

# Efectos del cambio de lengua en la comparación de voces mediante LTAS

---

ROSEANO, PAOLO; FERNÁNDEZ PLANAS, ANA MA.;  
ELVIRA-GARCÍA, WENDY; MARTÍNEZ CELDRÁN,  
EUGENIO

## 1 Introducción

A lo largo de las últimas décadas se ha ido definiendo un conjunto de parámetros acústicos que se consideran aptos para ser utilizados en la comparación de voces con finalidades judiciales (v. Gil, 2014). Entre ellos se encuentra el Espectro Medio a Largo Plazo (o LTAS) que, como es sabido, es una representación del espectro promediado de la amplitud en una gama de frecuencias dada calculado para un intervalo de tiempo.

Desde los primeros intentos de utilizar el LTAS para el reconocimiento automático del locutor (Pruzansky, 1963), se ha abierto un filón de investigación cuya finalidad es valorar la fiabilidad de esta técnica en la comparación de voces (Zalewski *et al.*, 1975; Pittam, 1987; Nolan, 1983, Molina de Figueiredo & Bernal, 1999; Cicres, 2011; i.a.). Uno de los aspectos que han recibido cierta atención es el de la efectividad de esta técnica en la comparación de muestras de voz del mismo locutor que difieren entre sí en cuanto a variaciones debidas a ajustes fonatorios o a efectos del canal de transmisión (Klingholz *et al.*, 1988, Nordenberg & Sundberg, 2003; i.a.). Una atención menor se ha dedicado a la aplicabilidad del LTAS en los casos donde las muestras de voz dubitada e indubitada son de lenguas distintas (a causa, por ejemplo, de que un sospechoso poco colaborativo se niegue a grabar en la misma lengua de la muestra dubitada).

Los trabajos que se centran en la aplicabilidad del LTAS en la comparación de muestras de voz producidas en lenguas distintas por hablantes bilingües, además de ser escasos, no coinciden en sus conclusiones. Mientras Harmegnies *et al.* (1991), por ejemplo, concluyen que las características de cada lengua no influyen sobre el LTAS, estudios como los de Bruyninckx *et al.* (1994) y Manwa *et al.* (2012) abogan que la lengua hablada tiene efectos sobre los LTAS. Cabe destacar también que ninguno de estos trabajos está enfocado a la fonética judicial.

## 2 Objetivos

Este trabajo tiene dos objetivos, de los cuales uno es de tipo teórico y el otro tiene una caracterización más bien aplicada. Desde una perspectiva más teórica, se pretende retomar el estudio de los efectos del cambio de lengua sobre el LTAS, filón de investigación que en los últimos 20 años ha sido objeto de una atención marginal.

El segundo objetivo, de carácter metodológico y aplicado, es desarrollar una propuesta de rutina de análisis del LTAS con fines judiciales. La rutina en cuestión, preparada en el entorno de Praat (Boersma & Weenink, 2013), permite medir las distancias entre muestras de habla basándose en el LTAS (Majewski & Hollien, 1974; Zalewski *et al.*, 1975). La automatización de la comparación entre los LTAS de dos muestras de voz no tiene como objetivo el de desautorizar al experto o de reducir la importancia de su papel, sino el de obviar algunos de los límites reconocidos de los métodos más tradicionales de comparación de muestras de habla, es decir su posible subjetividad y su alto coste en términos de tiempo necesario (Künzel, 2011: 39). Además, es importante subrayar como la literatura científica de los últimos años sugiere que la comparación automática del habla podría representar una ventaja justamente en los casos reales en los que se disponga de dos muestras de habla en lenguas diferentes una de las cuales sea poco conocida por el experto (Künzel, 2011: 41-42).

## 3 Metodología

### 3.1 Corpus

El corpus utilizado para este trabajo se compone de las grabaciones de 12 hablantes bilingües de italiano regional de Friul y del dialecto friulano del Bajo Valle de Gorto, 8 mujeres y 4 hombres, leyendo el texto de *El viento Norte y el sol* en ambas lenguas<sup>1</sup>. Las grabaciones se efectuaron en una sola sesión para cada hablante, en casa del mismo, en una habitación silenciosa, mediante una grabadora digital Marantz modelo PMD620 que llevaba conectado un micrófono direccional Shure SM58. Los archivos de sonido se grabaron en formato .wav mono con una frecuencia de muestreo de 44.100 Hz.

