Informe final publicable de proyecto
Codificación predictiva en el efecto de fiesta de coctel

Código de proyecto ANII: FCE_1_2019_1_155889

04/12/2023

CERVANTES CONSTANTINO, Francisco (Responsable Técnico - Científico)
CARBONI, Alejandra (Investigador)
SÁNCHEZ COSTA, Thaiz (Investigador)
BIURRUN, Cecilia (Investigador)
CABALLERO, Franco (Investigador)
CARAMÉS, Rodrigo (Investigador)

UNIVERSIDAD DE LA REPÚBLICA. FACULTAD DE PSICOLOGÍA (Institución Proponente) \\
UNIVERSIDAD DE LA REPÚBLICA. FACULTAD DE PSICOLOGÍA

Resumen del proyecto

El efecto de la fiesta de cóctel alude a la situación en la que la voz de varios hablantes se superponen, y de la mezcla resultante el cerebro necesita descifrar solamente una fuente para poder analizar su mensaje. Se desconoce cómo esto se logra, siendo de particular interés para la clínica y los sistemas artificiales. El presente proyecto tiene el objetivo de determinar cómo la experiencia previa sirve a los mecanismos de selección durante el efecto de la fiesta de coctel. Se evalúa la hipótesis de que al resolver el problema de hablantes simultáneos, el cerebro explota un modelo interno formado a partir de la experiencia que pueda ser de utilidad. Para probar cómo el conocimiento previo tiene efectos sobre la forma en que el cerebro filtra la información, utilizamos técnicas de codificación de estímulo mediante correlación reversa a través de la señal electroencefalograma. Encontramos que (1) la experiencia previa total modifica la forma en que la selección atencional ha lugar durante el efecto de la fiesta de coctel, en acuerdo con un estudio previo; (2) sin embargo, su influencia ocurre a través de la disminución de la actividad asociada a la parte del estímulo ya conocido; (3) ésta baja incide no antes de los 200 ms, sugiriendo un impacto mayormente sobre procesos de extracción de información lingüística. Además, investigamos si similarmente coadyuva en esta disminución el conocimiento previo de información parcial sobre un discurso, a decir la voz del interlocutor y, separadamente, el mensaje. Para estos casos, no encontramos una influencia de la experiencia previa cuando se trata de información parcial. Los resultados indican que durante la etapa posterior a la selección atencional (hasta 150 ms), que coincide con procesos clave de comprensión, existen representaciones de la información auditiva consistentes con el formato de error de predicción.

Ciencias Sociales / Psicología / Psicología especial (incluye terapia para el aprendizaje, habla, audición, visión, etc.) / Neurociencia cognitiva de la audición y el ha

Palabras clave: electroencefalografía del habla / modelado de la atención / predicción auditiva /

Introducción

La comunicación es la base de la vida social, y el habla es el medio de cambio más natural. El fenómeno de la "fiesta de cóctel" relata la habilidad natural que tiene un oyente para seleccionar la conversación de un orador entre múltiples, produciendo mensajes que son esencialmente similares desde el punto de vista acústico [1]. El fenómeno subraya el problema de asignación de recursos cognitivos y, al mismo tiempo, el problema de inferir un significado a partir del sonido. Ambos tienen un objetivo conjunto en transmitir un mensaje de interés con la mayor precisión posible.

La atención es la facultad a cargo de gestionar recursos cognitivos, basándose en reglas a determinar lo que es importante sin asumir los costos de procesar lo que no lo es [2]. Por otro lado, los sistemas de procesamiento del habla existentes en el cerebro cuentan con la experiencia acumulada a través del proceso de desarrollo y aprendizaje. Se cree que las partes más críticas de estos dos procesos dinámicos, atención e inferencia lingüística, están sustentadas por estructuras corticales cerebrales, y su dinámica temporal asociada continúa siendo un tema relevante de estudio.

