Ensayo - Reconocimiento de archivos de audio correspondiente a las voces de personas

Jeffersson Garzón Alzate – C.C 1088352340

Computación Blanda – Docente: Gilberto Vargas Cano

Universidad Tecnológica de Pereira – 2020

“Cuando escuchamos a alguien hablar, aunque no lo veamos ni conozcamos, obtenemos diferentes tipos de información a partir de su voz: su mensaje lingüístico, su estado emocional y también, por extraño que parezca, una imagen de sus características físicas, una especie de “cara auditiva”. Si el hablante es alguien conocido, podemos además saber quién es, es decir, podemos identificarlo únicamente por la idiosincrasia de su voz. Cuando voz y cara aparecen juntas, la información obtenida a partir de ambas modalidades sensoriales se integra para crear una imagen global y completa del hablante.”

Alba Mónica Casado Ramos

Programa de Doctorado en Psicología, Universidad de Granada, España

Introducción

Un sistema de reconocimiento de hablantes se resume de que este sea capaz de reconocer automáticamente que persona esta hablando de entre las personas pertenecientes en su especificación dado el caso a una base de datos de señales de audio previamente clasificadas, por lo cual en y para la construcción de este sistema se utiliza información conocida sobre el tracto vocal de una persona y no menos importante las propiedades físicas del sonido.

Desarrollo

La voz humana es definida por la Real Academia de la lengua como “sonido que el aire expelido de los pulmones produce al salir de la laringe, haciendo que vibren las cuerdas vocales” o “Calidad, timbre o intensidad de este sonido” [RAE]

El habla consiste en combinar las unidades fónicas (fonemas) que la voz humana es capaz de generar para formar una lengua. Estos símbolos no son idénticos de un individuo a otro, pero poseen características comunes que los hace descifrables dentro de una lengua o dialecto.

Voces masculinas y femeninas

hombres y mujeres poseen voces distintas, debido a que sus sistemas fonadores tienen un desarrollo durante la pubertad.

Generalmente la voz masculina es mas grave, con tonos mas bajos, debido a cuerdas vocales más gruesas (entre 17 y 25 mm de longitud, contra 12.5 y 17 mm de las mujeres).

Tipos de voz

La clasificación más común de las voces humanas aplica a sus matices vocales, esto dado sobre una clasificación de la ópera italiana

**Soprano**

Las voces más agudas, capaces de un registro de voz que en piano iría de un Do4 a Do6. Pueden ser sopranos ligeras (provistas de más agudeza), líricas (con más volumen, pero menos agudeza) o dramática (voz más grave y timbre brillante).

**Mezzo-soprano**

Voces notoriamente más graves que las sopranos, pero con esencialmente las mismas características, tanto así que pueden clasificarse del mismo modo: mezzo-ligeras, mezzo-líricas y mezzo-dramáticas.

**Contralto**

Las más graves de las voces agudas, apenas un 2% de las mujeres del mundo las poseen. Poseen una rica sonoridad y enorme amplitud en su registro, oscilando entre un fa3 y un fa5.

**Tenor**

Las más agudas de las voces graves, en un registro corto: entre Do3 y Do4. Se clasifican al igual que las sopranos en ligeras, líricas y dramáticos.

**Barítono**

Las más comunes de las voces masculinas son barítonos, se trata de voces poco ágiles, pero cuyo rango intermedio les permite alcanzar una versatilidad entre agudeza y gravedad.

**Bajo**

Las voces más graves y oscuras de todas, semejantes a las notas bajas de un violoncelo. Son casi totalmente masculinas y aparecen luego del desarrollo sexual del individuo.

Generalmente las tres primeras son voces femeninas y las tres últimas son voces masculinas.

Cualidades de la voz

No menos importantes debemos tener en cuenta que la voz humana tiene cuatro cualidades fundamentales:

**Articulación**

Los modos en que la voz se modifica a través de los movimientos de los órganos del aparato fonador, para conseguir sonidos determinados y así formar, por ejemplo, las palabras.

**Tono o entonación**

La modulación en el ritmo de producción de la voz, que permite atribuirle a los sonidos un cariz específico, como la emoción o el acento.

**Timbre**

El modo individual y característico con que la voz es producida, pudiendo ser graves (bajas) o agudas (chillonas).

