

APLICACIONES DEL ULTRASOUND SOUND IMAGE (UTI) A LOS ESTUDIOS DE FONÉTICA

Wendy Elvira-García¹, Paolo Roseano^{2,3}, Lourdes Romera Barrios³, Victoria Marrero Aguiar¹

¹Universidad Nacional de Educación a Distancia, ²Universitat de Barcelona, ³University of South Africa
welvira@flog.uned.es, paolo.roseano@ub.edu, lourdesromera@ub.edu, vmarrero@flog.uned.es

ABSTRACT

Este texto, presentado a la sesión monográfica de *Nuevas metodologías para el estudio fonético: qué nos aportan y qué nos exigen*, pretende ofrecer una panorámica de las características y de las aplicaciones de la visualización de la lengua por ultrasonido (UTI) al ámbito de la fonética. El trabajo presenta el equipo necesario para llevar a cabo una grabación de ultrasonido y algunos análisis posibles.

Palabras clave: ultrasound tongue imaging, técnicas fonéticas, instrumentos

This paper aims to give an overview of the ultrasound tongue imaging (UTI) characteristics and applications in phonetics. It is been submitted to the specific panel *Nuevas metodologías para el estudio fonético: qué nos aportan y qué nos exigen*. The paper presents the equipment needed to perform an ultrasonic recording and some of the possible analyses.

Keywords: ultrasound tongue imaging, phonetic techniques, tools

1. INTRODUCCIÓN

A lo largo de las décadas se han ido desarrollado varias técnicas de análisis articulatorio, como la pletismografía, la electromiografía, la fotoglografía, la electrolaringología, la resonancia magnética, la radiografía, la articulografía electromagnética, la aerometría, la endoscopia, la electropalatografía o la videoquimografía.

En los años 70 del siglo pasado se empezó a usar una nueva técnica para visualizar el aparato fonador, la ecografía. Al principio, se usó para visualizar las articulaciones faríngeas (Kelsey et al., 1969; Skolnick et al., 1975; Zagzebski, 1975; MacKay, 1977, etc.) y, más tarde, en los años 80 se aplicó a la visualización de la lengua (Morrish et al., 1984). Esta última técnica, se llamó *Ultrasound Tongue Imaging*, o 'visualización de la lengua mediante ultrasonido' (UTI, por sus siglas en inglés) (Wrench y Scobbie, 2006; Gick et al., 2013; Lulich et al., 2018).

Pese a que desde los 80 se venían efectuando estudios lingüísticos, el principal uso del UTI era el ámbito clínico. Son abundantes los trabajos sobre alteraciones de la deglución (por ejemplo, Shawker, 1983), la succión en bebés (Bosma et al., 1990) o la

posición de la lengua durante la anestesia (Abernethy et al., 1990), entre otros.

Sin embargo, en los 2000 se produjeron dos hechos que ayudaron a impulsar los estudios lingüísticos realizados con UTI. En primer lugar, nació una iniciativa, el UltraFEST, en la que se reúnen investigadores que usan el UTI para la investigación lingüística. En segundo lugar, la aparición de unidades de ultrasonido portátiles y de menor precio y la mejora de los softwares de análisis de imágenes han hecho que aumente su adquisición por parte de los laboratorios de fonética.

El UTI en la actualidad es una técnica no invasiva para realizar imágenes dinámicas y permite a los investigadores ver la forma, la posición y los movimientos de la lengua (desde la raíz hasta el ápice) en tiempo real durante el habla. Esta herramienta está disponible, hasta donde sabemos, en tres laboratorios de fonética del estado español: el de la Universidad de Navarra (UN), el del Institut d'Estudis Catalans (IEC) y el de la Universitat de Barcelona (UB).

El objetivo de este trabajo es mostrar como el uso de una técnica de análisis articulatorio de última generación puede arrojar nueva luz sobre temas controvertidos del ámbito fonético-fonológico, a la

vez que se anima a los asistentes habituales al CIFE a empezar por cuestiones aún abiertas en la lingüística española, y más en general, iberorrománica y románica. Estaríamos hablando, por ejemplo, de temas tan discutidos como la variación dialectal de las realizaciones de la /s/ en español (Alvar, 1996; Cano, 2005), el punto de articulación de la /tʃ/ en catalán y en español peninsular central (Fernández Planas, 2000; Martínez Celdrán y Fernández Planas, 2001; Recasens, 2019), el papel del rasgo ATR en la armonía vocálica del andaluz o del valenciano (Lloret y Jiménez, 2009; Herrero y Jiménez, 2013), entre otros.

