# Apéndice A – Cálculo de las variables acústicas

En este apéndice se detallan, en primer lugar, las variables acústicas calculadas para cada tarea de este estudio y, posteriormente, se describe la técnica de VAD empleada para la detección de voz y silencios en las grabaciones.

## Tareas de fluidez verbal

De cada grabación realizada por los participantes en las tareas de fluidez verbal se calcularon las siguientes 17 variables acústicas:

* El número total de palabras emitidas por el sujeto en el minuto que duraba la tarea (1 variable).
* La diferencia en segundos entre el instante de comienzo de la primera palabra emitida por el sujeto y el resto de palabras, contando como mucho 9 palabras (8 variables).
* La posición relativa en un intervalo [0, 1] del comienzo de las 9 primeras palabras emitidas por el sujeto, tomando como posición relativa cero el comienzo de la primera palabra y posición relativa uno el final de la grabación (1 minuto) (8 variables).

El recuento de palabras emitidas en cada grabación y el etiquetado del tiempo de inicio de las palabras se llevó a cabo de forma manual usando el software Audacity versión 2.1.0.

## Tareas de conteo hacia atrás y descripción de la lámina del Test de Boston

En las grabaciones realizadas para estas dos tareas se calculó, por un lado, las energías logarítmicas en decibelios (dBs) a nivel de trama (ventana de análisis tipo Hamming con una longitud de 20 ms solapadas 10 ms y un factor de preénfasis igual a 0.97) y, por otro, la frecuencia fundamental de la voz (F0), también a nivel de trama. Ambas medidas se calcularon usando el programa Wavesurfer versión 1.8.8 para MacOS.

En base a las medidas de energía logarítmica obtenidas se aplicó la técnica de VAD descrita más abajo para decidir qué tramas contenían voz y cuáles silencio (no voz). Por otro lado, el valor de F0 se utilizó para clasificar las tramas como segmentos periódicos (F0 > 0) o aperiódicos (F0 = 0), en función de vibración de las cuerdas vocales. Una vez clasificadas las tramas, se calculó de forma automática la duración en segundos de los segmentos del mismo tipo (voz, silencio, periódicos y aperiódicos). En resumen, para cada grabación se obtuvieron una lista con las duraciones de los segmentos de voz, segmentos de silencio, segmentos periódicos y, finalmente, segmentos aperiódicos.

A partir de la lista de duraciones de los segmentos se calcularon las siguientes 36 variables acústicas para cada grabación:

1. Número total de segmentos de cada tipo (voz, silencio, periódicos, aperiódicos) en cada grabación (4 variables por cada grabación).
2. Promedio de las duraciones en segundos para cada tipo de segmento (4 variables por grabación).
3. Mediana de las duraciones en segundos para cada tipo de segmento (4 variables por grabación).
4. Desviación estándar de las duraciones en segundos para cada tipo de segmento (4 variables por grabación).
5. Suma de las duraciones en segundos para cada tipo de segmento (4 variables por grabación).
6. Cociente entre las duraciones promedio para cada par de tipos de segmento, esto es, las siguientes 4 variables:
   1. duración promedio de los segmentos tipo voz/duración promedio de los segmentos tipo silencio,
   2. duración promedio de los segmentos tipo silencio/duración promedio de los segmentos tipo voz,
   3. duración promedio de los segmentos periódicos/duración promedio de los segmentos aperiódicos y
   4. duración promedio de los segmentos aperiódicos/duración promedio de los segmentos periódicos.
7. Lo mismo que el punto 6 pero con las medianas de las duraciones obtenidas en el punto 3 (4 variables por grabación).
8. Lo mismo que el punto 6 pero con las desviaciones típicas de las duraciones obtenidas en el punto 4 (4 variables por grabación).
9. Lo mismo que el punto 6 pero con las suma de las duraciones obtenidas en el punto 5 (4 variables por grabación).

## Detección de actividad de voz (VAD)

Para la detección de los segmentos de voz y no voz en las grabaciones se utilizó la técnica de VAD propuesta en (Bimbot et al., 2004) y que, de forma breve, se describe a continuación:

1. En primera lugar, se calcularon las energías logarítmicas en decibelios (dB) a nivel de trama (ventana de análisis tipo Hamming con una longitud de 20 ms solapadas 10 ms y un factor de preénfasis igual a 0.97) usando el programa Wavesurfer versión 1.8.8 para MacOS.
2. Se ajustó un modelo de mezcla de distribuciones normales univariadas con 2 componentes (modelo bi-Gaussiano) a los datos de energía logarítmica calculados en el paso 1 para cada grabación.
3. La componente del modelo con menor valor para la media se consideró que modelaba los segmentos de silencio (no voz), mientras que aquella con mayor valor absoluto en la media se consideró que modelaba los segmentos de voz.
4. La decisión sobre qué segmentos de la grabación se consideraban silencio o voz se basó en un umbral θ calculado de forma automática como el punto medio en donde las dos distribuciones de probabilidad (voz y silencio) se cortan. Este umbral se corresponde con la raíz real situada entre las medias de ambas distribuciones de la siguiente ecuación cuadrática:

donde

siendo y las distribuciones normales que modelan el silencio y la voz (paso 2), respectivamente.

1. Se consideraron como silencio aquellas tramas (segmentos de 10 ms) obtenidas en el paso 1 cuya energía fuese inferior al umbral θ y como voz las tramas con energía superior a este umbral.
2. Finalmente, sobre las decisiones tomadas en el paso 5 se aplicó un filtro de mediana de longitud 0.1 s centrado en cada muestra El objetivo de este filtro fue robustecer la clasificación voz/silencio ante posibles ruidos espontáneos.