Un enfoque de la función cortical postula al cerebro como un dispositivo de evaluación de hipótesis, basado en la formulación activa de predicciones y en mecanismos de actualización basados en datos [3], [4]. En este marco, la codificación de las señales sensoriales del habla se logra no completamente a través de la representación del estímulo, sino que incluye el cómputo de errores de predicción, es decir, la diferencia observada entre lo que es plausible esperar del estímulo y lo el estímulo en sí [5]. En otras palabras, esta estrategia subraya el ahorro en recursos en la medida en que las representaciones neuronales son sustituidas por representaciones de los reajustes internos que son necesarios para representar la evidencia de forma precisa [6]. Existe evidencia clara de la interacción de procesos atencionales como predictivosen el sistema visual [7], no está claro cómo se aprovechan los mecanismos de atención de en la situación de habla interferida o fenómeno de "fiesta de coctel". Se espera que el proceso requiera interacción a través de los niveles jerárquicos de análisis acústico, fonético y léxico de la jerarquía, sin quedar claro sin embargo, cómo funciona en términos de error de predicción neuronal.

Por ejemplo, no está claro cómo, en el fenómeno, el problema de extracción de significado original (que se tenía sin interferencia) interactúa con el problema de formación y delimitación de objetos auditivos determinados por los hablantes múltiples.

Es plausible que los mecanismos predictivos corticales que ayudan a reducir la incertidumbre en el problema original de extracción de significado, también puedan tener lugar durante la escucha de múltiples hablantes. Para esclarecer la interacción entre el proceso de recopilación de evidencia, y el de formación de objetos, es importante abordar si el filtrado de atención cortical está sujeto a mejoras (reducción de incertidumbre) introducidas por la experiencia con ciertos aspectos particulares de los hablantes. Una medida objetiva clave para proporcionar la estimación de dicha mejora incluye índices neuronales de seguimiento cortical efectivo [8] - [10]. La inspección de estos índices ha sido utilizado para estudiar cómo la experiencia previa con el hablante objetivo previsto ("Conocido atendido", "CA") en realidad causa una mejora en la representación posterior del mismo hablante durante la interferencia, lo que ayudaría a los procesos de formación de objetos y extracción de significado [11]. Para demarcar mejor esta distinción, también examinamos cómo la experiencia previa con el flujo de la competencia ("Conocido desatendido", "CD") influye en los efectos atencionales. Debido a que el objetivo del proceso de extracción de significado se da únicamente sobre la secuencia atendida, en este caso la evidencia de la codificación predictiva puede ser atribuible únicamente al proceso de formación de objetos. Se espera que los efectos observables sean en términos de representaciones neurales (empleando como en casos previos electroencefalografía, EEG) se vean aumentadas para los flujos atendidos en lugar de los desatendidos, y un error de predicción reducido en el flujo atendido.

El estudio de la diferencia entre estas dos contribuciones es conveniente porque los efectos de atención suelen estar indicados por el aumento de la actividad para el flujo asistido, sin embargo, en la codificación predictiva, los efectos del conocimiento previo son más bien opuestos, con un contenido representacional que se reduce como consecuencia de la experiencias. Tales reducciones se observan con frecuencia para estímulos repetidos, sin embargo, en un entorno coctel la información previa a menudo sólo es parcial, es decir, no escuchamos de nuevo la mezcla total, y en ocasiones sólo sabemos algo acerca de uno de los hablantes. Una condición de control sin conocimiento previo, ("Sin previo", "SP") sirve aquí para inspeccionar el contenido representativo en la actividad neuronal para las corrientes espaciales atendidas y no atendidas.

Por lo anterior, es importante establecer los efectos de generalización de la experiencia previa (y, por lo tanto, de la capacidad predictiva) sobre el filtrado de atención del habla en la “fiesta de coctel” durante la interferencia. En particular, se estudia aquí si la experiencia con la voz de un hablante, y por separado, con el contenido del mensaje deseado, influyen en la manera en que el córtex analiza el habla. Los datos experimentales abordan la primacía relativa entre dos efectos opuestos: el de la amplificación de la representación neural cuando es atendida, versus la atenuación de la representación neural cuando existe una predicción correcta acerca del contenido representado (disminución del error de predicción).

