estructura propuesta por Coutinho y Cangelosi (2011) resultaría, en términos predictivos, más práctica que una estructura donde la respuesta fisiológica presenta un papel mediador.

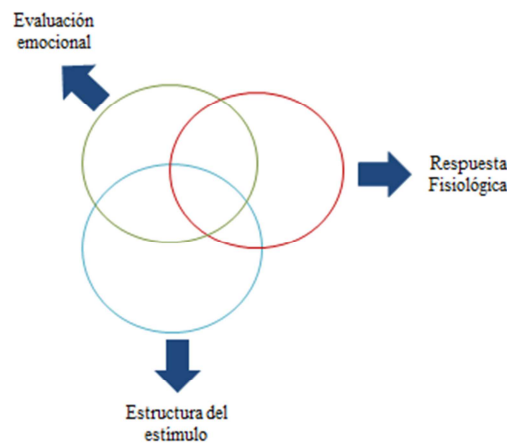


Figura 1: Diagrama sobre el porcentaje de varianza que explicaría un modelo en el que las variables fisiológicas actúan como mediadoras

Otra cuestión estructural interesante a la que se enfrenta la propuesta de Coutinho y Cangelosi (2011) frente a, por ejemplo, las propuestas estructurales de Korhonen (2004) y Saiz-Clar y Reales (2017), se refiere a si es más eficiente una estructura donde se empleen los mismos predictores para la activación y valencia emocional por un lado, y por otro, que el procesamiento de la red para dar una respuesta a las dimensiones de activación y valencia emocional sea en paralelo. De esta forma, mientras que en Saiz-Clar y Reales (2017) se presenta una estructura dual e independiente donde el procesamiento del estímulo musical se realiza utilizando dos rutas paralelas, una para la activación y otra para la valencia emocional,

la red propuesta por Coutinho y Cangelosi (2011) plantea un estructura más simple donde este procesamiento se produce en un único sistema donde los nodos para la activación y la valencia emocional se encuentran conectados a todas las características del input.

Si bien la estructura de Coutinho y Cangelosi (2011) es más simple porque implica un único sistema de procesamiento, la propuesta de Saiz-Clar y Reales (2017) cumple criterios estructurales basados, por un lado, en la ruta dual de procesamiento del estímulo propuesta en el modelo Neuropsicológico de la Percepción musical de Peretz y Coltheart (2003) y en la conexión entre este doble sistema de procesamiento y las dimensiones de activación y valencia emocional que propone la Teoría dual de Balkwill y Thompson (1999). Esta misma estructura es recientemente retomada también por Egermann, Fernando, Chuen y McAdams (2015) quienes, en un estudio sobre la transculturalidad de la respuesta fisiológica a la escucha musical, concluyen que las propiedades de bajo nivel tienen un reflejo sobre la respuesta subjetiva de activación y la activación fisiológica mientras que el efecto de la valencia emocional tiene un componente más aprendido y cultural.

Objetivo

El objetivo de este trabajo se centra por tanto en investigar, por un lado, si el factor temporal aumenta la precisión de las propuestas de redes neuronales realizadas por de Coutinho y Cangelosi (2011) y por Saiz-Clar y Reales (2017) determinando cómo afecta el estado previo del oyente sobre su respuesta en el continuo temporal. Por otro lado, y a partir de los resultados observados en Saiz-Clar et al. (2017b), queremos comprobar si la estructura paralela propuesta por Saiz-Clar y Reales (2017) obtiene mejor rendimiento predictivo que la propuesta estructural de Coutinho y Cangelosi (2011). Para ello, hemos construido una RNA de retropropagación que permite estudiar la relación entre la estructura del estímulo, la respuesta fisiológica y la evaluación emocional de la pieza pero bajo una concepción

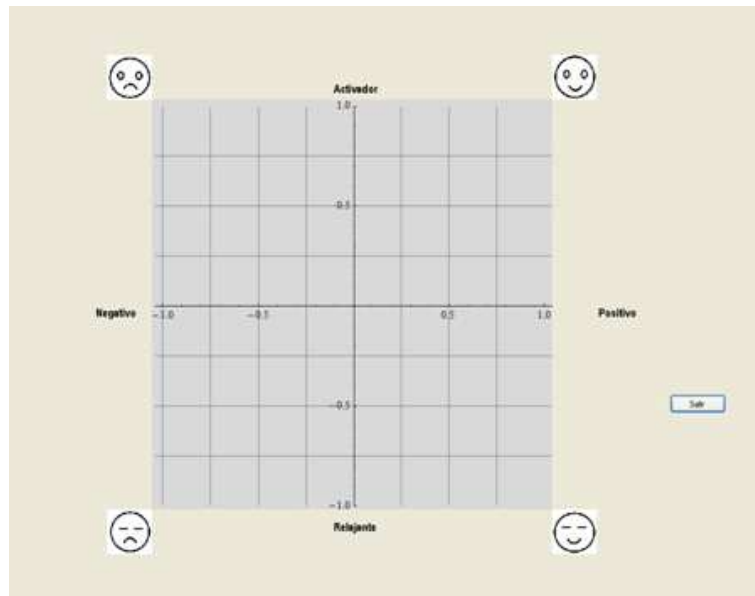
dinámica y cambiante del estímulo, es decir, teniendo en cuenta la **evolución temporal** de las variables.

Método

Construimos la red neuronal basándonos en los resultados obtenidos en Saiz-Clar et al. (2017b). Estos autores evaluaron experimentalmente la respuesta emocional continua de una muestra de 23 participantes (15 mujeres y 8 hombres) a una selección de piezas musicales extraídas del repertorio musical de Vieillard et al. (2008). Los participantes no tenían conocimientos formales en música. El rango de edad de la muestra era entre los 18 y 27 años con un promedio de 21.5 años ($DT = 3.10$). Todos ellos fueron voluntarios no recibiendo ningún tipo de recompensa material por su participación.

Los participantes realizaron el experimento en una habitación anecoica y dispuestos en posición sedente frente a una pantalla de ordenador de 18 pulgadas donde se iba sucediendo la presentación estimular en orden aleatorio. Durante la fase de colocación de los electrodos para el registro fisiológico los oyentes eran instruidos en la tarea por escrito y de forma oral (para una descripción más detallada del procedimiento de adquisición de la señal remitimos al lector interesado al artículo original). Una vez finalizada la fase de instrucción, los oyentes comenzaban la tarea de escucha musical informatizada. Esta tarea consistía en valorar cada pieza musical de forma continua en las dimensiones de activación (plano vertical) y la valencia emocional (plano horizontal) tal y como se observa en la Figura 2. Los promedios de valoración de las piezas musicales en las dimensiones de activación y valencia emocional, en las dos muestras experimentales (Vieillard et al., 2008 y Saiz-Clar et. al., 2017) fueron semejantes, demostrando que este *set* de piezas musicales presenta una buena consistencia inter-sujeto.

Figura 2: Pantalla de respuesta en el registro continuo de las dimensiones emocionales en el estudio de Saiz-Clar et al., (2017b).



Descripción de las variables

Medidas Fisiológicas

A pesar de que en el estudio de Saiz-Clar et al. (2017b) los autores emplearon un número considerable de medidas fisiológicas, en este estudio hemos decidimos emplear como predictores de la actividad del sistema nervioso autónomo 1) la **intensidad de la actividad electrodérmica** (EDA) durante la escucha de cada pieza musical medida en microSiemens (μ S), 2) la **frecuencia cardiaca** medida en función de las pulsaciones por minuto o los *beats per minute* (bpm). La justificación de esta elección responde a dos criterios complementarios: por un lado mantener la simplicidad del modelo incluyendo exclusivamente las medidas que tradicionalmente más efecto han mostrado para la predicción del juicio emocional (Coutinho

y Cangelosi, 2011); por otro, mantener la simplicidad del modelo reduciendo el número de parámetros innecesarios o redundantes.

Variables del estímulo musical

En el modelo de Countinho y Cangelosi (2011) las medidas usadas fueron extraídas del dominio Sonoridad, Tempo, Nivel del tono (medido mediante el centroide del espectro de potencia), Contorno melódico, Timbre (medido mediante la variable *sharpness*) y Textura (multiplicidad). En nuestro trabajo, hemos decidido usar las medidas empleadas en los trabajos de Saiz-Clar y Reales (2017) y Saiz-Clar et al. (2017) a la vista de los resultados obtenidos en sus respectivos trabajos. Desde este marco, hemos incluido como variables estáticas en el modelo el tempo, el compás musical, el promedio de notas por compás (PNC), el modo, la frecuencia mínima y la amplitud tonal absoluta (ATA).

Las medidas de la Curva de *Onsets* (CO) fueron calculadas a partir de los parámetros descritos previamente en Saiz-Clar et al. (2017), y ponderadas a partir de los pesos factoriales obtenidos en dicho trabajo. La obtención de las variables necesarias para obtener dichos parámetros se llevó a cabo conforme al procedimiento descrito en el trabajo de Saiz-Clar et al. (2017) y mediante el MIRToolbox de Matlab (Lartillot y Toivianen, 2007 y Lartillot, Toivianen y Eerola, 2008). Para una descripción más detallada de estas variables remitimos al lector a los trabajos citados.

Medida de la emoción percibida

Las dimensiones de Activación y Valencia Emocional fueron medidas para cada sujeto y para cada ítem de forma continua en el tiempo. El método empleado es semejante al usado por Nagel, Kopiez, Grewe, y Altenmüller (2007) mediante el software *EMuJoy*. Este software reproduce unos ejes cartesianos en la pantalla del ordenador, el eje horizontal representa la

activación y el eje vertical la valencia emocional. Como resultado del movimiento del ratón por parte del oyente se obtiene en cada instante de muestreo unas coordenadas X e Y que permiten situar el juicio emocional en uno de los cuatro cuadrantes resultantes.

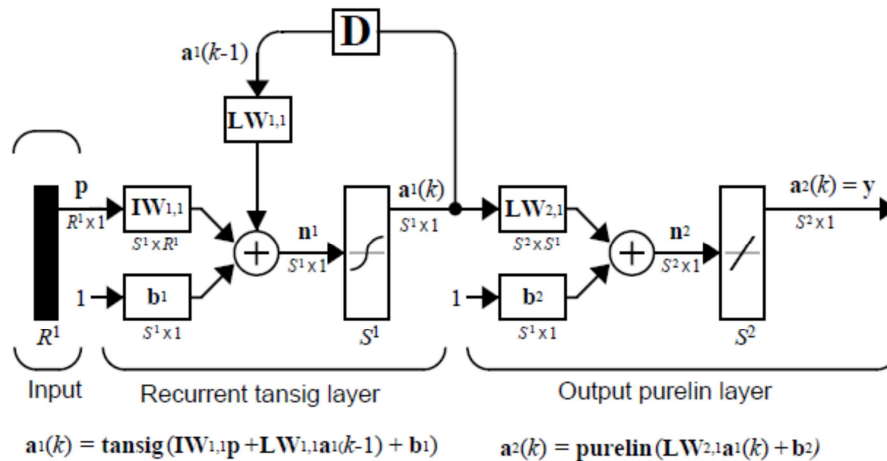
Descripción de la red

Empleamos en este trabajo el mismo tipo de red neuronal artificial (RNA) que el empleado por Coutinho y Cangelosi (2011), una RNA recurrente tipo Elman (1990). Como ya hemos indicado, las redes tipo Elman permiten establecer el estado de un sistema en un momento t , a partir del contexto estimular en dicho momento temporal (discreto), pero teniendo en cuenta además el estado del sistema en un momento $t-1$. Es decir, no sólo tiene en cuenta las entradas al sistema, sino el estado de activación en el que se encontraba dicho sistema en el momento de recibir la información. Este aspecto es extremadamente relevante cuando intentamos predecir la respuesta del organismo a un estímulo, dado que nos permite no sólo tener en cuenta en la predicción los factores ambientales (es decir, el *input* en la RNA), sino también el estado interno del organismo (unidades de contexto).

Las RNA tipo Elman tienen una arquitectura semejante a la del Perceptrón Multicapa (Figura 3). Presentan una capa de entrada con neuronas sensitivas que captan las características del estímulo. En este contexto, nos referimos a todas las propiedades que vamos a emplear para definir el estímulo musical. A continuación, este tipo de red presenta una capa de unidades ocultas cuya función es procesar esas propiedades del estímulo para emitir una salida a la capa siguiente. Es usual emplear la función de identidad para establecer la relación entre esta capa y la capa sensorial de forma que la información que llega a la capa oculta es el producto de la entrada por el peso de la conexión. El número de unidades en la capa oculta es variable y responde más a criterios de eficiencia en la predicción. En el modelo de Coutinho y Cangelosi (2011) el número de unidades ocultas era de cinco. Según Coutinho

y Cangelosi (2011), la función de transferencia entre la capa oculta y las salidas suele ser también la función de identidad aunque pueden emplearse otros tipos de funciones.

Figura 3: Arquitectura típica de una red recurrente tipo Elman



Extraído de Matlab Neural Network Toolbox User guide (2011).

En nuestro caso hemos decidido usar la función **tangente hiperbólica** entre la capa oculta y la capa de salida dado que permite suavizar los valores extremos de la distribución y normalizarlos. La ventaja que supone suavizar la tendencia afectará sin embargo a la precisión del modelo dado que será menos sensible a los valores extremos en virtud de ser más preciso para los valores centrales.

Hasta aquí tendríamos la arquitectura típica de un Perceptrón Multicapa, con una transferencia de información hacia delante. Sin embargo, la novedad que incluyen las redes Elman la encontramos en las unidades de contexto o de memoria. Estas unidades forman una capa de “memoria” que retiene la información del estado en el que se encuentra la red en el momento anterior a t . Estas neuronas memorizan la salida de las unidades ocultas en $t-1$ y se encuentran directamente relacionadas con ellas, de forma que la salida que emitirán las

neuronas de la capa oculta hacia las neuronas motoras será producto de las propiedades del estímulo en t y de lo retenido en $t-1$. La red fue entrenada con Matlab 2012, con una tasa de aprendizaje de 0.5 y en un entrenamiento de 20000 épocas. Las matrices de pesos iniciales para la conexión entre las neuronas sensitivas y la capa oculta (*Input Weight* o IW), y para las neuronas de la capa oculta y la capa de salida (*Layer Weight* o LW), fueron inicializadas a cero; las neuronas de sesgo (*bias*) existentes entre la capa sensitiva y la capa oculta y las existentes entre la capa oculta y la de salida fueron inicializadas a 1. Se llevaron a cabo varias simulaciones con diferentes arquitecturas y diferentes unidades en la capa oculta. El código anexo se refiere a la estructura que finalmente fue incluida en este trabajo por mostrar mejor rendimiento predictivo y menor error.

La muestra inicial de estímulos fue segmentada para crear un **set de entrenamiento** de la red mediante el cual la red aprende la asociación entre los patrones de entrada y las salidas deseadas, y otro **set de generalización** que nos permitiera determinar la precisión de la red en la predicción. El *set* de entrenamiento estaba formado por 10 ítems y comprendía un total de 121 entradas, mientras que el *set* de generalización estaba compuesto por los restantes 6 ítems y un total de 81 entradas.

Llevamos a cabo dos tipos de arquitecturas: una red que reproducía la arquitectura de Coutinho y Cangelosi (2009 y 2011) pero con las variables de entrada indicadas en la sección previa, y 2) dos redes que procesaban, por un lado, la activación y, por otro, la valencia emocional tal y como proponen Saiz-Clar y Reales (2017). Cada una de estas arquitecturas fue simulada con una variación en la capa de memoria: 5 neuronas como proponían Coutinho y Cangelosi (2009 y 2011) o una sola neurona de procesamiento buscando un modelo menos redundante y más parsimonioso.

Plan de análisis

Tras la fase de simulación descrita en el apartado anterior, el siguiente paso fue analizar la estructura más óptima para la predicción de la respuesta continua del oyente a las dimensiones emocionales. Para ello establecimos una serie de criterios que debía cumplir la red para ser considerada óptima. El primer criterio empelado fue el análisis del error cuadrático medio (ECM) que se trata de un método de cuantificación de la desviación entre la variable empírica y la predicción realizada por el modelo. Es decir, mediante el cálculo del ECM promediamos la diferencia entre el valor real y el valor estimado a partir de un modelo. El cálculo del ECM es un recurso muy común para comparar el rendimiento entre diferentes modelos de predicción pero no debe ser el único criterio a la hora de valorar el rendimiento ya que el ECM presenta promedios muy sensibles a la presencia de casos extremos y/o anómalos. Un modelo puede ser altamente predictivo para un conjunto de casos pero menos sensible a otros casos. Por ejemplo en la red propuesta por Saiz-Clar y Reales (2017), encontraron tasas de clasificaciones correctas diferentes en función del tipo de emoción que suscitaba el estímulo. Por otro lado, tampoco existe un criterio claro para cuantificar si la magnitud del ECM es aceptable o no ya que no es una medida estandarizada.

A la vista de la definición y dificultades que presenta el error cuadrático medio, consideramos insuficiente que la decisión sobre la idoneidad de la arquitectura se base exclusivamente en este criterio. Es también necesario determinar si este error de predicción se trata de un error aleatorio o sistemático y para ello debemos determinar 1) que se distribuye de acuerdo a la distribución Normal (lo evaluamos mediante el **test de Kolmogorov-Smirnov** para una muestra), 2) que se trata de un error aleatorio (lo evaluamos mediante el **test de Rachas**).

Pero tan sólo el error de predicción no es un criterio suficiente para determinar la efectividad de un modelo. Entendemos que un modelo es capaz de predecir de forma precisa una respuesta cuando la distribución de las predicciones es semejante a la distribución de los valores experimentales. Esto se encuentra relacionado con el concepto previamente indicado sobre la sensibilidad del ECM a los valores atípicos. El valor del ECM puede inflarse como consecuencia de la presencia de valores atípicos concretos o lo que podríamos denominar como “*outliers* en la predicción”

Desde otro punto de vista creemos que además del análisis del error y su distribución, una parte importante del estudio de la precisión es el análisis del ajuste de la distribución entre el *output* y el *target* de la red. Es decir, entendemos que la salida de la red (*output*) es semejante a su respectivo valor objetivo (*target*), no solo cuando el error es pequeño sino también cuando ambas variables son semejantes en su distribución. Para determinar que *output* y *target* son semejantes emplearemos tres estrategias: 1) determinar la magnitud de la relación entre ambas medidas mediante el **coeficiente de correlación de Pearson**, lo cual nos dará una estimación de la precisión en la predicción del *output* sobre el *target* (o en otras palabras, el porcentaje de variación explicado), 2) comprobar que existe un ajuste en la distribución a través del **test de Wilcoxon** para muestras relacionadas y por último 3) determinar si la distribución ambas medidas sigue una distribución conocida mediante el test de **Kolmogorov-Smirnov**. Una vez determinada la arquitectura RNA óptima a partir de los criterios descritos previamente, procedimos a analizar las conexiones y pesos de la red definitiva para determinar la dirección y el efecto sobre la capa oculta.

Resultados

En la Tabla 1 podemos comprobar los coeficientes de correlación entre el *output* y el *target* para las 6 simulaciones realizadas tanto en el entrenamiento como en la fase de generalización, el porcentaje global de varianza explicada y error cuadrático medio en cada una de las simulaciones.

Tabla 1: Magnitud de la asociación entre los valores *output* y *target* y ECM de las seis simulaciones.

