

LA INVARIACIÓN ACÚSTICA: INTRODUCCIÓN

E. MARTÍNEZ CELDRÁN, R. ESTAPÀ Y L. ROMERA
Laboratori de Fonètica, Facultat de Filologia
Universitat de Barcelona

RESUMEN

La evidente variación de las transiciones de los F2 para un mismo punto de articulación a causa de la vocal siguiente permitió a Liberman y colegas (1967) negar la posibilidad de parámetros acústicos invariantes en la percepción del habla. Ellos alegaron que la invariación existía en los mecanismos motores de la producción del habla. Pero la teoría de los "quanta" de Stevens (1972) y el fenómeno de la percepción categorial pusieron de manifiesto que las cavidades articulatorias producen patrones acústicos estables que pueden ser captados por el sistema perceptivo, lo cual restringe considerablemente el número de patrones fónicos que las lenguas pueden utilizar. Así, pues, esas restricciones apoyan la existencia de propiedades acústicas invariantes que el sistema perceptivo humano puede captar. Basándonos en esta hipótesis hemos realizado numerosos estudios buscando los parámetros acústicos invariantes en los sonidos del español, con la pretensión de que esto nos sirviera para la implementación del reconocimiento automático del habla basado en rasgos.

ABSTRACT

The obvious variation of F2 transitions for equal place of articulation, due to the influence of the following vowel, provided Liberman et col. (1967) with evidence against the existence of invariant acoustic parameters in speech perception. They claimed that there was invariance in the motor mechanisms of speech production. However, Stevens's *quantal theory* (1972) and the phenomenon of categorical perception made it clear that the articulatory cavities produce steady acoustic patterns which can be captured by the human perceptive system, thus limiting substantially the number of phonic patterns that can be used in a language. So then, these restrictions support the existence of acoustic invariant cues that can be processed by the human perceptive system. Having this hypothesis as basis, we have carried out various studies seeking invariant acoustic parameters for the sounds of Spanish, with the aim that this will serve us to implement an automatic speech recognition model based on phonetic features.

1. PRESENTACIÓN

En un artículo, considerado ya como un clásico, Alvin Liberman y sus colaboradores (1967) planteaban el porqué de la constancia en nuestra percepción del sonido [d] ante los cambios de las transiciones del F2 en las realizaciones de las sílabas [di] y [du] (Véase Fig. 1).

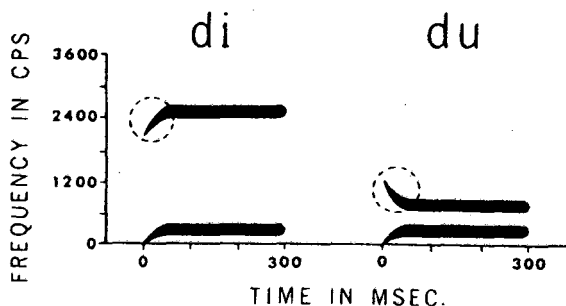


Fig. 1

Los autores concluyen que la invariación en nuestra percepción de estas consonantes no puede explicarse mediante el mero análisis de la señal acústica, ya que las pistas acústicas dependen del contexto en el que aparecen las constantes. Lo que se debe hacer es buscar algunos aspectos de los sonidos que permanezcan invariables con el cambio del contexto; pero esto no sucede en las ondas acústicas sino en la articulación. Es decir, el gesto articulatorio para producir un sonido permanece independiente del contexto, por tanto, nuestra percepción atiende más a la articulación que a la forma de las ondas sonoras (Teoría motora de la percepción del habla).

La conclusión de Liberman y sus colegas de los laboratorios Haskins permite una negación de la invariación acústica, a lo cual se oponen de forma rotunda Stevens y sus colegas del MIT, que, además, extienden su influencia a otros autores de diversas Universidades: S. E. Blumstein, A. Lahiri, S. J. Behrens, etc. Aunque es preciso indicar que esta no es la única corriente que se plantea el estudio y la defensa de la invariación; también lo harán B. Delgutte, D. Kewley-Port, H.M. Sussman, por ejemplo, desde otras perspectivas.

La cuestión planteada tiene una gran importancia desde diversos puntos de vista tanto teóricos como prácticos. Nuestro objetivo teórico es establecer un modelo de percepción de la cadena fónica por parte del oyente y validarlo mediante un modelo computacional de reconocimiento automático basado en rasgos; lo cual implica que en la señal sonora existen parámetros invariantes que se pueden capturar mediante reglas. Esos parámetros se asociarán con rasgos y estos, con fonemas. Si la variación acústica es tal que impide la asociación permanente de unos parámetros con unos rasgos, entonces el modelo computacional es imposible; pero estamos convencidos de lo contrario y, por ello, es fundamental llevar a cabo toda una investigación básica para buscar qué es realmente lo invariante en las ondas sonoras y qué se puede asociar con los rasgos fónicos.