La decisión de utilizar voces masculinas y femeninas, así como la de trabajar con grabaciones de alta calidad de habla de laboratorio realizadas en una sola sesión (todas condiciones diferentes de las que se darían en un caso judicial real) responden a la necesidad de testar, en esta primera fase, el funcionamiento del método automático de comparación de voces en una situación próxima a la ideal. Si el resultado fuera satisfactorio en estas condiciones, se podría pasar a otras fases de experimentación que consistirían, por ejemplo, en utilizar voces

<sup>1</sup> Los materiales se grabaron en diciembre de 2013 en el marco de un proyecto de investigación sobre las características rítmicas de lenguas en contacto (friulano e italiano) que el Laboratori de Fonètica de la Universitat de Barcelona está llevando a cabo en colaboración con la Universität Hamburg.

de hablantes del mismo sexo y con grados diferentes de alteración por efecto del canal de transmisión.

Los sistemas fonéticos del italiano y del friulano del Bajo Valle de Gorto presentan un grado de divergencia que podríamos definir moderado. En cuanto al sistema vocálico, ambas lenguas presentan siete timbres para las vocales tónicas [i, e, ε, a, ɔ, o, u], que en posición átona pasan a ser cinco en italiano [i, e, a, o, u] y seis en friulano [i, e, a, v, o, u]. En relación con los glides, ambas lenguas muestran los mismos, es decir [j, w]. Con respecto al sistema consonántico, las dos lenguas comparten una parte importante de elementos, es decir [m, n, ɲ, ŋ, p, b, t, d, k, g,  $\widehat{ts}$ ,  $\widehat{dz}$ ,  $\widehat{tʃ}$ ,  $\widehat{dʒ}$ , f, v, s, z, ʃ, r, l]. A esos, el italiano regional añade [ʎ]; el friulano del Valle de Gorto [ʒ, c, ʝ]. En cuanto a la presencia de elementos geminados, cabe recordar que en friulano, a diferencia del italiano, existen también vocales tónicas fonológicamente largas [i:, e:, ε:, a:, ɔ:, o:, u:]. Además, mientras que el italiano presenta un repertorio muy amplio de consonantes geminadas, en friulano solamente se pueden producir en algunos casos por fonética sintáctica.

## 3.2 Script

La comparación entre los LTAS de las diferentes muestras se ha llevado a cabo mediante un script de Praat (Elvira-García *et al.*, 2014) que efectúa dos operaciones: 1) calcula el LTAS de cada muestra y, a continuación, 2) calcula las distancias entre ellos y los guarda en un documento de texto que puede ser gestionado mediante Excel.

### 3.2.1 Cálculo del LTAS

Las versiones más recientes de Praat permiten calcular, además del LTAS tradicional, también el pitch-corrected LTAS, que es una versión del LTAS que intenta corregir los efectos de las variaciones de F0 (Boersma y Kovačić, 2006) y que es la que se ha utilizado para este trabajo. El script que se ha predispuesto calcula el LTAS para un rango frecuencial que llega hasta los 8000 Hz, divididos en pasos de 100 Hz. Como resultado, el LTAS de cada muestra resulta ser una secuencia de 80 valores numéricos (figura 1), ordenados según la frecuencia a la cual corresponden. Los 80 valores numéricos en cuestión se guardan en una base de datos desde la cual se recuperarán en la fase siguiente del análisis.

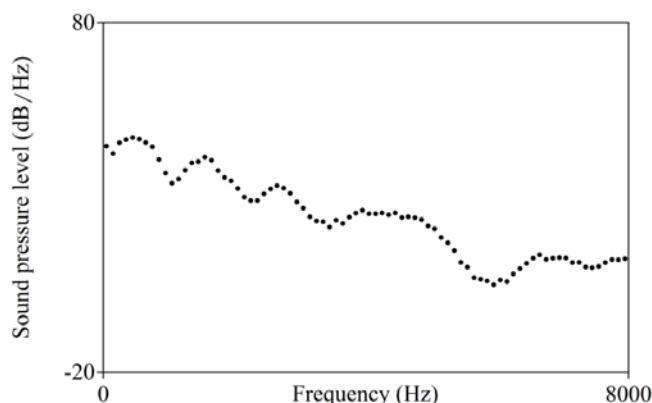


Figura 1. LTAS de un archivo que contiene la grabación de *El viento norte y el sol*