		Activación / Valencia Emocional		Activación		Valencia Emocional	
Neuronas en la capa oculta		5	1	5	1	5	1
Entrenamiento	Correlación	.661***	.661***	.778***	.778***	.787***	.787***
	ECM	.523	.523	.322	.322	.395	.395
Generalización	Correlación	.038	.020	.831***	.831***	.643***	.641***
	ECM		1.341	.298	.322	.559	.561
R²		.218	.153	.642	.642	.511	.510

A partir de la Tabla 1 llegamos a determinar que la estructura más eficiente en términos de predicción es la aquella que separa el procesamiento de la activación y la valencia emocional del estímulo en dos redes diferentes. Tanto el error cuadrático medio como la magnitud de la asociación es menor en este tipo de arquitectura en comparación con la estructura de una sola red. Este rendimiento es mejor tanto en la fase de entrenamiento como en la generalización. Sin embargo, no existen muchas diferencias en el rendimiento entre una estructura de 5 neuronas en la capa oculta o de una única neurona (Tabla 2).

A la vista de la similitud entre ambas estructuras a partir de los resultados de la Tabla 2, debemos considerar preferible aquella que tenga un menor número de parámetros, en este caso las arquitecturas con una única neurona en la capa oculta (Figura 3). Empleando este tipo

de arquitectura y a partir de la Figura 4 y los resultados de la Tabla 3, observamos un buen ajuste entre las salidas de ambas redes (*outputs*) y sus respectivos valores objetivo o *targets* tanto en la fase de entrenamiento como en la de generalización.

Tabla 2: Estudio del ajuste de la distribución de los valores del *output*, el *target* y del error en la RNA de Elman.

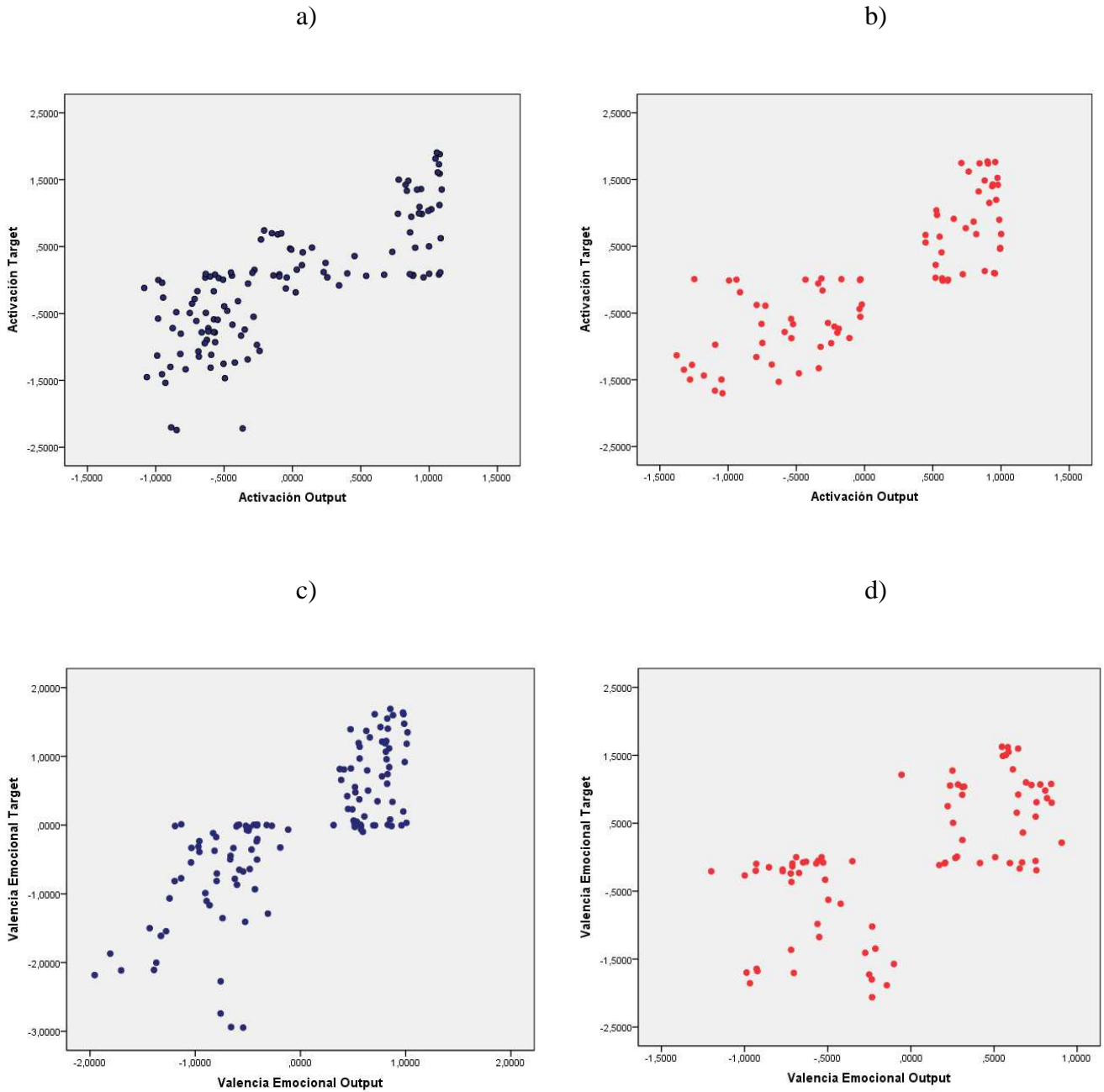
		Estructura de 5 neuronas de memoria		Estructura de 1 neurona de memoria	
		Activación	Valencia Emocional	Activación	Valencia Emocional
Test de Wilcoxon	Test	Z[121] = -.374	Z[121] = -.847	Z[121] = -.371	Z[121] = -.839
		p = .709	p = .397	p = .711	p = .401
	Generalización	Z[81] = -.864	Z[81] = -.600	Z[81] = -.873	Z[81] = -.619
		p = .388	p = .548	p = .382	p = .536
Error	Test de Rachas	Z[202] = -1.662	Z[202] = 2.077	Z[202] = -1.662	Z[202] = 2.077
		p = .097	p = .038	p = .097	p = .038
	K-S	Z[202] = .768	Z [202]= 1.256	Z[202] =.770	Z [202]= 1.264
		p = .597	p = .085	p =.593	p = .082

Nota. K-S: test de Kolmogorov-Smirnov.

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$.

En cuanto al error que comete la red (véase la Tabla 3), este es similar en ambas redes para la fase de entrenamiento aumentando en la fase de generalización para la valencia emocional, donde precisamente la correlación entre *target* y *output* reduce su magnitud. Así y todo, el error de las predicciones de la Activación emocional se distribuye de acuerdo a la distribución normal en ambos casos y aleatoriamente.

Figura 4: Gráficos de dispersión para los valores de a) *output* y *target* en Activación en la fase de entrenamiento, b) *output* y *target* en Activación la fase de generalización, c) *output* y *target* en Valencia emocional para la fase de entrenamiento y d) *output* y *target* en Valencia emocional para la fase de generalización.



A la vista de estos resultados no detectamos un patrón sistemático de error. Además, este error no está relacionado con los *outputs* de activación ($r = .032, p > .05$) y valencia emocional respectivamente ($r = -.002, p > .05$) aunque sí existe relación entre la

distribución del error y el *target* para la activación ($r = .624, p < .001$) y para la valencia emocional ($r = .680, p < .001$).

Tabla 3: Pesos sinápticos en las redes de Activación y Valencia Emocional tipo Elman

ACTIVACIÓN		VALENCIA EMOCIONAL	
Compás	-.0601	Modo	-.2809
Tempo	-1.943	Frecuencia Mínima	-.2660
Promedio de notas por compás	-.2371	Amplitud tonal Absoluta	-.2151
Factor 1 CO	.8057	Factor 1 CO	.0490
Factor 2 CO	-.8767	Factor 2 CO	-.2864
Factor 3 CO	-.9332	Factor 3 CO	-1.0202
EDA	-.2222	EDA	-.4772
HR	-.5045	HR	-.0403
Sesgo 1	-.3505	Sesgo 1	-.9950
LW	13.393	LW	-4.3186
Sesgo 2	-.2235	Sesgo 2	-3.0881

EDA=Electrodermal Activity; HR= Heart Rate; LW=Layer Weight.

Por último analizamos los pesos sinápticos del modelo en la Tabla 3. En ella comprobamos que para la predicción de la activación, las variables con mayor fuerza de conexión resultaron ser Tempo musical, los factores de la CO y la Frecuencia Cardíaca. Mientras que para la predicción de la Valencia emocional, las variables con mayor peso son el factor 3 de la CO, la intensidad de la EDA y el Modo de la pieza musical.

Para finalizar, señalar que el test de Kruskal-Wallis para k muestras independientes empleando como factor el tipo de emoción y como variables dependientes el error de predicción para ambas redes, muestra un efecto significativo para la activación ($Z[202] = 25.253, p < .001$) pero no para la valencia emocional ($Z[202] = 7.594, p > .05$). Estos resultados muestran que el error de predicción se distribuye de forma semejante para la

valencia emocional a lo largo de las cuatro categorías emocionales. En cambio, observamos que la red para la activación comete mayor error de predicción para la emoción de tranquilidad infra-estimando ligeramente el poder activador de estos estímulos sobre el oyente.

Conclusiones

En base a la concepción de Schubert (1999) que consideraba el estímulo musical como un estímulo dinámico, nos planteamos cómo podíamos modelar la respuesta emocional para apresar el efecto que tenía la evolución temporal de la pieza sobre el juicio emocional del oyente. Para ello nos basamos en dos arquitecturas diferentes. Por un lado, Coutinho y Cangelosi (2009 y 2011) presentaron una estructura dinámica mediante las RNA recurrentes tipo Elman cuya ventaja principal es la inclusión de una capa de memoria que permite tener en consideración en el modelo el estado de activación del sistema en el instante $t-1$. Por otro lado, Saiz-Clar y Reales (2017) presentaron una arquitectura Adaline, basada en el modelo neuropsicológico de la percepción musical de Peretz y Coltheart (2003) y de la Teoría Dual de Balkwill y Thompson (1999), donde se separa el procesamiento de la estructura de la pieza musical en dos rutas (Peretz y Coltheart, 2003): una ruta temporal que procesa básicamente las propiedades rítmicas de la pieza, y otra ruta de procesamiento tonal donde se procesan aspectos relacionados con el dominio de la frecuencia. Esta doble vía de procesamiento, a su vez, afecta de forma diferencial a las dimensiones emocionales planteadas por Russell (1980) de forma que el procesamiento de las propiedades métricas afecta prioritariamente al juicio sobre la activación, mientras que el procesamiento del tono afecta en mayor medida al juicio emocional (Balkwill y Thompson, 1999).

Si bien, la estructura de Saiz-Clar y Reales (2017) es más simple (y por tanto, más parsimoniosa) que la propuesta por Coutinho y Cangelosi (2009 y 2011), y además está

basada en un criterio teórico más que un criterio de predicción, la estructura de Coutinho y Cangelosi (2009 y 2011) es dinámica y se adapta al estado del oyente. Por otro lado, la estructura de Saiz-Clar y Reales (2017) es estática y se basa en las respuestas globales del oyente tras la escucha sin tener en cuenta el dominio tiempo. A partir de estas cuestiones decidimos mezclar los componentes teóricamente justificables de ambas arquitecturas y construir una RNA recurrente que pusiera a prueba la estructura más eficiente para la predicción de la respuesta emocional del oyente en el dominio del tiempo.

Si comparamos el rendimiento de nuestro modelo final frente a los resultados experimentales de Saiz-Clar et al. (2017b), observamos que en su trabajo obtuvieron, en un modelo lineal por pasos sucesivos y empleando las mismas entradas, un porcentaje de varianza explicada para respuesta de activación del 63%, y de la valencia emocional del 55%, mientras que un modelo que tiene en consideración la capa recurrente explica en la **fase de entrenamiento** un porcentaje de varianza para la activación del 60.5% y del 62% para la valencia emocional. La limitación que presenta nuestro modelo frente a la propuesta de Saiz-Clar et al. (2017b) es que empleamos un menor número de piezas musicales para obtener los parámetros del modelo pero, en su lugar, sí que dedicamos una muestra de los mismos al estudio de la generalización, obteniendo índices de ajuste similares. Sin duda una ventaja fundamental del uso de las redes tipo Elman sobre los modelos lineales se observa en la predicción de la valencia emocional. Este resultado muestra la importancia que tiene el estado previo del sistema a la hora de emitir un juicio emocional sobre la entrada sensorial que recibe en un momento concreto

Tras varias simulaciones donde hemos probado el rendimiento de diferentes arquitecturas, hemos podido comprobar que la estructura óptima en términos de rendimiento implica introducir como *inputs* la respuesta fisiológica del oyente junto con la estructura del

estímulo tal y como planteaban Coutinho y Cangelosi (2009 y 2011). Esto tiene sentido si en lugar de una perspectiva estructural de la relación entre los sistemas **estímulo – SNA – respuesta emocional**, lo planteamos desde una perspectiva funcional donde la respuesta voluntaria del sujeto viene determinada por un lado por las **características del *input* que percibe**, por otro lado de la **respuesta interna que este *input* le genera** y por último, del **estado de activación previo** del sistema.

Por otro lado, la arquitectura obtenida confirma el planteamiento de Saiz-Clar y Reales (2017) basado en los trabajos de Balkwill y Thompson (1999), en donde se conciben como sistemas separados la activación y la valencia emocional. A la vista de nuestros resultados resultan más eficientes dos redes independientes que procesan algunos parámetros del estímulo de forma común a como se planteaba en Saiz-Clar et al. (2017) y otros parámetros de forma independiente: factores relacionados con el ritmo para la activación y factores relacionados con la tonalidad para la valencia emocional. La concepción separada de la activación y la valencia emocional en los modelos de predicción es coherente también con el enfoque ortogonal de las dimensiones emocionales tal y como fueron concebidas por Russell (1980).

En términos de varianza explicada, nuestra arquitectura mantiene un rendimiento similar al encontrado en otros trabajos (ej. Coutinho y Cangelosi, 2009 y 2011; Korhonen; 2004, Saiz-Clar y Reales, 2017). Las salidas de la red se han mostrado semejantes a los *targets* tanto en la fase de entrenamiento como en la fase de generalización, comprobando que los *outputs* se ajustan a la distribución de sus respectivos *targets* y obteniendo coeficientes de correlación de magnitud elevada a pesar de que la función de transferencia de nuestra red entre la capa oculta y la salida no era una función de identidad. Recordemos que tal y como hemos explicado en la introducción, la función de transferencia empleada permite suavizar los

valores extremos en la salidas lo cual reduce la precisión del modelo cuando el criterio para evaluarla se basa en la comparación con el *target*.

En este sentido, las redes Elman se han mostrado eficientes para simular esta respuesta del oyente. Sin embargo, nos hemos enfrentado a algunos problemas con este tipo de redes neuronales. Uno de ellos surge al tener que introducir los *inputs* de forma continua sin ninguna alteración o separación entre ítems, es decir, no tenemos en cuenta el inicio o el final de un ítem. Esta es una limitación a la hora de programar la capa de memoria dado que en el comienzo de una pieza se considera el estado previo al final del ítem anterior. En la siguiente arquitectura podría incluirse espacios de ruido blanco entre piezas musicales eliminando la valoración global del estímulo que ya hemos comprobado en este experimento. Los espacios de ruido blanco nos permitirán además de solventar esta limitación, estudiar el efecto de la escucha en el tiempo una vez termina la presentación del estímulo. Es decir, en próximos diseños podrá incluirse una línea base inter-estimular.

Hoy en día, las arquitectura tipo Elman suelen emplearse con fines de investigación y docentes debido a que su arquitectura ha sido superada por otros modelos más modernos como las NARX (*Nonlinear AutoRegressive neural network*). Sería conveniente estudiar el rendimiento de este tipo de arquitecturas más actuales con el rendimiento de nuestra red para comprobar si incrementan de forma significativa la precisión en la predicción. Además, debemos incluir en un futuro trabajo más piezas musicales de forma que podamos tener una muestra mayor de elementos para introducir en la red y de esta forma, explorar el efecto sobre la predicción de medidas fisiológicas que han mostrado un buen rendimiento en otras simulaciones.

CAPÍTULO 6. Conclusiones

Esta tesis se iniciaba presentando el modelo Cognitivo-Neuropsicológico del procesamiento musical (véase la Figura 1) de Peretz y Coltheart (2003) y, dentro de este amplio modelo construido a partir de muchos años de investigación en el área y el trabajo de multitud de autores, nosotros centramos nuestra atención en una pequeña parte del mismo, aquella que denominamos la “caja negra” de la emoción. Planteamos al comienzo su presencia como una caja negra porque nuestro conocimiento sobre ella se limitaba a una función de recepción del *input* musical, y de emisor de información procesada hacia el núcleo del modelo, lo que Peretz y Coltheart denominan “repertorio” que es el lugar donde se acumula la experiencia musical.

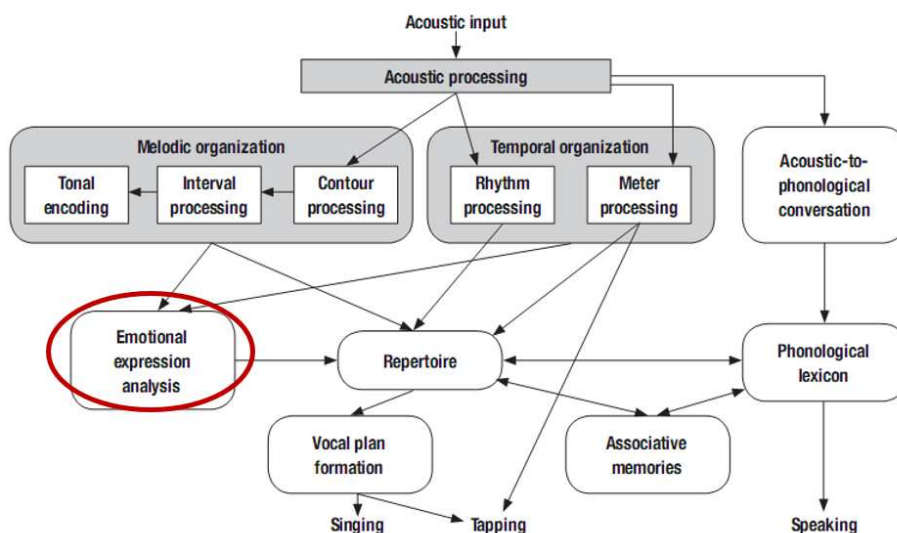


Figura 1: Modelo Neuropsicológico de Peretz y Coltheart (2003)

Una vez identificado nuestro constructo de interés en un modelo teórico de referencia en el campo musical, nos planteamos el primer problema conceptual: qué entendemos por emoción y qué paradigma de modelización podíamos emplear que profundizara en el modelo de bloques planteado por Peretz y Coltheart (2003). Señalamos la existencia de, al menos, dos

cuestiones importantes en torno al concepto de emoción. Por un lado, debimos decidir si basábamos nuestro trabajo en un enfoque discreto de la emoción acorde a la postura de Kreibig (2010) o si por el contrario debíamos emplear un enfoque dimensional del concepto de emoción como propone Feldman-Barret (2006). En nuestros tres trabajos de modelización y experimentales hemos observado una importante consistencia inter-sujeto en la categorización del estímulo musical. En nuestro primer modelo simulamos el procesamiento emocional del estímulo musical culminando el mismo en un mapa auto-organizado que era capaz de clasificar la categoría emocional del estímulo musical de forma similar a la categorización obtenida experimentalmente por Vieillard et. al (2008).

