Máster en Metodología en las Ciencias del Comportamiento y de la Salud

Consideraciones prácticas respecto del uso de variables acústicas en evaluación

Borja Artiñano Arizmendi.

Modalidad: Prácticas Externas.

Tutor Interno: Francisco José Abad.

Tutor Externo: David Aguado García.

Centro: Instituto de Ingeniería del Conocimiento.

Periodo: Octubre – Diciembre (2023)

Junio, 2024.

Resumen

Se desarrollaron unas prácticas profesionales en el Instituto de Ingeniería del Conocimiento, durante el periodo Octubre – Diciembre de 2023. Los objetivos planteados se adscribieron a un proyecto de investigación relacionado con el empleo de AVIs en evaluación automática de la personalidad, a través de aspectos de la prosodia humana. Dado el incipiente estado del proyecto, alrededor de una tercera parte de las prácticas se centró en la identificación de las variables relevantes, y en aprender a manejar y automatizar el uso de software especializado para su extracción. El resto del periodo se dedicó a explorar cuestiones como son la longitud del audio a analizar, la redundancia de las variables, y su posible estructura factorial. Se consiguió extraer satisfactoriamente un set de variables relevante a través del software *openSmile*, y se creó un *script* en Python para automatizar su uso. Sin embargo, respecto a las otras cuestiones planteadas, se concluyó que el volumen de datos era insuficiente para extraer conclusiones sólidas. Debido a esto, se decidió tratar el trabajo realizado como un 'proyecto piloto'. Finalmente se recomendó a la organización repetir las pruebas tras aumentar la muestra, y contar con un experto en acústica para interpretar los resultados adecuadamente.

Abstract

The internship took place at the *Instituto de Ingeniería del Conocimiento*, covering the 2023 October – December period. The stated goals were part of a research project related to the use of AVIs in automatic personality assessment. Given the nascent state of the project, roughly a third of the internship's period was focused on identifying the relevant variables and learning to employ and automate the use of specialized software for their extraction. The rest of the period was devoted to exploring research questions such as the length of the audio to be analyzed, the redundancy of the set of variables, and its potential factor structure. A relevant set of variables was successfully identified and extracted through the openSmile software. In addition, a Python script was created to automate its use. At the same time, it was concluded that the data volume available at the time was insufficient to appropriately answer the laid-out research questions. This led to considering the project as a sort of 'pilot study'. The internship finalized with a recommendation to the organization to gather more data before repeating the analyses, and to count on an acoustics expert to better interpret the results.

Tabla de Contenidos

Introducción

Antecedentes generales

Evaluación de la personalidad en RRHH

La evaluación de la personalidad es un tema de gran interés en la psicología organizacional, en tanto que, durante las últimas décadas, ha mostrado ser un constructo útil para predecir multitud de indicadores deseables en el trabajo, como son la satisfacción laboral, el rendimiento, el compromiso organizacional, la realización de comportamientos contraproducentes, etc (Hough y Dilchert, 2017). Huffcutt et. al (2001), por ejemplo, muestran que la personalidad es la categoría de constructo más evaluada en las entrevistas (aunque nótese que su estudio tiene algo más de dos décadas; sin embargo, Hough y Dilchert (2017) parecen opinar que la tendencia se mantiene y va al alza). Es evidente, sin embargo, que la entrevista es un proceso costoso e imposible de aplicar de forma masiva, como puede desearse para filtrar inicialmente a los candidatos, por ejemplo.

Una alternativa popular es el uso de cuestionarios psicométricos, principalmente debido a su mayor eficiencia (Ryan et. al, 2015). Sin embargo, a pesar de las ventajas de estos instrumentos, existe una preocupación generalizada, tanto entre psicólogos organizacionales como entre profesionales de RRHH, sobre la distorsión intencionada de la respuesta en contextos de selección o promoción profesional (Mueller-Hanson et. al, 2003; Ryan y Ployhart, 2014; Ryan et. al, 2015), si bien algunos autores aseguran que esta preocupación, aunque legítima, es exagerada y puede resolverse por medio de otras vías (véase Hough y Johnson, 2013; Hough y Dilchert, 2017).

En cualquier caso, está claro que existe un interés generalizado por parte de compañías y profesionales de RRHH por adoptar una alternativa económica, informativa y eficaz que guíe los primeros pasos del proceso de selección. Una opción que parece cada vez más popular, especialmente en respuesta al 'boom' actual de la IA, es el uso de las denominadas Asynchronous Video Interviews (AVIs) (Lukacik et al, 2022). En una AVI, el candidato responde a una serie de preguntas grabándose en vídeo a través de un portal online. Como el nombre indica, esta comunicación es asíncrona, sin que exista una interacción con un entrevistador. Posteriormente la AVI es evaluada, bien por un profesional, bien por un algoritmo informático. Como es de esperar, esta última opción puede interesar mucho a las empresas por sus muchas ventajas (abaratamiento de costes, rapidez, potencial ausencia de sesgos, etc). De hecho, en el mercado ya se han introducido proveedores de soluciones de reclutamiento basadas en AVIs, que producen informes automáticos (HireValue, 2018), si bien

es necesaria más investigación sobre las garantías psicométricas de estos instrumentos (Lukacik et al, 2022). Estos algoritmos pueden tomar información de múltiples canales (contextual, visual, sonoro o semántico). Es precisamente en el canal sonoro en el que se ha centrado el trabajo realizado durante estas prácticas profesionales, en el que se tratarán cuestiones variadas sobre los datos procedentes de variables acústicas.

Personalidad y prosodia del discurso

En nuestro día a día, muchas veces sin darnos cuenta, realizamos juicios sobre otras personas a partir de múltiples señales, entre ellas indudablemente su forma de hablar, reír o expresarse vocalmente en general (Hall et al., 2016). Nestler y Back (2013) conceptualizan este proceso a través del Modelo de Lente, en el que no nos detendremos por no alargar la extensión del trabajo. Baste decir que, según este modelo, para poder realizar juicios válidos de otros necesitaremos disponer de señales (variables) que mantengan una relación real con el fenómeno, y utilizarlas (ponderarlas) adecuadamente. En el caso de la personalidad, estos juicios que realizan observadores no entrenados, a partir de muy diversos tipos de señales, y en situaciones en las que no se conoce previamente a la persona, han mostrado tener una cierta correlación con puntuaciones obtenidas por autoinforme (de entre 0.1 y 0.3; Back y Nestler, 2016).

Entre estas señales que las personas emitimos, se encuentran las acústicas, que a su vez suelen ser divididas en tres categorías: semánticas, prosódicas y espectrales (Wu et al, 2011). Las primeras, que no contemplaremos en este trabajo, hacen referencia al significado de los términos que las personas emplean. Las segundas hacen referencia a aquellas características del lenguaje hablado que le proporcionan un ritmo y melodía particulares, y lo hacen sonar 'natural', incluyendo algunas variables como la variabilidad del tono, la intensidad, o la duración de los fonemas (Chittaragi et al., 2017). Las terceras, en cambio, hacen referencia a la distribución de la energía del sonido en diferentes rangos de frecuencia, e incluyen variables como los Coeficientes Cepstrales en Frecuencias de Escala Mel, que han mostrado ser útiles en multitud de tareas relacionadas con la predicción de criterios a través de la voz (Eyben et al., 2016).

La posibilidad de realizar inferencias sobre la personalidad de los hablantes a través de la prosodia de la voz, no es un campo de estudio nuevo en Psicología. El lector interesado puede consultar a Polzehl (2015) para ver un resumen de la literatura desde sus comienzos cerca de la mitad del siglo XX. Si algo llama la atención de los primeros trabajos es la dificultad para

operativizar de forma clara y objetiva las variables estudiadas, el escaso número de indicadores empleados, y la divergencia en el uso de estos de un autor a otro. En la actualidad, es indudable que la investigación se ha visto muy influenciada por los avances computacionales, dando lugar a algoritmos de extracción masiva y semiautomática de parámetros acústicos de forma precisa, si bien no siempre estandarizada (ver, por ejemplo, Schuller et al., 2012). Si bien esto ha tenido como resultado una mejora en la predicción, también ha implicado, necesariamente, una pérdida de la interpretabilidad de los mecanismos que expliquen la relación entre prosodia y personalidad (Schuller et al., 2009), así como una preocupación porque los modelos se sobreajusten a los (comparativamente pequeños) conjuntos de datos usados para el entrenamiento (Schuller et al., 2010). Por otra parte, la investigación es en conjunto escasa, y el uso de diferentes sets de parámetros dificulta la comparación de resultados.