### 3.2.2 Cálculo del LTAS

Una vez determinado el LTAS de todas las muestras que se quieren comparar, el paso siguiente es el cómputo de las diferencias entre cada una de ellas y las demás. El cálculo de la diferencia entre dos secuencias de 80 números, que llamamos  $LTAS_1$  y  $LTAS_2$ , es una operación matemática sencilla. La diferencia en cuestión, que denominamos  $\Delta_{LTAS}$ , es la suma de los valores absolutos de las diferencias entre cada pareja de números que ocupan la misma posición en la secuencia dividida por el número de observaciones. En términos más formales:

$$\begin{aligned} LTAS_1 &= \{x_1, x_2, x_3, x_4, x_5 \dots x_{80}\} \\ LTAS_2 &= \{y_1, y_2, y_3, y_4, y_5 \dots y_{80}\} \\ \Delta_{LTAS} &= \sum(|x_1 - y_1| + |x_2 - y_2| + |x_3 - y_3| + |x_4 - y_4| \dots + |x_{80} - y_{80}|) / 80 \end{aligned}$$

Cuanto más grande es  $\Delta_{LTAS}$ , mayor es la diferencia entre los LTAS de las dos muestras. Si fuera cierto que, en hablantes bilingües, el cambio de lengua no produce ningún efecto apreciable sobre el LTAS (que podemos considerar nuestra hipótesis de trabajo o  $H_p$ ), el  $\Delta_{LTAS}$  entre las dos muestras del mismo sujeto hablando lenguas distintas debería ser inferior al  $\Delta_{LTAS}$  entre cada una de esas muestras y cualquier otra muestra de sujetos distintos. En términos más formales, si definimos como  $\Delta_{id}$  la distancia entre los LTAS de la muestra indubitada y la dubitada y como  $D$  el conjunto de las distancias  $\Delta$  entre todas las muestras del corpus (con exclusión, obviamente, de  $\Delta_{id}$ ):

$$H_p \Leftrightarrow \forall \Delta \in D : \Delta_{id} < \Delta$$

No obstante, antes de verificar esta hipótesis a partir de los datos de nuestro corpus, es preciso efectuar algunas correcciones al modelo matemático – demasiado sencillo – que se acaba de presentar. Estas correcciones son requeridas por la naturaleza misma del LTAS. Recordamos, de hecho, que el LTAS es función, en primer lugar, de la *intensidad* de la señal en cada banda de frecuencia. Esto puede tener consecuencias nefastas a la hora de comparar los LTAS de dos muestras, como se demuestra en el ejemplo a continuación. Imaginemos que se graben dos sujetos (el hablante A y el hablante B) leyendo el mismo texto en castellano (*El viento norte y el sol*). El hablante A es grabado con una grabadora conectada con un micrófono direccional puesto a 30 cm de su boca. La muestra así obtenida se llamará  $\alpha_{30}$ . El hablante B es grabado con dos grabadoras idénticas conectadas con dos micrófonos del mismo modelo, puestos a 30 y a 70 cm de su boca. Las muestras así obtenidas se llamarán  $\beta_{30}$  y  $\beta_{70}$  (figura 2).

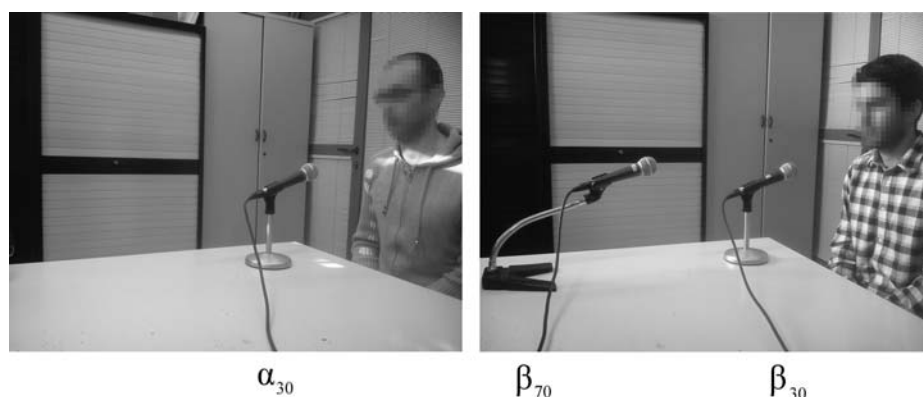


Figura 2. Condiciones de grabación del hablante A (panel de la izquierda) y del hablante B (panel de la derecha)

Los LTAS de las tres muestras, calculados con el método expuesto anteriormente, son las que se representan en la figura 3. Las dos líneas discontinuas representan la muestra  $\beta_{30}$  (línea de rayas, la más alta de las tres), la  $\beta_{70}$  (línea de puntos, la más baja) y la  $\alpha_{30}$  (línea continua).