Posteriormente, en nuestro tercer trabajo y con nuestra propia muestra tarea experimental, obtuvimos resultados semejantes en la clasificación de las piezas musicales que la obtenida por Vieillard et al. (2008). Estos resultados invitan a pensar que la categorización emocional de la pieza musical en una categoría emocional discreta es un proceso estable inter-sujetos dado que existe un consenso elevado a la hora de clasificar las piezas musicales en alguna de las cuatro emociones básicas consideradas, pero con dos matizaciones. En primer lugar, esta estabilidad sólo se ha observado en algunos estímulos, no en todos. En la categorización en algunas piezas musicales hemos obtenido índices de acuerdo muy bajos entre los oyentes. Ante este mismo problema se encuentran Vieillard et al. (2008) quienes descartan de su trabajo 16 piezas de las 56 piezas musicales de las que partían por obtener tasas de acuerdo excesivamente bajas. Por otro lado, estos resultados pueden contener un elevado componente cultural ya que las muestras que hemos analizado son de participantes occidentales y las composiciones se adecúan a las reglas de composición occidental.

Tal y como señalábamos en la introducción, un enfoque discreto de la emoción es poco práctico y la mayoría de trabajos sobre música y emoción optan por enfoques

dimensionales donde el juicio categórico puede ser predicho a partir del juicio dimensional del oyente. Tal y como señala Carretié (2009), los resultados en investigación distan de haber identificado un patrón fisiológico de respuesta claro y unívoco con respecto a las categorías emocionales. En nuestro trabajo tampoco hemos hallado ni constatado dicho patrón. De hecho, los análisis exploratorios realizados muestran diferencias muy sutiles en la respuesta fisiológica en función de la categoría emocional del estímulo. Eso no significa indudablemente que no haya un efecto de la actividad fisiológica sobre la emoción del oyente, como hemos podido comprobar en el tercer artículo de nuestro trabajo, pero no hemos identificado un patrón claramente diferenciado y generalizable a todos los sujetos que permita diferenciar una categoría emocional de otra. Podemos plantear que, quizá, el mismo intento de encontrar este patrón común esté desencaminado. Ya desde el principio de economía señalado por Norris y cols. (2010 y 2011), se planteaba que la existencia de un patrón único de respuesta fisiológica a cada emoción sería muy poco práctico dado el número de emociones posibles que puede categorizar el ser humano.

Por este motivo, podemos señalar que el pragmatismo del enfoque dimensional de Wundt supera el enfoque categórico de James-Lange resultando mayoritarios los trabajos realizados desde el enfoque dimensional en relación al enfoque categórico. Esto nos conduce ineludiblemente a la cuestión del concretar el número de dimensiones necesarias para representar de forma precisa el universo emocional. En este trabajo hemos comprobado no sólo el elevado acuerdo entre oyentes a la hora de calificar una pieza musical como más o menos activadora o como más o menos placentera sino el que las piezas categorizadas como alegres y tensas resultan significativamente más activadoras que las piezas tranquilas y tristes, y por ende, que las piezas alegres y tranquilas resultan más placenteras que las piezas tristes y/o tensas.

A lo largo de los cuatro trabajos realizados en esta tesis, y sin entrar a cuestionar la existencia de una tercera dimensión para conceptualizar la emoción, nos remitimos a las palabras de Russell (1980) y consideramos que, si bien pueden existir tres o más dimensiones para definir el espacio emocional, las dos dimensiones propuestas por este autor resultan suficientes para obtener una representación precisa de las cuatro emociones básicas.

Aunque los modelos tridimensionales se encuentran actualmente vigentes en la investigación en nuestro campo y compiten con el enfoque bidimensional de Russell, el aporte encontrado por esta tercera dimensión resulta tan pequeño en trabajos como Bigand et al. (2005) o Eerola (2009, 2010 y 2012) y el rendimiento de los modelos bidimensionales es tan bueno y estable entre sujetos, que basándonos en el principio de economía debemos elegir el modelo más parsimonioso.

Otro hallazgo interesante en nuestro tercer artículo hace referencia a la latencia de respuesta observada para la respuesta conductual en el espacio bidimensional de Russell (1980). Hemos estimado que el oyente emplea entre 0,5 y 2 segundos, dependiendo de la dimensión emocional juzgada, para comenzar a emitir un juicio emocional. Este hallazgo merece un trabajo centrado exclusivamente en analizar qué sucede en el oyente durante aproximadamente estos dos segundos en los que además se produce su respuesta de orientación fisiológica. Quizás la clave para entender la relación entre la respuesta fisiológica y la respuesta cognitiva se encuentren justo en estos dos segundos iniciales.

Sin embargo, consideramos que existen aún muchas cuestiones por resolver sobre la categorización del estímulo musical. Si bien es cierto que nuestros modelos alcanzan buenas tasas de ajuste y éxito en la clasificación, por el momento no explican la ambigüedad en el juicio observada para algunos estímulos. Tanto en el trabajo de Vieillard et al. (2008) como en nuestros dos trabajos experimentales (Saiz-Clar y Reales, 2017 y Saiz-Clar et al. 2017)

hemos encontrado que algunos estímulos no generan una respuesta homogénea entre los oyentes mientras que otros estímulos alcanzan tasas de acuerdo cercanas al 100% en la clasificación. Por el contrario, detectamos otras piezas que resultan ambiguas para el oyente con tasas de categorización correcta en torno al 50% o inferior. Será necesario próximamente abrir este campo de investigación apenas explorado en la actualidad y entender que tienen estos estímulos para que el juicio del oyente presente tanta variabilidad inter-sujetos.

Este fenómeno podría explicarse considerando el reciente trabajo de Kawakami et al. (2013). Estos autores abordan un resultado recurrente en los diferentes trabajos sobre emoción y música y que nosotros mismos hemos comprobado en nuestros dos experimentos. Según el modelo dimensional de Russell (1980), la tristeza se caracterizaría por una evaluación emocional negativa del estímulo y una baja activación. Esto lo hemos observado en estímulos como la Marcha Fúnebre de Mendelshon o la de Chopin: nuestros oyentes categorizaban como tristes estos estímulos y sus promedios en activación y valencia emocional resultaban los esperados para un estímulo calificado como triste. Sin embargo, esto no sucede en la mayoría de los estímulos con intención comunicativa de tristeza que presentamos a los oyentes en nuestras investigaciones. En general su evaluación hedónica es más placentera de lo que cabría esperar según el modelo de Russell. Kawakami et al. (2013) plantean que algunas de estas piezas musicales pueden generar en el oyente respuestas emocionales de nostalgia o romanticismo pero que las tareas experimentales que utilizamos fuerzan a los participantes a elegir la categoría más parecida a las emociones que experimentan. Y esta emoción es la tristeza. Por tanto, una posible hipótesis sobre el origen de la variabilidad en la categorización emocional de algunas piezas musicales puede ser que cuatro categorías emocionales no sean ya suficientes para predecir y simular la respuesta emocional del oyente. Quizás sea el momento de ampliar el abanico de categorías emocionales en nuestros estudios permitiendo al oyente más margen para precisar sus respuestas.

Sin entrar a posicionarse entre los postulados cognitivistas vs. emotivistas e independientemente de la variabilidad observada para algunos estímulos concreto, un aspecto que queda claro a partir de los resultados de esta tesis es que es la estructura del estímulo, y no factores internos del sujeto, lo que permite explicar la mayor parte de la varianza del juicio emocional del oyente. Es decir, son las características estructurales de la pieza musical las que permiten al oyente clasificar la pieza musical dentro de una categoría emocional básica u otra.

Sobre la estructura del estímulo

Cuando esta tesis comenzó a plantearse, los trabajos que recalcan la importancia de la estructura del estímulo sobre el juicio emocional del oyente eran escasos. El desarrollo de algoritmos como los empleados para determinar la Curva de *Onsets* (Kapluri, 1999; Scheirer, 1998) y el de *software* específicos para la detección precisa de propiedades estructurales de la pieza musical como MARYSAS (Tzanetakis y Cook, 2002), PsySound (Cabrera, 1999) o el MIRToolbox (Lartillot, Toiviainen y Eerola, 2008) han permitido incluir estos parámetros dentro de los modelos teóricos y estudiar el efecto de la manipulación de los mismos en la respuesta del oyente.

Tal y como hemos señalado en una sección anterior, antes de este desarrollo tecnológico tuviera lugar, multitud de trabajos presentaban un gran déficit en el control estimular de forma que cada investigador empleaba en sus experimentos las piezas musicales que a su juicio generaban una cualidad emocional concreta en el oyente. En un gran número de ellos, se empleaban piezas propias del género de Música Clásica Occidental indicando el nombre de la pieza musical pero no el intérprete, versión o el segmento mismo de la exposición. En otros trabajos, ni tan siquiera se llega a nombrar la pieza musical indicando exclusivamente a que género musical pertenecía. De esta forma observamos imprecisiones metodológicas importantes en trabajos como en Labbé, Schmidt y Babin (2007), quienes

presentan conclusiones tan generalistas como que la música *Heavy Metal* resulta menos placentera para el oyente que la Música Clásica, y que ésta última es más relajante que la primera.

Este tipo de trabajos no tienen en cuenta aspectos tan básicos como la heterogeneidad de las piezas musicales que coexisten en un género musical concreto. Basta con remitirse a nuestro trabajo para apreciar que diferentes piezas musicales del repertorio clásico generan diferentes juicios emocionales (tabla 1 del material suplementario, capítulo 2). A nivel práctico esto resulta extremadamente relevante porque las discrepancias en los resultados que se observan en trabajos de investigación aplicados al campo de la salud pueden tener su origen en un pobre control del estímulo. Por consiguiente, un aspecto relevante de nuestro trabajo es que muestra la relevancia que tiene la estructura física del estímulo sobre el juicio emocional del oyente y como las conclusiones obtenidas pueden generalizarse a otros corpus estímulares siempre y cuando se mantenga la estructura estimular.

Otra conclusión importante de esta tesis referida a su aplicabilidad es comprensible si imaginamos el proceso de escucha musical como una situación comunicativa con sus diversos componentes (emisor, receptor, canal). En este caso, el emisor podría ser el intérprete de la pieza musical (que puede ser también el compositor de la misma), mientras que el receptor sería el oyente. En este sentido, emisor y receptor pueden compartir espacio y tiempo en el contexto comunicativo si hablamos de un concierto en directo; o no compartir ninguno de esos elementos si estamos escuchando una pieza musical en nuestro reproductor de música. Otro aspecto indispensable en el acto comunicativo es el **canal** por el cual se realiza la comunicación. En el lenguaje verbal disponemos fundamentalmente de dos canales de comunicación: el lenguaje hablado y el lenguaje escrito. Pero un aspecto importante para nuestra área es que en el contexto musical también disponemos de estos dos canales: podemos

escuchar música cuando ésta es interpretada bien en directo o bien de forma asincrónica, o también podemos leer la música a partir de las partituras o de la notación musical mediante tablaturas⁹. Esta analogía justifica las conclusiones de nuestro trabajo referidas a la necesidad de incorporar en nuestro modelo medidas propias de la estructura física del estímulo y, por tanto, relativas a la interpretación de la pieza musical, así como medidas propias de la escritura musical y por tanto, propias de la composición.

Es indudable que para que exista una comunicación eficaz entre un emisor y un receptor, ambos deben compartir además un **código**. Podemos disponer de todos los elementos de la comunicación necesarios, pero si el emisor emite su mensaje en un idioma desconocido para el receptor, hará la comunicación inviable. Es decir, si emisor y receptor no comparten el código en el que se está emitiendo el mensaje, la comunicación será ineficaz. Este aspecto es muy relevante cuando lo trasladamos al contexto musical y nos planteamos el efecto que tiene sobre el oyente una pieza musical interpretada en un lenguaje desconocido para el oyente. Asimismo es importante este concepto de código compartido cuando retomamos el modelo de redundancia de la señal de Balkwill y Thompson (1999).

Ambos ejemplos están introduciendo ya el contenido del último elemento de la comunicación que nos faltaba por introducir, el **mensaje**. Cuando hablamos de comunicación verbal es fácil entender el concepto de mensaje como una idea o concepto que pretende ser transmitido al receptor. Sin embargo cuando nos centramos en la comunicación musical el concepto de mensaje no es tan obvio. Teniendo por ahora en cuenta la música instrumental y teniendo presente el modelo neuropsicológico de Peretz y Coltheart (2003), el mensaje que se transmite en el contexto musical es eminentemente emocional. Un autor escribe una obra o pieza instrumental con el fin de transmitir un mensaje emocional a la audiencia. A partir de

⁹ Tablatura: notación musical simplificada que se basa en la representación de la posición de la nota en la cuerda o tecla del instrumento con que esta se toca.

los resultados de este trabajo podemos confirmar que en la mayoría de las piezas musicales existe una congruencia entre la intención comunicativa del compositor en los fragmentos musicales de Vieillard y cols. (2008) y el juicio emocional que experimenta el oyente.

Además, hemos podido generalizar los resultados experimentales obtenidos por Vieillard y cols. (2008) con sus piezas musicales, con otra muestra experimental obteniendo índices de categorización y evaluación emocional similares a los obtenidos en su trabajo. Esta similitud de la respuesta apoyaría la hipótesis de la estabilidad inter-sujeto en el reconocimiento emocional para piezas musicales dentro de una misma cultura (Balkwill y Thompson, 1999; Fritz y cols. 2009).

Centrándonos propiamente en la estructura del estímulo, a lo largo de esta tesis hemos empleado dos fuentes de variables para definir al estímulo musical. En el segundo capítulo empleamos medidas estáticas propias de la escritura musical para predecir la respuesta global del oyente a la pieza musical utilizando una red neuronal (véase la Figura 2) congruente con el modelo neuropsicológico de Peretz y Coltheart (2003).

Esta arquitectura nos ha permitido demostrar que bastan 6 propiedades fácilmente extraíbles de la partitura musical para poder predecir de forma precisa la respuesta global del oyente a la pieza emocional. Sin embargo, nos encontramos con una limitación importante y es que una misma pieza musical puede ser interpretada por artistas diferentes y generar respuestas diferentes en el oyente. En este caso, las propiedades extraídas de la partitura musical son idénticas pero no así el estímulo sonoro. Por este motivo en el capítulo 3 de esta tesis buscamos propiedades que extraídas directamente de la onda sonora nos permitieran capturar el efecto de la interpretación sobre el juicio emocional.

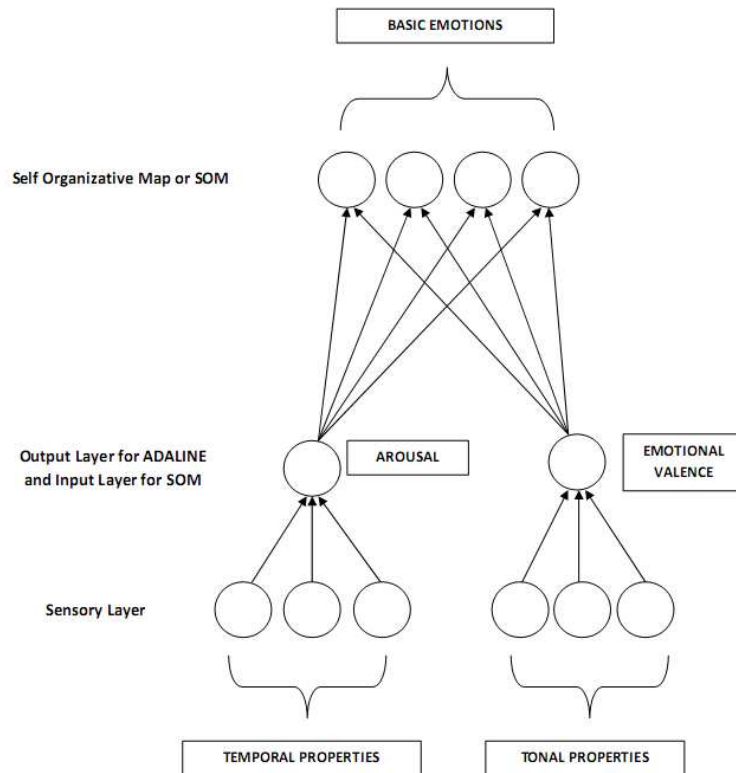


Figura 2: Estructura RNA obtenida en el capítulo 2 compuesta por dos ADALINE para el procesamiento de la activación y valencia emocional y una SOM para el juicio categórico

En este punto nos encontramos durante el desarrollo de la tesis con un gran problema puesto que el número de propiedades del estímulo musical bien en el dominio tiempo, o bien en el dominio de la frecuencia, que habían sido usadas tanto en ingeniería como en psicología era muy elevado. Esta multiplicidad de medidas era también un problema a la hora de juzgar algunos de los modelos propuestos por las razones indicadas en la introducción del tercer capítulo de esta tesis: 1) por un lado la gran mayoría de estos modelos no aportaban información sobre la importancia o el peso que cada una de estas medidas del estímulo tienen sobre el juicio emocional, 2) por otro lado, estos modelos obtenían altos coeficientes de determinación como consecuencia del sobreajuste, la redundancia y la colinealidad al introducir tantos predictores simultáneamente en el modelo.

Para solventar el problema de la colinealidad, decidimos utilizar el Análisis Factorial exploratorio como una técnica de reducción de la dimensionalidad de la estructura estimular en el dominio del tiempo debido a su buen funcionamiento. Esta técnica en esencia permite combinar variables linealmente relacionadas en una magnitud elevada en una única componente que simplifica el espacio dimensional intentando maximizar la varianza común de los elementos. Ante problemas de colinealidad en el modelado de procesos psicológicos debería ser una técnica de uso más frecuente.

Sobre la respuesta fisiológica

La **respuesta autónoma** es una respuesta involuntaria del organismo. Muchos son los trabajos que han empleado este tipo de medidas para estudiar la respuesta emocional al estímulo musical (Arheim, Mülberg, Pauli y Wiedemann, 2004; Baumgartner y cols. 2006; Blood y Zatorre, 2001; Burns, Labbé, Arke, Capeless, Cooksey, Steadman y Gonzales, 2002; Coutinho y Cangelos, 2011, Dellacherie y cols. 2010; Dousty, Daneshvar y Haghjoo, 2011; Etzel y cols. 2006; Gomez y Danuser, 2004 y 2007, Iwanaga y cols. 2005; Kennter-Mabiala y cols. 2007; Khalifa y cols. 2002; Khalifa y cols. 2008; Kolesch, 2015; Koelsch, Kilches, Steinbeis y Schelinski; 2008; Krumhansl, 1997; Labbé y cols. 2007; Nakahara, Furuya, Masuko, Francias y Kinoshita, 2011, Nakahara, Furuya, Obata, Masuko, y Kinoshita, 2009; Nater y cols. 2006; Roy y cols. 2008; Roy, Mailhot, Gosselin, Paquette y Peretz, 2009; Sammler y cols. 2007; Scheufeler, 2000, entre otros) siendo las medidas más usadas la conductancia eléctrica de la piel, medidas cardiovasculares y medidas respiratorias.

Sin embargo, tal y como señalamos en el cuarto capítulo de esta tesis, los resultados son heterogéneos entre investigaciones y aún no logramos entender completamente esa variabilidad en la experimentación. En futuros trabajos será necesario estudiar con mayor precisión la relación entre la respuesta autónoma del oyente en diferentes parámetros

(respiratorios, cardiacos y de conductancia eléctrica) y la respuesta emocional del oyente, así como precisar la relación entre la estimulación del sistema y dicha respuesta, especialmente en el dominio del tiempo donde hemos encontrado un aporte mayor de la respuesta autónoma a la predicción del juicio emocional.