Trasladando esta información al contexto de las AVIs, se puede concluir que existen medios técnicos para implementar la extracción de variables prosódicas y espectrales de las respuestas del candidato, además de investigación que avala (parcialmente) una relación entre predictores de esta clase y la personalidad. Sin embargo, existen aún múltiples interrogantes que no permiten el uso de esta tecnología en un contexto aplicado, y mucho menos en un contexto tan delicado como es la selección de personal.

Contexto de las prácticas profesionales

Estas prácticas se realizan en el área de Talento del Instituto de Ingeniería del Conocimiento (IIC). Las prácticas se enmarcan en un proyecto de investigación en AVIs, que se encuentra en un estado muy incipiente, y dentro del cual este trabajo ha supuesto un primer acercamiento al tratamiento de datos relacionados con la voz. Esto explica que, dentro de los objetivos que expondré a continuación, el principal ha sido siempre revisar el estado del arte y transmitir ese conocimiento de vuelta a la organización.

Se dispone de un total de 42 videos de estudiantes de Psicología en la UAM, que responden de forma asíncrona a una pregunta genérica en el contexto AVI (*"Cuenta una historia de entre 1 y 2 minutos de duración, utilizando las palabras que se indican a continuación: 'fuegos artificiales', 'accidente aéreo', 'camarera', 'supermercado' y 'Edad Media' ."*). Los videos son grabados por los propios participantes, con su equipo y en el lugar que han deseado, y con una duración variable.

Inevitablemente, surge la necesidad de transmitir la dificultad de extraer conclusiones válidas con tan limitado volumen de datos, especialmente ante algunas de las demandas de la

organización (realizar un Análisis Factorial, usar técnicas de *Machine Learning* etc) y en un ámbito en el que se suele trabajar con un gran número de variables independientes. A pesar de esto, se anima al alumno en prácticas a continuar con los objetivos que se comentarán a continuación. Por ello, finalmente se decidió tratar el resultado de esta actividad como un estudio o proyecto piloto, para explorar la cuestión y, si es prometedora, hacer análisis más rigurosos en el futuro, al disponer de un mayor volumen de datos.

Objetivos

Se plantean los siguientes objetivos. Más allá de los dos primeros, el ritmo es algo improvisado, siendo cada objetivo una pregunta de investigación que surge según se trabaja con el material:

1) Aprender a realizar la extracción de variables acústicas de una AVI, y automatizarla y facilitarla todo lo posible.
2) Buscar un set de variables a extraer que sea estándar, potencialmente interpretable, y apropiado para la tarea. La elección debe basarse en la literatura previa sobre la relación prosodia – personalidad, dentro de lo posible.
3) Explorar si las variables extraídas son estables a lo largo del mismo audio.
4) Explorar cómo la duración de un segmento de audio afecta a la estabilidad en la recuperación de cada variable.
5) Dado que el número de variables a tratar es elevado, y su contenido parece redundante, detectar y eliminar variables redundantes.
6) Realizar una reducción de dimensionalidad e interpretar los factores resultantes, en la medida en que esto sea posible con el volumen de datos disponible.
7) Explorar relaciones entre estas variables y puntuaciones en cuestionarios de personalidad.

Alcance del trabajo realizado

Este trabajo ha supuesto una primera aproximación al estudio de aspectos relacionados con la prosodia y la personalidad de los hablantes en el IIC. Como tal, ha aportado herramientas y conocimiento para facilitar la investigación en esta línea por parte del equipo. De forma más general, ha servido de excusa para reflexionar sobre diversos aspectos en relación al uso de las AVIs como herramienta de evaluación. Los resultados se presentarán en forma de póster en el congreso de la International Test Commission en 2024.

Principales resultados obtenidos

En relación con los objetivos planteados previamente, los resultados fueron:

- Se consiguió realizar de forma satisfactoria la extracción de variables acústicas de los videos de la muestra, y se creó un notebook interactivo que automatiza el proceso. También se creó documentación para explicar su uso y la configuración de los distintos argumentos.
- Se identificó un set de parámetros, relativamente parsimonioso y potencialmente útil, para emplearse en evaluación de la personalidad.
- Se exploró la estabilidad de las variables, así como el efecto de la duración de los segmentos analizados en esta estabilidad.
- Se detectó la presencia de variables redundantes a través del método UVA, y se procedió a eliminar aquellas que, conceptualmente, se vieron como redundantes. El empleo de técnicas con esta finalidad parece ser nuevo en este ámbito de investigación, si bien es difícil de decir, porque esta es muy escasa.
- Se realizó un Análisis Factorial con el conjunto de variables resultante. Los resultados, sin embargo, son muy poco concluyentes. Se informó a la organización de que es necesario conseguir un mayor volumen de datos, además de contar con un experto en acústica, para realizar adecuadamente esta tarea.

Descripción del trabajo

Las actividades realizadas se describen por semanas. Cada Martes hay una reunión con el tutor profesional en la que este aporta retroalimentación sobre el trabajo de la semana pasada, y se plantean metas para la siguiente.

Semana 1: 03/10 – 06/10

Se introduce al alumno a las actividades del departamento, y se le permite explorarlas. En esta semana se informa al alumno de que, en la actualidad, el IIC no permite a los alumnos en prácticas trabajar con datos de clientes, por lo que automáticamente varias actividades más tradicionales (ajuste de modelos psicométricos, por ejemplo), quedan descartadas. Dado que parece ser la actividad que mejor concuerda con los conocimientos obtenidos durante el máster, se decide aprovechar una incipiente línea de investigación en AVIs para estudiar posibles aplicaciones de la prosodia a estas. En este momento se plantean los dos primeros objetivos: 1) Aprender a extraer variables acústicas de audios, para transmitir el conocimiento de vuelta al instituto, y 2) Hacer una revisión de la investigación en esta temática, y buscar un conjunto de variables apropiado para predecir el criterio.

No se define una temporalidad para estos dos objetivos, debido al desconocimiento inicial de ambas partes sobre los mismos, pero se asume que pueden ser la actividad principal durante las prácticas.

Semanas 2 y 3: 9/10 - 20/10

Se identifica el software open source openSmile (Eyben et al., 2022), que permitirá cumplir el primer objetivo. El trabajo con openSmile, sin embargo, resulta ser poco *user-friendly* por varias razones, entre ellas, su falta de una interfaz gráfica de usuario, que requiere operar desde una terminal.

Como el objetivo principal de las prácticas es, en este momento, facilitar el uso de estas herramientas por parte de la organización, se crea un script que permite ejecutar el programa y realizar algunas operaciones previas de tratamiento de vídeo través de Python, de forma sencilla. Las librerías empleadas son '*moviepy*' y '*pydub*', para tratamiento de vídeo y audio, y '*os*' y '*subprocess*', para interactuar con el sistema operativo y la terminal, respectivamente. Además, se crea un documento *.ypinb* que permite ejecutar el script en un Jupyter Notebook (ver 'Python_Notebook_Prosodia.ipynb' en Github).

Es también en este momento cuando se toma contacto con el material, y se recomienda obtener una muestra mayor antes de intentar extraer conclusiones sólidas.

Semana 4: 23/10 - 27/10

Tras haber dado por satisfactorio, el primer objetivo, se decide continuar con el segundo: elegir un conjunto de variables que podamos emplear como predictores. Tras explorar algunas alternativas, se decide emplear el set de parámetros acústicos eGeMAPS v2, basándonos en su parsimonia y fundamentación teórica (Eyben et al., 2016), y en la existencia de literatura previa que relaciona la personalidad con este set, si bien el campo de investigación es, en general, escaso (Solera-Ureña et al., 2016; Koutsombogera et al., 2020; Rangra et al., 2023). Se consideró que, de cara a los objetivos de la organización, era más relevante elegir un set con un número relativamente pequeño de dimensiones (88) con cierta fundamentación, frente a otros sets más populares. La razón fue que estos últimos tienen una dimensionalidad enorme (más de 6000 variables en algunos casos), lo que complicaría la explicabilidad de los algoritmos, especialmente teniendo en cuenta el reducido tamaño muestral del que se disponía en este momento, e incluso del que se podría esperar disponer en el futuro.

Las variables que incluye eGeMAPS parten de 16 Descriptores de Bajo Nivel, como se les llama en la literatura. Esto es, propiedades acústicas del audio, como el tono, volumen, pendiente espectral etc. A estos descriptores se les aplican funciones, que permiten indexarlos a través del audio (media, coeficiente de variación, rango percentil etc). A petición de la organización, se hace una breve descripción del significado de cada variable, y sus correlatos perceptivos en los casos en los que esto tiene sentido (ver documento 'Descripción variable eGeMAPS.pdf' en Github).