1. PRESUPUESTOS TEÓRICOS

Es bien conocido el hecho de que los sonidos del habla comparten un conjunto limitado de características fundamentales. Específicamente, se ha hipotetizado que el ser humano nace con una predisposición a percibir y producir sonidos que son biológicamente significativos. Se ha sugerido que las propiedades definitorias del habla están determinadas por ciertas constricciones impuestas por el sistema articulatorio (véase, por ejemplo, la teoría cuántica de Stevens (1972 y 1989) por una parte y por el sistema perceptivo por otra, en la percepción categorial (B.H. Repp, 1984).

En un nivel, es obvio que los sistemas articulatorio y auditivo poseen ciertas limitaciones intrínsecas; esto es, un sonido no puede ser producido si cae fuera de la capacidad fisiológica del sistema articulatorio, o ser percibido si sus características acústicas están fuera del alcance de la capacidad del sistema auditivo. No obstante, el conjunto de los sonidos del habla parece que está restringido de un modo teórica y fundamentalmente más interesante. En particular, se ha demostrado que pequeños cambios en algunas configuraciones articulatorias producen cambios considerables en ciertos atributos de la forma de onda acústica resultante. En contraste, cambios similares en otras configuraciones articulatorias apenas afectan a los atributos de la onda sonora. Por tanto, sólo un conjunto limitado de configuraciones articulatorias produce patrones acústicos estables. Estos patrones acústicos son no sólo relativamente insensibles a las perturbaciones en las configuraciones articulatorias sino que además son distintivos en el sentido de que poseen atributos acústicos que son diferentes de las propiedades producidas por otras configuraciones. Es este conjunto de

configuraciones acústicas estables con sus propiedades distintivas el que define teóricamente el conjunto finito de los sonidos del habla usado en las lenguas naturales.

Stevens [1972] lo expresa mediante la figura siguiente, donde la región II posee diferencias acústicas considerables, pero mínimas desde el punto de vista articulatorio y en III sucede todo lo contrario:

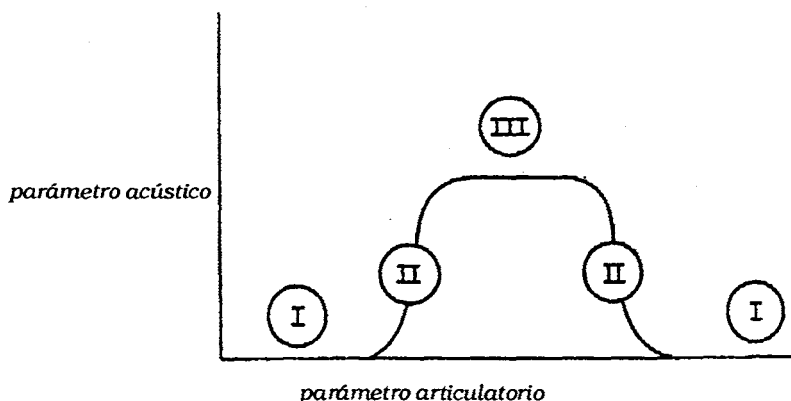


Fig. 2

Por tanto, es muy posible que una diferencia articulatoria como la mantenida por dentales y alveolares por ejemplo, apenas posea una repercusión acústica como muestra la región III. Lo cual no significa que no exista la diferencia articulatoria. Simplemente, no será relevante ni en la perspectiva acústica ni en la perceptiva. Stevens et alii ya sugirieron que, aunque los rasgos fonéticos de las lenguas naturales están caracterizados en términos de propiedades invariantes, estas propiedades necesitan emerger sólo cuando el correspondiente rasgo fonético cumple una función pertinente en la lengua.

La teoría cuántica pone de manifiesto que existen regiones naturales que propician una serie de categorías fonéticas y una de esas regiones parece estar formada por la zona dentoalveolar sin divisiones marcadas en ella. En las lenguas del mundo, es muy difícil que la región dentoalveolar se divida en dos para establecer distinciones pertinentes: "very few languages have phonologically contrastive pairs at the alveolar and dental places -they will have one or the other" (Henton et alii, 1992).

En cambio otra región próxima como la alveolar y la prepalatal propicia diferencias llamativas entre las fricativas que muchas lenguas utilizan de forma pertinente.

De modo similar, parece que existen constricciones en el sistema auditivo humano que también restringe el abanico posible de propiedades de los sonidos que pueden ser utilizados en las lenguas naturales. Esas limitaciones se ponen de manifiesto en el conocido fenómeno de la percepción categorial (E. Martínez Celdrán, 1993). Existen algunos continua acústicos en los que el oyente no es capaz de discriminar ni identificar todos los items que son equidistantes perceptivamente; antes bien, tiene dificultad de discriminar los pares adyacentes de sonidos en una parte del continuum, mientras que discrimina fácilmente los pares adyacentes en otra parte del continuum. La escala física en la que los sonidos difieren en distancias iguales no es la misma que la escala utilizada por el sistema perceptivo para juzgar las diferencias existentes entre los sonidos. La escala de ese sistema está deformada: en una parte de la escala se necesitan diferencias enormes para producir diferencias perceptivas mínimas, mientras que en otra parte de la escala, con pequeñas diferencias físicas se producen grandes diferencias perceptivas.