Sin duda el aspecto más relevante a comentar en esta sección sobre nuestros resultados en el ámbito fisiológico se concentra en los dos segundos que se suceden entre el inicio de la presentación estimular y la emisión de un juicio emocional por parte del oyente. Es en estos dos segundos de exposición donde, de forma sistemática, mayor variabilidad hemos encontrado entre los oyentes (Figuras 3 a 10 de los anexos). Por tanto en un futuro, nuestro interés se centrará en el estudio de la respuesta de orientación y seguramente, en la introducción en el modelo de medidas de la actividad del Sistema Nervioso Central como el EEG o los potenciales evocados.

Sobre la aplicación de los modelos de redes neuronales en la investigación

También consideramos importante dedicar un apartado en esta sección a la metodología utilizada y a la aplicación del modelado mediante redes neuronales artificiales al estudio de los procesos psicológicos. Una de las dificultades con las que se encontraron Schubert (1999) y Korhonen (2004) es que a pesar de que los modelos lineales son los más simples y parsimoniosos, implican la asunción de supuestos muy restrictivos en el análisis que limitan la eficacia de los modelos en la predicción. Esta fue la sugerencia de planteó Korhonen (2004) en sus conclusiones y que fue retomada por Coutinho y Cangelosi (2009) y por nosotros mismos (Saiz-Clar y Reales, 2017). Este autor proponía el uso de las redes neuronales artificiales por la flexibilidad que plantean estas arquitecturas, completamente programables estructural y funcionalmente, pero que exigen por parte del usuario un conocimiento previo de las variables a utilizar para elegir una estructura apropiada.

Una ventaja principal de las RNA es que no estamos sujetos al cumplimiento de determinados supuestos sobre la distribución de las variables o sobre su relación para su aplicación. Tan solo empleamos supuestos paramétricos en la comparación entre las salidas de la red y los valores experimentales obtenidos aunque muchos trabajos juzgan la eficiencia de sus modelos basándose exclusivamente en el error de predicción a través del error cuadrático medio (ECM) o de la raíz cuadrada del mismo (RECM). El modelado por RNA también permite suponer relaciones entre los elementos más allá de la relación lineal a pesar de que el modelo lineal suele ser el más eficiente en términos de ausencia de sobreajuste y economía de parámetros.

De hecho, tanto en las ADALINE del segundo capítulo como en las conexiones entre la capa de *input* y la capa oculta de las redes Elman del quinto capítulo hemos preferido emplear como función de transferencia de las unidades de entrada de nuestras RNA la función de identidad. Esta función reproduce en esencia un modelo lineal donde las salidas son calculadas como sumatorio de las entradas por sus pesos, es decir, el mismo procedimiento que en cualquier modelo lineal. Sin embargo, la función de transferencia empleada en el segundo nivel de las redes Elman es la función tangente hiperbólica que supone una relación distinta a la lineal entre las salidas de la capa de computación y la salida de la red.

Otra gran ventaja de la aplicación de las RNA al estudio de los procesos psicológicos es que la gran variedad de arquitecturas posibles permite simular dichos procesos con un parecido muy razonable al procesamiento humano. Por ejemplo, en el capítulo 2 creamos una red tipo SOM para emitir el juicio categórico al estímulo musical a partir de la información procesada por el sistema de dos ADALINEs para la Activación y la Valencia Emocional respectivamente. En este sentido, amoldamos la estructura de la red neuronal para que fuera similar al tipo de relación propuesta teóricamente.

Al inicio del entrenamiento, cuando la red aún no ha aprendido a responder al estímulo, cualquier input provoca la activación de todas las neuronas de la capa SOM. La salida que emite la red es comparada con el valor correcto que debía haber emitido la red y la neurona que se ha estimulado correctamente incrementa su peso de conexión con la neurona sensorial a la que está conectada, mientras que toda neurona incorrectamente activada recibe un “castigo” reduciendo la fuerza de conexión. Con un número suficiente de ensayos la red aprende a emitir una respuesta con un mínimo error de predicción ajustando los pesos a los diferentes patrones de entrada presentados en las sucesivas épocas de entrenamiento. Tal y como señala Kallinen (2005), las SOM resultan un tipo de arquitectura perfecta para simular procesos de aprendizaje por ensayo y error en los que median los refuerzos positivos y el castigo.

También la arquitectura Elman empleada en el quinto capítulo de esta tesis nos aporta un enfoque novedoso en el modelado de procesos psicológicos y es que nos permite tener en consideración como fuente de variabilidad en la respuesta el estado del sistema en el momento previo a la entrada de información. Retomando el modelo neuropsicológico de Peretz y Coltheart (2003), el resultado de las redes Elman y el efecto que tiene la capa de memoria sobre la respuesta actual del oyente obliga a plantearse que la relación entre el componente emocional presente en el modelo y el repertorio de fragmentos musicales memorizados sea bidireccional y no exclusivamente desde la emoción al repertorio.

Por último señalar que en los sucesivos capítulos de esta tesis se observa que el mejor rendimiento en la predicción del juicio emocional del oyente, se obtiene cuando separamos el procesamiento de la activación del de la valencia emocional. Esto no es solo coherente con los modelos ya comentados de Peretz y Coltheart (2003) y Balkwill y Thompson (1999), sino con la investigación neurocognitiva sobre los déficits afásicos musicales. Por ejemplo, Peretz y

Herbert (1995) encuentran evidencias para distinguir dos tipos de afasias musicales o amusias: una para el procesamiento temporal del estímulo y otra afasia (más común en la población) para el tono.

A partir de todo lo expuesto en esta sección, podemos concluir que el modelado de procesos cognitivos mediante las Redes Neuronales Artificiales es un enfoque interesante para la comprensión del procesamiento humano de la información, pero también es un recurso importante en la investigación psicológica dado que nos permite no solamente simular procesos, sino también predecir el resultado de los mismos lo cual, a su vez, nos permite comparar la ejecución observada en la experimentación. Aunque los tres grupos de fuentes de variación consideradas en esta tesis han resultado suficientes para conseguir buenos índices de predicción (estructura del estímulo, variables fisiológicas y respuesta emocional), existen sin embargo otras variables que modulan la respuesta emocional del oyente a la escucha musical y que tendrán que ser introducidas paulatinamente en versiones posteriores de nuestro modelo.

Otros factores internos del sujeto

Durante el experimento descrito en el cuarto capítulo de esta tesis, recogimos información sobre algunas de las variables que modulan la respuesta emocional del oyente a la escucha musical como son determinados rasgos de personalidad, los hábitos de escucha musical o la preferencia musical. Sin embargo el volumen de datos obtenidos no resultaba manejable para esta tesis y decidimos posponer su análisis a futuros trabajos. En relación con este último conjunto de variables, es interesante comentar brevemente en qué medida pueden influir en la respuesta del oyente. En la bibliografía consultada para esta tesis hemos podido comprobar un recurrente interés en la investigación por relacionar determinados factores de personalidad con la escucha musical y en especial con la preferencia musical ya que esta

compone un factor fundamental en el grado en que un fragmento musical puede emocionar a un participante concreto.

Sin embargo, en muchas ocasiones este interés ha considerado negativamente determinados géneros musicales como el *Rock*, el *Heavy Metal* o el *Rap* frente a otros géneros musicales más neutros como el *Pop* y ha considerado otros géneros como la Música Clásica o el *Jazz* exponentes de un estándar preferible. De este tipo de trabajos podemos encontrar conclusiones como la de Carpentier y cols. (2003) quienes afirman que la preferencia musical por la música con contenido “desafiante” o “*defiant*” (v.g., *Heavy Metal* o *Punk*) es propia de personalidades rebeldes, hostiles y desinhibidas. O también en Coyne y Padilla-Walker (2015) quienes concluyen que los chicos en edad adolescente que escuchaban música con contenido agresivo y sexual presentaban más conductas antisociales y un menor número de conductas prosociales, a pesar de que en su trabajo tienen en cuenta factores tradicionalmente mucho más predictivos de estas conductas como la situación sociofamiliar del menor. También en Scheel y Westefeld (1999) concluyen que las personas con afinidad por el *Heavy Metal*, presentaban menos razones para vivir (a pesar de que la magnitud de la correlación era muy baja y semejante a la obtenida ante géneros como el *pop/rock* o música *Indie*), menor responsabilidad familiar y un menor respeto a las normas sociales. En este estudio por ejemplo, consideran como relevantes coeficientes de determinación del orden de .06.

Aunque el efecto realmente es pequeño, sí se ha establecido una relación entre algunos rasgos de personalidad como la búsqueda de sensaciones, la apertura mental o la extraversión sobre algunas características relacionadas con la escucha musical como la variabilidad en la preferencia musical (Pinilla, Saiz-Clar y Serrano, 2007). De forma que las personas con una puntuación elevada en el perfil de personalidad para la Apertura Mental o la Búsqueda de Sensaciones presentan preferencia por un mayor número de géneros musicales.

Otro factor muy interesante que modula la preferencia musical y como consecuencia, el agrado por la pieza musical en cuestión es la familiaridad de la pieza para el oyente. Spuznar, Schellenberg y Pliner (2002) observan que bastan alrededor de ocho exposiciones a una pieza musical concreta para que el agrado por la misma se incremente significativamente con respecto a la exposición inicial. Lo más interesante es que observan un efecto no solo de habituación al estímulo pasado este número de exposiciones donde éste pierde capacidad para generar el mismo placer en el oyente, sino que hacia la trigésima segunda exposición los oyentes generan una respuesta de rechazo y desagrado hacia la pieza musical. Este trabajo fue posteriormente replicado por Schellenberg, Peretz y Vieillard (2008) obteniendo semejantes resultados numéricos. Otros trabajos relacionan los efectos de la exposición sobre el agrado por la pieza musical con algunos factores de personalidad anteriormente señalados. Por ejemplo, Hunter y Schellenberg (2011) observan que existe un patrón diferente de habituación a una pieza musical en función de variables intrínsecas de la persona como su tendencia a la búsqueda de estímulos novedosos, es decir, la elevación en el rasgo de Apertura Mental. Pero también por otro lado, las personas con una menor necesidad de estimulación novedosa presentan una mayor resistencia a la habituación. Es importante tener presente en el ámbito terapéutico estas interacciones ya que el efecto emocional o de agrado por una pieza musical va a incrementar con las primeras exposiciones pero pasado un umbral y alcanzado el punto de sobreexposición, los efectos beneficiosos de la pieza musical se revierten generando incluso aversión por la misma.

También hay otros trabajos que identifican una relación entre determinados rasgos de personalidad y la respuesta fisiológica del oyente ante la escucha musical. Por ejemplo en Zuckerman, Simons y Como (1987), observan que los buscadores de sensaciones presentan una respuesta de orientación caracterizada por un descenso de la frecuencia cardiaca con el aumento de la intensidad sonora, mientras que los oyentes con menor puntuación en este

rasgo presentan una respuesta defensiva, caracterizada por un incremento de la frecuencia cardiaca a medida que aumento la intensidad sonora del estímulo. No consiguen observar este patrón con la intensidad de la EDA pero señalan otros trabajos (v.g., Ridgeway y Hare, 1981 o Stelmack, Plouffe y Falkenberg, 1983) donde si identifican dicho patrón. A la vista de estos resultados, todas las fuentes de variación señaladas en este apartado resultan importantes para reducir el error de predicción del modelo e incrementar su precisión. Por tanto en futuras revisiones de nuestro modelo será necesario dedicar un espacio en el mismo a los factores de personalidad y a la preferencia musical.

Aplicaciones de nuestro trabajo

Como indicábamos en la introducción, el objetivo de este trabajo consistía en investigar cómo funcionaba el mecanismo implicado en la respuesta emocional del oyente a la música. Aún queda mucho por entender pero estamos más cerca de comprenderlo. Este trabajo ha conseguido desvelar algo más sobre el bloque emocional que aparece en el modelo de Peretz y Coltheart (2003) aportando una idea más clara de qué sucede a nivel emocional cuando procesamos un estímulo musical. Pero más allá de las aplicaciones descritas en este capítulo podemos extender la aplicabilidad de estos resultados a los ámbitos clínico y computacional.

Aplicaciones en el ámbito clínico

Tal y como hemos señalado durante esta tesis, a lo largo de la investigación se ha podido comprobar que la música es capaz de generar emociones en el oyente y que estas emociones son relativamente consistentes a lo largo del tiempo (Bigan y cols. 2005) y entre diferentes oyentes (Eerola y cols., 2009; Hunter, Schellenberg y Schimmack, 2010; Vieillard y cols. 2008). Asimismo, se ha podido comprobar que el organismo responde de forma involuntaria a determinadas características del estímulo musical. En Gomez y Danuser (2007)

observamos, por ejemplo, que propiedades como el tempo o la intensidad sonora incrementan la activación del organismo y que esta respuesta se manifiesta a su vez en cambios en la respuesta cardiaca (Iwanaga y cols. 2005; Baumgartner y cols. 2006), cambio en los parámetros respiratorios (Gomez y Danuser, 2007) o cambios en la respuesta de conductancia eléctrica (Nater y cols. 2006). Es decir, ante la exposición a estímulos musicales se ha encontrado que determinadas propiedades de dichos estímulos modulan la respuesta autónoma del oyente y el juicio emocional sobre el mismo.

Este potencial de la música para inducir estados placenteros o de tranquilidad se hace patente en los trabajos sobre el miedo y la ansiedad. En este sentido, Roy y cols. (2008) y Zhao y Chen (2009) demostraron que la sensación subjetiva de miedo se reducía significativamente en una tarea de dolor inducido cuando los participantes estaban expuestos a música placentera. Es decir, que cuando la música era agradable para el oyente, la intensidad del miedo a meter una mano en agua caliente era menor que cuando no había música o ésta era desagradable (Roy y cols. 2008). Estos mismos efectos se han visto fuera de las tareas de laboratorio y en el campo clínico, concretamente en el pánico y ansiedad ante intervenciones quirúrgicas (Alred, Byers y Sole, 2010; Costa y cols. 2010; Lee, Chao, Yiin, Chiang y Chao, 2011; Miller y cols. 2002; Voss, y cols., 2004; Wang y cols., 2002), pruebas médicas invasivas (Agku y Okoye, 2007; Smole, Topp y Singer, 2002) o en tratamientos como radioterapia (Chen, Wang, Shih y Wu, 2013). Por ejemplo en el trabajo de Voss y cols. (2004) observaron que los pacientes que eran expuestos a música sedativa antes de una operación a corazón abierto tenían menores niveles de ansiedad y miedo tanto antes como después de la cirugía.

Asimismo, en muchos de estos estudios se observa también una reducción de las medidas fisiológicas de la ansiedad ante la exposición a estímulos musicales relajantes (Mok

y Wong, 2003; Wang y cols., 2002) como son la frecuencia cardiaca (Mok y Wong, 2003), parámetros respiratorios, de respuesta de conductancia eléctrica de la piel o nivel de catecolaminas u oxitocina en sangre (Nilsson, Unosson y Rawal, 2005; Nilsson y cols., 2009), entre otras medidas. Incluso en algunos de estos trabajos han conseguido no solo demostrar que el miedo previo puede reducirse de forma efectiva con música y por tanto mejorar el confort del paciente, sino que esto supone un ahorro en términos de dosis de sedantes, hipnóticos u opioides (Miller y cols., 2002; Nilsson, Kokinsky, Nilsson, Sidenvall y Enskär, 2009; Pellino y cols. 2005) y analgésicos (Huang, Good y Zauszniewsky, 2010; Miller y cols., 2002 y Voss y cols., 2004)

La prevalencia aproximada que los trastornos de ansiedad tiene en los países Occidentales es del 30% y frecuentemente se encuentran asociados a otros tipos de trastornos como los relacionados con el sueño. En España se estima que aproximadamente 1 de cada 5 personas mayores de 15 años cumplen criterios diagnósticos para el insomnio (Ohayon y Sagales, 2010). La capacidad de la música para reducir la respuesta fisiológica de activación así como su capacidad para estimular las ondas Delta y Gamma cerebrales (Hauk, Metzner, Rohlffs, Lorenz y Engel, 2013) y su capacidad para inducir estados placenteros pueden ser usados como tratamiento coadyuvante en el insomnio (De Niet, Tiemens, Lendemeijer y Hutschemaekers, 2009). La mejora en la calidad del sueño en pacientes insomnes mediante la exposición a estímulos musicales ha sido descrita por diversos autores (Chang, Lai, Chen Hsieh y Lee, 2012 o Ziv, Rotem, Arnon y Haimov, 2008).

Así por ejemplo, en el citado trabajo de Miller y cols. (2002), se observó que los pacientes expuestos a una terapia musical previa a la operación manifestaron no solo menos miedo y ansiedad sino que los pacientes autoinformaban de una mayor calidad del sueño la noche previa a la operación que el grupo control. Por ejemplo, en el trabajo de Chang y cols.

(2012) se observó que en los pacientes insomnes expuestos a música relajante durante 45 minutos antes de acostarse, la duración de la fase II del sueño era más corta mientras que la fase REM en este grupo era más larga que en el grupo control; es decir, que los pacientes insomnes expuestos a música relajante entraban antes en la fase REM y ésta era más duradera que en los pacientes insomnes que no eran expuestos a música relajante. Además los autores señalan que este efecto podía observarse tanto con la música autoseleccionada por los pacientes como por la música seleccionada por los investigadores.

Aplicaciones en ingeniería y computación

Sin embargo no sólo la psicología puede obtener un beneficio del enfoque computacional. Este aporte al corpus científico puede resultar productivo también en el ámbito de la ingeniería y la computación, concretamente en los denominados “*music retrieval systems*”. Yang y Chen (2012) en su revisión sobre reconocimiento automático de la emoción musical ponen de manifiesto la relevancia que tiene la escucha musical en las personas en su día a día, acompañándolo en actividades tan cotidianas como conducir, estudiar, hacer deporte o realizar las tareas del hogar (Juslin y Sloboda, 2001). Yang y Chen (2012) mencionan un estudio de la importante plataforma de música *online* **LastFm** en la cual, la etiqueta “emoción” fue la tercera más empleada por los usuarios de la misma (Lamere, 2008).

La importancia que en estas plataformas otorgan los usuarios a la emoción a la hora de seleccionar la música que quieren escuchar, ha centrado el interés tanto del ámbito científico como el de la industria (Huq, Bello, Sarroff, Berger y Rowe, 2009; Lu et al. 2006; Yang y Chen, 2011). En estas plataformas se han desarrollado algoritmos que **basados en los modelos psicológicos de la emoción** pretenden ofrecer al usuario listas de canciones que se ajusten a su estado de ánimo y/o a la tarea que están realizando. Por ejemplo, un usuario que durante su sesión de estudio desee escuchar una “*playlist*” acorde a esa tarea, espera una lista

de canciones que no interfirieran con dicha actividad. Dicha *playlist* por tanto tendría que estar compuesta por música placentera y relajante para el oyente para generar una sensación de tranquilidad que favorezca el estudio, y por supuesto, esta música debería ser instrumental para no generar un efecto de interferencia del contenido verbal. Si por ejemplo el usuario está realizando una actividad física, la *playlist* que ofrezca la plataforma, debería estar compuesta por piezas que activen y motiven al oyente siendo adecuada la música con un ritmo marcado y rápido. El uso de la música para modular el comportamiento del oyente no es, sin embargo, una corriente actual. Ya Smith y Curnow (1966) en la década de los 60 estudiaban las propiedades que debían tener las piezas musicales para estimular en el oyente la conducta de consumo.

