Consideraciones prácticas respecto del uso de variables acústicas en evaluación

MATERIAL COMPLEMENTARIO

(Descripción EGEMAPS)

Borja Artiñano Arizmendi

A continuación, se definen brevemente los parámetros incluidos en el set eGeMAPS, junto con una breve descripción de los mismos, las funciones aplicadas, y posibles asociaciones entre ellos y los rasgos del modelo Big Five, todas ellas extraídas de Polzhel (2015). Es importante mencionar que Polzhel (2015) pertenece a la tradición que selecciona número escaso de predictores para evaluar el impacto de cada uno en la evaluación de la personalidad, por lo que hay pocas características de eGeMAPS que sean mencionadas.

En total hay 16 tipos de descriptores de bajo nivel (LLD, por *Low Level Descriptors*):

Tono (pitch). Indica como de grave o agudo se percibe un sonido. Depende directamente de la frecuencia, por lo que a veces se dice que el tono es la frecuencia percibida del sonido. En concreto, la relación entre tono y frecuencia es aproximadamente logarítmica, donde un aumento exponencial de la frecuencia provoca un aumento lineal en el tono. Se operativiza a través de los semitonos, incrementándose un semitono cada vez que se dobla la frecuencia física. El semitono 0 se da en 27.5 Hz.

Se ha asociado con la Apertura y la Extraversión, especialmente las funciones que miden variación y tasa de cambio. Este último aspecto también se ha relacionado con el Neuroticismo. El valor medio del tono se ha asociado con la Responsabilidad.

Fluctuación del retardo. Diferencia en segundos entre los periodos de dos frecuencias fundamentales consecutivas. Es decir, representa, en cierto modo, la estabilidad de la señal.

Una fluctuación del retardo elevada puede asociarse con varios cambios fisiológicos, como tensión en las cuerdas vocales, cambios en la presión pulmonar, y en general, se puede usar para evaluar la estabilidad y control de la persona sobre sus cuerdas vocales.

Frecuencia de los formantes 1, 2 y 3. La numeración hace referencia a la región de frecuencias concretas en las que se concentra la energía. El formante 1 tiene frecuencias más bajas que el formante 2, y el 2 tiene frecuencias más bajas que el 3. Para operativizarlos, se asigna a cada uno el centro del espectro de frecuencias, que se corresponde con el punto central de la región de frecuencias donde la intensidad es máxima.

Los formantes en general y diversas funciones derivadas de ellos se han asociado con la Apertura y la Amabilidad.

Ancho de banda de los formantes 1 , 2 y 3. Indica en ancho de banda de los tres primeros formante, es decir, la anchura de la región en torno a la frecuencia central en la que la intensidad es elevada. Nos indica como de concentrada está la energía en torno al centro de los formantes.

Se han asociado con la Amabilidad.

Shimmer. Representa cómo la intensidad varía de una frecuencia fundamental a otra consecutiva. Es decir, como de regular es la intensidad.

Volumen. Función logarítmica de la intensidad, expresada en decibelios (dB), relacionada con la percepción de un sonido como fuerte o débil.

Se ha asociado con la predicción de Responsabilidad, la Amabilidad y la Extroversión.

***Harmonics-to-Noise Ratio* (HNR)**. Representa como de armónica es la señal, esto es, como las distintas frecuencias de la señal son múltiplos de F0, lo que confiere al sonido una cualidad musical y agradable. Cuanto menor sea el ratio, más ruido habrá. Esto se suele interpretar como que la señal es de peor calidad. También se usa para diagnosticar patologías de la voz.

Se ha asociado con la predicción de la Amabilidad.

Alpha ratio. Representa el ratio de la suma de la energía en frecuencias 50-1000 Hz frente a la suma de la energía en frecuencias 1-5kHz. Nos indica si en la señal dominan las frecuencias altas o bajas, es decir, para estudiar la distribución de la energía.

