1. Sobre el consenso entre expertos en música

El artículo de este mes versa sobre un problema que me he encontrado con cierta frecuencia en el campo de la teoría musical y de la musicología. Ese problema es el del consenso entre expertos en música. A la hora de evaluar un fenómeno musical, ¿cómo se ponen de acuerdo los expertos? ¿Son capaces de formalizar los criterios por los cuales toman su decisión? Si hay desacuerdo entre ellos, ¿cómo se formula tal desacuerdo? ¿Qué metodología usan para evaluar el fenómeno y poner en común una evaluación final? ¿Cómo se matiza tal evaluación? ¿Cuántos expertos es recomendable tener para una evaluación mínimamente fiable? Estas preguntas aparecen en el transcurso de la investigación en música. Por asombroso que parezca, en numerosas ocasiones he visto evaluaciones hechas por un único experto y que el resto de la comunidad ha dado por buena o al menos con muy pocas voces discordantes. De que ese único experto tenía un conocimiento y experiencia formidables no cabía ninguna duda; pero incluso los expertos cometen errores de juicio; pero además no es riguroso aceptar la opinión de un solo experto, por muy prestigioso que este sea. He visto también, por ejemplo, que un experto prestigioso ha basado su evaluación en pequeñísimo número de piezas musicales, a veces tres, pero en otros casos no más de una decena. También aquí parece que falta rigor. Lo observable en un número tan pequeño de piezas puede no ser generalizable al resto y si así lo es habría que justificarlo adecuadamente (normalmente tal justificación está ausente).

En el transcurso de mis investigaciones me he encontrado con ejemplos de esta situación, tanto al estudiar artículos como en los proyectos de investigación en que he participado. Por ejemplo, en el caso del flamenco no hay consenso en cuanto a cómo se tiene que transcribir, si bien creando una nueva notación, posiblemente partiendo de la notación occidental, o bien tomando la notación occidental como método único de transcripción. La notación occidental se creó para escribir una música cuyas características no coinciden totalmente con las del flamenco. Además, hay diferencias entre proponer un sistema de transcripción para la guitarra y otro para la voz. La guitarra es un instrumento de afinación fija, pero la voz y menos en el flamenco, no lo es. Donnier para la voz propone un sistema que parte del cante gregoriano [Don11
,
Don96
], pero otros autores como los hermanos Hurtado abogan rotundamente por la notación occidental para todo el flamenco; véase [
HH02
]. El guitarrista y musicólogo Rafael Hoces, en su tesis doctoral
La transcripción para guitarra flamenca
[
Hoc13

], apoya la idea del uso de la notación occidental solo para la transcripción de la guitarra. Entre los flamencólogos, cuando se presenta este debate, algunos llegan a decir es mejor seguir con la notación occidental pues no se alcanzaría acuerdo en diseñar una nueva notación que se adecuase a las peculiaridades del flamenco. De nuevo, aquí estamos en presencia del problema del consenso entre expertos.

En los últimos años se está investigando con fuerza los mecanismos que subyacen en la improvisación. Hay dos escuelas de pensamiento al respecto, una que propone que la improvisación se configura a partir de reglas, al estilo de las gramáticas generativas de Chomsky, o trasladado al ámbito músical, al estilo de la teoría generativa de la música de Fred Lerdahl y Ray Jackendoff [LJ03] (véase la serie correspondiente en esta columna [Góm14]). Cada estilo (jazz, flamenco, etc.) tiene sus reglas precisas que hacen que una improvisación se vea dentro del estilo o fuera de él. La otra escuela mantiene que la improvisación se hace a base de patrones, que pueden ser de todo tipo: melódicos, armónicos, rítmicos, formales; y que entonces la calidad de la improvisación está en función de la combinación acertada de esos patrones. Probablemente, la improvisación venga dada por una combinación de ambas. No se sabe, empero, para qué parámetros musicales y en qué grado se produce tal combinación. Investigadores de ambas escuelas de pensamiento han escrito programas que toman, por ejemplo, un corpus de solos de un trompetista de jazz (Parker, Coltrane u otros) y a partir de ese corpus, bien por reglas [GKT10] o por patrones [NSM13], componen solos en su estilo. A la hora de evaluar los resultados del programa, esto es, cuán fielmente se reflejan las características del músico en cuestión, con frecuencia nos encontramos que es la opinión de los autores del artículo el único criterio de evaluación. Los autores afirman que los solos son buenos porque “suenan al trompetista”, o porque “reflejan su pensamiento musical”, pero no aportan razones que sostengan estas afirmaciones. Y no dudo de la honestidad intelectual de estos investigadores, pero desde el punto de vista del rigor metodológico, en ciencia (y la musicología lo es) es difícil aceptar esas afirmaciones.

2. ¿Qué pueden hacer las matemáticas?