Hoy en día se están desarrollando también los denominados sistemas “*Music Emotion Recognition*” o MER. Estas aplicaciones informáticas (“apps”), buscan poder ofrecer la mejor *playlist* al oyente pero lo hacen recogiendo información del usuario bien a nivel psicofisiológico o mediante el reconocimiento de patrones de expresión facial (v.g., Kim y André, 2008 o Lin et al. 2009). Retomando los ejemplos anteriores, los MER ayudarían al estudiante a recuperar la activación cuando detectaran una bajada en su actividad autónoma ofreciéndole de manera automática música más activadora. El deportista por su parte podría regular su actividad física a unos parámetros fisiológicos concretos, por ejemplo, detener la escucha o reducir algunos parámetros como el tempo musical, cuando alcanzara una frecuencia cardiaca demasiado elevada para su salud.

Este último ejemplo nos sirve para unir los dos ámbitos de aplicación señalados en estas conclusiones, es decir, combinar los sistemas MER con el *biofeedback*. Esto puede tener importantes aplicaciones en terapia dado que podríamos desarrollar un sistema de *biofeedback* que detecte, por ejemplo, parámetros fisiológicos de activación en el paciente y que

automáticamente, le presente música con unas características apropiadas para inducir un estado de relajación en el mismo. La exposición musical se iría regulando a medida que se va registrando y procesando en tiempo real dichos parámetros fisiológicos del paciente. Por este motivo es sumamente importante relacionar los tres sistemas planteados en el cuarto capítulo de esta tesis: las propiedades del estímulo, la respuesta fisiológica del oyente y su evaluación emocional. Todo ello, además, en tiempo real.

Los sistemas MER se encuentran a su vez íntimamente ligados a los denominados sistemas MIR o “*Music Information Retrieval*” dedicados a la construcción de algoritmos y modelos que son capaces de reconocer y percibir de forma automática la emoción en la música. La diferencia entre los sistemas MER y MIR radica en que los primeros reconocen la emoción en el oyente mientras que los segundos buscan identificar la emoción desde la estructura del estímulo. Algunos de estos sistemas MIR (como el nuestro) se basan para su desarrollo en los sistemas de aprendizaje, memoria, percepción y/o procesamiento de la información humanos.

El último campo de aplicación que mencionan Yang y Chen (2012) donde la ciencia psicológica puede aportar aplicaciones tecnológicas dirigidas a mejorar la salud y ejecución de las personas es en el ámbito de la Inteligencia Artificial (IA) y la robótica. El conocimiento que podemos extraer de los sistemas, modelos y algoritmos MIR como el nuestro pueden ser aplicados en IA para simular procesos de reconocimiento emocional en el campo de la robótica. Estos algoritmos podrán ser más eficaces y útiles en el día a día, cuanto más se aproximen al sistema de procesamiento humano, y mejor puedan predecir nuestras necesidades. Si unimos los sistemas MIR y MER en el ámbito de la IA, podremos desarrollar sistemas robóticos que puedan extraer información emocional del entorno, reconocer nuestra emoción y/o ofrecernos estimulación capaz de modificar dicho estado de ánimo.

La incursión de la Psicología en la IA también redundará en un beneficio para el avance en la investigación en nuestro campo. El desarrollo de sistemas de IA similares al procesamiento humano de la información, con tasas de aprendizaje, acierto y error parecidas a las constatadas a nivel experimental, nos permitirá poder predecir comportamientos humanos en situaciones complejas para la investigación en laboratorio. Podremos por ejemplo, predecir la respuesta del oyente en situaciones de ambigüedad de información o también ante condiciones extremas (v.g., intensidad sonora elevada o escucha musical continua), y determinar el alcance de situaciones de riesgo para la salud del oyente sin necesidad de experimentar directamente en humanos. Por ejemplo, para determinar el umbral de discomfort en zonas urbanas donde coexisten zonas de ocio nocturno. Obtener una evidencia científica sobre el efecto que pueden tener factores ya tratados en esta sección como la sobreexposición sobre la respuesta del organismo tanto a nivel emocional como a nivel fisiológico puede resultar útil también en este contexto legal.

Como hemos expuesto en estas conclusiones, el estudio de la respuesta del oyente al estímulo musical no es solo importante porque se trate de un estímulo presente constantemente en nuestra vida diaria, en la publicidad, en los comercios, en nuestras actividades de ocio. No lo es tampoco porque la industria musical mueva cada año millones de euros en beneficios y sea una de las actividades de ocio más populares en la población. Es importante por la aplicabilidad que tienen los resultados expuestos en diferentes contextos como la computación o el ámbito clínico. Solo hemos dado unas pinceladas de estas aplicaciones habiendo múltiples campos específicos de actuación, y pudiendo conectar este conocimiento con otros procesos cognitivos como el lenguaje, la memoria o el aprendizaje.

Sin embargo, para poder avanzar en este ámbito es necesario superar tanto las limitaciones presentadas en algunos de los capítulos de esta tesis, como la dificultad de

encontrar cabida en un contexto investigador en el que el estudio de la música y sus efectos sobre el oyente se encuentra en un segundo plano de interés. En un sistema que concibe la formación y educación musical como un conocimiento secundario y prescindible en la formación reglada, es difícil encontrar cabida en el contexto investigador. Por fortuna en otros contextos como en la ingeniería o en la medicina, cada año se publican nuevos artículos sobre “música” y “emoción”. Porque parece que ambos conceptos están íntimamente unidos, porque es difícil concebir el uno sin el otro, y porque tal y como señaló Nietzsche porque “sin música, la vida sería un error”.

Bibliografía

- Agku, K.K. y Okoye, I.J. (2007). The effect of music on the anxiety levels of patients undergoing hysterolapingography. *Radiography*. 13, 122-125. <https://doi.org/10.1016/j.radi.2005.12.002>
- Alred, K.D., Byers, J.F. y Sole, M.L. (2010). The effect of music on postoperative pain and anxiety. *Pain management Nursing*. 11 (1), 15-25. <https://doi.org/10.1016/j.pmn.2008.12.002>
- Amrhein, C., Mühlberger, A., Pauli, P. y Wiedemann, G. (2004). Modulation of event-related brain potentials during affective picture processing: a complement to startle reflex and skin conductance response?. *Journal of Psychophysiology*. 54, 231-240. <https://doi.org/10.1016/j.ijpsycho.2004.05.009>
- Balkwill, L. L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music perception: An interdisciplinary journal*, 17(1), 43-64. <https://doi.org/10.2307/40285811>
- Balkwill, L.L., Thompson, W.F., y Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research*, 46(4), 337-349. Doi: 10.1111/j.1468-5584.2004.00265.x
- Barrett, P. (1986). Factor comparison: An examination of three methods. *Personality and Individual Differences*, 7(3), 327-340. DOI: [https://doi.org/10.1016/0191-8869\(86\)90008-5](https://doi.org/10.1016/0191-8869(86)90008-5)
- Baugmgartner, T., Esslen, M. y Jäncke, L. (2006). From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *International Journal of Psychophysiology*. 60, 34-43.
- Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., & Sandler, M. B. (2005). A tutorial on onset detection in music signals. *IEEE Transactions on speech and audio processing*, 13(5), 1035-1047.
- Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*, 19(8), 1113-1139. doi: 10.1080/02699930500204250
- Blood, A.J. y Zatorre, R.J. (2001) Intensely pleasurable responses to music correlate with activity in brain regions implicated with reward and emotion. *Proceedings of the National Academy of Sciences*, 98, 11818-11823. Doi: 10.1073/pnas.191355898
- Boltz, M. G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception & Psychophysics*, 60 (8), 1357-1373. DOI: 10.3758/BF03207998

- Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion, 1*(3), 276. <http://dx.doi.org/10.1037/1528-3542.1.3.276>
- Burns, J.L., Labbé, E., Arke, B., Capeless, K., Cooksey, B., Steadman, A., y Gonzales, C. (2002). The effects of different Types of music on perceived and physiological measures of stress. *Journal of Music Therapy, 39* (2), 101-116. Doi: <https://doi.org/10.1093/jmt/39.2.101>
- Cabrera, D. (1999, November). PSYSOUND: A computer program for psychoacoustical analysis. In *Proceedings of the Australian Acoustical Society Conference* (Vol. 24, pp. 47-54).
- Cameron, D. J., & Grahn, J. A. (2014). Neuroscientific investigations of musical rhythm. *Acoustics Australia, 42* (2), 111.
- Carpentier, F. D., Knobloch, S., & Zillmann, D. (2003). Rock, rap, and rebellion: comparisons of traits predicting selective exposure to defiant music. *Personality and individual differences, 35*(7), 1643-1655. [https://doi.org/10.1016/S0191-8869\(02\)00387-2](https://doi.org/10.1016/S0191-8869(02)00387-2)
- Carretié, L. (2009). Psicofisiología. Ediciones Pirámide.
- Cattell, R. B. (1978). The scientific use of factor analysis in behavioral and life sciences. New York: Plenum.
- Chang, E. T., Lai, H. L., Chen, P. W., Hsieh, Y. M., & Lee, L. H. (2012). The effects of music on the sleep quality of adults with chronic insomnia using evidence from polysomnographic and self-reported analysis: a randomized control trial. *International journal of nursing studies, 49*(8), 921-930. <https://doi.org/10.1016/j.ijnurstu.2012.02.019>
- Chen, L. C., Wang, T. F., Shih, Y. N., & Wu, L. J. (2013). Fifteen-minute music intervention reduces pre-radiotherapy anxiety in oncology patients. *European Journal of Oncology Nursing, 17*(4), 436-441. <https://doi.org/10.1016/j.ejon.2012.11.002>
- Coath, M., Denham, S. L., Smith, L. M., Honing, H., Hazan, A., Holonowicz, P., & Purwins, H. (2009). Model cortical responses for the detection of perceptual onsets and beat tracking in singing. *Connection Science, 21*(2-3), 193-205. DOI: 10.1080/09540090902733905
- Costa, A., Montalbano, L.M., Orlando, A., Ingoglia, C., Linea, C.,Giunta, M., Mancuso, A., Mocciano, F., Bellingardo, R., Tiné, F. y D'Amico, G. (2010). Music for colonoscopy: A single-blind randomized controlled trial. *Digestive and Liver Disease, 42*, 871-876. <https://doi.org/10.1016/j.dld.2010.03.016>
- Coutinho, E., & Cangelosi, A. (2009). The use of spatio-temporal connectionist models in psychological studies of musical emotions. *Music Perception, 27*(1), 1-15. doi: 10.1525/mp.2009.27.1.1.

- Coutinho, E., & Cangelosi, A. (2011). Musical emotions: Predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. *Emotion, 11*(4), 921-937. DOI:10.1037/a0024700
- Coyne, S. M., & Padilla-Walker, L. M. (2015). Sex, violence, & rock n'roll: Longitudinal effects of music on aggression, sex, and prosocial behavior during adolescence. *Journal of adolescence, 41*, 96-104. <https://doi.org/10.1016/j.adolescence.2015.03.002>
- Dalla-Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition, 80*(3), B1-B10. doi: 10.1016/S0010-0277(00)00136-0.
- De Niet, G., Tiemens, B., Lendermeijer, B., & Hutschemaekers, G. (2009). Musicassisted relaxation to improve sleep quality: meta-analysis. *Journal of advanced nursing, 65*(7), 1356-1364. Doi: 10.1111/j.1365-2648.2009.04982.x
- Dellacherie, D., Roy, M., Hugueville, L., Peretz, I. y Samson, S. (2010) The effect of musical experience on emotional self-reports and psychophysiological responses to dissonance. *Psychophysiology* , 48, 337-349. Doi: [10.1111/j.1469-8986.2010.01075.x](https://doi.org/10.1111/j.1469-8986.2010.01075.x)
- Dousty, M., Daneshvar, S. y Haghjoo, M. (2011). The effects of sedative music, arousal music, and silence on electrocardiography signals. *Journal of Electrocardiology. 44* (3), 396 e.1- 396-e6. Doi: <http://dx.doi.org/10.1016/j.jelectrocard.2011.01.005>
- Droit-Volet, S., Ramos, D., Bueno, J. L., & Bigand, E. (2013). Music, emotion, and time perception: The influence of subjective emotional valence and arousal?. *Frontiers in Psychology, 4*, 4-12. doi: 10.3389/fpsyg.2013.00417
- Eerola, T. (2012). Modeling listeners' emotional response to music. *Topics in Cognitive Science, 4*(4), 607-624. doi: 10.1111/j.1756-8765.2012.01188.x
- Eerola, T., & Vuoskoski, J.K. (2010). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music, 39*(1), 18-49. doi: 10.1177/0305735610362821
- Eerola, T., Lartillot, O., & Toiviainen, P. (2009, October). Prediction of multidimensional emotional ratings in music from audio using multivariate regression models. *Proceedings of 10th International Conference on Music Information Retrieval. ISMIR 2009* (pp. 621-626).
- Egermann, H., Fernando, N., Chuen, L., & McAdams, S. (2015). Music induces universal emotion-related psychophysiological responses: comparing Canadian listeners to Congolese Pygmies. *Frontiers in Psychology. 7*. doi: 10.3389/fpsyg.2014.01341
- Ekman, P. (1976). Movements with precise meanings. *Journal of Communication, 26*(3), 14-26. Doi: 10.1111/j.1460-2466.1976.tb01898.x

- Etzel, J.A., Jhonsen, E.L., Dickerson, J., Tranel, D. y Adolphs, R. (2006). Cardiovascular and respiratory responses during musical mood induction. *International Journal of Psychophysiology*. 61 (1), 57-69. <https://doi.org/10.1016/j.ijpsycho.2005.10.025>
- Fausett, L. (2006). Simple neural nets for pattern classification. En: *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*. Ed: Pearson Education.
- Feldman-Barrett, L. & Russell J.A. (1999). The structure of current affect: Controversies and emerging consensus. *Current directions in psychological science*. 8, 10-14. <http://dx.doi.org/10.1111/1467-8721.00003>
- Feldman-Barrett, L. (2004). Feelings words? Understanding the content in self-report ratings of emotional experience. *Journal of Personality and Social Psychology*, 87, 266-281. <http://dx.doi.org/10.1037/0022-3514.87.2.266>
- Feldman-Barrett, L. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and social psychology review*, 10(1), 20-46. [10.1207/s15327957pspr1001_2](http://dx.doi.org/10.1207/s15327957pspr1001_2)
- Feldman-Barrett, L. , Quigley, K. S., Bliss-Moreau, E., & Aronson, K. R. (2004). Interoceptive sensitivity and self-reports of emotional experience. *Journal of personality and social psychology*, 87(5), 684. <http://dx.doi.org/10.1037/0022-3514.87.5.684>
- Feng, Y., Zhuang, Y., & Pan, Y. (2003, October). Music information retrieval by detecting mood via computational media aesthetics. In *Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on* (pp. 235-241). IEEE.
- Fernández-Sotos, A., Fernández-Caballero, A., & Latorre, J. M. (2016). Influence of tempo and rhythmic unit in musical emotion regulation. *Frontiers in Computational Neuroscience*, 10, 1-13. doi: 10.3389/fncom.2016.00080.
- Flom, R., Gentile, D.A. y Pick, A.D. (2008). Infant's discrimination of happy and sad music. *Infant Behavior and Development*. 31, 716-728. <https://doi.org/10.1016/j.infbeh.2008.04.004>
- Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A.D., & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology*, 19, 1-4. DOI: 10.1016/j.cub.2009.02.058
- Gabrielsson, A. y Lindström, E. (2010). The role of structure in the musical expression of emotions. En: Juslin, P.N. y Sloboda, J.A. (Ed). *Handbook of music and emotion: Theory, Research, Applications*. Oxford University Press.
- Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to “happy-sad” judgements in Equitone melodies. *Cognition and Emotion*, 17(1), 25-40. doi: 10.1080/02699930302279

- Gainza, M., Lawlor, B., Coyle, E., & Kellecher, A. (2004). Onset Detection and Music Transcription for the Irish tin Whistle. *Irish Signals and Systems Conference, ISSC, Belfast, Northern Ireland*.
- Ghezeljeh, T. N., Ardebili, F. M., Rafii, F., & Haghani, H. (2016). The effects of patient-preferred music on anticipatory anxiety, post-procedural burn pain and relaxation level. *European Journal of Integrative Medicine*. <https://doi.org/10.1016/j.eujim.2016.12.004>.
- Gomez, P, Stahel, W.A. y Danuser, B. (2004). Respiratory responses during affective picture viewing. *Biological Psychology*. 67(3), 359-373.
- Gomez, P. y Danuser, B. (2004). Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology*. 53 (2), 91-103.
- Gomez, P. y Danuser, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion*. 7 (2), 377-387.
- Gomez, P., Zimmermann, P., Guttormesen-Schär, S. y Danuser, B. (2005). Respiratory responses associated with affective processing of film stimuli. *Biological Psychology*. 68, 3, 223-235.
- Gómez-Romero, M., Jiménez-Palomares, M., Rodríguez-Mansilla, J., Flores-Nieto, A., Garrido-Ardila, E. M., & González-LópezArza, M. V. (2014). Beneficios de la musicoterapia en las alteraciones conductuales de la demencia. Revisión sistemática. *Neurología*. <https://doi.org/10.1016/j.nrl.2014.11.001>
- Gosselin, N., Peretz, I., Johnsen, E., & Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. *Neuropsychologia*, 45, 236-244. doi: 10.1016/j.neuropsychologia.2006.07.012
- Gosselin, N., Peretz, I., Noulhiane, M., Hasboun, D., Beckett, C., Baulac, M. y Samsom, S. (2005). Impaired recognition of scary music following unilateral temporal lobe excision. *Brain*. 128, 628-640.
- Hauck, M., Metzner, S., Rohlfs, F., Lorenz, J., & Engel, A. K. (2013). The influence of music and music therapy on pain-induced neuronal oscillations measured by magnetencephalography. *Pain*, 154(4), 539-547.
- Hein, G., Lamm, C., Brodbeck, C., & Singer, T. (2011). Skin conductance response to the pain of others predicts later costly helping. *PLoS One*, 6(8), 1–6. doi: 10.1371/journal.pone.0022759 .
- Hevner, K. (1935). The affective character of the major and minor modes in music. *American Journal of Psychology*, 47(1), 103-118. doi: 10.2307/1416710
- Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology*, 48(2), 246-268. doi: 10.2307/1415746

- Hevner, K. (1937). The affective value of pitch and tempo in music. *American Journal of Psychology*, 49(4), 621-630. doi: 10.2307/1416385
- Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185. DOI:10.1007/BF02289447
- Hsu, K. C., Chen, L. F., & Hsieh, P. H. (2016). Effect of music intervention on burn patients' pain and anxiety during dressing changes. *Burns*, 42(8), 1789-1796. <https://doi.org/10.1016/j.burns.2016.05.006>.
- Huang, S.T., Good, M. y Zauszniewsky, J.A. (2010). The effectiveness of music in relieving pain in cancer patients: A randomized controlled trial. *International Journal of nursing Studies*. 47, 1354-1362. <https://doi.org/10.1016/j.ijnurstu.2010.03.008>
- Hui, C.M., Fok, H.K., y Bond, MH. (2009). Who feels more ambivalence?: Linking dialectical thinking to mixed emotion. *Personality and Individual Differences*. 46, 493-498.
- Hunter, P. G., Schellenberg, E. G., y Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition and Emotion*, 22, 327-352.
- Hunter, P.G. y Schellenberg, E.G. (2011). Interactive effects of personality and frequency of exposure on liking for music. *Personality and Individual Differences*. 50, 175-179.
- Hunter, P.G., Schellenberg, E.G. y Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: similarities, differences, and mixed emotions. *Psychology of aesthetics, Creativity, and the Arts*. 4 (1), 47-56.
- Huq, A., Bello, J. P., Sarroff, A., Berger, J., & Rowe, R. (2009). Sourcetone: An automated music emotion recognition system. In *Proceedings of the International Conference on Music Information Retrieval*.
- Husain, G., Thompson, W.F., & Schellenberg, E.G. (2002). Effects of musical tempo and mode on arousal, mood and spatial abilities. *Music Perception*, 20(2), 151-171. doi: 10.1525/mp/2002.20.2.151
- Ilie, G. y Thompson, W.F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*. 23, 319-329.
- Iwanaga, M., Kobayashi, A. y Kawasaki, C. (2005). Heart Rate variability with repetitive exposure to music. *Biological Psychology*. 70 (1), 61-66. <https://doi.org/10.1016/j.biopsycho.2004.11.015>
- Jausovec, N.; Jausovec, K y Gerlic, I. (2006). Influence of Mozart's music on brain activity in the process of learning. *Clinical Neurophysiology*. 117, 2703-2714.