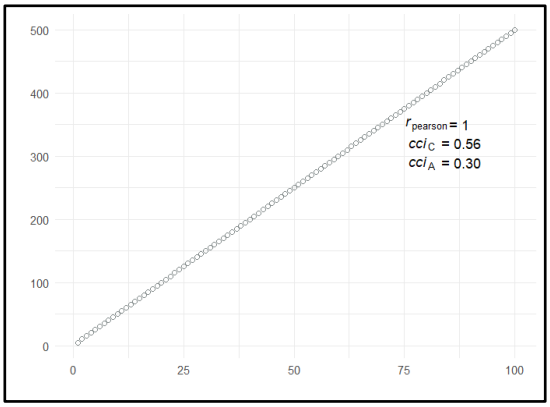
Semanas 5 y 6: 30/10 - 10/11

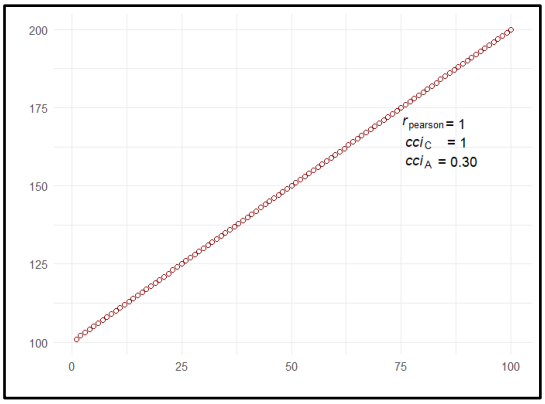
Cumplidos los objetivos principales, pasamos al desarrollo del 'estudio piloto', en el que nos plantemos distintas preguntas de cara a la futura implementación de AVIs.

La primera de estas preguntas es: ¿son estables las variables extraídas, al menos a nivel intra-sesión? Es decir, ¿tendrá una misma variable valores similares a lo largo de distintos segmentos de un mismo audio?. Esto es importante, en tanto que si no tenemos ciertas garantías de la estabilidad temporal de las variables, difícilmente se podrán usar como predictores válidos en un modelo. Nos acercamos, por tanto, al concepto de fiabilidad, si bien sería adecuado estudiar la estabilidad de estas variables a lo largo de distintos días y en distintos vídeos, o dicho de otra manera, a nivel inter-sesión.

Dado que la duración de cada vídeo es distinta, se decide tomar como material de referencia un tramo de 1 minuto para cada respuesta: del segundo 10 al 70. Este tramo es divido en dos segmentos de igual longitud (30s). Para ver el grado de acuerdo entre ambos segmentos, se decide emplear el Coeficiente de Correlación Intraclase de dos vías de efectos mixtos, en su versión de acuerdo absoluto con múltiples jueces. Siguiendo la notación de Shrout y Fleiss (1979), se usará la expresión $CCI_{(3,\,k)}$ en adelante, si bien $k = 2$ a lo largo de todo el trabajo.

Este coeficiente representa la proporción de varianza total que debida a los sujetos. Se ha escogido frente a, por ejemplo, el coeficiente de correlación de Pearson, en base a McGraw y Wong (1996). A modo de resumen, las razones son: que este coeficiente asume que las variables estudiadas pertenecen a la misma clase (sus varianzas y su métrica son similares), y que, en su versión de acuerdo absoluto, tiene en cuenta la varianza que corresponde al efecto de los jueces (segmentos, en este caso), que en otros indicadores de correlación se considera una fuente de variación irrelevante. Como consecuencia de esto, cualquier diferencia en valor absoluto entre las medidas emparejadas reduce el valor del coeficiente. Los efectos prácticos de estas diferencias entre coeficientes quizás se entiendan mejor con un ejemplo:





Ambos gráficos muestran asociaciones lineales perfectas. En el de la izquierda, disponemos de una variable uniforme *X* con valores entre 0 y 100, y una variable *Y* tal que $Y = 2X - 30$. Al ser *Y* una transformación lineal de *X*, el coeficiente de correlación de Pearson vale 1. Sin embargo, el CCI por consistencia toma un valor de 0.56, al ser la desviación típica de *Y* el doble de la de *X*. El CCI por acuerdo absoluto, que vale 0.30, es aún menor al incluir en el denominador la varianza correspondiente a las columnas. En el gráfico derecho, la variable *X* la misma, y la variable *Y* resulta de sumarle una constante de 80. En este caso, además de ser *Y* una transformación lineal de *X*, es una transformación puramente aditiva. En este caso, el CCI por consistencia juzga que ambas variables están en perfecto acuerdo, mientras que el CCI por acuerdo absoluto sigue siendo sensible a las diferencias entre los valores emparejados, lo cual resulta de interés especial para nuestros propósitos.

Este coeficiente puede basarse en un único juez/segmento, o en la media de los *k* segmentos. Se elige trabajar con el basado en la media, para extrapolar los resultados al que sería el coeficiente basado en un único segmento del doble de longitud.

Partiendo de la lógica del ANOVA, podemos obtener el $CCI_{(3,\,k)}$ muestral como:

$$CCI_{(3,k)} = \frac{MS_R - MS_E}{MS_E + \frac{MS_C - MS_E}{n}}$$

Donde MCS es la media cuadrática por sujetos, MCE la media cuadrática del error, y MCJ la media cuadrática por segmentos.

Para facilitar la comparación de resultados con Stegmann et al. (2020), se obtiene también el Coeficiente de Variación Intrasujeto (CVI). Mientras que el CCI tiene en cuenta tanto la varianza intersujeto como la intrasujeto, el CVI tan solo toma en cuenta la varianza intrasujeto. Es decir, nos informa de cómo las sucesivas medidas de una persona se parecen entre sí, en

relación a su media (Quan y Shih, 1996). Esto puede ser relevante también de cara a la comparación de los resultados actuales con otros basados en una muestra más heterogénea. Así, por ejemplo, si en un futuro estudio se dispone de una muestra perteneciente a una población con mayor variabilidad intersujeto en las variables de interés, manteniendo la variabilidad intrasujeto, el valor del CCI sería mayor, mientras que el CVI se mantendría igual. El cálculo de esta medida se omite en variables que pueden tomar tanto valores negativos como positivos, dado que la interpretación en estos casos pierde su sentido.

Una vez escogidos los análisis a realizar, se escribe un script en R para realizarlos. La interpretación del CCI como medida de buena o mala fiabilidad debe basarse en su contexto de aplicación. Basándose ambos estudios en la guía genérica de Koo y Li (2016), Stegmann et al., (2020) consideran que la estabilidad es buena si el CCI toma valores superiores a 0.75, mientras que Feng et al., (2024) toman como referencia el punto de corte 0.5. Nuestros resultados no serán exactamente comparables, dadas las diferentes características de las tareas empleadas, el uso de distintos sets de variables acústicas, y la ausencia en los estudios anteriores del uso del contraste de hipótesis para saber si los coeficientes obtenidos son significativos.

Se utiliza la librería *'irr'* de R para obtener el $CCI_{(3,\,k)}$ y se realiza, para cada variable, un contraste de hipótesis con tal que $H_0 : \rho_{CCI_{(3,k)}} \leq r_0$, donde r_0 puede tomar los valores 0.5, 0.75 y 0.9. Esto da lugar a un total de 244 contrastes. Para abordar el problema de las comparaciones múltiples se utiliza el procedimiento de Benjamini-Hochberg, que se basa en el control de la tasa de falsos descubrimientos (Benjamini y Hochberg, 1995), que en este caso fijamos a un máximo de 0.05. Debido al escaso tamaño muestral sabemos que la potencia será muy escasa, por lo que los resultados son, si acaso, orientativos, y deben replicarse en mejores condiciones.

Semana 7: 13/11 - 24/11

En esta semana surge la siguiente pregunta de investigación: ¿afectará la longitud del segmento analizado a los resultados anteriores? Los análisis previos reportan una medida de la estabilidad de la media de dos segmentos de 30 segundos. Sin embargo, pongamos que en un contexto aplicado se desea analizar datos que provienen de una respuesta de duración mucho menor. ¿Seguiremos obteniendo los mismos resultados? Se hipotetiza que, a mayor longitud del audio, mayor será la estabilidad de la variable extraída.