2. EQUIPO NECESARIO

Como casi todas las técnicas de fonética articuladora, el UTI requiere un equipo específico. Básicamente, se trata de un ecógrafo y una sonda, un elemento de sujeción de la sonda, así como de los controladores y del programario correspondientes.

2.1. El ecógrafo

Los laboratorios de fonética han pasado en los últimos tiempos del uso de grandes equipos (como los Digital Sona-Graph 7800 y DSP Sona-Graph 5500) al uso de un ordenador personal. La historia de los equipos de ecografía ha mostrado un desarrollo similar.

En la actualidad, se ha pasado de la existencia de grandes ecógrafos de pantalla integrada (equivalentes en tamaño a los Sona-Graph) a equipos que se pueden conectar a un ordenador mediante USB (Figura 1).



Figura 1: Ecógrafo de bolsillo de Telemed y sonda.

2.2. La sonda

A este equipo se conecta una sonda, que es la encargada de producir el sonido de alta frecuencia que sirve para crear la imagen del ecógrafo.

Las características de la sonda dependen del lugar donde se quiera aplicar el ecógrafo. En los estudios de fonética, la sonda se suele colocar sobre la parte

baja de la mandíbula, orientada de detrás-adelante para conseguir una imagen sagital de la lengua. Como es una zona pequeña, se suelen usar para ello sondas pediátricas como la MC4-2R20S-3 de TELEMED. Esta sonda en concreto, disponible para el uso de investigadores en el laboratorio de fonética de la UB (Figura 1), tiene una frecuencia de 2-4MHz, 20mm de radio y un campo máximo campo de visión de 92º, perfecto para cortes sagitales.

2.3. La monitorización o sujeción de la sonda

Existen varios modos de colocar la sonda al informante. La primera manera es que el propio informante sujete la sonda sin que haya un sistema de sujeción. Este método es el más cómodo de usar y es el utilizado para efectuar, por ejemplo, demostraciones sobre articulaciones en clase (Mozafari et al., 2008). Sin embargo, tiene inconvenientes. El principal entre ellos es que el ultrasonido ofrece imágenes de la lengua, pero no su posición respecto una referencia absoluta. Para calcular la posición de la lengua en la boca, se usa como referencia el centro de la imagen que se graba. Por lo tanto, para mantener esa referencia fija, la sonda no se puede mover durante la grabación.

Por esa razón, aunque es cierto que se puede confiar en el pulso del informante para que él mismo sujete la sonda, lo más probable es que en algún momento de la grabación el informante mueva la cabeza o la sonda se mueva. Si eso ocurre, la grabación queda inutilizable.

Para evitar perder los puntos de referencia durante las grabaciones, se han creado diferentes métodos de monitorización del movimiento y sujeción de la sonda.

Para evitar que el informante se mueva demasiado cuando está sujetando la sonda, hay investigadores que usan un puntero láser que apunta desde la sonda a un cuadro en una pared cercana y ordenan al informante que mantenga el puntero dentro del cuadro mientras habla. Este método hace que tanto el investigador como el informante puedan monitorizar cuando se ha movido la posición de la sonda y, si esto ocurre, la grabación se debería desechar y empezar de nuevo, ya que se habrían movido los puntos de referencia para el análisis.

Un método que requiere menos confianza en la habilidad del informante es la monitorización del movimiento de la cabeza respecto a la sonda. Si se dispone de la información sobre los movimientos que ha efectuado el informante, después se puede reorientar la imagen de ultrasonido usando esa información. Para implementar este sistema, existen

estándares como el de Haskins Optically Corrected Ultrasound System (HOCUS), en el que se usan diodos emisores de infrarrojos en la sonda y en diversos puntos de la cara del informante (Whalen et al., 2005).

Por último, existen diversos modelos de cascos de sujeción para la sonda. Estos cascos, aunque son la solución más limpia en términos de análisis, tienen algunos detractores, ya que la inmovilización parcial de la mandíbula que se lleva a cabo con el casco puede tener efectos en la producción con naturalidad de los estímulos.