- Johnson, B., Raymond, S., & Goss, J. (2012). Perioperative music or headsets to decrease anxiety. *Journal of PeriAnesthesia Nursing*, 27(3), 146-154. <https://doi.org/10.1016/j.jopan.2012.03.001>
- Jones, S.J. y Zigler, E. (2002). The Mozart effect not learning from history. *Applied Developmental Psychology*. 23, 355-372.
- Kallinen, K. (2005). Emotional ratings of music excerpts in the western art music repertoire and their self-organization in the Kohonen neural network. *Psychology of Music*, 33(4), 373-393. doi: 10.1177/0305735605056147
- Kallinen, K., Saari, T., Ravaja, N., & Salminen, M. (2006). The Effects of Pre-existing Moods on the Emotional Responses to Music. Paper presented at the International Conference on Music Perception and Cognition, Alma Mater Studiorum University of Bologna.
- Kawakami, A., Furukawa, K., Katahira, K., & Okanoya, K. (2013). Sad music induces pleasant emotion. *Frontiers in psychology*, 4, 311. doi: 10.3389/fpsyg.2013.00311
- Kenealy, P., y Monseth, A. (1994). Music and IQ tests. *The Psychologist*, 7, 346.
- Kenntner-Mabiala, R., Gorges, S., Alpers, G.W., Lehmann, A.C., y Pauli, P. (2007). Musically induced arousal affects pain perception in females but not in males: a psychophysiological examination. *Biological Psychology*. 75 (1), 19-23. <https://doi.org/10.1016/j.biopsycho.2006.10.005>
- Khalifa, S., BELLA, S. D., Roy, M., Peretz, I., & Lupien, S. J. (2003). Effects of relaxing music on salivary cortisol level after psychological stress. *Annals of the New York Academy of Sciences*, 999(1), 374-376. Doi: 10.1196/annals.1284.045
- Khalifa, S., Peretz, I., Bolndin, J.P. & Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience letters*, 328(2), 145-149. [https://doi.org/10.1016/S0304-3940\(02\)00462-7](https://doi.org/10.1016/S0304-3940(02)00462-7)
- Khalifa, S., Roy, M., Rainville, P., Dalla Bella, S., & Peretz, I. (2008). Role of tempo entrainment in psychophysiological differentiation of happy and sad music?. *International Journal of Psychophysiology*, 68(1), 17-26. <https://doi.org/10.1016/j.ijpsycho.2007.12.001>
- Kim, J., & André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, 30(12), 2067-2083. Doi: 10.1109/TPAMI.2008.26
- Klapuri, A. (1999, March). Sound onset detection by applying psychoacoustic knowledge. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on* (Vol. 6, pp. 3089-3092).

- Klassen, J.A., Liang, Y., Tjosvold, L. (2008). Music for pain and anxiety in children undergoing medical procedures: A systematic review of randomized controlled trials. *Ambulatory Pediatrics*, 8, 117-128. <https://doi.org/10.1016/j.ambp.2007.12.005>
- Ko, C. H., Chen, Y. Y., Wu, K. T., Wang, S. C., Yang, J. F., Lin, Y. Y., ... & Hsieh, M. H. (2016). Effect of music on level of anxiety in patients undergoing colonoscopy without sedation. *Journal of the Chinese Medical Association*. <https://doi.org/10.1016/j.jcma.2016.08.010>.
- Koelsch, S. (2015). Music-evoked emotions: Principles, brain correlates, and implications for therapy. *Annals of the New York Academy of Sciences*, 1337(1), 193-201. doi: 10.1111/nyas.12684
- Koelsch, S., Kilches, S., Steinbeis, N. y Schelinski, S. (2008). Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. *PLoS ONE*. 3 (7), e2631.
- Korhonen, M. (2004). *Modeling continuous emotional appraisals of musical experience* (unpublished master's thesis). University of Waterloo, Canada.
- Korhonen, M. D., Clausi, D. A., & Jernigan, M. E. (2005). Modeling emotional content of music using system identification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 36(3), 588-599. doi: [10.1109/TSMCB.2005.862491](https://doi.org/10.1109/TSMCB.2005.862491)
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological psychology*, 84 (3), 394-421. doi:10.1016/j.biopsycho.2010.03.010
- Krumhansl, C.L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology*. 51 (4), 336-353. <http://dx.doi.org/10.1037/1196-1961.51.4.336>
- Labbé, E., Schmidt, N., Babin, J. y Pharr, M. (2007). Coping with stress: the effectiveness of different Types of music. *Applied Psychophysiology and Biofeedback*. 32 (3-4), 163-168.
- Lamere, P. (2008). Social tagging and music information retrieval. *Journal of new music research*, 37(2), 101-114. <http://dx.doi.org/10.1080/09298210802479284>
- Lartillot O., Toiviainen P., Eerola T. (2008) A Matlab Toolbox for Music Information Retrieval. In: Preisach C., Burkhardt H., Schmidt-Thieme L., Decker R. (eds) *Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Berlin, Heidelberg. Doi: 10.1007/978-3-540-78246-9_31

- Lartillot, O., & Toiviainen, P. (2007). MIR in Matlab (II): A toolbox for musical feature extraction from audio. In *Proceedings of the International Conference on Music Information Retrieval*. 127–130. DOI: <http://users.jyu.fi/~lartillo/mirtoolbox/>.
- Laurier, C., Lartillot, O., Eerola, T. y Toiviainen, P. (2009). Exploring relationships between audio features and emotion in music. 7th Triennial conference of European society Science o Music (ESCOM 2009).
- Lee, K. C., Chao, Y. H., Yiin, J. J., Chiang, P. Y., & Chao, Y. F. (2011). Effectiveness of different music-playing devices for reducing preoperative anxiety: a clinical control study. *International journal of nursing studies*, 48(10), 1180-1187. <https://doi.org/10.1016/j.ijnurstu.2011.04.001>
- Lee, W. C., & Kuo, C. C. J. (2006, July). Musical onset detection based on adaptive linear prediction. In *Multimedia and Expo, 2006 IEEE International Conference on* (pp. 957-960). IEEE.
- Lee, W. P., Wu, P. Y., Lee, M. Y., Ho, L. H., & Shih, W. M. (2017). Music Listening Alleviates Anxiety and Physiological Responses in Patients Receiving Spinal Anesthesia. *Complementary Therapies in Medicine*. 31, 8-13. <https://doi.org/10.1016/j.ctim.2016.12.006>
- Lerdahl, F., & Jackendoff, R. (2003). *Teoría generativa de la música tonal*. Madrid, España: Akal.
- Levitin, D.J. (2011). *Tu cerebro y la Música: El estudio científico de una obsesión humana*. Barcelona, España: RBA Libros.
- Lin, Y. P., Wang, C. H., Wu, T. L., Jeng, S. K., & Chen, J. H. (2009, April). EEG-based emotion recognition in music listening: A comparison of schemes for multiclass support vector machine. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on* (pp. 489-492). IEEE. Doi: [10.1109/ICASSP.2009.4959627](https://doi.org/10.1109/ICASSP.2009.4959627)
- Lu, L., Liu, D., & Zhang, H. J. (2006). Automatic mood detection and tracking of music audio signals. *IEEE Transactions on audio, speech, and language processing*, 14(1), 5-18. Doi: [10.1109/TSA.2005.860344](https://doi.org/10.1109/TSA.2005.860344)
- Lu, L., Liu, D., & Zhang, H. 2006. Automatic mood detection and tracking of music audio signals. *IEEE Trans. Audio, Speech Lang. Process.* 14, 1, 5–18.
- Lundqvist, L.O., Andersson, G. y Viding, J. (2009). Effects of vibroacoustic music on challenging behaviors with autism and developmental disabilities. *Research in Autism Spectrum Disorders*. 3 (2), 390-400.
- Marchi, E., Ferroni, G., Eyben, F., Gabrielli, L., Squartini, S., & Schuller, B. (2014, May). Multi-resolution linear prediction based features for audio onset detection with bidirectional LSTM

- neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on* (pp. 2164-2168). IEEE.
- Mathur, A., Vijayakumar, S. H., Chakrabarti, B., & Singh, N. C. (2015). Emotional responses to Hindustani *raga* music: The role of musical structure. *Frontiers in Psychology*, *6*, 513. doi: 10.3389/fpsyg.2015.00513
- Mejía-Rubalcava, C., Alanís-Tavira, J., Mendieta-Zerón, H., & Sánchez-Pérez, L. (2015). Changes induced by music therapy to physiologic parameters in patients with dental anxiety. *Complementary therapies in clinical practice*, *21*(4), 282-286. <https://doi.org/10.1016/j.ctcp.2015.10.005>
- Mejía-Rubalcava, C., Alanís-Tavira, J., Mendieta-Zerón, H., & Sánchez-Pérez, L. (2015). Changes induced by music therapy to physiologic parameters in patients with dental anxiety. *Complementary therapies in clinical practice*, *21*(4), 282-286. <https://doi.org/10.1016/j.ctcp.2015.10.005>.
- Miller, K., & Bernatzky, G. (2005). Purpose of music and relaxation for health promotion after bariatric surgical procedure—prospective, randomized study. *Surgery for Obesity and Related Diseases*, *1*(3), 286. <http://dx.doi.org/10.1016/j.soard.2005.03.183>
- Mok, E., & Wong, K. Y. (2003). Effects of music on patient anxiety. *AORN journal*, *77*(2), 396-410. [https://doi.org/10.1016/S0001-2092\(06\)61207-6](https://doi.org/10.1016/S0001-2092(06)61207-6)
- Nagel, F., Kopiez, R., Grewe, O., & Altenmüller, E. (2007). EMuJoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods*, *39*(2), 283-290. doi: 0.3758/BF03193159
- Najafi, M., Motamed, F., Kiani, M. A., Sabbagh, M., Attaei, P., Ghayour Mobarhan, M., ... & Javan, A. (2014). The Effect of Music on Pain, Anxiety and Vital Signs of Children during Colonoscopy. *International Journal of Pediatrics*, *2*(2.1), 26-26. Doi: [10.22038/IJP.2014.2479](https://doi.org/10.22038/IJP.2014.2479)
- Nakahara, H., Furuya, S., Masuko, T., Francis, P.R. y Kinoshita, H. (2011). Performing music can induce greater modulation of emotion-related psychophysiological responses than listening to music. *International Journal of Psychophysiology*. *81* (3), 152-158.
- Nakahara, H., Furuya, S., Obata, S., Masuko, T. y Kinoshita, H. (2009). Emotion-related changes in heart rate and its variability during performance and perception of music. *Annals of New York Academic of Science*. *1169*, 359-362.
- Nantais, K.M. y Schellenberg, E.G. (1999). The Mozart effect: An artifact of preference. *Psychological science*. *10*, 370-373.

- Naranjo, C., Kornreich, C., Campanella, S., Noël, X., Vandriette, Y., Gillain, B., Longuerville, X., Delatte, B., Verbanck, P. y Canstant, E. (2011). Major depression is associated with impaired processing of emotion in music as well as in facial and vocal stimuli. *Journal of Affective Disorders*. 128, 243-251.
- Nater, U.M., Abbruzzese, E., Krebs, M. y Ehlert, U. (2006). Sex differences in emotional and psychophysiological responses to musical stimuli. *International Journal of Psychophysiology*. 62 (29), 300-308. <https://doi.org/10.1016/j.ijpsycho.2006.05.011>
- Nieto-Romero, R. M. (2017). Efectos de la musicoterapia sobre el nivel de ansiedad del adulto cardiópata sometido a resonancia magnética. *Enfermería Universitaria*, 14(2), 88-96. <https://doi.org/10.1016/j.reu.2017.02.002>
- Nilsson, S., Kokinsky, E., Nilsson, U., Sidenvall, B., & Enskär, K. (2009). School-aged children's experiences of postoperative music medicine on pain, distress, and anxiety. *Pediatric Anesthesia*, 19(12), 1184-1190. Doi: 10.1111/j.1460-9592.2009.03180.x
- Norris, C.J., Gollan, J., Bernston, G.G. y Cacioppo, J.T. (2010). The current status of research on the structure of evaluative space. *Biological Psychology*. 84, 422-436. The current status of research on the structure of evaluative space
- Norris, C.J., Larsen, J.T., Crawford, E. y Cacioppo, J.T. (2011). Better (or worse) for some than others: Individual differences in the positivity offset and negativity bias. *Journal of Research in Personality*. 45, 100-111.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods*, 32(3), 396-402. DOI: 10.3758/BF03200807
- Ohayon, M. M., & Sagales, T. (2010). Prevalence of insomnia and sleep characteristics in the general population of Spain. *Sleep Medicine*, 11(10), 1010-1018. <https://doi.org/10.1016/j.sleep.2010.02.018>
- Osgood, C.E., Suci, G. y Tannenbaum, P.H. (1957). The measurement of meaning. Urbana: University of Illinois Press.
- Pearce, M., & Rohrmeier, M. (2012). Music cognition and the cognitive sciences. *Topics in Cognitive Science*, 4, 468-484. doi: 10.1111/j.1756-8765.2012.01226.x
- Pellino, T. A., Gordon, D. B., Engelke, Z. K., Busse, K. L., Collins, M. A., Silver, C. E., & Norcross, N. J. (2005). Use of nonpharmacologic interventions for pain and anxiety after total hip and total knee arthroplasty. *Orthopaedic Nursing*, 24(3), 182-190. Doi: [10.1097/00006416-200505000-00005](https://doi.org/10.1097/00006416-200505000-00005)

- Peretz, I. (2001). Brain specialization for music. *Annals of the New York Academy of Sciences*, 930(1), 153-165. DOI: 10.1111/j.1749-6632.2001.tb05731.x
- Peretz, I., & Gagnon, L. (1999). Dissociation between recognition and emotional judgements for melodies. *Neurocase*, 5(1), 21-30. <http://dx.doi.org/10.1080/13554799908404061>
- Peretz, I., & Hébert, S. (1995). Music processing after brain damage: The case of rhythm without melody. In *Music and the mind machine* (pp. 127-137). Springer Berlin Heidelberg. Doi: 10.1007/978-3-642-79327-1_13
- Peretz, I., Champod, S. & Hyde, K. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, 999, 58-75. doi: 10.1196/annals.1284.006
- Peretz, I., Gagnon, L. & Bouchard, B. (1998) Music and emotion: perceptual determinants, immediacy and isolation after brain damage. *Cognition* , vol. 68, pp. 111-141.
- Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, 6, 688-691. DOI:10.1038/nn1083
- Pérez, M.L. y Martín, Q. (2003). Algunos modelos de redes neuronales artificiales. En: Aplicaciones de las redes neuronales artificiales a la estadística. ED: La muralla.
- Pinilla, F., Saiz, E. & Serrano, M.A. (Marzo, 2007). Relación entre la búsqueda de sensaciones y la apertura mental con las preferencias musicales y cinematográficas: Un estudio piloto. Comunicación presentada en el IV Congreso Nacional de Estudiantes de Psicología.
- Punkaenen, M., Eerola, T. y Erkkilä, J. (2011). Biased emotional recognition in depression: Perception of emotions in music by depressed patients. *Journal of Affective Disorders*. 130, 118-126.
- Quim, S., & Watt, R. (2006). The perception of tempo in music. *Perception*, 35, 267-280. doi: 10.1068/p5353
- Rauscher, F.H., Shaw, G.L. y Ky, K.N. (1993). Music and spatial task performance. *Nature*. 365, 611.
- Rauscher, F.H., Shaw, G.L. y Ky, K.N. (1995). Listening Mozart enhances spatial-temporal reasoning: Towards a neurophysiological basis. *Neuroscience Letters*. 185, 44-47.
- Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology*, 84(6), 1236.
- Rickard, N.,S., Wong, W.W. y Velik, L. (2012). Relaxing music counters heightened consolidation of emotional memory. *Neurobiology of Learning and memory*. 97, 220-228.
- Ridgeway, D. & Hare, R.D. (1981). Sensation seeking and Psychophysiological responses to auditory stimulation. *Psychophysiology*. 18, 613-618. doi: 10.1111/j.1469-8986.1981.tb01833.x

- Rosenblatt, F. (1958). The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65 (6), 386-408. <http://dx.doi.org/10.1037/h0042519>
- Roy, M., Mailhot, J. P., Gosselin, N., Paquette, S., & Peretz, I. (2009). Modulation of the startle reflex by pleasant and unpleasant music. *International Journal of Psychophysiology*, 71(1), 37-42. <https://doi.org/10.1016/j.ijpsycho.2008.07.010>
- Roy, M., Peretz, I., & Rainville, P. (2008). Emotional valence contributes to music-induced analgesia. *Pain*, 134(1), 140-147. doi: 10.1016/j.pain.2007.04.003
- Russell, J.A. (1978). Evidence of convergent validity on the dimensions of affect. *Journal of Personality and Social Psychology*, 36(10), 1152-1168. DOI:10.1037/0022-3514.36.10.1152
- Russell, J.A. (1979). Affective space is bipolar. *Journal of Personality and Social Psychology*, 37(3), 345-356. doi: 10.1037/0022-3514.37.3.345
- Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178. doi: 10.1037/h0077714
- Saiz-Clar, E, Serrano, M.A., Reales, J.M. (2017). Prediction of emotional response of music using principal component factors extracted from the onset curves. *Submitted*.
- Saiz-Clar, E. & Reales, J.M. *Predicting emotional response to music through a compound neural network*. Psychology of Music. Prepublished May, 18, 2017. DOI: 10.1177/0305735617705452.
- Saiz-Clar, E., Tortosa-Abad, D., Serrano, M.A. y Reales, J.M. (2017b). Predicción de la respuesta emocional continua del oyente a la escucha musical: Efecto de la estructura del estímulo y de la respuesta fisiológica. *En preparación*.
- Sammler, D., Grigutsch, M, Fritz, T. y Koelsch, S. (2007). Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology*. 44 (2), 293-304.
- Samson, S., Zatorre, R. J., & Ramsay, J. O. (1997). Multidimensional scaling of synthetic musical timbre: Perception of spectral and temporal characteristics. *Canadian Journal of Experimental Psychology* 51(4), 307. DOI: [10.1037/1196-1961.51.4.307](http://dx.doi.org/10.1037/1196-1961.51.4.307)
- Scheel, K. R., & Westfeld, J. S. (1999). Heavy metal music and adolescent suicidality: an empirical investigation. *Adolescence*, 34(134), 253.
- Scheirer, E. D. (1998). Tempo and beat analysis of acoustic musical signals. *The Journal of the Acoustical Society of America*, 103(1), 588-601. DOI: <http://dx.doi.org/10.1121/1.421129>