Esta pregunta puede ser relevante en el contexto aplicado, si bien nuestro material puede no ser del todo apropiado para estudiarla. Dunlop et al., (2020) muestran que, en promedio, los

proveedores de AVIs permiten respuestas de una duración de hasta 2 minutos (si bien la variabilidad es alta), y que los participantes suelen aprovechar entre el 57% y el 79% del tiempo disponible (aunque nótese que su estudio se limita a Australia). Las longitudes de audio que sería más interesante estudiar, por tanto, son mayores a la del material que disponemos, ya que parece que, en una situación real, habría pocos participantes que emitan respuestas mucho menores a 30 segundos. Aun así, se realiza una exploración piloto con el material disponible.

El único que estudio que se encontró de una temática parecida fue Almaghrabi et al. (2022), que encuentra que la reproducibilidad de un número importante de variables disminuye de forma relevante al reducir la duración de los segmentos analizados de 60s a 30s.

Para explorar esta pregunta con nuestro material, se repite el procedimiento anterior con los siguientes niveles de duración del segmento: 5s, 10s, 15s, 20s, 25s y 30s. Se mantiene constante *k* (número de jueces/segmentos) para todos los niveles de la variable independiente, y se realiza nuevamente un contraste de hipótesis para cada variable, fijando r0 a 0.75, y utilizando de nuevo el método de Benjamini-Hochberg para mantener la tasa de falsos descubrimientos inferior o igual a 0.05.

Semanas 8 y 9: 13/11 - 24/11

Alrededor de esta semana se tiene una tutoría con el tutor académico, que sirve para orientar los próximos objetivos.

Se plantea por primera vez la meta de hacer una reducción de dimensionalidad de las variables predictoras. Si bien los datos de los que se dispone probablemente no permitirán alcanzar una solución estable, se considera que puede ser un objetivo que merezca la pena empezar a explorar, de cara a futuro. Esto es por tres razones principales: extraer factores interpretables puede mejorar la aceptabilidad de un producto basado en variables acústicas, puede ser útil para estudiar los mecanismo que puedan explicar las asociaciones que se presumen, y utilizar un número reducido de variables como predictores puede llegar a ser deseable, especialmente dada la dificultad de conseguir datos anotados en este contexto, así como de cara a trabajar con modelos más parsimoniosos.

Se comienza realizando una exploración de la literatura. Los únicos artículos que, a conocimiento del autor de la memoria, hay al respecto son Feng et al. (2024) y Van Rijn y Larrouy-Maestri (2023). En ambos casos se detectan decisiones discutibles a nivel metodológico, desde un tamaño muestral muy escaso (aunque no somos quién para criticar en ese aspecto) a confusiones entre el ACP y el AFE, y la extracción de factores débiles y de

dudosa interpretabilidad. Llama la atención que, si bien puede ser comprensible la dificultad para obtener datos adecuados, no se discutan las limitaciones que ello tiene en un contexto como este, en el que muchas veces $p \gg n$.

Una reflexión que surge a raíz de este proceso es que, de hecho, las variables de eGeMAPS (y de cualquier set de parámetros acústicos) parecen mostrar elevada redundancia entre sí, a nivel conceptual. Se decide que, de forma previa al Análisis Factorial, es necesario hacer una selección de las variables predictoras que se emplearán, para evitar estimaciones incorrectas de los parámetros.

De forma previa a cualquier otra prueba, se decide eliminar las siguientes variables:

- El rango del percentil 20 al 80, al tener ya cada uno de esos percentiles por separado.
- Aquellas variables referidas al percentil 50, ya que muestran tener una correlación superior a 0.99 con aquellas referidas a la media.
- Todas aquellas variables calculadas sobre segmentos sin voz, ya que no son relevantes, excepto aquellas referidas a la duración de los mismos y a su número.

Se plantea como objetivo, además, encontrar un método estadístico que informe de la posible redundancia que exista en el set de datos. Por su fácil interpretación (y en parte, por interés del alumno sobre el mismo), se elige emplear el Unique Variable Analysis (UVA) de Christensen et al. (2023), proveniente de la perspectiva de redes psicométricas.

En una red psicométrica al uso, cada variable se representa mediante un nodo, y sus uniones representan asociaciones únicas (correlaciones parciales) entre ellas (Christensen et al., 2020). Cada uno de estos nodos, si se quiere considerar un componente del sistema, debe ser 'causalmente autónomo', en tanto que su patrón de relaciones con el resto de nodos del sistema debe ser único. En cambio, si dos variables muestran un patrón de relaciones similar, pueden considerarse como dos expresiones diferentes de un mismo componente y, por tanto, redundantes (Cramer et al., 2012). Una métrica que nos informa del grado en el que el patrón de relaciones de dos nodos de la red es similar es el Weighted Topological Overlap (wTO), que pone en relación el conjunto de pesos compartidos por dos nodos con el total de pesos que conectan a uno de los nodos con la red. En concreto, se calcula como:

$$\omega_{ij} = \frac{\sum_u a_{iu} \cdot a_{uj} + a_{ij}}{\left\{k_i, k_j\right\} + 1 - a_{ij}}$$

Representando $\sum_u a_{iu} \cdot a_{uj}$ la suma de las conexiones que comparten ambos nodos, a_{ij} la conexión entre ambos nodos, y k la suma de todas las conexiones de un nodo.

El método UVA consiste en establecer una red psicométrica con el conjunto de variables y emplear el wTO para analizar la redundancia de las variables. Christensen et al. (2023) comprueban que, en comparación con otros métodos contemporáneos, resultó el mejor método para la detección de dependencia local en multitud de condiciones (número de factores, tipo de datos, variables por factor, etc). Siguiendo su recomendación, se empleó el valor 0.25 como punto de corte para determinar que dos variables pueden ser redundantes.

Para establecer la red psicométrica, es necesario obtener la matriz de correlaciones parciales de las variables. Esta matriz se estima empleando el método de regularización *EBICglasso*, que en primer lugar introduce un parámetro λ que busca penalizar la log-verosimilitud en tanto que la suma de los coeficientes estimados sea mayor (Epskamp y Fried, 2018). Generalmente se estiman varias redes, desde una totalmente conectada a una totalmente desconectada, según varía el valor de λ, y se busca minimizar el EBIC para elegir un modelo. Si se desea profundizar en el proceso que se emplea para seleccionar los parámetros λ del LASSO y γ del EBIC pueden consultarse Tay et al. (2023) y Christensen et al. (2023), respectivamente. Un supuesto de este procedimiento es que las variables empleadas sigan una distribución normal multivariante. Cuando esto no ocurre, Epskamp y Fried (2018) recomiendan aplicar previamente una transformación no-paranormal, que enlaza cada valor único de una variable con un valor único de una variable normal estándar, que se asume es latente a la observada, a partir de las distribuciones cumulativas de ambas variables. Para más información sobre este proceso, ver Liu et al. (2009).

Para comprobar si se cumple esta hipótesis, se empleó un test de Royston (recomendado por Mecklin y Mundfrom (2005) en tamaños muestrales pequeños, si bien el test de Henze-Zirkler podría haber sido más apropiado, pero no se exploró por falta de tiempo). Se rechazó la hipótesis de normalidad, por lo que se aplicó la transformación no-paranormal de forma previa a la estimación de la red. Para comprobar que, efectivamente, el patrón de correlaciones se mantiene, se empleó un test de Mantel, que ordena por pares los coeficientes de las matrices de correlación de ambos conjuntos y calcula la correlación entre ellos. El coeficiente obtenido se compara con una distribución de coeficientes obtenidos de manera similar, previa permutación de una de las matrices empleadas, para obtener un indicador de la significación estadística (Bakker, J. D., 2024, capítulo 18).

Los análisis se realizan en R, usando los paquetes '*MVN'*, '*huge'*, '*vegan*', y '*EGAnet'*,

Semanas 10 y 11: 27/11 - 15/12

Se decide continuar con el objetivo de realizar un Análisis Factorial para explorar la estructura interna de las variables.

El tamaño muestral recomendado para la realización de un Análisis Factorial varía según condiciones como las saturaciones de las variables o el número de indicadores por factor (véase Lloret-Segura et al., 2014), si bien, en general, el empleo de esta técnica en muestras pequeñas está completamente desaconsejado (algo que, como ya se ha comentado, no ha impedido que se utilice en este área de investigación). Ferrando y Anguiano-Carrasco (2010), por poner un ejemplo, recomiendan un tamaño muestral mínimo de 200 personas incluso en condiciones 'óptimas'. Por supuesto, algo así está fuera de nuestro alcance ahora mismo, por lo que es necesario operar recordando la filosofía inicial con la que se plantearon las prácticas: hacer un ejercicio que tiene como principal objetivo transmitir conocimiento práctico a la organización, sobre cómo operar al disponer de datos de calidad.

