



Elvis resucita

¿Nunca te has engominado el tupé y has imitado a Elvis frente al espejo, lamentando que tu voz no se parezca a la del Rey? Ahora, con proyectos como el que un equipo del MTG desarrolló en Barcelona, lo tienes más fácil

Elvis Presley nació en una familia de granjeros humildes en la ciudad de Tupelo, Mississippi, en enero de 1935, y murió en agosto de 1977 en Graceland, habiendo sido coronado como el rey del rock 'n' roll. El mito sobrevivió y aparecieron miles de imitadores que se dispersaron por todo el mundo copiando su manera de peinarse, de vestirse, de mover la pelvis, y lo más importante, también su manera de cantar. Japón, por motivos poco estudiados, concentra gran cantidad de estos imitadores. En la ciudad de Tokyo, por ejemplo, Mori Yasumasa (conocido también como J' Elvis) sube habitualmente al escenario a imitar a su cantante favorito.

Sin movernos de país, aunque veinte años atrás, en una ciudad llamada Kobe sucedió que el propietario de un bar, ante la noticia de que el guitarrista de la banda que venía a tocar aquella noche había enfermado, decidió sustituir al músico por algunas cintas grabadas que tenía en el bar. El cantante de la banda actuó esa noche acompañado de esas cintas y el público se entusiasmó con la idea. Poco más tarde el karaoke arrasó Japón.

Es a partir del conocimiento de estas dos historias que uno puede empezar a entender por qué Yamaha, una de las empresas con mayor dominio del sector del karaoke en Japón, decidió hace cosa de cuatro años iniciar un proyecto sobre transformación de la voz cantada para que los usuarios de sus máquinas de karaoke pudieran imitar la voz y la expresión de sus cantantes favoritos con ayuda de las nuevas tecnologías. Lo que no deja sorprender en esta historia

es que la empresa nipona encargara el proyecto a un grupo de investigación y desarrollo de tecnología musical ubicado en Barcelona (ver cuadro "El Grupo de Tecnología Musical (MTG) de la Universitat Pompeu Fabra").

El objetivo del proyecto Elvis era desarrollar un sistema de transformación de voz pensado para integrarse en los sistemas comerciales ya existentes de karaoke que permitiera al usuario convertir su voz en la de su cantante favorito. La declaración de tal propósito levanta muchas dudas sobre cuáles son las

posibilidades del sistema. Por eso quizás lo mejor sea hacer un repaso a las preguntas más frecuentes que se han hecho al respecto.

¿Puedo cantar *Satisfaction* tal y como lo hubiera hecho Bing Crosby?

No. Para poder imitar la voz y expresión de un cantante son condición indispensable dos cosas. La primera es disponer de una grabación del cantante al que se quiere imitar interpretando la canción que el usuario del karaoke haya elegido.

La segunda es que esta grabación debe ser "seca", es decir, sin ningún tipo de efecto (reverberación, compresión...) ni acompañamiento musical. Así pues, para poder hacerlo se debería disponer de una grabación totalmente seca de Bing Crosby interpretando *Satisfaction*.

¿Puedo cantar *Yesterday* como Paul McCartney?

Si consigues una grabación seca de su voz, sin guitarras ni violines, entonces sí puedes. Una opción es contactar con la



La interfaz del programa Elvis. El texto corresponde a la canción (en japonés, como habrás podido apreciar) que se está interpretando. A la derecha, el Elvis animado reacciona con sonrisas o muecas de tristeza según lo bien o mal que cantes.

discográfica y pedirles el master. Otra posibilidad, más real, es confiar en las nuevas técnicas de análisis de audio para, a partir de la mezcla final, separar cada uno de los instrumentos, obtener así la pista de voz y des-reverberarla si es necesario. No obstante, dado el estado actual de la cuestión, lo mejor que puedes hacer es hibernar una temporada y confiar que cuando te despiertes ya se habrá inventado un des-reverberador y un separador de pistas profesional.

¿Entonces por qué dices que podría cantar como mi cantante favorito?