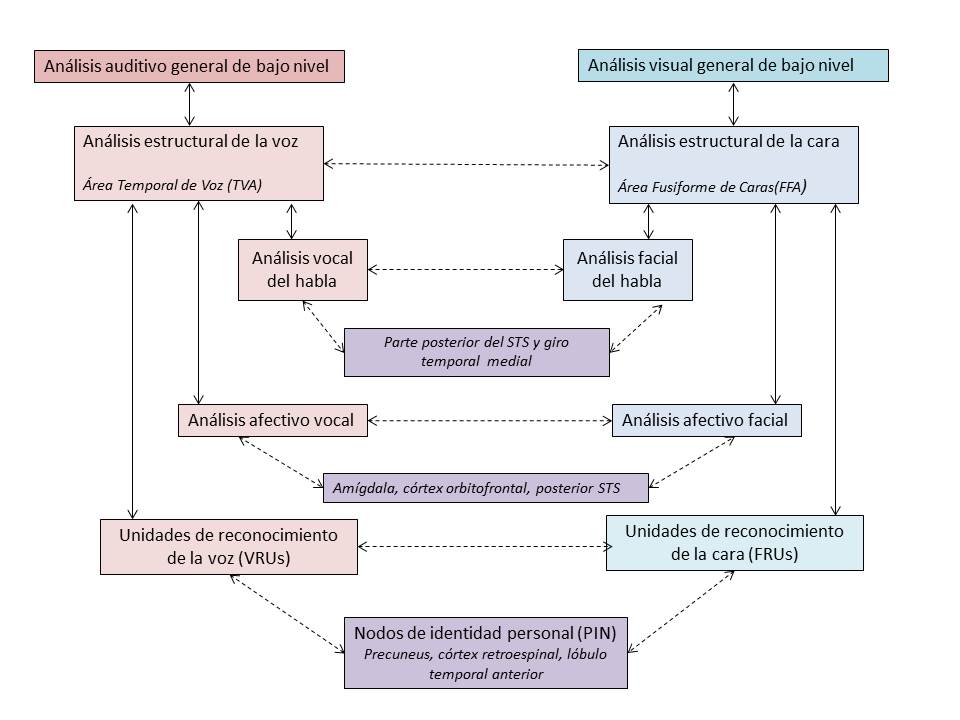
**Intensidad**

La cantidad de fuerza con que se obtienen los sonidos, pudiendo recaer en algunas partes de la cadena sonora, como en las distintas sílabas de una palabra.

La voz, junto con información lingüística, transmite también información ligada a las características físicas y emocionales del hablante. Por ello, la voz no es sólo el principal y más primitivo vehículo del lenguaje, sino que además es nuestra “**cara auditiva**”. Cada hablante tiene unas características acústicas específicas que vienen marcadas por la forma de su aparato vocal. Cuando el aire pasa a través de las cuerdas vocales de la laringe, éstas vibran produciendo un sonido complejo con una frecuencia fundamental (F0) que depende de la longitud y el grosor de las mismas. Este sonido básico va resonando a través de las cavidades superiores -faríngea, oral y nasal-, concentrando la energía sonora en determinadas bandas de frecuencias y dando como resultado los formantes (F1, F2, F3, …). Por lo general, la voz de las mujeres presenta un patrón de frecuencias o timbre más alto que el de los hombres, tanto en su F0 (entre 165 y 225 Hz para las mujeres, frente a entre 85 y 180 Hz para los hombres) como en sus correspondientes formantes. Por lo tanto, analizando estos parámetros podemos identificar una serie de características físicas del hablante (edad, género, altura, peso…) y generar una imagen visual a partir de la voz.

Si el hablante es una persona conocida la podemos además identificar a través de su voz, activando la imagen visual que de ella tenemos almacenada en la memoria. Esta habilidad está presente desde que somos pequeños, aproximadamente a los 7 meses de edad, y se da también en otras especies animales, como los macacos (Belin y col., 2011). La habilidad de reconocer a las personas a través de la voz puede verse afectada por un déficit neurológico denominado fono agnosia, que consiste en la incapacidad de discriminar entre diferentes voces y la incapacidad de reconocer voces familiares. Este déficit es análogo a la prosopagnosia, la incapacidad de reconocer a las personas a partir de su cara, y demuestra que el procesamiento de la identidad de las personas puede realizarse por diferentes vías independientes (Belin y col., 2011).

La mayoría de las interacciones sociales comunicativas, la información auditiva y visual se presentan juntas, dadas por varios profesionales en conjunto presenta un modelo de reconocimiento de las personas a través de la voz.



*Modelo de percepción de la voz de Belin, Fecteau y Bedard (2004). El procesamiento de la información vocal estaría disociado en tres sistemas independientes: el que analiza la información semántica, el que analiza la información afectiva y el sistema que analiza la identidad del hablante. Estos sistemas entrarían en constante interacción con sus homólogos en el procesamiento de las caras. En la figura se mencionan también las principales áreas cerebrales implicadas en cada una de estas fases de procesamiento.*

Conclusión

En conclusión, con la información acústica propia de cada persona, que depende de su aparato fonatorio, creamos su “cara auditiva”, es decir, obtenemos información de cómo es físicamente, además de información acerca de su estado emocional. Cuando esa persona es conocida, escuchar su voz permite también acceder a la información almacenada sobre ella e identificarla, aunque no la veamos. Además, en los procesos de comunicación habituales integramos la información visual y auditiva disponible sobre el hablante, lo que redunda en una mayor eficacia y rapidez en su identificación.

**Referencias**

**Main Web: http://www.cienciacognitiva.org/?p=854**

Belin, P., Fecteau, S., y Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. Trends in Cognitive Sciences, 8, 129–135.

Belin, P., Bestelmeyer, P. E. G., Latinus, M., y Watson, R. (2011). Understanding voice perception. British Journal of Psychology, 102, 711-725.

Bruce, V., y Young, A. (1986) Understanding face recognition. British Journal of Psychology, 77, 305–327.

Campanella, S., y Belin, P. (2007). Integrating face and voice in person perception. Trends in Cognitive Sciences, 11, 535–543.

Quiñones-González, I., Bobes-León, M. A., Belin, P., Martínez-Quintana Y., Galán-García, L., y Sánchez-Castillo, M. (2012). Person identification through faces and voices: An ERP study. Brain Research, 1407, 13-26.