En otros campos ya ha surgido el problema de alcanzar consenso entre expertos. En medicina, por ejemplo, es un problema que aparece con frecuencia. ¿Cómo lo resuelven en medicina? Hay varios métodos, pero uno de ellos, que goza de cierta popularidad, es el llamado método Delphi. La técnica Delphi es un método para recoger información de expertos y construir consenso a partir de dicha información. Vamos a describir ese método y ver cómo se podría aplicar a la teoría de la música.

Jorm [Jor15], en un artículo titulado *Using the Delphi expert consensus method in mental health research* , investiga la aplicación del método Delphi al acuerdo entre expertos en el campo de la salud mental. El primer paso en la implementación del método es la selección de los expertos. Basándose en el trabajo de Surowiecki [Sur04], el famoso libro *The wisdom of crowds: why the many are smarter than the few* , propone las siguientes condiciones para elegirlos:

1. Diversidad de expertos. Un grupo heterogéneo de expertos previsiblemente producirá resultados de mayor calidad que un grupo fuertemente homogéneo.
2. Independencia. Los expertos han de tomar sus decisiones de modo independiente y sin influencia externa.
3. Descentralización. Los expertos trabajan de manera autónoma en la producción de sus resultados.
4. Coordinación. Para los resultados finales existe un mecanismo de coordinación entre los expertos.

Aunque no en todas las circunstancias el trabajo de un grupo de expertos da buenos resultados, se han estudiado las condiciones bajo las cuales esto ocurre. Hay una gran variedad de contextos en que dicho trabajo es útil y valioso; para más detalles, véanse las referencias del artículo de Jorm [Jor15] (página 888).

Hay muchas variantes del método Delphi, sobre todo en función de la aplicación particular, pero se puede describir de forma general como una serie de rondas en que el coordinador del método manda a los expertos unos cuestionarios. Los expertos han de responder a estos cuestionarios y devolverlos al coordinador, quien a su ve estructura la información y los vuelve a mandar a los expertos, quienes, a su vez, han de revisar y criticar sus respuestas anteriores. Este proceso se repite hasta que se alcanza el máximo número de rondas establecido o se alcanza consenso. Asociado al método suele haber tratamiento estadístico de los datos, tanto cuantitativo como cualitativo.

Veamos más en concreto cómo se implementa el método Delphi; seguimos aquí el trabajo de Jorm. Los pasos que este establece son los siguientes:

1. **Establecimiento de la pregunta de investigación**. Como en toda investigación, hay

una serie de pregunta o preguntas que se esperan responder en este caso a partir del consenso entre los expertos.

2. **Selección del panel de expertos**. Más arriba se describió cómo elegirlos.
3. **Determinación del tamaño** del panel de expertos. Esta cuestión es delicada y depende en gran medida de la disponibilidad de los expertos y del problema en concreto. Obviamente, un número excesivamente pequeño de expertos no proporciona buenos resultados, pues la opinión de cada experto tendría mucha influencia. Lo ideal es encontrar el número mínimo de expertos que garanticen la estabilidad en los resultados. Algunos autores recomiendan un número alrededor de 23 expertos. En ciertos contextos, esto no es posible porque no hay un número tan alto de expertos o porque los expertos no siguen la metodología Delphi fielmente (y entonces hay que descartar su aportación).

4. **Diseño del cuestionario**. El cuestionario se basa en una fase previa de documentación, la cual se hace mediante una revisión de la bibliografía existente. Es importante hacer preguntas que sean de máxima relevancia (estamos usando el precioso tiempo de los expertos). Cuanto mejor esté formulada la pregunta de investigación, más relevantes serán las preguntas en el cuestionario. Existen metodologías específicas para redactar los cuestionarios; véanse las referencias citadas en [Jor15].

5. **Información previa** proporcionada al panel de expertos. En algunos casos, los expertos reciben información sobre cómo puntuar las preguntas (si estas así lo exigen, típicamente en una escala de Likert), el formato de las preguntas o la justificación de las respuestas. Es importante que las instrucciones de cómo contestar a los cuestionarios sean muy claras de modo que los expertos contesten correctamente.
6. **Distribución del cuestionario**. Los expertos no tienen que reunirse para contestar a los cuestionarios. Los medios para distribuir son variados, desde una encuesta por vía de un formulario web hasta el clásico correo electrónico.
7. **Análisis y crítica** de las información recogida en las rondas. El método Delphi requiere una definición de consenso. Una definición general y aplicable a cada no existe. Cada equipo de investigadores tiene que construir su propia definición y ponerla a prueba durante el proceso. Tras la primera ronda, el equipo de investigadores analiza los resultados y en función de ellos vuelve a mandar una segunda ronda de cuestionarios. Los expertos reciben críticas y comentarios a las respuestas de su primera ronda y se les pide que contesten a esta segunda ronda. Este proceso se repite cierto número de veces. Algunos autores recomiendan que sea tres o cuatro veces. De nuevo, depende de la investigación, pero no puede ser muy alto ya que se produce cansancio intelectual y psicológico en los expertos. El tiempo entre ronda y ronda no debería ser muy alto, pues de lo contrario se pierde interés en el proceso. Si la naturaleza del problema lo permite, se pueden tomar medidas cuantitativas y cualitativas y llevar a cabo análisis estadísticos.
8. **Informe de los resultados**. El informe de resultados puede adoptar muchas formas. Puede consistir simplemente en un recuento de los puntos en los que hubo acuerdo o puede llegar a ser algo muy complejo que se puede describir en términos de grafos, mapas conceptuales, análisis de agrupamientos, entre otros. En la figura 1 se ve un ejemplo