- Schellenberg, E. G., Peretz, I., & Vieillard, S. (2008). Liking for happy-and sad-sounding music: Effects of exposure. *Cognition & Emotion*, 22(2), 218-237. <http://dx.doi.org/10.1080/02699930701350753>
- Scherer, K.R. & Zenter, M.R. (2001). Emotional effects of music production rules. In P.N. Juslin & J.A. Sloboda (ed.), *Music and emotions: Theory and research* (pp. 71-104). Oxford: Oxford University Press.
- Scherer, K.R. (1972). Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences. In S. Weitz (ed.), *Nonverbal communication* (pp. 105-111). New York, NY: Oxford University Press.
- Scheufele, P.M. (2000). Effects of progressive relaxation and classical music on measurements of attention, relaxation and stress responses. *Journal of Behavioral Medicine*, 23 (2), 207-228
- Schubert, E. (1996). Enjoyment of negative emotions in music: An associative network explanation. *Psychology of music*, 24(1), 18-28.
- Schubert, E. (1999). *Measurement and time series analysis of emotion in music* (unpublished thesis). Faculty of Arts & Social Sciences. University of New South Wales. Australia.
- Schubert, E. (2013). Emotion felt by the listener and expressed by the music: Literature review and theoretical perspectives. *Frontiers in psychology*, 4, 4-18. doi: 10.3389/fpsyg.2013.00837
- Schulze, H.H. (1978). The detectability of local and global displacements in regular rhythmic patterns. *Psychological Research*, 40(2), 173-181. doi: 10.1007/BF00308412
- Schulze, H.H. (1989). Categorical perception of rhythmic patterns. *Psychological Research*, 51(1), 10-15. doi: 10.1007/BF00309270
- Sethares, W.A. (2007). *Rhythm and transforms*. London, UK: Springer-Verlag.
- Shao, B., Wang, D., Li, T., & Ogihara, M. (2009). Music recommendation based on acoustic features and user access patterns. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(8), 1602-1611. doi:10.1109/TASL.2009.2020893
- Sloboda, J. A., & Juslin, P. N. (2001). Psychological perspectives on music and emotion. In P.N. Juslin & J.A. Sloboda (ed.), *Music and emotions: Theory and research* (pp. 361-392). Oxford: Oxford University Press.
- Sloboda, J.A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music*, 19(2), 110-120. doi: 10.1177/0305735691192002
- Smith, P. C. y Curnow, R. (1966). "Arousal hypothesis" and the effects of music on purchasing behavior. *Journal of Applied Psychology*, Vol 50(3), Jun 1966, 255-256.

- Smolen, D., Topp, R., & Singer, L. (2002). The effect of self-selected music during colonoscopy on anxiety, heart rate, and blood pressure. *Applied Nursing Research*, 15(3), 126-136. <https://doi.org/10.1053/apnr.2002.34140>
- Steele, K. M., Bass, K. E., y Crook, M. D. (1999). The mystery of the Mozart effect: Failure to replicate. *Psychological Science*, 10 (4), 366-369.
- Stelmack, R.M., Plouffe, L. & Falkenberg, W. (1983). Extraversion, sensation seeking and electrodermal response: Probing a paradox. *Personality and Individual Differences*, 4, 607-614. [https://doi.org/10.1016/0191-8869\(83\)90115-0](https://doi.org/10.1016/0191-8869(83)90115-0)
- Sturn, I., Treder, M., Miklody, D., Purvins, H., Dähne, S., Blankertz, B. & Curio, G. (2015). Extracting the neural representation of tone onsets for separate voices of ensemble music using multivariate EEG analysis. *Psychomusicology: Music, Mind, and Brain*, 25 (4), 366-379. <http://dx.doi.org/10.1037/pmu0000104>
- Szpunar, K.K., Schellenberg, E.G. y Pliner, P. (2004). Linking and memory for musical stimuli as a function of exposure. *Journal of Experimental Psychology. Learning, Memory and Cognition*. 30(2). 370-381. <http://dx.doi.org/10.1037/0278-7393.30.2.370>
- Tabachnick, B. G., and L. S. Fidell. (1989). *Using Multivariate Statistics* (2nd ed.). New York: Harper and Row.
- Thompson, W., Schellenberg, E. y Husain, G. (2001). Arousal, mood and the Mozart effect. *Psychological Science*. 12 (3), 228-251.
- Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5), 293-302. doi:10.1109/TSA.2002.800560
- VanderArk, S. D., & Ely, D. (1992). Biochemical and galvanic skin responses to music stimuli by college students in biology and music. *Perceptual and Motor Skills*, 74, 1079–1090.
- Vieillard, S., Peretz, I., Gosselin, N., Khalifa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition and Emotion*, 22(4), 720-752. doi: 10.1080/02699930701503567
- Vos, P.G., van Assen, M., & Franek, M. (1997). Perceived tempo change is dependent on base tempo and direction of change: Evidence for a generalized version of Schulze's (1978) internal beat model. *Psychological Research*, 59(4), 240-247. doi: 10.1007/BF00439301
- Voss, J. A., Good, M., Yates, B., Baun, M. M., Thompson, A., & Hertzog, M. (2004). Sedative music reduces anxiety and pain during chair rest after open-heart surgery. *Pain*, 112(1), 197-203. <https://doi.org/10.1016/j.pain.2004.08.020>

- Vuilleumier, P. (2002). Facial expression and selective attention. *Current Opinion in Psychiatry*, 15, 291–300.
- Wagner, H. (1989). The peripheral physiological differentiation of emotions. **Wagner, H.** (Ed); Manstead, Antony (Ed). (1989). Handbook of social psychophysiology., (pp. 77-98). Oxford, England: John Wiley & Sons
- Wang, S. M., Kulkarni, L., Dolev, J., & Kain, Z. N. (2002). Music and preoperative anxiety: a randomized, controlled study. *Anesthesia & Analgesia*, 94(6), 1489-1494. Doi: 10.1213/00000539-200206000-00021
- Webster, G.D., & Weir, C.G. (2005). Emotional responses to music: Interactive effects of mode texture and tempo. *Motivation and Emotion*, 29(1), 19-39. doi:10.1007/s11031-005-4414-0
- Wedin, L. (1972). A multidimensional study of perceptual-emotional qualities in music. *Scandinavian Journal of Psychology*. 13 (4), 241-257.
- Widrow, B. y Hoff, M.E. (1960) Adaptive switching circuits. IRE WESCON Convection Record. 4, 96-104.
- Witvliet, C. V. O., & Vrana, S. R. (2007). Play it again, Sam: Repeated exposure to emotionally evocative music polarizes liking and smiling responses, and influences other affective reports, facial EMG and heart rate. *Cognition and Emotion*, 21, 3–25.
- Wong P.C.M., Ciocca V., Chan A.H.D., Ha, L.Y.Y., Tan, L-H, et al. (2012). Effects of Culture on Musical Pitch Perception. *PLoS ONE*, 7(4), 334-324. doi:10.1371/ journal.pone.0033424
- Wrigley, C. & Neuhaus, J. O. The Matching of Two Sets of Factors. *American Psychologist*, 1955, 10, 418-419.
- Wundt, W.M. (1874). *Grundzüge der physiologischen Psychologie*. Leipzig. Engelmann. <http://www.jstor.org/stable/25109930>
- Yang, Y. H., & Chen, H. H. (2011). Ranking-based emotion recognition for music organization and retrieval. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(4), 762-774. [10.1109/TASL.2010.2064164](https://doi.org/10.1109/TASL.2010.2064164)
- Yang, Y. H., & Chen, H. H. (2012). Machine recognition of music emotion: A review. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(3), 40. doi: 10.1145/2168752.2168754
- Yang, Y. H., Lin, Y. C., Cheng, H. T., & Chen, H. H. (2008, October). Mr. Emo: Music retrieval in the emotion plane. In *Proceedings of the 16th ACM international conference on Multimedia* (pp. 1003-1004). ACM.

- Yang, Y. H., Liu, C. C., & Chen, H. H. (2006, October). Music emotion classification: A fuzzy approach. In *Proceedings of the 14th ACM international conference on Multimedia* (pp. 81-84). ACM.
- Yang, Y.H., & Chen, H. H. (2012). Machine recognition of music emotion: A review. *ACM Trans. Intell. Syst. Technol.*, 3, 3, Article 40 (May 2012). DOI: 10.1145/2168752.2168754.
- Yang, Y.H., Lin, Y.C., Su, Y.F., & Chen, H.H. (2008). A Regression Approach to music emotion recognition. *IEEE Transactions on audio speech and language processing*, 16(2), 448-457. doi: 10.1109/TASL.2007.911513
- Zacharopoulou, K. y Kyriakidou, A.D. (2009). A cross-cultural comparative study of the role of musical structural features in the perception of emotion in Greek traditional music. *Journal of Interdisciplinary Music Studies*. 3, 1-15.
- Zenter, M., Grandjean, D., & Scherer, K.R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8(4), 494-521. doi: 10.1037/1528-3542.8.4.494
- Zhao, H., & Chen, A. C. (2009). Both happy and sad melodies modulate tonic human heat pain. *The Journal of Pain*, 10(9), 953-960. <https://doi.org/10.1016/j.jpain.2009.03.006>
- Ziv, N., Rotem, T., Arnon, Z., & Haimov, I. (2008). The effect of music relaxation versus progressive muscular relaxation on insomnia in older people and their relationship to personality traits. *Journal of music therapy*, 45(3), 360-380. <https://doi.org/10.1093/jmt/45.3.360>
- Zuckerman, M., Simons, R. F., & Como, P. G. (1988). Sensation seeking and stimulus intensity as modulators of cortical, cardiovascular, and electrodermal response: A cross-modality study. *Personality and Individual Differences*, 9(2), 361-372. [https://doi.org/10.1016/0191-8869\(88\)90099-2](https://doi.org/10.1016/0191-8869(88)90099-2)

Material Suplementario

Capítulo 2

ANN Training phase

The standardised values of each one of the six properties obtained from the scores provided in Vieillard et al.'s (2008) study were used as inputs for the Adaline networks. The purpose of the standardisation was so that all the variables were in the same scale in order for the network not to overestimate the connection strength of some properties in detriment of the rest due to a scale effect. Afterwards, the 40 stimuli of the battery were divided into two parts. Each part had a total of 20 stimuli, five stimuli for each of the four basic emotions. The aim was to create two balanced sets of stimuli, one used for the ANN training phase and the other group for the generalization phase. In each of these phases, the order of presentation of stimuli was also randomized in order to avoid the networks learning systematic presentation patterns.

The construction and training of the networks have been performed using Matlab v. 7.8, with the identity function as the transfer function of all units. The weights of both Adalines were initialized by assigning real random values prior to training with a range of -3 to 3. During the Adaline training, the Widrow-Hoff Delta learning rule was used for the modification of the weights. The error was calculated by comparing the outputs of the Adaline networks with the valence and arousal values provided by Vieillard et al. (2008). The network's learning stopped adequately when the error reached a lower value than .001 or when the weights were unchanged after three consecutive iterations. The value of the learning rate (η) was maintained at an initial value of .5 with a reduction rate of .001 per epoch.

Once the two Adaline networks had reached the learning criterion, their weights were fixed and the SOM network training was carried out using the outputs obtained from the Adaline as inputs. The SOM was programmed to have four units in the competitive layer and that these were set as a square in space with the "*gridtop*" topological feature of Matlab. The "*dist*" function was chosen as the distance function to calculate the distance between neurons. The learning rule used by SOM networks is known as the Kohonen Learning Rule (function "*learnk*" in Matlab). Given proper recognition, its goal is that weights of the winning neuron are maximized and the rest are minimized. Both Adaline and SOM were subjected to 20,000 epochs during the training phase.

During the generalization phases for both the Adaline and the SOM, there is no learning as the task is to evaluate the ability of the network to estimate activation and valence of and classify each stimulus in its emotional category. In both cases, a simulation was performed during the generalization phase with the set of stimuli specifically prepared for this phase and with the estimated weights from the learning phase. The result of this simulation was subjected to the same statistical analysis as the results obtained in the training phase in order to test the predictive and generalizing power of the model.

Stimulus Variables

The **rhythm** of a musical piece is a complex variable, thus the psychological literature often uses for its measurement properties like regularity or frequency of the signal in the time domain (Sethares, 2007). When a pattern of sounds and silences is recurring within a musical piece, it can be classified as regular and this makes the rhythm predictable to the listener. Therefore, the value of the autocorrelation function of the acoustic signal in the time domain is often used as a measure of rhythmic regularity (e.g. Eerola, 2012). Maintaining the

simplicity criterion, we decided to measure the rhythm regularity by computing the Average Number of Notes per Bar (ANNB) measuring it from the musical scores.

The **tempo** is used to indicate to performers the speed at which a musical piece must be performed. We measured it as the number of musical notes that occur in each time unit (beats per minute or b.p.m.). We have introduced the value of tempo that was referenced in the musical score. Finally, we also used the **time signature** as a metric measure of the number of beats per bar (Levitin, 2008). In Western music, the most frequent time signatures are 3/4 (triple meter) and 4/4 (quadruple meter), although they are not the only ones used. We included time signature as a metric variable within our model because of its relevance in music theory and its effect on cognitive judgments.

The overall tonality of a piece of music is defined by multiple variables within which we have selected three of them. The **mode** indicates the distribution of tones and semitones within a scale. It generates two exclusive categories (major mode and minor mode) which allow one to classify the musical piece in an easy and clear way. The **Absolute Tonal Range (ATR)**, another variable used to measure the range of the pieces, is defined as the difference between the frequency of the highest note and the frequency of the lowest note and shows the frequency range of the staff. The range of variation affects the hedonic judgment thereof so the pieces with a smaller range of frequencies are associated with positive emotions, such as joy, while pieces with a greater range of frequencies are associated with negative emotions, such as fear/tension (Scherer, 1972). Finally, we have included the **Frequency of the Lowest Note (FLN)** as another measure of range because Eerola et al.'s work (2009) showed that the percentage of low frequency energy in the acoustic signal is a relevant variable in the study of musical properties such as articulation and predict mood states of the listener (Feng, Zhuang, & Pan, 2003).

Tabla 1. Descripción de las piezas musicales empleadas en el segundo experimento del capítulo 2.

Title of the track	Composer	Fragment	Duration	Tempo	Time Signatura	Mode	Activation	Emotional Valence	Emotion	Index Agreement
Grande Valse Brillante, Opus 18	Chopin	1 to 12	10	210	1	M	61.13	68.69	J	81
Grande Valse Brillante, Opus 18	Chopin	77 to 84	6	210	1	M	67.38	61.63	J	81
Piano Sonata D960	Schubert	79 to 74	11	99	2	M	55.56	56.94	J	75
Polonaise in A-flat Opus 53	Chopin	17 to 23	12	100	1	M	78.25	70.88	J	75
Piano Sonata N°16 K545 (first movement)	Mozart	22 to 29	11	137	2	M	77.56	69.44	J	75
Nocturne N°2, Opus 27 D flat	Chopin	2 to 4	13	42	1	M	25.75	63	C	100
Nocturne N°2, Opus 27 D flat	Chopin	65 to 67	11	50	1	M	33.19	55.5	C	75
Sonata Opus 35, N°1	Chopin	42 to 51	12	96	2	m	30.75	54.75	C	75
Preludes, Opus 28, N°13	Chopin	21 to 22	14	48	1	M	25.38	61.56	C	75
Piano Sonata N°16 K545 (second movement)	Mozart	1 to 5	13	55	1	M	49.38	61.38	C	56
Etude N° 11 "Winter wind", Opus 25	Chopin	13 to 17	9	130	2	m	74.38	33.81	F	88
Sonata 27 E minor Opus 90	Beethoven	189 to 198	11	156	1	m	67.13	41.94	F	75
Polonaise in A-flat Opus 53	Chopin	1 to 7	11	110	1	M	69.06	40.63	F	75
N° 8-8 Etudes tableaux, Opus 33	Rachmaninov	9 to 12	12	64	2	m	62	36.69	F	75
Sonata Opus 35 Funeral March N°4	Chopin	1 to 7	8	202	2	m	73.63	34.25	F	75
N°3 Funeral March Songs without words	Mendelssohn	7 to 9	12	67	2	m	47.88	31.19	S	69

book 5, Opus 62

Sonata Opus 35 Funeral March N°3	Chopin	6 to 8	12	55	2	m	43.13	36.69	S	75
Preludes, Opus 28, N°2	Chopin	3 to 5	14	50	2	m	28.25	37.88	S	69
Preludes, Opus 28, N°2	Chopin	21 to 22	13	46	2	m	25.69	41.63	S	63
Preludes, Opus 28, N°4	Chopin	1 to 4	12	54	2	m	37.19	52.38	S	69

Note:

“Fragment” is first and last bar considered for each track. Emotion: J = Joy, C = Calm, F = Fear; S = Sadness. Mode: M = Mayor mode; m = Minor mode; Time Signature: 1 = triple meter; 2 = quadruple meter.

Capítulo 3

Annexe 1: MIRToolbox script used to extract CO properties of musical stimuli

```
Onsets = mironsets(audio);  
[c d] = mirgetdata(mironsets(audio));  
[e] = mirgetdata(mirattacktime(onsets));  
[f] = mirgetdata(mirattackleap(onsets));  
[g] = mirgetdata(mirattackslope(onsets));  
[ASR] = mirgetdata(mirlowenergy(audio.'ASR'));
```

Annexe 2: SPSS 22 syntax to compute CO properties showed in Figure 1

```
COMPUTE valle_x = onsets_x - onsets_tiempo_ataque.
VARIABLE LABELS valle_x 'Posición temporal del valle'.
FORMATS valle_x (F8.6).
VARIABLE LEVEL valle_x (SCALE).
EXECUTE.

COMPUTE valle_y = onsets_y - leap.
VALUE LABELS valle_y 'Magnitud del valle'.
FORMATS valle_y (F8.6).
VARIABLE LEVEL valle_y (SCALE).
EXECUTE.

SHIFT VALUES VARIABLE = onsets_x RESULT = retardo LAG = 1.
EXECUTE.

COMPUTE distancia_onsets = onsets_x - retardo.
VARIABLE LABELS distancia_onsets 'Distancia entre onsets'.
FORMATS distancia_onsets (F8.6).
VARIABLE LEVEL distancia_onsets (SCALE).
EXECUTE.

COMPUTE Magnitud_Ataque = onsets_tiempo_ataque * onsets_pendiente.
VARIABLE LABELS Magnitud_Ataque 'Magnitud del Ataque'.
FORMATS Magnitud_Ataque (F8.6).
VARIABLE LEVEL Magnitud_Ataque (SCALE).
EXECUTE.

SHIFT VALUES VARIABLE = valle_y RESULT = retardo_valle LEAD = 1.
EXECUTE.

COMPUTE magnitud_declive = onsets_y - retardo_valle.
VARIABLE LABELS magnitud_declive 'Magnitud del Declive'.
FORMATS magnitud_declive (F8.6).
VARIABLE LEVEL magnitud_declive (SCALE).
EXECUTE.

SHIFT VALUES VARIABLE = valle_x RESULT = retardoX_valle LEAD = 1.
COMPUTE tiempo_declive = retardoX_valle - onsets_x.
```

EXECUTE.