Pendiente espectral en los rangos 0–500 Hz y 500–1500 Hz. Representa la tasa de cambio entre la potencia y frecuencia (escala logarítmica), en los rangos de frecuencia indicados. Es decir, es la pendiente en una regresión que muestra cómo cambia la energía al incrementarse la frecuencia. Si es positivo, la energía aumenta al aumentar la frecuencia, y si es negativo, la energía disminuye al aumentar la frecuencia.

Hammarberg index. Ratio del pico de energía entre 0-2kHz y el pico de energía entre 2-5kHz. De nuevo, se usa para estudiar la distribución de la energía en las frecuencias de la señal.

Energía relativa de los formantes 1, 2, y 3. Es un ratio de la intensidad del pico de cada uno de los tres primeros formantes entre la intensidad en el pico de la frecuencia fundamental. Indica qué proporción de la intensidad total de la señal se da en cada formante.

Diferencia de la energía entre harmónicos H1-H2. Diferencia entre la energía contenida en el primer harmónico y el segundo. El primer harmónico sería el primer múltiplo presente de la frecuencia fundamental, y el segundo, el segundo múltiplo presente.

Diferencia de la energía entre H1-F3. Similar al anterior, pero es la diferencia entre la energía del primer harmónico y la energía del harmónico que coincida con el tercer formante.

Flujo espectral. Diferencia entre dos distribuciones de frecuencias de dos consecutivas.

Coeficientes MFC. Los MFCC (Mel Frequency Cepstral Coefficients) son una representación de la distribución de la energía de la señal de sonido en distintas frecuencias. Esta distribución se calcula para intervalos cortos (20-40ms) que se sobreponen los unos a los otros, para obtener la distribución de frecuencias del audio completo.

Es sabido que el sistema auditivo humano no procesa todas las frecuencias de forma similar, sino que, perceptivamente, unas son más importantes que otras. La escala MFC pondera las frecuencias del audio según la responsividad a las mismas del oído humano.

Resumidamente, el cálculo de los coeficientes en escala MFC es el siguiente:

- Se obtiene la distribución de la energía en las distintas frecuencias del audio, a partir de frames que representan intervalos muy cortos y que se sobreponen unos a otros, usando la transformada de Fourier.

- Estas frecuencias se pasan por un filtro que las pondera de acuerdo a la responsividad del oído humano.

- Posteriormente, se aplica una función logarítmica, ya que el oído humano procesa la cualidad de 'loudness' (que un sonido sea más alto que otro) logarítmicamente.

- Finalmente, se aplica una Discrete Cosine Transform, un algoritmo de reducción de dimensionalidad, para volver la solución más simple y estable, y reducir la correlación entre los coeficientes. (Müller, 2007)

La interpretación de estos coeficientes no es clara, pero como ya se ha mencionado, resultan útiles en un gran número de tareas relacionadas con el procesamiento de audio.

Respecto a la predicción de la personalidad, han sido asociados con la Apertura. En eGeMAPS se incluyen los MFCC 1-4.

Nivel de sonido equivalente. Indexa la energía producida a lo largo de todo el audio.

Se incluyen además el número de máximos relativos por segundo en la intensidad, y la duración de aquellos segmentos que presentan o no voz de forma continua.

Funciones aplicadas a cada LLD

	Media	Coeficiente de Variación	Percentiles (20, 50, 80)	Rango percentil (20-80)	Media de la pendiente	Desviación típica de la pendiente	¿Separación voz/sin voz?
Tono	✓	✓	✓	✓	✓	✓	
Fluctuación del retardo	✓	✓					
Frecuencia de formantes 1 a 3	✓	✓					
Ancho de banda formantes 1 a 3	✓	✓					
Shimmer	✓	✓					
Volumen	✓	✓	✓	✓	✓	✓	
HNR	✓	✓					
Alpha ratio	✓	✓					
Hammarberg index	✓	✓					
Pendiente espectral 0 – 500 Hz	✓	✓					
Pendiente espectral 5 – 1500 Hz	✓	✓					
Energía relativa formantes 1 a 3	✓	✓					
Diferencia energía H1 – H3	✓	✓					
Diferencia energía H1 – F3							
Flujo espectral	✓	✓					✓
MFCC 1 a 4	✓	✓					✓