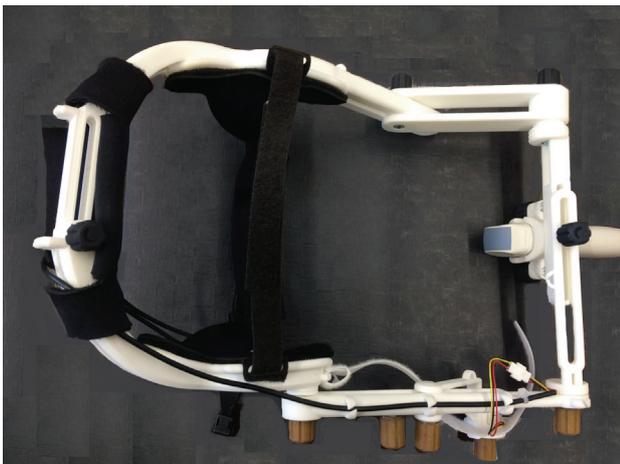


Figura 2. Casco de sujeción UltraFit.

Sin embargo, el diseño de los cascos también ha mejorado en los últimos años y se ha pasado de modelos de metal (Stone, 1995; Scobbie et al., 2008) a modelos de nylon como el UltraFit, que pesa 350gr (Spreafico et al., 2018) (Figura 2).

2.4. La sincronización

Para que las imágenes obtenidas con el ultrasonido puedan tener una interpretación lingüística, el investigador necesita oír el sonido que está sonando a la vez que la imagen que se reproduce.

Este hecho, obvio para los fonetistas, en realidad no es una cuestión técnica trivial, ya que la imagen y el sonido se transmiten a diferente velocidad. Quienes hayan estudiado procesamiento del lenguaje saben que los humanos tardamos 30ms en percibir un estímulo sonoro y 80-95ms en procesar un estímulo visual y, sin embargo, cuando vemos caer un vaso al suelo oímos como se rompe a la vez; esto es gracias a los cómputos y predicciones que hace nuestro cerebro.

Pero, como es habitual, programarlo en una máquina no es fácil. Para ello, los *settings* de ultrasonido para fonética usan diferentes sistemas de sincronización en los que una unidad externa se

encarga de que el ordenador reciba prácticamente a la vez la señal de video y audio. Todos ellos son aproximaciones lo más exactas posible, pero ninguno puede ofrecer una sincronización perfecta (al menos en un laboratorio ordinario, con un ordenador común y con la técnica disponible en la actualidad) (Wrench y Scobbie, 2006).

2.5. Controladores y software

El software de análisis del UTI más usado es el Advanced Articulate Analysis (AAA; también útil para analizar datos de EPG) (Figura 3). Este programa permite realizar grabaciones, cambiar los ajustes de grabación del ultrasonido (como la frecuencia de muestreo de la grabación o la profundidad del ultrasonido), anotar los datos y exportarlos en diferentes formatos.

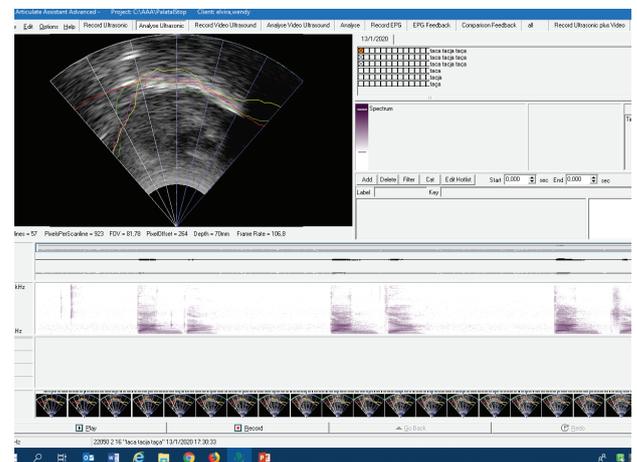


Figura 2: Interfaz del software de grabación y análisis AAA.

Una grabación estándar se realiza usando una profundidad de 70mm y un campo de visión de 90%, con la potencia al mínimo posible para reducir la exposición del informante al ultrasonido y la ganancia al máximo.

Si el objetivo es obtener un video para reproducir a modo de demo (por ejemplo, en una clase de fonética articulatoria) es necesario tener en cuenta que las pantallas europeas reproducen un máximo de 24 frames por segundo. Ahora bien, para un análisis estadístico en R, por ejemplo, podemos grabar a mucho más muestreo y captar movimientos que incluso serían imposibles de ver para el ojo humano. La sincronización entre las señales de audio y video se realiza la primera vez que se analiza la grabación. Para ello, el usuario solo necesita ordenar al software que sincronice y él lo hará automáticamente a partir de una pequeña señal que se emite durante la sesión de grabación.