No tenemos grabaciones de la voz seca y sin acompañamiento de, por ejemplo, Frank Sinatra. Pero sí tenemos este tipo de grabaciones tomadas de la voz de un buen imitador de Frank Sinatra. Así pues, en realidad, en una definición más precisa del sistema, deberíamos decir de éste que es capaz de hacerte cantar como el imitador de tu cantante favorito.

¿Y qué pasa si un chico quiere cantar como Aretha Franklin?

El sistema es capaz de hacer cantar un chico como una chica y viceversa. En la mayoría de estos casos el proceso de transformación comprenderá un cambio de octava, además de otras transformaciones un poco más complicadas.

¿Qué pasa si desafino?

Que cantas mal. El sistema es capaz de ajustar tu voz a la afinación del cantante profesional. Eso no sólo implica cantar con la afinación del cantante imitado sino también añadir vibrato a las notas largas o poder marcar las transiciones entre notas tal y como lo haría un buen cantante.

¿Qué pasa si no canto la letra de la canción?

El sistema incluye un sistema de reconocimiento del habla que ha sido entrenado para poder aplicar la transformación adecuada a cada uno de los fonemas de la voz de entrada. Si el usuario no canta la letra de la canción, el sistema de reconocimiento se confunde, se reconoce erróneamente el fonema y por lo tanto la transformación a aplicarle, y la voz sintética que resulta del proceso parece un trabalenguas. Desgraciadamente para la gente que no habla el japonés, el sistema de reconocimiento ha sido entrenado para trabajar exclusivamente con fonética japonesa.

¿Cómo funciona el sistema?

Los dos pilares que sostienen el algoritmo son:

- ◆ la alineación en tiempo real de los fonemas del usuario y del cantante profesional
- ◆ la interpolación de modelos, también en tiempo real, de las dos voces cantadas, la del usuario y la del cantante profesional.

Mientras el usuario canta, el sistema



Jonas Vidal

Álex Loscos, uno de los artifices de Elvis, haciendo una demostración del programa. Quizá no se aprecia en la pantalla del ordenador, pero el pobre Elvis está llorando a lágrima viva...

localiza el fragmento que éste está interpretando en la grabación del cantante profesional. Una vez localizado, se interpolan los modelos de las dos voces cantadas de una manera flexible y musical. Se podría definir como un sistema de análisis-morph-síntesis de voz cantada a tiempo real.

¿Qué significa morph?

El *morph* (abreviatura de *metamorphosis*, o sea, transformación no abrupta de la forma de un ser vivo, objeto, etc.) es una técnica que persigue, a partir de dos o más elementos, la generación de nuevos elementos con propiedades híbridas. En el campo del

OTROS PROYECTOS DEL MTG

El control de emisiones musicales en la radio convencional siempre se ha basado en las listas que las emisoras elaboran cada mes para que la Sociedad General de Autores determine a quién le corresponde percibir derechos de pública difusión, y en qué magnitud. Desgraciadamente, ese sistema no está exento de picaresca (y no daremos nombres ni pondremos ejemplos, claro). Con la proliferación de emisoras via Internet el panorama se complica aún más, ya que ejercer algún tipo de control sobre las emisiones resulta muy difícil... ¿O no? Al fin y al cabo, bastaría con disponer de muchos "oyentes virtuales", cada uno sintonizado a una emisora (convencional o no), que fueran anotando qué canción es la que se está emitiendo en cada momento. Eso es, más o menos, lo que persigue el proyecto RAA (Recognition and Análisis de Audio). RAA es un proyecto parcialmente

financiado por la Comunidad Europea, en el que participan, además del Grupo de Tecnología Musical de la UPF, Joanneum Research (Austria), HS-ART Digital Service (Austria), TaurusMediaTechnik, Kirch Media (Alemania), Filmkunst-Musikverlag (Alemania), Nederlands Omroepproductie (Holanda), y Radio Flaixbac (España). En la actualidad se ha implementado un sistema informático experimental que es capaz de reconocer, a los pocos segundos de empezar a emitirse, prácticamente todas las canciones que emite la emisora participante en el proyecto. El sistema es tan robusto que poco importa que el locutor hable encima, que la canción se emita a partir de un punto diferente del comienzo, que se la haya comprimido en dinámica, o que se le haya hecho una compresión de tiempo (¡para conseguir meter más canciones en una hora, claro!).