Este balance se inspecciona a en diferentes niveles de la jerarquía de procesamiento que existe en las transformaciones de sonido a significado. Para ello, el diseño controla la experiencia (mediante una presentación previa) de uno de los hablantes en la mezcla ("Voz conocida", "VK"). Por otro lado, los efectos del conocimiento previo también se examinan en los casos en que el contenido del mensaje ya se conociera de antemano. Esta segunda condición se lleva a cabo mediante presentaciones, de un orador diferente, del mismo mensaje articulado de manera diferente. Asimismo se incluye una condición "Sin conocimiento previo" como control, con un análisis similar al de la primera parte del proyecto.

Metodología/diseño del estudio

Diseño de investigación y metodología

Sujetos humanos: En esta serie de experimentos los participantes elegibles se reclutan de una base de datos en línea, con edades entre 18 y 35 años, y con una visión normal o corregida. No informan haber tenido trastornos auditivos o neuropsicológicos, ni el uso de sustancias psicotrópicas antes del estudio. Los sujetos elegibles proporcionan el consentimiento informado de acuerdo con la Declaración de Helsinki.

Experimento 1: Modulaciones por conocimiento previo sobre los efectos atencionales durante la “fiesta de coctel”

Diseño de la tarea: 36 participantes elegibles realizan una tarea computarizada que consta de 240 ensayos. En cada prueba, primero se les presenta una oración hablada, de 4 a 6 segundos en duración, consistente en un discurso libre de ruido ("Pre-listening", vea la Figura 1), seguida posteriormente por una mezcla de discursos superpuestos de 2 oradores de

la misma duración ("Cocktail-party listening"). En dos tercios de los ensayos (108 en total), la mezcla contiene el mismo discurso que se escuchó libre de ruido anteriormente; en la mitad de estos, se requiere que el oyente atienda dicho discurso (condición de "atendida conocida", 'AK'), mientras que en la otra mitad, tal oración escuchada durante Pre-listening será la que el interlocutor efectivamente tenga que ignorar ("desatendida conocida", 'UK'). En el tercio de ensayos restantes, la primera oración limpia no formará parte de la mezcla ("Sin conocimiento previo", "NP"), como control. Aparte de la estructura de repetición controlada de los datos, los ensayos se generan de novo, y no se repiten.

Experimentos 2a y 2b: Influencia jerárquica de los conocimientos previos durante la escucha de "cóctel"

Diseño de la tarea: 38 participantes elegibles realizan una tarea computarizada repartida en dos experimentos, con una estructura de ensayo similar a la del Experimento 1. En 1/3 de los ensayos, la mezcla de "fiesta de coctel" contiene una oración hecha por la voz del mismo orador que durante la fase de Pre-listening, aunque transmitiendo un mensaje diferente ("Voz conocida"," VK '; Experimento 2a). En otro tercio de los ensayos, la información conocida es únicamente acerca de la voz del interlocutor que se necesita ignorar.

Separadamente, presentamos un experimento similar con la información parcial obtenida únicamente a partir de parfáfrasis del mensaje ("Mensaje conocido", "MK"; Experimento 2b). 1/3 de ensayos consisten en una presentación de un orador durante la fase de Pre-listening que no se involucra más en la mezcla de escucha de "cocktail party", pero que sin embargo transmite un mensaje idéntico utilizando una estructura de composición diferente. Otro tercio de los ensayos corresponde con información conocida del mensaje a ignorar. Para ambos experimentos, un tercio de los ensayos queda reservado para servir como condición de control, como en "NP" en el Experimento 1, arriba. Los Experimentos 2a y 2b incluyen la construcción de base de datos de voz controlada, y para lo cual se utilizan Grabaciones nuevas basadas en una base pública de paráfrasis (Microsoft Research Paraphrase Corpus).