De Winter et al. (2009) exploran la estabilidad de los pesos factoriales en diversas condiciones (número de variables, saturaciones, número de factores) y con un tamaño muestral muy escaso, tanto en estructuras factoriales simples como en estructuras afectadas por diversos tipos de distorsión. A pesar de que sus resultados no son exactamente trasladables a nuestro contexto (no estudian más de una 'distorsión' a la vez, mientras que nosotros si las encontramos), son usados para guiar la interpretación.

Para decidir el número de factores a extraer se emplea el método de Análisis Paralelo a partir de la matriz R no reducida (Lim y Jahng, 2019), comparando los autovalores empíricos con la media y con el percentil 95 de los simulados. Aunque, el Análisis Paralelo es considerado uno de los criterios por excelencia para decidir el número de factores a extraer, es sensible al tamaño muestral, y puede infraestimar el número de factores cuando estos están altamente correlacionados. Auerswald y Moshagen (2019) muestran que el Análisis Paralelo es el que mejor funciona a lo largo de una gran variedad de condiciones, y es más resistente que otros a la presencia de factores menores, algo que esperamos que se de en nuestro caso. Por aprender un método no visto en el Máster, probaremos también el método Comparison Data, que en el estudio de Auerswald y Moshagen (2019) fue más sensiblea la presencia de factores menores. Este método comienza formando una matriz poblacional con una estructura factorial conocida (comenzando por un factor) que reproduce la distribución multivariada de los datos originales.

De esta matriz, se seleccionan N casos y se calculan sus autovalores. Posteriormente, se obtiene la raíz del error cuadrático medio respecto de los autovalores reales. Este proceso se repite 500 veces, para cada par de número de factores evaluados, obteniendo 2 distribuciones de RMSEs; por ejemplo, una calculada basándonos en una estructura de 1 factor, y otra basándonos en una estructura de 2 factores. Las dos distribuciones se comparan empleando un test U de Mann-Whitney, repitiendo el proceso hasta que aumentar el número de factores no resulte en una distribución de RMSEs menor (Ruscio y Roche, 2012).

En cualquier caso, es necesario tener en cuenta que elegir el número correcto de factores será muy desafiante con la escasa muestra disponible (De Winter et al., 2009).

Para la extracción de los factores se emplea el método Unweighted Least Squares (ULS) por ser más apropiado para tamaños muestrales pequeños y no tener supuestos sobre la distribución de los datos (Lloret-Segura et al., 2014), y rotación Oblimin para permitir a los factores correlacionar.

Antes de continuar con el Análisis Factorial, se decide realizar bootstrap no paramétrico con 500 réplicas, tanto en las técnicas para decidir el número de factores como en el Análisis Factorial, de cara a evaluar la precisión de los parámetros estimados. La utilidad de esta técnica será limitada, de nuevo, dada la inadecuación de la muestra (ver Osborne, 2014, capítulo 5), pero si observamos intervalos de confianza relativamente estrechos, obtendremos evidencias a favor de la estabilidad de la solución en muestras con características similares . Realizar bootstrap para las saturaciones factoriales puede ser problemático, en tanto que cada réplica hay una cierta arbitrariedad en el orden de los factores y el signo de las saturaciones. La librería '*psych*' resuelve este problema rotando las soluciones por bootstrap hacia la solución original (Revelle, 2024).

Todos los análisis estadísticos se realizan con el software R, en concreto las librerías: '*boot*', '*psych'*, '*EFAtools'*, '*EFA.dimensions*'.

Semana 12: 18/12 - 21/12

Esta semana se realizó una presentación al equipo de Talent del IIC sobre las actividades desarrolladas durante las prácticas (hasta el momento, las reuniones habían sido casi en exclusiva con el tutor). Con el objetivo de hacer algo más interesante la presentación, se acordó con el tutor un nuevo objetivo: explorar relaciones entre las variables acústicas tratadas hasta ahora y las puntuaciones de los sujetos en cada uno de los rasgos del Big Five, a través del cuestionario MINI IPIP (Donnellan et al., 2006). No se usaron las puntuaciones factoriales

porque no se tuvo confianza en las estimaciones del AF. El método usado fue un análisis de correlaciones de Pearson.

Poco después de finalizar las prácticas se envía un *abstract* describiendo las actividades realizadas al congreso ITC de 2024, siendo este aceptado para su presentación en póster.

La siguiente tabla representa el desarrollo de las prácticas de acuerdo a los objetivos descritos en la Introducción:

	Semana											
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Objetivo 1	✓	✓	✓									
Objetivo 2				✓								
Objetivo 3					✓	✓						
Objetivo 4							✓					
Objetivo 5								✓	✓			
Objetivo 6										✓	✓	
Objetivo 7												✓

Resultados y Discusión

A continuación se presentan los resultados de los análisis realizados, por orden cronológico. Dado que las tablas de resultados al completo ocuparían alrededor de la mitad de la extensión máxima, se han incluido en el documento 'Tablas Material Suplementario.pdf', disponible en Github.

Coeficientes de Correlación Intraclase: primera prueba

Se reporta en primer lugar el número de variables que superan el punto de corte establecido, del mismo modo que se hace en la literatura previa:

Tabla 1. Número de variables cuyo $CCI_{(3,k)}$ no supera r_0

Punto de corte	Nº de features
$\boldsymbol{r_0 = 0.5}$	9 (13)
$\boldsymbol{r_0 = 0.75}$	13 (20)
$\boldsymbol{r_0 = 0.9}$	21 (39)

Nota: Entre paréntesis se incluyen los mismos resultados, aplicando contraste de hipótesis.

También se incluye el CCI obtenido, el intervalo al 95% de confianza, y los valores del CVI, para cada variable:

Tabla 2. Ejemplo de reporte de resultados CCI y CVI

Variable	ICC	ICC_l	ICC_u	WSCV
F0semitoneFrom27.5Hz_sma3nz_amean	0.99	0.97	0.99	0.02

Nota: ICC =Intraclass Correlation Coefficient, ICC_l = límite inferior ICC, ICC_u = límite superior ICC, WSCV = Within-Subjects Coefficient of Variation. Los nombres de las variables son los asignados por defecto por openSmile.

La comparación con resultados de diferentes estudios es difícil, ya que se emplean distintos sets de parámetros, no se reporta la forma de obtenerlos, y el número de variables suele ser muy grande para detenerse individualmente en cada una. El reciente trabajo de Feng et al. (2024) encuentra también que resultan estables las siguientes variables que comparte nuestro estudio: promedios de la frecuencia fundamental (F0), los formantes 1 y 2, la fluctuación del retardo (*jitter*), el *shimmer,* el Harmonics to Noise ratio, la duración de las pausas; y desviaciones típicas de F0 y las pausas, en tareas intrasesión. Aunque esto no se cumple para todas las variables, otros estudios confirman parcialmente estos resultados en diseños intersesión: véase Afacan et al., (2022), para la fluctuación del retardo y el *shimmer*; González et al., (2002) para parámetros relacionados con la frecuencia fundamental, la amplitud y sus variaciones, entre otros; o Almaghrabi et al. (2022), para los MFCC, parámetros espectrales en general, y promedio y variación de los dos primeros formantes.