Además, el AAA permite al usuario efectuar la grabación mientras en pantalla aparecen las palabras que tiene que pronunciar el informante.

Con el mismo software se pueden analizar datos mediante la creación de abanicos que delimitan las zonas del paladar. También es en este software donde el investigador deberá reseguir el perfil de la lengua a mano alzada para crear los perfiles de la lengua en los diferentes *frames*. Estos perfiles (llamados *splines*) son los que normalmente se exportan y usan posteriormente para el análisis cuantitativo de los datos. AAA también permite la importación de TextGrids de Praat y el uso de los espectrogramas incluidos en la visualización (v. Figura 2).

3. USOS DEL ULTRASONIDO EN FONÉTICA

3.1. Determinación del punto de articulación

El uso más transparente del UTI es la determinación del punto de articulación de un sonido. Para efectuar ese tipo de análisis, se crean unos abanicos que se superponen a la imagen de la lengua en los que se definen los límites de cada zona del paladar (alveolar, palatal, velar) al modo de los estudios de EPG.

En el caso del catalán, Recasens y Rodríguez (2016, 2017) establecen los límites entre estas zonas a partir de las articulaciones de los siguientes sonidos del propio informante:

- Límite alveolar-palatal: [r]
- Límite velar-palatal [iç]
- Zona velar: 1,26cm o 1,51cm (según sexo, basado en Fitch & Giedd, 1999)

Una vez conseguidos esos puntos de referencia, se realiza la articulación que se quiere estudiar y se mide en el momento central de la articulación el punto donde se produce la máxima constricción. Así se obtendría de manera estática el punto de articulación de un sonido.

3.2. Comparación de dos articulaciones

Cuando el problema que se plantea es si dos articulaciones son iguales o diferentes (por ejemplo, consonante en ataque y en coda), se puede recurrir a un análisis que determina si la forma de la lengua es diferente en las dos condiciones.

Para ello se analizan dos grupos de *splines*¹ de los que se obtiene el perfil medio mediante Smoothing spline SSANOVA (Davidson, 2006). Después, se puede determinar si hay diferencias significativas entre los *splines* que se ajustan mejor (*fit*) a dos conjuntos de datos que se comparan.

La interacción puede ser significativa o no, pero en el caso de que lo sea, queda la duda de qué parte de la lengua ha hecho que la articulación sea diferente. Para saber qué parte (sección de la lengua) ha cambiado, se usan intervalos de confianza bayesianos.

Aunque este es posiblemente uno de los análisis más extendidos, no es el único posible: otros métodos como el Análisis de Componentes principales se han usado para ver, por ejemplo, las diferencias entre [l] y [ɫ] (Turton, 2015).

3.3. Comparación de dos articulaciones en el tiempo

Hasta ahora los análisis que se han planteado son estáticos. Es decir, de cada producción de un sonido se ha elegido el momento central (como una fotografía) y se han comparado sets de fotos. Sin embargo, los datos que proporciona el ultrasonido son de video y, por tanto, dinámicos.

El análisis de datos dinámicos en fonética no es nuevo. De hecho, las trayectorias de los formantes en diptongos lo son, y precisamente para ello se aplicó por primera vez el tipo de análisis que vamos a exponer a continuación: el Generalised Additive Modelling (GAM) (Sóskuthy, 2017). Además de a trayectorias de formantes, estos modelos se han usado en EMA, a partir de lo cual se han aplicado a datos de imagen de ultrasonido. El GAM crea una función para predecir cómo van a cambiar los datos usando como predictores las relaciones entre los datos que varían en el tiempo. Sin embargo, como en todos los modelos, de ahí no se puede obtener un valor de significación “p” directo, sino que hay que efectuar una comparación de dos modelos usando *Likelihood Ratios*.

4. CONCLUSIONES

Este trabajo quiere mostrar el funcionamiento y las posibles aplicaciones de un equipo de UTI al campo de la fonética, con la voluntad de que en un futuro próximo se puedan ver más estudios con UTI realizados en España.

¹ Hay que tener en cuenta que el UTI no permite ver el primer centímetro del ápice de la lengua.