Configuración y registro de EEG: las presentaciones de sonido se construyen a partir de repositorios de audio disponibles públicamente con MATLAB® (Natick, MA) y se entregen con el software de presentación PsychoPy [14] disponible en el laboratorio EEG del Centro de Investigación Básica en Psicología (CIBPsi). La secuencia visual, incluidas las instrucciones, la fijación y las presentaciones de avatar, se presentan a través de un monitor CRT (E. Systems, Inc., CA) con una visualización máxima de 23,5 cm, resolución de ~ 83 ppp y frecuencia de actualización de 60 Hz. Las grabaciones de EEG se realizan con un sistema de canales activos BioSemi para dos, 64 con un electrodo de referencia para la nariz. La actividad del electrooculograma vertical y horizontal se registran con cuatro electrodos como referencia. La frecuencia de muestreo es de 512 Hz. Los experimentos pueden duran alrededor de 60 minutos en total.

Preprocesamiento de EEG: la serie temporal de grabaciones sin procesar de la matriz de sensores de EEG se envía a una implementación rápida del análisis de componentes independientes [15], desde el cual se seleccionan componentes independientes dominados por artefactos de actividad no neuronal. Estos componentes independientes, combinados con los canales de referencia físicos, se tratan como fuentes de ruido ambiental que surgen de señales eléctricas no deseadas no relacionadas con la actividad cerebral de interés, y se eliminan utilizando el análisis de componentes principales desplazado en el tiempo [16]. Las fuentes específicas de sensores de señales no relacionadas con la actividad cerebral se reducen mediante la supresión del ruido del sensor [17]. Para analizar la actividad neuronal de seguimiento al habla, las grabaciones se filtran entre 1 y 8 Hz.

Análisis de datos: una técnica de separación de fuente ciega, Denoising Source Separation [18] se usa para construir componentes clasificados en orden de reproducibilidad de ensayo a ensayo para los análisis de época a lo largo de los sujetos, basados en un conjunto de datos registrado con equipo de características similares [10]. Para medir los efectos de atención, utilizaremos la correlación inversa. La relación entrada-salida entre una representación de la entrada del estímulo auditivo y la respuesta cortical provocada se modela mediante una función de respuesta temporal (TRF) [19], [20]. En la configuración de cóctel, este modelo lineal predice la respuesta evocada a una característica acústica del habla (el ataque de la envolvente en este caso) de la secuencia de voz atendida y, separadamente, de la desatendida [12]. La respuesta neural a analizar es el primer componente DSS fijado en todos los sujetos [21]. El objetivo será estimar las funciones de respuesta a impulsos correspondientes a cada hablante y compararlas en magnitud. El efecto de atención sobre el seguimiento cortical se mide con esta medida de contraste [13], [22]. Una vez que se obtienen los efectos de atención, abordamos cómo son modulados por la configuración de conocimientos previos. La condición "Sin conocimiento

previo" (NP) sirve como referencia.

Resultados, análisis y discusión

La fuente de los efectos inducidos por la experiencia sobre las modulaciones atencionales se restringe a casos de idéntica repetición de un estímulo auditivo. El efecto consiste en la reducción de la actividad neuronal de seguimiento al estímulo objetivo, cuando este ya ha sido presentado anteriormente de forma fidedigna. La reducción no ocurre en todo momento en general, sino que está contenida entre los 200 y 300 ms, modificando el procesamiento asociado con la etapa P2. No hay evidencia de una reducción cuando hay experiecia del estímulo desatendido, ni en la condición control. A diferencia de otro reporte [11] tampoco encontramos un incremento de la actividad neuronal en la etapa de procesamiento N1 o anterior. Esto resuelve que el contenido representacional en las modulaciones atencionales abarca tanto representación de información de entrada (en etapas temprana) como representaciones de error de predicción (en etapas tardías donde se encuentra interacción). El único nivel donde se encuentra cambios en la selección atencional basados en el conocimiento se dan cuando el conocimiento es total. Estos cambios ocurren tardíamente, sin asistir la selección sino reduciéndola, sugiriendo una dinámica de supresión. Los resultados contrastan con reportes donde el conocimiento previo amplifica la representación del objeto atendido [11]. Por otra parte, El conocimiento de características parciales de un discurso, como la voz, o el mensaje, no demuestra incidir en el proceso de selección atencional.