Coeficientes de Correlación Intraclase: segunda prueba

Se reporta ahora una tabla similar a la Tabla 1, pero tomando como punto de corte $r = 0.75$, y variando la longitud del segmento empleado:

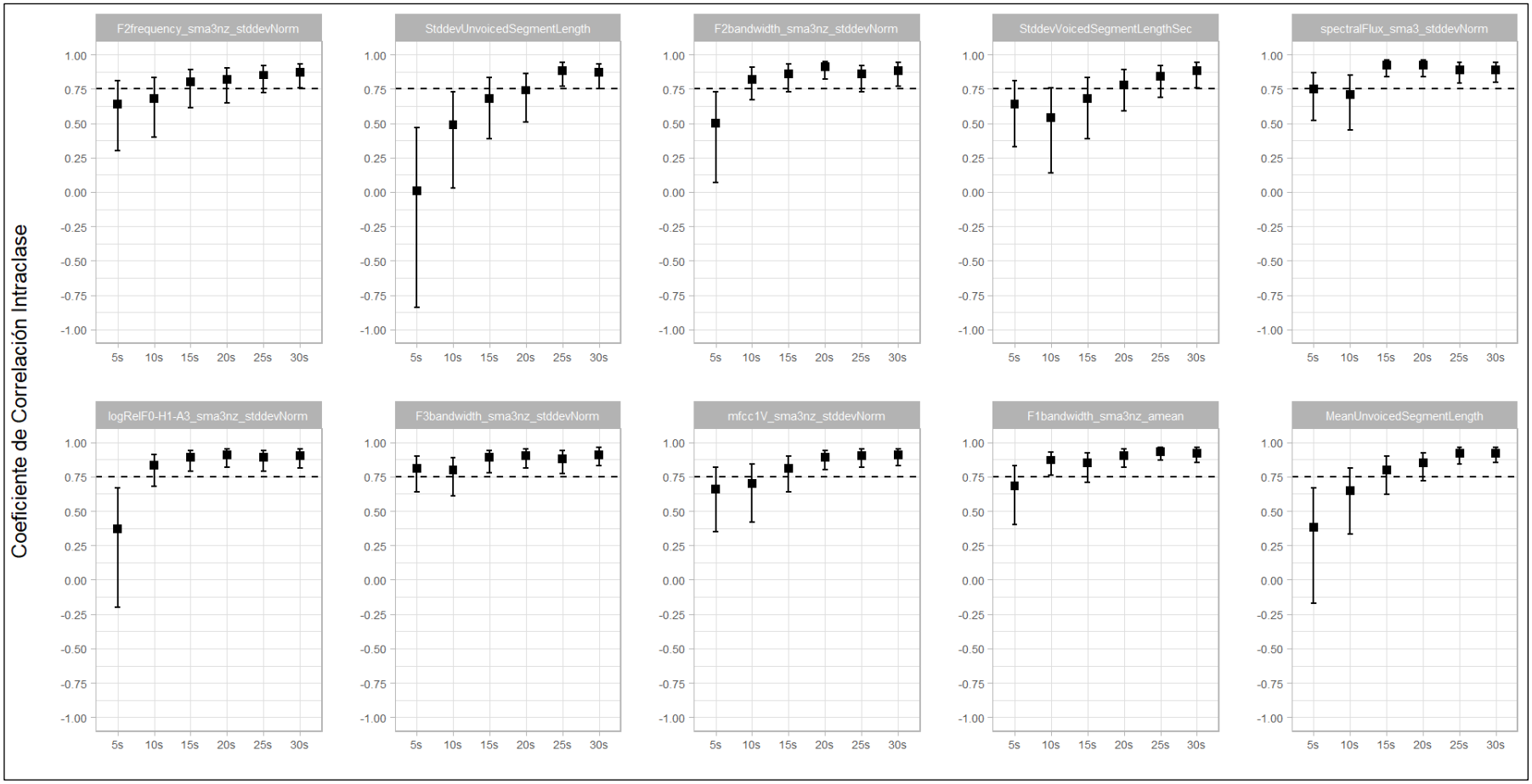
Tabla 3. Número de variables cuyo $CCI_{(3,k)}$ no supera 0.75.

Duración del segmento	Número de variables con $CCI < 0.75$
5s	41 (64)
10s	25 (48)
15s	13 (32)
20s	15 (24)
25s	14 (21)
30s	13 (20)

Nota: Entre paréntesis se incluyen los resultados al aplicar contraste de hipótesis.

En general el número de variables que resultan estables aumenta al incrementarse la duración del segmento seleccionado, si bien entre los 15 y los 30 segundos la magnitud de las diferencias

Figura 1. Ejemplos de evolución del CCI según la longitud de los segmentos



comienza a descender. Estos resultados contrastan con los de Almaghrabi et al. (2022) que, trabajando con el CCI muestral, encuentran grandes diferencias al pasar de segmentos de 30 segundos a segmentos de 60 segundos, incluso a pesar de trabajar de forma similar (empleando un único audio por sujeto, dividiéndolo en segmentos y calculando el ICC entre estos segmentos). Es posible que, aunque las diferencias de 5 segundos con las que trabajamos nosotros sean lo bastante pequeñas parano producir un efecto en el CCI muestral, doblar la longitud del audio sí que produzca resultados visibles. La Figura 1 puede ayudar a visualizar mejor el cambio en el CCI al aumentar la duración.

Unique Variable Analysis

El test de Royston devuelve un estadístico $H = 150.78$ $(p < 0.0001)$, con lo que rechazamos la hipótesis nula de normalidad multivariante. Se realiza la transformación no paranormal, tras la cual se repite el test, obteniendo un estadístico $H = 92 \cdot 10^{-8}$ $(p \approx 1)$.

Por último, se reportan los wTOs de aquellos pares en los que este es superior a 0.25:

Tabla 3. Ejemplo de reporte de resultados wTO

Variable 1	Variable 2	wTO
mfcc3_sma3_amean	mfcc3V_sma3nz_amean	0.61

Se utiliza el criterio de eliminación que implementan Golino y Christensen (2024): en pares de variables, se elimina aquella variable con mayor wTO máximo con cualquier otra variable del set que no sea aquella con la que es redundante. Si tres o más variables son redundantes entre sí, se retiene aquella con mayor wTO promedio con todas las que son redundantes entre sí.

Existen otras variables que podrían ser eliminadas siguiendo este criterio pero que, sin embargo, para el estudiante en prácticas, parecen representar conceptos sustantivos distintos. Surge el dilema entre retener las variables, sabiendo que hacerlo conlleva a un error metodológico (la matriz de covarianzas no es positiva definida, y los pesos estimados serán más inestables), o seguir las recomendaciones del UVA, eliminar variables, y arriesgarnos a tener una mala representación del problema. Se decide tomar la primera decisión, bajo la idea de que es mejor cometer un error conocido que uno cuyas consecuencias no sabremos detectar, y de que es mejor no olvidar variables que podrían ser útiles posteriormente. En total, retenemos 50 variables de las 88 originales.

Análisis Factorial Exploratorio

Se presentan en primer lugar los resultados para decidir el número de factores.

Tabla 4. Proporción de réplicas que recomiendan cada número de factores

Método	Nº factores recomendado										
	4	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***	***13***	***14***
AP 0.5	0.012	0.310	0.620	0.056	0.002	0	0	0	0	0	0
AP 0.95	0.060	0.484	0.442	0.014	0	0	0	0	0	0	0
CD	0.008	0.048	0.170	0.208	0.178	0.154	0.126	0.088	0.010	0.008	0.002

Vemos que cualquiera de los métodos ofrece resultados inestables, en tanto que pueden hay, en todos los casos, un porcentaje relevante de soluciones que no coinciden con la más frecuente. Esto es particularmente llamativo en el caso del método Comparison Data, que ofrece soluciones en un rango de 4 a 14 factores. Merece la pena mencionar también que este último método tuvo un coste computacional mucho mayor que el Análisis Paralelo, requiriendo más 1 minuto por réplica.

Ante la petición de continuar con el Análisis Factorial, se decidió extraer una solución de 6 factores, que explicaron un 73% de la varianza total. La matriz de saturaciones con los la media y el rror típico obtenidos a través del bootstrap, pueden consultarse en el documento 'Tablas Material Suplementario.pdf' de Github. 41 de las soluciones obtenidas por bootstrap contuvieron al menos un caso Heywood. Si bien no se exploró en detalle por qué pudo ocurrir esto, no es algo que sorprenda, sabiendo que, en general, la prevalencia de casos Heywood es mayor cuando el tamaño muestral es insuficiente (Cooperman y Waller, 2022) Se puede ver también que los errores típicos son, a menudo, inaceptables, siendo mayores que el propio estimador. Dada la dificultad de interpretar la solución, tanto en términos sustantivos como psicométricos, se decidió detener aquí el trabajo relacionado con este objetivo.

Correlaciones entre variables acústicas y rasgos de personalidad

La única correlación significativa tras corregir el valor *p* fue:

Variable	Extraversión	Amabilidad	Responsabilidad	Estabilidad	Apertura
F2amplitudeLogRelF0_ sma3nz_stddevNorm	-0.01	0.23	0.07	0.28	0.45*

**Significativo para* $\alpha = 0.05$ *tras corregir el valor p por el método de Benjamini-Hochberg.*

Si bien debe notarse la escasísima potencia del análisis, teniendo en cuenta que el tamaño muestral es de 42 y el número de variables de 50, y se calculan 5 correlaciones por

variable. Se observaron algunas correlaciones elevadas (pero no significativas) que intuitivamente parecen tener sentido (mayor volumen, y variabilidad del mismo, asociados a mayor extraversión, por ejemplo). Sin embargo, es necesario repetir los análisis con una mayor muestra antes de extraer conclusiones.