Las técnicas de fonética articuladora, como el UTI, contribuyen revitalizar los laboratorios de fonética como espacios físicos, pero sobre todo permiten un avance significativo en nuestro estado de conocimientos sobre los procesos necesarios para la emisión de sonidos, en condiciones cada vez más naturales, con un coste menor y más fácil interpretación.

5. BIBLIOGRAFÍA

- Alvar, M. (1996). *Manual de dialectología hispánica. El español de España*. Barcelona: Ariel Lingüística.
- Cano, R. (2005). *Historia de la lengua española*. Barcelona: Ariel Lingüística.
- Fernández Planas, A.M. (2000). Estudio electro-palatográfico de la coarticulación vocálica en estructuras VCV en castellano, Tesis doctoral, Universitat de Barcelona.
- Gick, B., Wilson, I. y Derrick, D. (2013). *Articulatory Phonetics*. Wiley: Blackwell.
- Herrero, R. y Jiménez, J. (2013). De la coarticulación a la armonía vocálica en valenciano. En: F. Sánchez Miret, D. Recasens (eds.): *Studies in phonetics, phonology and sound change in Romance* (pp. 43-63). München: Lincom,
- Kelsey, C.A., Woodhouse, R.J. y Minifie, F.D. (1969). Ultrasonic observations of coarticulation in the pharynx. *Journal of the Acoustical Society of America* 46(4), 1016-1018.
- Lloret, M.R. y Jiménez, J. (2009). Un análisis óptimo de la armonía vocálica del andaluz. *Verba. Anuario Galego de Filoloxía*, 36, 293-325.
- Lulich, S. M., Berkson K.H., y de Jong, K. (2018). Acquiring and visualizing 3D/4D ultrasound recordings of tongue motion, *Journal of Phonetics*, 71, 410-424.
- Mackay, I.R.A. (1977). Tenseness in vowels: An ultrasonic study. *Phonetica*, 34, 325- 351.
- Martínez Celdrán, E. y Fernández Planas, A.M. (2001). Propuesta de transcripción para la africada palatal sonora del español, *Estudios de Fonética Experimental*, 11, p. 173-190.
- Mozaffari, M.H., Guan, S., Wen, S., Wang, N., & Lee, W. (2018). Guided Learning of Pronunciation by Visualizing Tongue Articulation in Ultrasound Image Sequences. En: *Proceedings of IEEE 2018*, 1-5.
- Recasens, D. (2019). *Configuració lingual de vocals, consonants i grups de consonants del català: dades d'ultrasons*. Barcelona: Universitat Autònoma de Barcelona.
- Scobbie, J.M., Wrench, A.A., van der Linden, M. (2008). Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. En: *Proceedings of the 8th International Seminar on Speech Production*, 373-376.
- Skolnick, M.L., Zagzabski, J.A. y Watkin, K.L. (1975). Two dimensional ultrasonic demonstration of lateral pharyngeal wall movement in real time: A preliminary report. *Cleft Palate Journal*, 12, 299-303.
- Spreafico, L., Pucher, M. y Matosova, A. (2018). UltraFit: A speaker-friendly headset for ultrasound recordings in speech science. En: *Proceedings of Interspeech 2018, 19th Annual Conference of the International Speech Communication Association*, Hyderabad, India, 2-6 September 2018.
- Stone, M. y Davis, E.P. (1995). A head and transducer support system for making ultrasound images of tongue/jaw movement. *The Journal of The Acoustical Society of America*, 98(6), 3107-3112.
- Turton, D. (2015). Determining categoricity in English /l/-darkening: A principal component analysis of ultrasound spline data. En: *Proceedings of ICPhS 2015*.
- Whalen, D.H., Iskarous, K., Tiede, M.K. y Ostry, D.J. (2005). The Haskins Optically Corrected Ultrasound System (HOCUS). *Journal of Speech, Language, and Hearing Research*, 48, 543-553.
- Wilson, I. (2014). Using ultrasound for teaching and researching articulation. *Acoustical Science and Technology*, 35(6), 285-289.
- Wrench, A.A. y Scobbie, J.M. (2006). Spatio-temporal inaccuracies of video-based ultrasound images of the tongue, En: *Proceedings of the 7th International Seminar on Speech Production*. Ubatuba, Brazil, December 2006.
- Zagzabski, J.A. (1975). Ultrasonic measurement of lateral pharyngeal wall motion at two levels in the vocal tract. *Journal of Speech and Hearing Research*, 18, 308-318.