Es de notar también la identificación de modulaciones atencionales previos a la etapa de procesamiento N1, lo cual no es reportado usualmente en estudios de fiesta de coctel auditivo. Se propone que estos efectos son sutiles y requieren de tamaños de muestra sustanciales, como es el caso de este en este proyecto (N=73).

Conclusiones y recomendaciones

Se concluye con la identificación de la etapa posterior a la selección atencional (a partir de 200 ms) como una donde existen representaciones de la información auditiva consistentes con el formato de error de predicción. Los resultados son consistentes con esta etapa como una donde tiene lugar el análisis de un objeto auditivo ya segregado. En el caso del conocimiento previo de dicho objeto, entonces no necesita ser re-analizado, dando pie a la supresión en la actividad neuronal. Además, se confirma que la etapa de selección atencional (entre 100 y 200 ms) registra representaciones diferenciadas del objeto atendido versus desatendido, de acuerdo con reportes previos [22], [23]. Sin embargo, no se encuentra que esta etapa sea modulada por la presencia de información parcial o total sobre los discursos intervinientes durante la escucha de "fiesta de coctel". Es de notar la modulación temprana la cual sugiere modulaciones tempranas por la vía top-down en el caso auditivo a diferencia de lo que se observa en visión [24].

En conjunto, esperamos que este conocimiento mejore el marco de codificación predictivo existente en relación a fenómenos de interferencia, y que defina el dominio de los mecanismos predictivos que operan en el proceso de extracción de significado, en el proceso de formación de objetos, y en la interacción de ambos. Se valora en función de su aplicabilidad por razones de salud y tecnológicas. En el caso de la salud, puede dar indicaciones para que los regímenes de terapia basados en entrenamiento y experiencia puedan atenderse y cuantificarse con medidas neurales objetivas. En la instancia tecnológica, debido a las características del habla humana natural, y la gran similitud que existe entre dos hablantes cualesquiera (en términos de estructura acústica), son problema actual difícil de resolver para los sistemas artificiales. El desarrollo de los mismos ha visto por mecanismos bioinspirados en el dominio de la ingeniería neuromórfica.

Referencias bibliográficas

[1] E. C. Cherry, 'Some Experiments on the Recognition of Speech, with One and with Two Ears', J. Acoust. Soc. Am., vol. 25, no. 5, pp. 975–979, Sep. 1953.

[2] J. Fawcett, A. Kingstone, and E. Risko, The Handbook of Attention. MIT Press, 2015.

[3] K. Friston, 'A theory of cortical responses', Philos. Trans. R. Soc. B Biol. Sci., vol. 360, no. 1456, pp. 815–836, Apr. 2005.

[4] C. Wacongne, E. Labyt, V. van Wassenhove, T. Bekinschtein, L. Naccache, and S. Dehaene, 'Evidence for a hierarchy of predictions and prediction errors in human cortex', Proc. Natl. Acad. Sci., vol. 108, no. 51, pp. 20754–20759, Dec. 2011.

[5] H. E. M. den Ouden, P. Kok, and F. P. de Lange, 'How Prediction Errors Shape Perception, Attention, and Motivation', Front. Psychol., vol. 3, Dec. 2012.

[6] E. Sohoglu and M. H. Davis, 'Perceptual learning of degraded speech by minimizing prediction error', Proc. Natl. Acad. Sci. U. S. A., vol. 113, no. 12, pp. E1747-1756, Mar. 2016.