Conclusiones y Recomendaciones

El trabajo realizado ha tratado sobre aspectos relacionados con el tratamiento de variables acústicas. Juzgar su importancia o sus consecuencias es complicado en el contexto en el que nos encontramos. Por un lado, se ha conseguido proveer al equipo de Talento de una mayor independencia a la hora de trabajar en un futuro con este tipo de datos, a través del código que automatiza el procesamiento de los audios, y se ha promovido la discusión de cuestiones prácticas, como son la longitud de la AVI o la redundancia de los predictores. Es importante promover la involucración de psicólogos y psicómetras en el desarrollo de productos de este tipo, y dotarles de herramientas para ello, ya que a menudo somos los únicos preocupados por cuestiones como la fiabilidad o la validez. Aunque esto parezca una obviedad, en mi impresión (puramente personal), parece que estas tareas suelen encargarse a científicos o analistas de datos, dejando al equipo de psicólogos relegado a un segundo plano. Por ello creo que este primer paso puede ser muy importante, si bien es muy pronto para saber si lo será realmente.

Por otro lado, se han investigado algunos aspectos relacionados con la estabilidad y la interpretación de las variables prosódicas y espectrales. Los trabajos mencionados a lo largo de esta memoria demuestran que existe interés por estas cuestiones, si bien la investigación es escasa. El trabajo realizado llama la atención sobre cuestiones que no parecen haberse planteado previamente y presenta mejoras metodológicas respecto a trabajos previos. Ninguna de estas aportaciones es original en el ámbito de la metodología en general, pero sí que parecen serlo dentro de este contexto. Por último, la futura presentación del trabajo, como póster, supone un resultado tangible de estas prácticas.

Ahora bien, este trabajo ha presentado muchas limitaciones que no se pueden olvidar, y que han marcado su desarrollo. La primera de estas es la falta de tiempo, que ha provocado que en algunos momentos se avance de una forma que percibí como acelerada o ‘improvisada’. Una segunda limitación, incluso más preocupante, ha sido la falta de un volumen de datos adecuado para responder a las preguntas de investigación planteadas

(si bien este problema parece común, en general, a casi toda la literatura revisada). Esto se ha resuelto asumiendo desde el principio que las conclusiones que extrajéramos serían provisionales y limitadas, repitiendo los análisis con un mayor volumen de datos en cuanto sea posible. Finalmente, es innegable que faltaba en el equipo una persona con conocimientos especializados en acústica humana, que guiara la investigación y ayudara a interpretar los resultados.

Estas limitaciones han afectado también al contenido de esta memoria de prácticas. Puede haberse notado que, más allá de comparaciones con otros estudios, apenas se han ofrecido interpretaciones sustantivas de los resultados, y las tablas se han omitido (aunque pueden consultarse online). En gran medida esto se explica por la limitada extensión de la memoria, y por el hecho de haber trabajado con un número tan elevado de variables. Pero también es porque tiene poco sentido detenerse a interpretar un Análisis Factorial o una red psicométrica, partiendo de una muestra de apenas 40 casos, y sin contar con el conocimiento sustantivo necesario para ello.

Entre las recomendaciones hechas a la organización, estuvo la de emplear audios de alguna base de datos pública, para aumentar la muestra en aquellas pruebas que no requieran trabajar con el criterio. De hecho, se identificaron algunas candidatas (Mozilla Common Voice, LibriSpeech), pero finalmente no se avanzó por esa ruta. También se comentaron aspectos como la calidad de los datos empleados durante el entrenamiento, en particular en lo relativo al control de variables como el sexo y la edad, que sin duda se relacionan con la voz. Estos factores deberían ser incluidos como covariables, algo que, aunque pueda parecer obvio, en el caso de la edad al menos no parece haberse discutido en la literatura. Otro aspecto que habría sido interesante de explorar es el impacto de la calidad del audio analizado (especialmente importante, al ser las AVIs usualmente grabadas a través de un equipo personal), pero el tiempo no lo permitió. En cualquier caso, la principal recomendación fue no fiarse en exceso de los resultados obtenidos, y repetir los análisis en el fututo

Valoración final

A nivel personal, he valorado de forma bastante positiva la experiencia de prácticas. Me ha gustado verme obligado a trabajar con Python y Jupyter Notebooks, y he sentido que he mejorado en mucho mis habilidades de programación con R. También ha sido muy interesante poder leer sobre temas como el análisis de redes psicométricas, o el bootstrap,

si bien es innegable que no ha sido posible profundizar demasiado en ellos. En cualquier caso, la sensación final que me llevo es de una mayor capacidad de autosuficiencia para aprender cosas nuevas.

Como lado negativo, me habría gustado trabajar en tareas más 'típicas' en psicometría o *HR Analytics*, como podría ser el ajuste de modelos de TRI o trabajar en problemas de predicción de rotación, desempeño, etc. No me quejo, porque el trabajo que he realizado me ha gustado, pero es cierto que no ha sido lo que esperaba cuando vi la descripción de la oferta de prácticas. Por otro lado, las limitaciones que ya he comentado han hecho que, a veces, el trabajo realizado haya sido poco satisfactorio. Un tratamiento adecuado de este problema requiere mucho más tiempo, recursos y conocimientos de los que se han podido disponer aquí, y pienso que podría haber alcanzado resultados más interesantes de haber dispuesto de ello. En cierto modo, la mayor aportación que he podido hacer es la de recomendar escepticismo ante las conclusiones que se puedan extraer al no disponer datos de calidad y falta de conocimiento experto. Si algo nos permite aportar valor como metodólogos frente a otros perfiles con los que trabajamos es, además de nuestro conocimiento en psicometría, nuestra preocupación por cuidar la validez de las interpretaciones que realizamos.

A pesar de todo esto, como primera toma de contacto con el mundo empresarial, la experiencia ha sido muy positiva.

*(*Nota: en la actualidad el IIC dispone de un volumen de datos de respuestas a AVIs mucho mayor al disponible durante la realización de las prácticas, si bien el alumno, ahora miembro del equipo, aún no ha podido disponer de ellos. Por ello, se espera poder repetir y mejorar los análisis aquí realizados, aplicando así el conocimiento obtenido durante el periodo de prácticas).*

Referencias

Almaghrabi, S. A., Thewlis, D., Thwaites, S., Rogasch, N. C., Lau, S., Clark, S. R. y Baumert, M. (2022). The reproducibility of bio-acoustic features is associated with sample duration, speech task, and gender. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 167-175.

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods, 24*(4), 468–491. https://doi.org/10.1037/met0000200

Back, M. D. y Nestler, S. (2016). Accuracy of judging personality. En Hall, J. A., Schmid, M. y West, T. V. (Eds.), The social psychology of perceiving others accurately. Cambridge University Press.

Bakker, J., D. (2024). *Applied Multivariate Statistics in R*. University of Washington.

Chittaragi, N.B., Prakash, A., y Koolagudi, S.G. (2017). Dialect Identification Using Spectral and Prosodic Features on Single and Ensemble Classifiers. *Arabian Journal for Science and Engineering, 43*, 4289 - 4302. doi:10.1007/s13369-017-2941-0

Cho, S.-J., Li, F., & Bandalos, D. (2009). Accuracy of the parallel analysis procedure with polychoric correlations. *Educational and Psychological Measurement, 69*(5), 748–759. https://doi.org/10.1177/0013164409332229

Christensen, A. P., Garrido, L. E., y Golino, H. (2023). Unique Variable Analysis: A Network Psychometrics Method to Detect Local Dependence. *Multivariate Behavioral Research, 58*(6), 1165-1182. https://doi.org/10.1080/00273171.2023.2194606.