tomado del artículo de Jorm donde se esquematiza el proceso de las rondas y se informa del número de ítems incluidos en una investigación médica.

Figure 1. Example of a flowchart showing the number and outcomes of items in each Delphi round (C

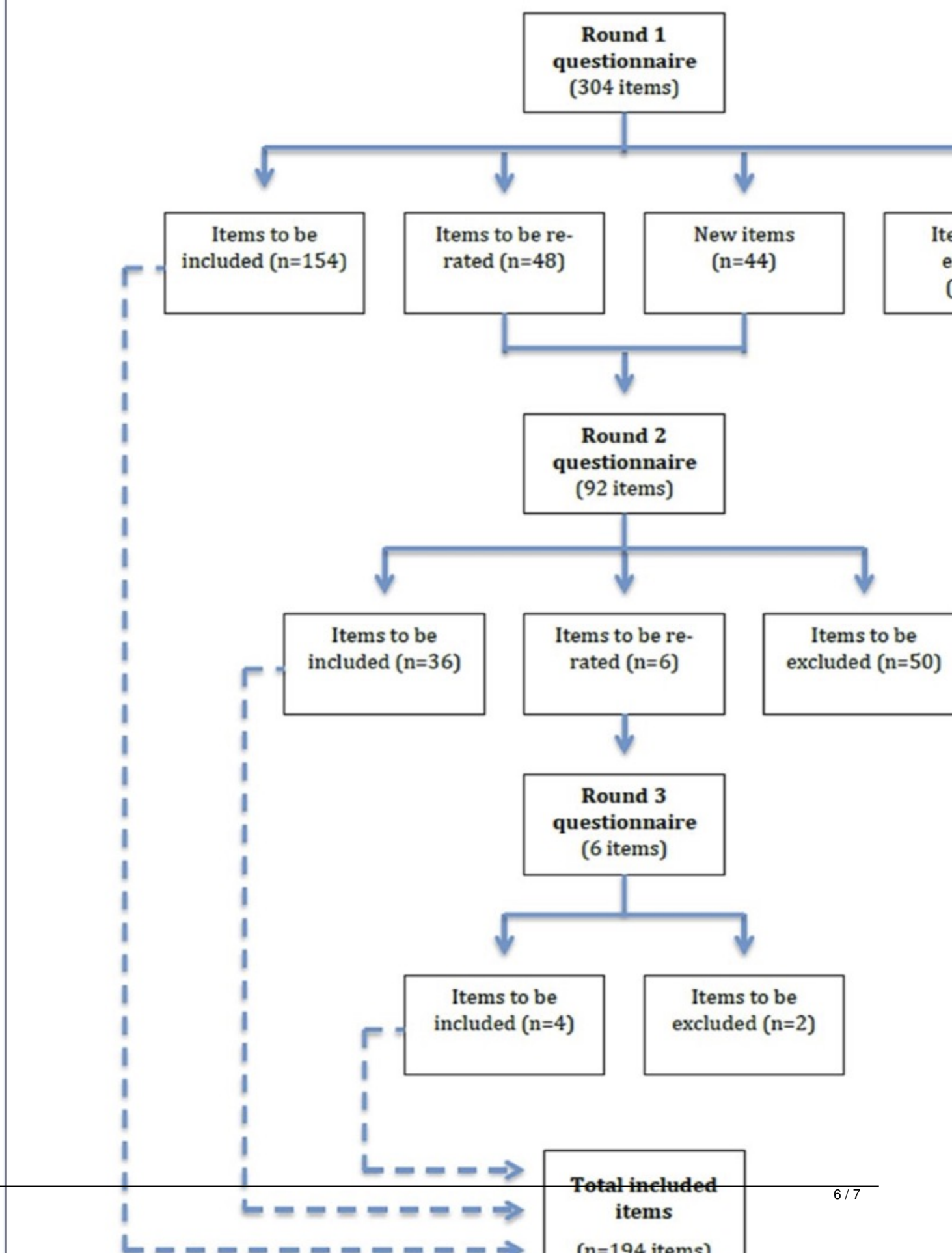

Figura 1: Diagrama de flujo asociado a un proceso Delphi (figura tomada de [Jar15])

[redacted]