[7] P. Kok, D. Rahnev, J. F. M. Jehee, H. C. Lau, and F. P. de Lange, 'Attention reverses the effect of prediction in silencing sensory signals', Cereb. Cortex N. Y. N 1991, vol. 22, no. 9, pp. 2197–2206, Sep. 2012.

[8] F. C. Constantino and J. Z. Simon, 'Dynamic cortical representations of perceptual filling-in for missing acoustic rhythm', Sci. Rep., vol. 7, no. 1, p. 17536, Dec. 2017.

[9] F. Cervantes Constantino and J. Z. Simon, 'Restoration and Efficiency of the Neural Processing of Continuous Speech Are Promoted by Prior Knowledge', Front. Syst. Neurosci., vol. 12, p. 56, 2018.

[10] J. Vanthornhout, L. Decruy, J. Wouters, J. Z. Simon, and T. Francart, 'Speech Intelligibility Predicted from Neural Entrainment of the Speech Envelope', J. Assoc. Res. Otolaryngol., vol. 19, no. 2, pp. 181–191, Apr. 2018.

[11] Wang Y, Zhang J, Zou J, Luo H, Ding N. Prior Knowledge Guides Speech Segregation in Human Auditory Cortex. Cereb Cortex. 2019 Apr 1;29(4):1561-1571. doi: 10.1093/cercor/bhy052. PMID: 29788144.

[14] J. W. Peirce, 'PsychoPy—Psychophysics software in Python', J. Neurosci. Methods, vol. 162, no. 1, pp. 8–13, May 2007.

[15] A. Hyvarinen, 'Fast and robust fixed-point algorithms for independent component analysis', IEEE Trans. Neural Netw., vol. 10, no. 3, pp. 626–634, May 1999.

[16] A. de Cheveigné and J. Z. Simon, 'Denoising based on time-shift PCA', J. Neurosci. Methods, vol. 165, no. 2, pp. 297–305, Sep. 2007.

[17] A. de Cheveigné and J. Z. Simon, 'Sensor noise suppression', J. Neurosci. Methods, vol. 168, no. 1, pp. 195–202, Feb. 2008.

[18] A. de Cheveigné and J. Z. Simon, 'Denoising based on spatial filtering', J. Neurosci. Methods, vol. 171, no. 2, pp. 331–339, Jun. 2008.

[19] F. E. Theunissen, S. V. David, N. C. Singh, A. Hsu, W. E. Vinje, and J. L. Gallant, 'Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli', Netw. Bristol Engl., vol. 12, no. 3, pp. 289–316, Aug. 2001.

[20] F. Cervantes Constantino, M. Villafañe-Delgado, E. Camenga, K. Dombrowski, B. Walsh, and J. Z. Simon, 'Functional significance of spectrotemporal response functions obtained using magnetoencephalography', bioRxiv, p. 168997, Jul. 2017.

[21] A. de Cheveigné and L. C. Parra, 'Joint decorrelation, a versatile tool for multichannel data analysis', NeuroImage, vol. 98, pp. 487–505, Sep. 2014.

[22] C. Brodbeck, L. E. Hong, and J. Z. Simon, 'Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech', Curr. Biol. CB, vol. 28, no. 24, pp. 3976-3983.e5, Dec. 2018.

[23] N. Ding and J. Z. Simon, 'Emergence of neural encoding of auditory objects while listening to competing speakers', Proc. Natl. Acad. Sci., vol. 109, no. 29, pp. 11854–11859, Jul. 2012.

[24] J. Alilovi?, B. Timmermans, L. C. Reteig, S. van Gaal, H. A. Slagter, No Evidence that Predictions and Attention Modulate the First Feedforward Sweep of Cortical Information Processing, Cerebral Cortex, Volume 29, Issue 5, May 2019, pp. 2261–2278, https://doi.org/10.1093/cercor/bhz038

Licenciamiento