Christensen, A. P., Golino, H., y Silvia, P. J. (2020). A psychometric network perspective on the validity and validation of personality trait questionnaires. *European Journal of Personality*, *34*(6), 1095–1108. doi: 10.1002/per.2265

Cooperman, A. W., & Waller, N. G. (2022). Heywood you go away! Examining causes, effects, and treatments for Heywood cases in exploratory factor analysis. *Psychological Methods*, *27*(2), 156–176. https://doi.org/10.1037/met0000384

Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., Kendler, K. S., y Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, *26*, 414–431. https://doi.org/10.1002/per.1866

De Winter, J. C. F., Dodou, D., y Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research, 44*(2), 147–181. https://doi.org/10.1080/00273170902794206

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment*, *18*(2), 192–203. https://doi.org/10.1037/1040-3590.18.2.192

Dunlop, P. D., Holtrop, D., & Wee, S. (2022). How asynchronous video interviews are used in practice: A study of an Australian-based AVI vendor. *International Journal of Selection and Assessment*, 30, 448–455. https://doi.org/10.1111/ijsa.12372

Epskamp, S., y Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*(4), 617–634. https://doi.org/10.1037/met0000167

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., Devillers, L., Epps, J., Laukka, P., Narayanan, S. S., y Truong, K. P. (2016). The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. *IEEE Transactions on Affective Computing*, *7*(2), 190-202. https://doi.org/10.1109/TAFFC.2015.2457417

Eyben, F., Weninger, F., Wöllmer, M., y Schuller, B. (2022). openSmile audio feature extraction. Recuperado de: https://audeering.github.io/opensmile/

Feng, F., Zhang, Z., Tang, L., Qian, H., Yang, L.-Z. y Jiang, H., Li, H. (2024). Test-retest reliability of acoustic and linguistic measures of speech tasks. *Computer Speech and Language*, *83*, 101547. https://doi.org/10.1016/j.csl.2023.101547.

Ferrando, P. J. y Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en psicología. *Papeles del Psicológo*, *31*(1), 18-33.

Golino, H., & Christensen, A. P. (2024). EGAnet: Exploratory Graph Analysis – A framework for estimating the number of dimensions in multivariate data using network psychometrics. R package version 2.0.5. Retrieved from https://r-ega.net

González, J., Cervera, T., y Miralles, J. L. (2002). Análisis acústico de la voz: Fiabilidad de un conjunto de parámetros multidimensionales. *Acta Otorrinolaringológica Española*, *53*(4), 256-268. https://doi.org/10.1016/S0001-6519(02)78309-X.

Hall, J. A., Schmid, M. y West, T. V. (2016). Accurate interpersonal perception: Many traditions, one topic. En Hall, J. A., Schmid, M. y West, T. V. (Eds.), *The social psychology of perceiving others accurately.* Cambridge University Press.

Hirevalue (2018). The Ultimate Guide to Candidate Experience. Recuperado de: https://www.hrgrapevine.com/resources/register/the-ultimate-guide-to-a-candidate-experience

Hough, L. M., y Johnson, J. W. (2013). Use and importance of personality variables in work settings. En N. W. Schmitt, S. Highhouse, y I. B. Weiner (Eds.), *Handbook of psychology: Industrial and organizational psychology* (2ª Ed., pp. 211–243). John Wiley y Sons, Inc.

Hough, L. y Dilchert, S. (2017). Personality: Its measurement and validity in employee selection. En Farr, J. L. y Tippins, N. T. (Eds), *Handbook of Employee Selection* (2ª Ed., pp. 298-325). Routledge.

Huffcutt, A. I., Conway, J. M., Roth, P. L., y Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology, 86*(5), 897–913. https://doi.org/10.1037/0021-9010.86.5.897

Koo, T. K., y Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155–163. doi:10.1016/j.jcm.2016.02.012

Koutsombogera, M., Sarthy, P., & Vogel, C. (2020). Acoustic features in dialogue dominate accurate personality trait classification. *Proceedings of the 2020 IEEE International Conference on Human-Machine Systems, ICHMS 2020.* doi*:* 10.1109/ICHMS49158.2020.9209445

Lim, S., y Jahng, S. (2019). Determining the number of factors using parallel analysis and its recent variants. *Psychological Methods, 24*(4), 452–467. https://doi.org/10.1037/met0000230

Liu, H., Lafferty, J. y Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research, 10*(80), 2295–2328.

Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., y Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología, 30*(3), 1151-1169.

Lukacik, E.-R., Bourdage, J. S., y Roulin, N. (2022). Into the void: A conceptual model and research agenda for the design and use of asynchronous video interviews. *Human Resource Management Review, 32*(1), Article 100789. https://doi.org/10.1016/j.hrmr.2020.100789

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

Mecklin, C., y Mundfrom, D. (2005). A Monte Carlo comparison of the Type I and Type II error rates of tests of multivariate normality. *Journal of Statistical Computation and Simulation, 75*(2), 93–107. doi:10.1080/0094965042000193233

Mueller-Hanson, R., Heggestad, E. D., y Thornton, G. C. III. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*(2), 348–355. https://doi.org/10.1037/0021-9010.88.2.348

Nestler, S. y Back, M. D. (2013). Applications and extensions of the lens model to understand interpersonal judgments at zero acquaintance. *Current Directions in Psychological Science, 22(5),* 374-379. https://doi.org/10.1177/0963721413486148

Osborne, J. W. (2014). *Best Practices in Exploratory Factor Analysis*. CreateSpace Independent Publishing.

Polzhel, T. (2015). *Personality in Speech: Assessment and Classification.* Springer.

Quan, H., y Shih, W. J. (1996). *Assessing Reproducibility by the Within-Subject Coefficient of Variation with Random Effects Models. Biometrics, 52*(4), 1195. doi:10.2307/2532835

Rangra, K., Kadyan, V., & Kapoor, M. (2023). Exploring the role of prosodic features to perform personality classification from labelled speech data. En *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, 1527-1531.

Revelle, W. (2024). *psych: Procedures for Psychological, Psychometric, and Personality Research (Version 2.4.3)*. Retrieved from https://CRAN.R-project.org/package=psych

Ruscio, J., y Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment, 24*(2), 282–292. https://doi.org/10.1037/a0025697

Ryan, A. M., Inceoglu, I., Bartram, D., Golubovich, J., Grand, J., Reeder, M., Derous, E., Nikolaou, l., y Yao, X. (2015). Trends in testing: Highlights of a global survey. In I. Nikolaou y J. K. Oostrom (Eds.), *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice* (pp. 136–153). Routledge/Taylor y Francis Group.

Ryan, A. M., y Ployhart, R. E. (2014). A century of selection. *Annual Review of Psychology, 65,* 693–717. https://doi.org/10.1146/annurev-psych-010213-115134

Schuller, B., Steidl, S., Batliner, A., Nöth, E., Vinciarelli, A., Burkhardt, F., Son, R.v., Weninger, F., Eyben, F., Bocklet, T., Mohammadi, G., y Weiss, B. (2012). The INTERSPEECH 2012 Speaker Trait Challenge. En *Proceedings Interspeech 2012*, 254-257. doi: 10.21437/Interspeech.2012

Schuller, B., Steidl, S., y Batliner, A. (2009). The INTERSPEECH 2009 Emotion Challenge. En *Proceedings Interspeech 2009*, 312-315. doi: 10.21437/Interspeech.2009-103

Schuller, B., Vlasenko, B., Eyben, F., Wöllmer, M., Stuhlsatz, A., Wendemuth, A., y Rigoll, G. (2010). Cross-corpus acoustic emotion recognition: Variances and strategies. En *IEEE Transactions on Affective Computing, 1*(2), 119–131.

Shrout, P. E., y Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Solera-Ureña, R., Moniz, H., Batista, F., Fernández Astudillo R., Campos, J., Paiva, A., y Trancoso, I. (2016). Acoustic-Prosodic Automatic Personality Trait Assessment for Adults and Children. Springer International Publishing. https://doi.org/10.1007/978-3-319-49169-1_19

Stegmann, G. M., Hahn, S., Liss, J., Shefner, J., Rutkove, S. B., Kawabata, K., Bhandari, S., Shelton, K., Duncan, C. J. y Berisha, V. (2020). Repeatability of commonly used speech and language features for clinical applications. *Digital biomarkers*, *4*(3), 109-122.

Tay, J. K., Narasimhan, B. y Hastie, T. (2023). Elastic Net Regularization Paths for All Generalized Linear Models. *Journal of Statistical Software*, *106*(1), 1–31. https://doi.org/10.18637/jss.v106.i01

Van Rijn, P., y Larrouy-Maestri, P. (2023). Modelling individual and cross-cultural variation in the mapping of emotions to speech prosody. *Nature Human Behaviour*, *7*(3), 386-396. https://doi.org/10.1038/s41562-022-01505-5

Weiner, J., Angrick, M., Umesh, S., y Schultz, T. (2018). Investigating the Effect of Audio Duration on Dementia Detection using Acoustic Features. In *INTERSPEECH 2018 – 19th Annual Conference of the International Speech Communication Association*.

Wu, S., Falk, T. H., y Chan, W.-Y. (2011). Automatic speech emotion recognition using modulation spectral features. *Speech Communication*, *53*(5), 768-785. https://doi.org/10.1016/j.specom.2010.08.013