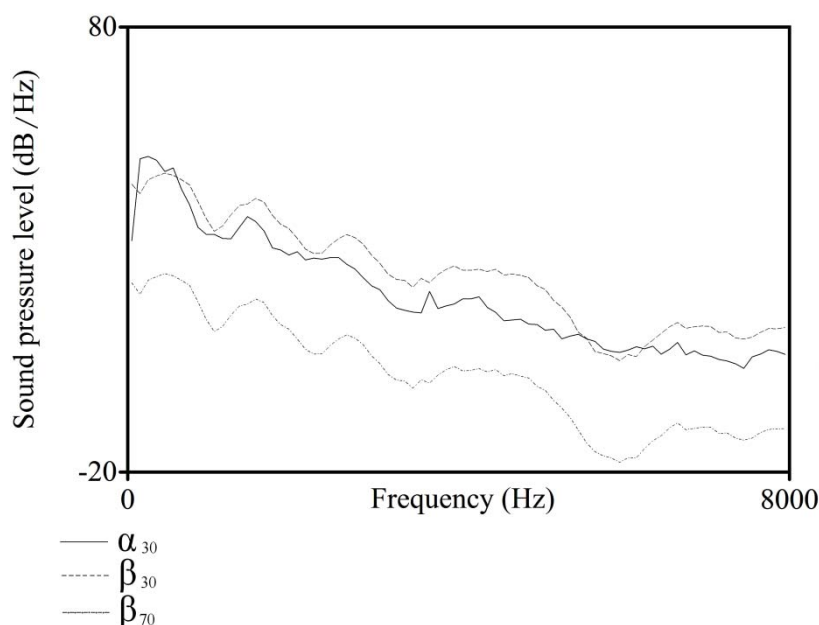


Figura 3. LTAS de las muestras  $\alpha_{30}$ ,  $\beta_{30}$  y  $\beta_{70}$

De acuerdo con las fórmulas para el cálculo de la distancia que se han presentado anteriormente, la diferencia entre el LTAS de  $\beta_{30}$  y  $\alpha_{30}$  sería mucho menor de la que hay entre los LTAS de  $\beta_{30}$  y  $\beta_{70}$ , a pesar de ser estas dos grabaciones de la misma emisión de voz. Por lo tanto, a partir del resultado de la fórmula en cuestión en un contexto judicial se tendría que rechazar la posibilidad de afirmar que  $\beta_{30}$  y  $\beta_{70}$  sean emisiones realizadas por el mismo hablante, ya que son las más diferentes del grupo. Esta conclusión es contraria a lo que sabemos que es cierto (y el ojo de un experto efectivamente reconoce que los LTAS de  $\beta_{30}$  y  $\beta_{70}$  tienen el mismo perfil, a pesar de encontrarse a alturas diferentes). El error deriva de la incapacidad de la fórmula antedicha para comparar eficazmente los perfiles de dos LTAS que, a pesar de tener un perfil

idéntico, se sitúan a alturas diferentes debido a la mayor/menor intensidad de grabación de la emisión.

La solución matemática que permite obviar errores tan evidentes consiste en calcular la media aritmética ( $\bar{x}$ ) de los LTAS que se comparan y utilizarla para neutralizar las diferencias *globales* de intensidad entre las muestras, mientras las diferencias *locales*, es decir los picos y valles del LTAS, se mantienen. Por ejemplo, si comparamos los LTAS de  $\beta_{70}$  y  $\beta_{30}$ , observamos que  $\bar{x}$  (LTAS  $\beta_{70}$ ) = 26 y  $\bar{x}$  (LTAS  $\beta_{30}$ ) = 4, por lo tanto la diferencia entre las dos es  $\Delta_{\bar{x}} = \bar{x}$  (LTAS  $\beta_{70}$ ) -  $\bar{x}$  (LTAS  $\beta_{30}$ ) = 22. Si, a continuación, sumamos a cada valor de la segunda secuencia del LTAS (en este caso  $\beta_{30}$ ) el valor de la de  $\Delta_{\bar{x}}$  neutralizamos el efecto de la diferencia global de intensidad. Si se quiere visualizar gráficamente esta corrección, en el caso de las curvas de los LTAS  $\beta_{30}$  y  $\beta_{70}$  de la figura 3 se puede decir que se desliza hacia arriba la línea discontinua inferior ( $\beta_{30}$ ) de 22 unidades, hasta sobreponerla a la superior ( $\beta_{70}$ ), con el resultado que se observa en la figura 4.