VARIABLE LABELS tiempo_declive 'Duración Temporal del Declive'.

FORMATS tiempo_declive (F8.6).

VARIABLE LEVEL tiempo_declive (SCALE).

EXECUTE.

COMPUTE valle_pendiente = magnitud_declive/tiempo_declive.

VARIABLE LABELS valle_pendiente 'Pendiente del valle'.

FORMATS valle_pendiente (F8.6).

variable levelvalle_pendiente (SCALE).

EXECUTE.

COMPUTE D_declive = SQRT((TIEMPO_DECLIVE*TIEMPO_DECLIVE) +
(magnitud_declive*magnitud_declive)).

VARIABLE LABELS D_declive 'Distancia Euclídea del Declive'.

FORMATS D_declive (f8.6).

Variable Level D_declive (SCALE).

EXECUTE.

COMPUTE D = SQRT((leap*leap) + (onsets_tiempo_ataque * onsets_tiempo_ataque)).

VARIABLE LABELS D 'Distancia Euclídea del Ataque'.

FORMATS D (F8.6).

VARIABLE LEVEL D (SCALE).

EXECUTE.

Tabla 2. Descripción de las piezas musicales empleadas en el segundo experimento del capítulo 3

Title	Author	Fragment	Activation	Emotional Valence	Duration	Tempo	Time Signature	Mode	Scale
Nº5 Black Key Etudes, Opus 10	Chopin	1 to 8	76.47	42.80	8	113	2	2	G flat
<i>Grande Valse Brillante</i> , Opus 18	Chopin	1 to 12	60.60	73.60	10	210	1	2	E flat
<i>Grande Valse Brillante</i> , Opus 18	Chopin	77 to 84	67.33	71.40	6	210	1	2	E flat
<i>Grande Valse Brillante</i> , Opus 18	Chopin	220 to 227	60.80	37.40	7	210	1	2	E Sharp
Piano sonata D960	Schubert	30 to 34	66.20	65.00	12	100	2	2	B flat
Piano sonata D960	Schubert	45 to 49	55.93	29.40	12	100	2	2	B flat
Piano sonata D960	Schubert	79 to 74	53.47	58.60	11	99	2	2	B flat
Piano sonata D960	Schubert	120 to 124	57.13	22.00	12	110	2	2	B flat
Polonaise Heroica, Opus 53	Chopin	1 to 7	67.93	35.00	11	110	1	2	A flat
Polonaise Heroica, Opus 53	Chopin	17 to 23	75.20	57.20	12	100	1	2	A flat
Études Nº3, Opus 25	Chopin	1 to 7	70.87	63.80	9	145	1	2	F
Preludes, Opus28, Nº5	Chopin	1 to 9	60.00	54.80	8	105	1	2	D
Winter winds Etudes, Opus 25, Nº11	Chopin	13 to 17	69.27	19.80	9	130	2	1	A
Revolutionary Etudes, Opus 10, Nº12	Chopin	9 to 14	64.33	41.40	12	120	2	1	C
Piano Sonata 27 in E minor, Opus 90	Bethoven	189 to 198	63.13	23.80	11	156	1	1	E
Etudes-tableaux, Opus 33, Nº8-8	Rachmaninov	9 to 12	57.67	22230.80	12	64	2	1	C sharp

Etudes-tableaux, Opus 33, N°8-8	Rachmaninov	27 to 29	60.67	34.00	10	64	2	1	C sharp
Funeral march, Opus 35, N°4	Chopin	1 to 7	70.07	47.20	8	202	2	1	B flat
Funeral march, Opus 35, N°4	Chopin	25 to 24	70.80	45.20	10	117	2	1	B flat
Funeral march, Opus 35, N°3	Chopin	6 to 8	39.60	26.80	12	55	2	1	B flat
Funeral march, Opus 35, N°1	Chopin	1 to 8	44.47	15.40	12	70	2	1	B flat
Funeral march, Opus 35, N°1	Chopin	42 to 51	29.20	60.60	12	96	2	1	B flat
Nocturne N°2, Opus 27	Chopin	2 to 4	24.40	65.40	13	42	1	2	D
Nocturne N°2, Opus 27	Chopin	65 to 67	33.13	58.60	11	50	1	2	D
Nocturne, Opus 9, N°2	Chopin	1 to 2	23.80	53.40	14	45	2	2	E flat
Preludes, Opus 28, N°13	Chopin	21 to 22	24.27	63.80	14	48	1	2	F sharp
Preludes, Opus 28, N°6	Chopin	4 to 7	33.13	48.20	13	52	1	1	B
Preludes, Opus 28, N°9 -	Chopin	1 to 2	37.93	60.20	13	35	2	2	E
Opus 19, N°4	Mendelssohn	5 to 8	44.33	37.40	14	60	2	2	A
Tango España, Opus 165	Albéniz	1 to 6	39.13	65.00	10	63	2	2	D
Ballade in G minor, Opus 23	Chopin	6 to 10	31.53	48.80	14	73	1	1	G
Funeral March, Opus 62, N°3	Mendelssohn	7 to 9	47.80	33.20	12	67	2	1	E
Funeral March, Opus 62, N°3	Mendelssohn	15 to 17	35.33	36.60	13	67	2	1	E
Nocturne N°1, Opus 27	Chopin	11 to 13	34.40	47.40	11	60	2	1	C
Nocturne N°1, Opus 27	Chopin	98 to 100	33.27	43.60	12	56	2	1	C

Preludes, Opus 28, N°2	Chopin	3 to 5	26.80	26.00	14	50	2	1	A
Preludes, Opus 28, N°2	Chopin	21 to 22	21.80	35.20	13	46	2	1	A
Preludes, Opus 28, N°4	Chopin	1 to 4	35.53	54.40	12	54	2	1	E
Piano Sonata K54, N°16	Mozart	22 to 29	74.33	77.20	11	137	2	2	C
Piano Sonata K54, N°16	Mozart	1 to 5	47.07	59.20	13	55	1	2	G

Note:

“Fragment” is first and last bar considered for each track. Duration: in seconds Mode: M=Mayor Mode; m=minor mode; Time Signature: 1= triple meter; 2= quadruple meter

Capítulo 4

Gráfico 3: Grado de activación autoinformado por el oyente segundo a segundo de las piezas tranquilas

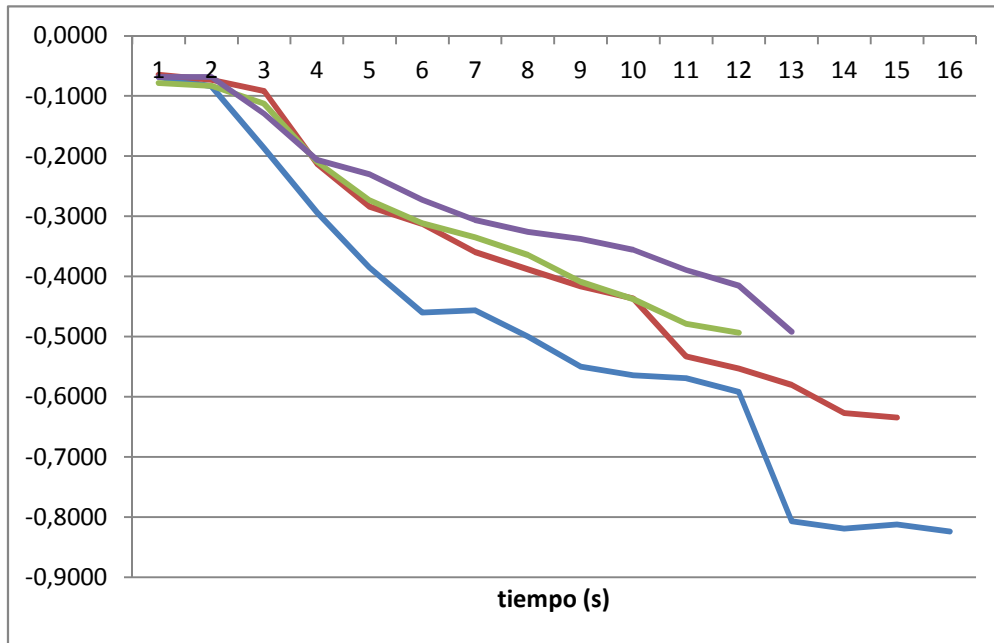


Gráfico 4: Grado de activación autoinformado por el oyente segundo a segundo de las piezas alegres

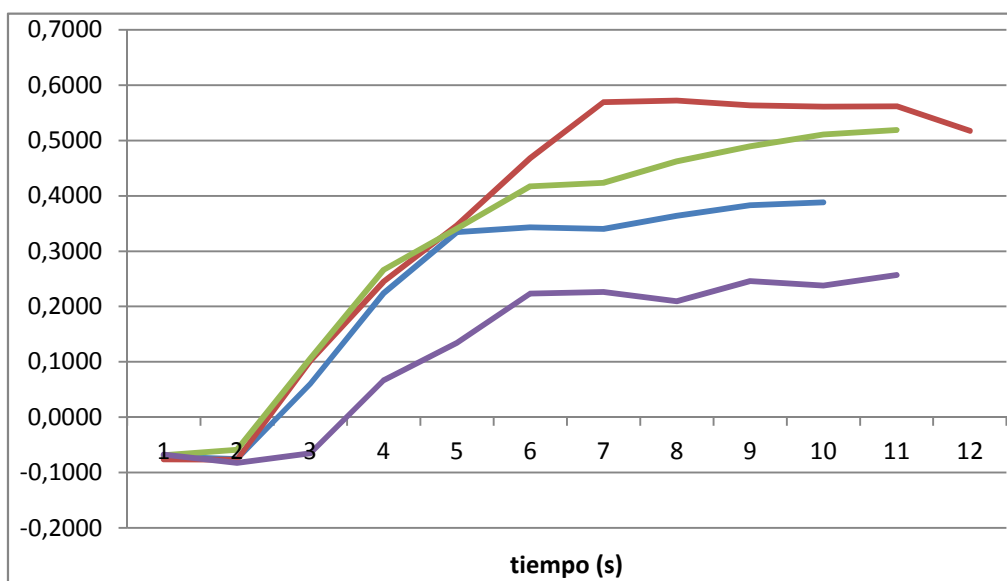


Gráfico 5: Grado de activación autoinformado por el oyente segundo a segundo de las piezas tensas

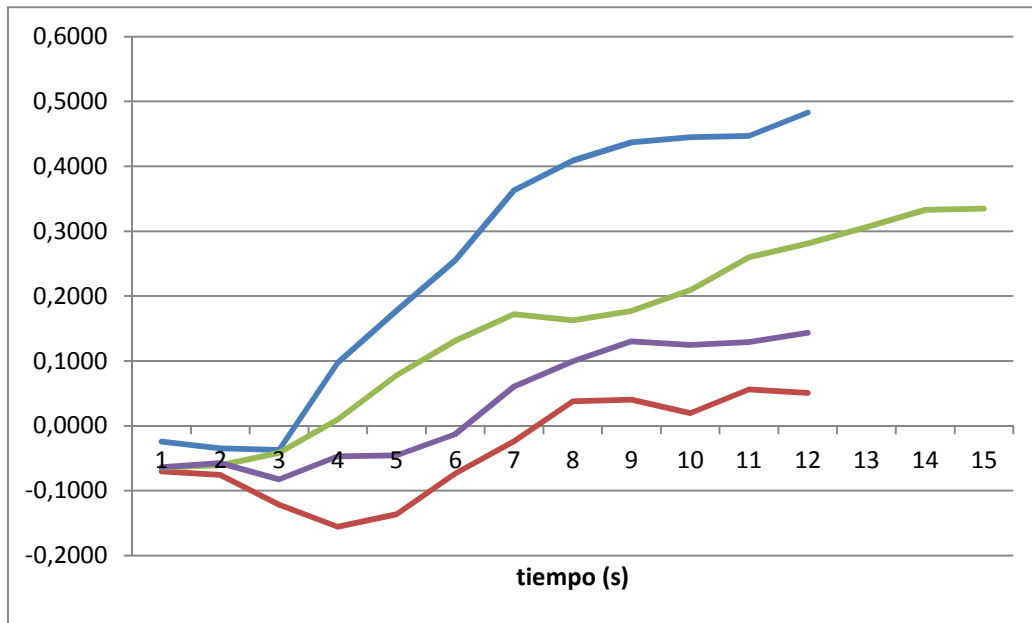


Gráfico 6: Grado de activación autoinformado por el oyente segundo a segundo de las piezas tristes

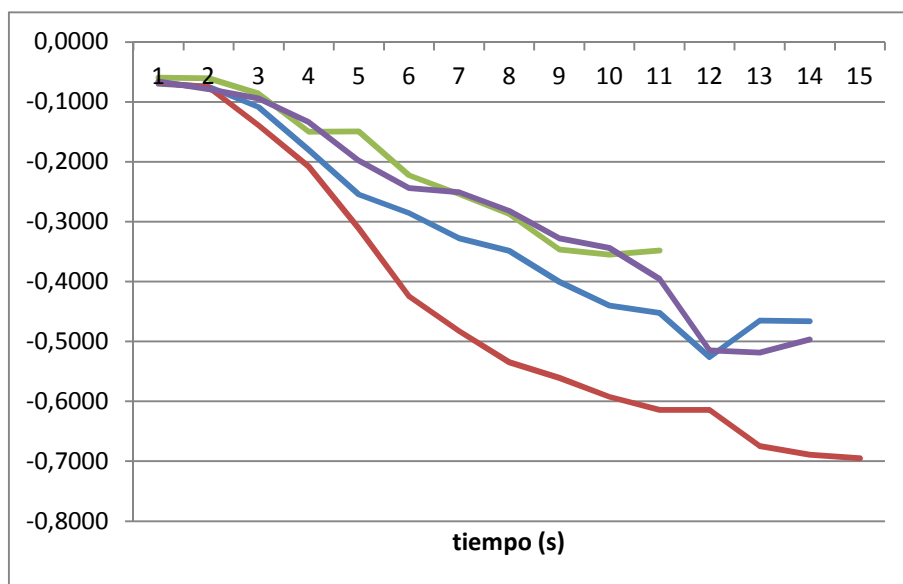


Gráfico 7: Grado de VE autoinformado por el oyente segundo a segundo de las piezas tranquilas

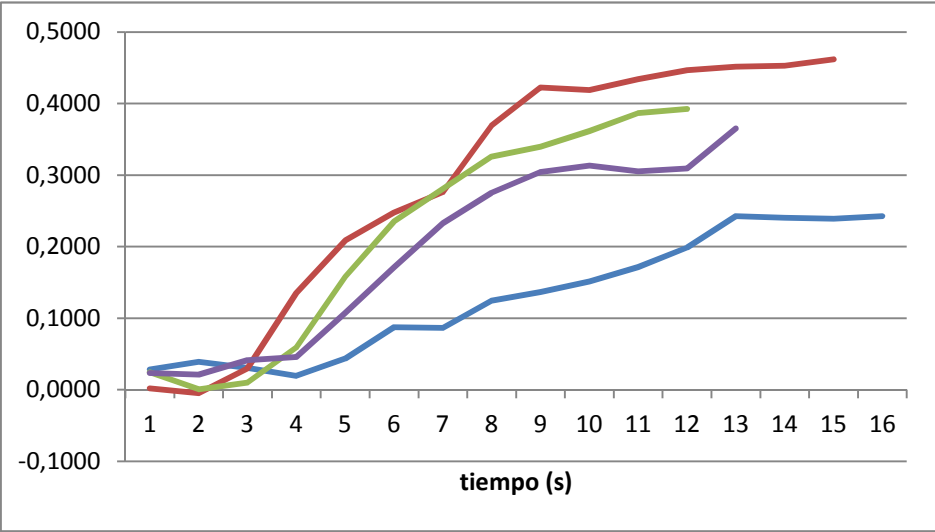


Gráfico 8: Grado de VE autoinformado por el oyente segundo a segundo de las piezas alegres

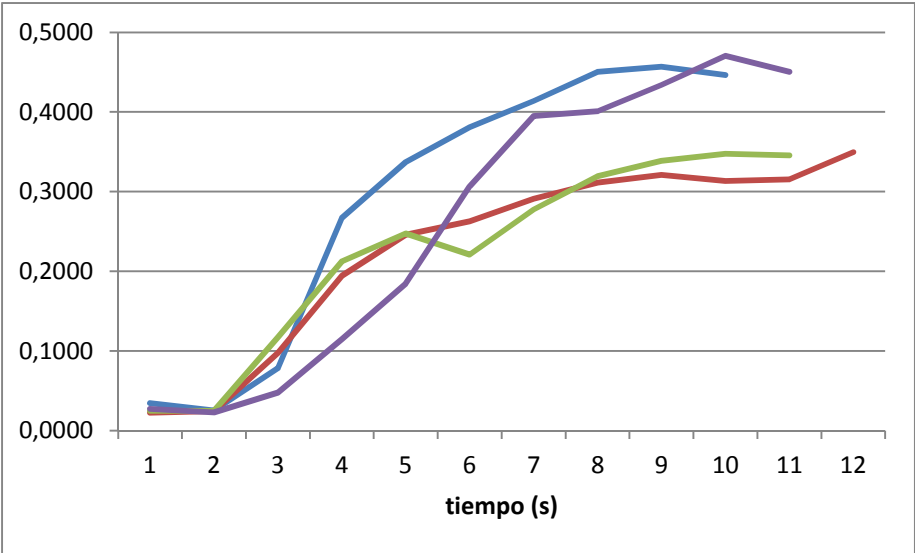


Gráfico 9: Grado de VE autoinformado por el oyente segundo a segundo de las piezas tensas

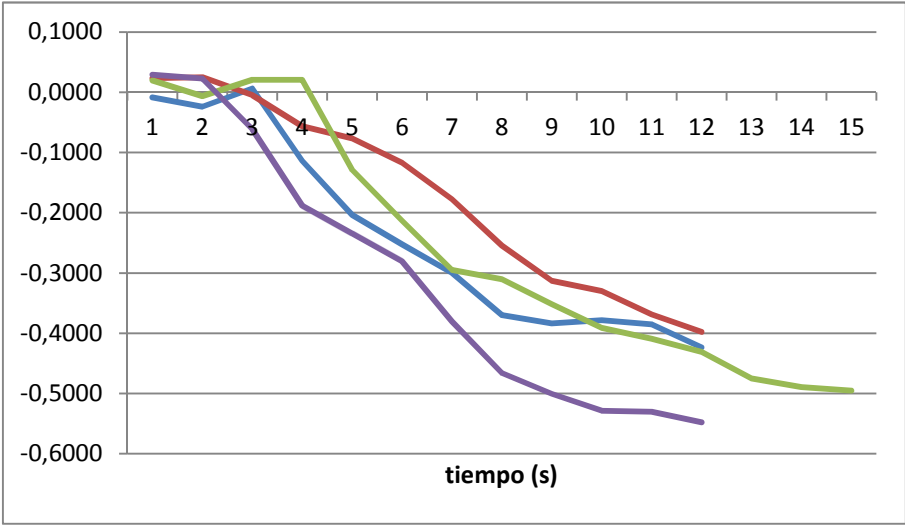


Gráfico 10: Grado de VE autoinformado por el oyente segundo a segundo de las piezas tristes