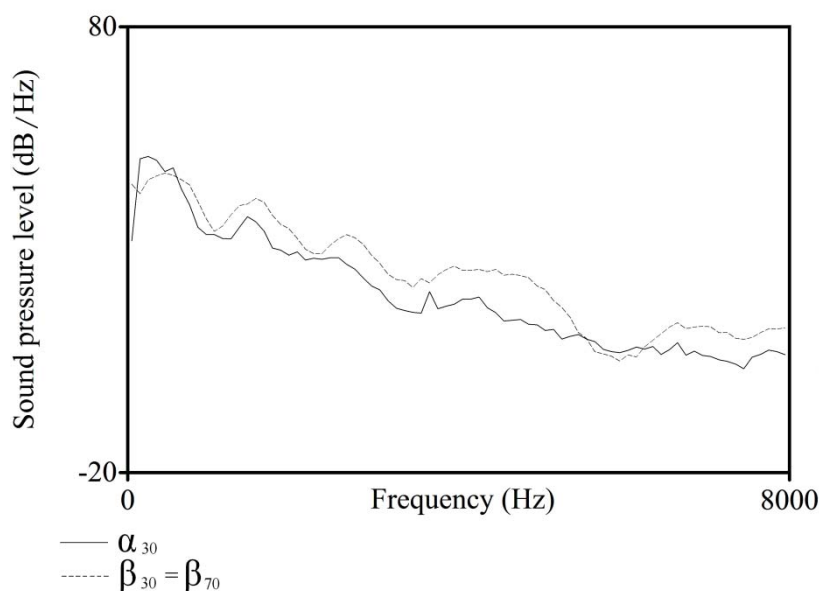


Figura 4. LTAS de las muestras  $\alpha_{30}$ ,  $\beta_{30}$  y  $\beta_{70}$  después de corregir los efectos de la intensidad global de la grabación

## 4 Resultados

Una vez calculadas las distancias entre el LTAS de cada muestra del corpus y los de todas las demás, se puede proceder a valorar si los resultados permiten confirmar nuestra hipótesis de trabajo, según la cual si es cierto que en hablantes bilingües el cambio de lengua no produce ningún efecto apreciable sobre el LTAS (o, en otras palabras, que el LTAS es un parámetro resistente al cambio de lengua), el  $\Delta$ LTAS entre las muestras del mismo sujeto hablando dos lenguas distintas debe ser inferior al  $\Delta$ LTAS entre cada una de las dos muestras del sujeto en cuestión y cualquier otra muestra de sujetos distintos. Esta hipótesis se cumple solo para 6 de los 12 informantes de nuestro corpus (figura 5).



contra de la hipótesis de la fiscalía (es decir que la muestra dubitada y la indubitada provienen del mismo hablante)<sup>1</sup>.

## 5 Conclusiones

Este trabajo, tal como se ha expresado en el apartado 2, tenía dos objetivos, de los cuales uno de carácter más teórico y el otro de tipo metodológico-aplicado. En cuanto a la pregunta de investigación teórica, se ha tenido que rechazar la hipótesis de trabajo, es decir que la diferencia entre los LTAS de las dos muestras del mismo sujeto *hablando lenguas distintas* debería ser inferior a la diferencia entre cada una de las esas muestras y cualquier otra muestra de sujetos distintos. Ya que esta hipótesis se ha rechazado, se ha tenido que concluir que el LTAS no parece un parámetro determinante en la comparación de muestras en lenguas distintas a la hora de avalar la hipótesis de la fiscalía, ni siquiera en condiciones experimentales ideales como las que se han utilizado (grabaciones de alta calidad, efectuadas en una sola sesión, de un número no especialmente grande de hablantes). No obstante, se ha observado como los datos proporcionados por la comparación automática de los LTAS, a pesar de no poder constituir un elemento a favor de la hipótesis de la fiscalía, podrían restarle credibilidad.

En cuanto al objetivo metodológico-aplicado, se ha podido definir un protocolo de análisis basado en un script de Praat que permite calcular automáticamente las distancias entre los LTAS de varias muestras. El script en cuestión, además de ser aplicable al caso de estudio concreto que se ha presentado, se puede utilizar en cualquier otra situación en que se quiera utilizar el LTAS como parámetro en la comparación de voces.

---

<sup>1</sup> Determinar la fuerza de esta evidencia contraria (es decir cuánto en contra de la hipótesis de la fiscalía están los datos del LTAS) y determinar la manera más adecuada para expresar esta conclusión constituyen un tema que trasciende las finalidades de este trabajo (v. Gil (2014) para una síntesis del debate sobre la presentación de los resultados de las comparaciones judiciales de muestras de habla).



## 6 Bibliografía

- Boersma, P., & Kovačić, G. (2006). Spectral characteristics of three styles of Croatian folk singing. *Journal of the Acoustical Society of America*, 119, 1805-1816.
- Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by compute. Version 5.3.57. Descargado el 27 de octubre de 2013, de <http://www.praat.org/>
- Bruyninckx, M., Harmegnies, B., Llisterri, J., & Poch, D. (1994). Language-induced voice quality variability in bilinguals. *Journal of Phonetics*, 22, 19-31.
- Cicres, J. (2011). Los sonidos fricativos sordos y sus implicaciones forenses. *Estudios Filológicos*, 48, 33-48.
- Elvira García, W., Roseano, P., & Fernández Planas, A. M. (2014). Ltas differences. Praat script. Descargado el 23 de setiembre de 2013, de <http://stel.ub.edu/labfon/en>
- Gil, J. (2014). Más allá del "efecto CSI": avances y metas en fonética judicial. En Y. Congosto, M. L. Montero Curiel, & A. Salvador (eds.), *Fonética Experimental, Educación Superior e Investigación*. Madrid: Arco Libros, vol. I, 63-112.
- Harmegnies, B., Bruyninckx, M., Llisterri, J., & Poch, D. (1991). Effects of language change in voice quality in bilingual speakers. Corpus content effects. En *Eurospeech 1991. Proceedings of the 2nd European conference on speech communication and technology*. Genova, Italy. 24-26 September, 1991. Vol. 1, 165-8.
- Klingholz, F., Penning, R., & Liebhart, E. (1988). Recognition of low-level of alcohol intoxication from speech signal. *Journal of the Acoustical Society of America*, 84(3), 929-935.
- Künzel, H. (2011). La prueba de voz en la investigación criminalística. *Ciencia Forense*, INACIPE-Academia Iberoamericana de Criminalística y Estudios Forenses, 1(1), 37-50.
- Majewski, W., & Hollien, H. (1974). Euclidean distance between long-term speech spectra as a criterion for speaker identification. *Speech Communication Seminar*, 1-3.
- Manwa, L. Ng, Chen, Y., & Chan, E. Y. K. (2010). Cantonese and English Produced by Proficient Cantonese-English Bilingual Speakers – A Long-Term Average Spectral Analysis. *Journal of Voice*, 26(4), 171-176.
- Molina de Figueiredo, R., & Bernales, M. (1999). Reconocimiento de hablantes basado en el espectro a largo tiempo. En *Actas del VI Simposio Internacional de Comunicación Social*. Santiago de Cuba: Centro de Lingüística Aplicada, 1372-1378.
- Nolan, F. J. (1983). *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.
- Nordenberg, M., & Sundberg, J. (2003). Effect on LTAS of vocal loudness variation. *TMH Quarterly Progress Status Report*, 1/2001.
- Pittam, J. (1987). The long term spectral measurement of voice quality as a social and personality marker: A review. *Language and Speech*, 30, 1-12.
- Pruzansky, S. (1963). Pattern-matching Procedure for Automatic Talker Recognition. *Journal of the Acoustical Society of America*, 35, 354-358.
- Zalewski, J., Majewski, W., & Hollien, H. (1975). Cross-correlation between Long-Term Speech Spectra as a criterion for speaker identification. *Acoustics*, 34, 20-24.

## 7 Agradecimientos

Queremos expresar nuestro agradecimiento al Institut für Romanistik de la Universität Hamburg, que ha puesto a disposición las grabaciones, y la Dra. Machuca de la Universitat Autònoma de Barcelona, puesto que gracias a una conversación mantenida con ella en el marco del posgrado en Estudios Fónicos del CSIC surgió la primera idea para este trabajo.