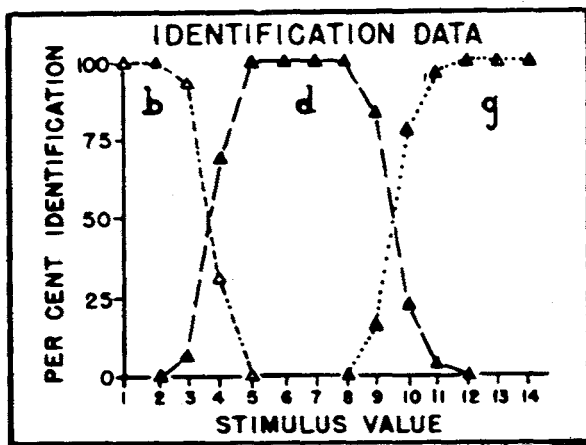


Fig. 3. Liberman et alii (1957) presentaron en un test 14 estímulos diferentes, pero los oyentes los categorizaron en tres conjuntos identificándolos como [b,d,g]. Ej.: entre 5 y 8 no perciben diferentes, mientras que entre 8 y 11 se produce un cambio de categoría.

Es, pues, este complejo "interface" de las constricciones articulatorias y perceptivas el que restringe teóricamente las propiedades de los sonidos del habla en los lenguajes naturales. No obstante, aún no se está seguro de cuál es el conjunto de las propiedades de los sonidos utilizadas tanto en la producción como en la percepción.

Tras la exposición de estos hechos, es evidente que existe una relación muy estrecha entre patrón acústico y configuración articulatoria, por una parte, y entre patrón acústico y percepción, por otra. Cuando un oyente no tiene ninguna dificultad para asignar un estímulo a una determinada categoría fonética, suele ser porque existe un patrón acústico estable, esto es, invariante, típico de esta clase fonética. Y si este sonido tiene patrones estables es porque se ha producido por una configuración articulatoria determinada. Por tanto, Stevens, Blumstein y sus colegas parten de dos hipótesis de trabajo en sus investigaciones fonéticas:

1ª Existe invariación acústica en la señal fónica que se corresponde con los rasgos fonéticos de las lenguas naturales. Y se mantiene que las pistas acústicas permanecen invariantes sean cuales sean los hablantes, los contextos y las lenguas.

2ª El sistema perceptivo humano es sensible a estas propiedades invariantes.

2. LA BÚSQUEDA DE INVARIANTES

En el período 92-94, hemos desarrollado el proyecto PB91-0278 sobre "la invariación acústica y el procesamiento del habla"; en él, después de replicar para el español los trabajos sobre el Punto de Articulación de las oclusivas que desarrollaron fundamentalmente para el inglés Blumstein y Stevens (1979, 1980, etc), resultó que sus conclusiones no eran válidas para nuestra lengua, por lo que tuvimos que investigar nuevos métodos: cálculo del locus mediante la regresión (Sussman, 1991, 1993...) y también mediante aplicación de ecuaciones de segundo y tercer grado a los formantes vocálicos que rodean a la consonante, lo cual produce sendas curvas que se intersectan en un punto que será característico para cada punto de articulación. En ambos casos hemos obtenido resultados positivos que señalan la posibilidad de ser utilizados en el reconocimiento automático.

No ha habido problema alguno en obtener los índices invariantes de las vocales basándonos en la frecuencia de los dos primeros formantes

como ya se había establecido en la tradición fonética desde los años cincuenta.

En el caso de las fricativas el estudio se centró en la amplitud de sus bandas ruidosas, partiendo del trabajo de Behrens y Blumstein (1988) para el inglés y adaptándolo a nuestra propia lengua (incluimos la fricativa [x]). Se trataba de contar con dos grupos de intensidad: por un lado [s] y [x] y, por otro, [f] y [θ] y manipular las secuencias de tal modo que las intensidades se intercambiaran entre ambos grupos para establecer justamente la intensidad, medida en dB, como el parámetro fundamental invariante acústicamente y reconocible e identificable perceptivamente. Los resultados fueron positivos al añadirle los índices de anchos de banda y transiciones vocálicas.

Respecto a las nasales, hay que decir que se consideró el murmullo como índice del modo de articulación y las transiciones hacia la vocal siguiente como índice del punto de articulación. En este sentido, nuestros estímulos (inspirados en Kurowski y Blumstein 1984) buscaban aislar los dos parámetros para ver cuál de ellos incidía más en el reconocimiento perceptivo posterior que nos llegaría a confirmar los índices acústicos invariantes y hacerlo con distintas medidas de tiempo. En este estudio tuvimos que incluir nuestra nasal palatal (no la habían incluido los autores americanos). Los resultados mostraron que en realidad es la conjunción de los dos parámetros estudiados la que proporciona los índices necesarios para percibir correctamente tanto el punto y el modo de articulación, como el timbre de la vocal siguiente, en muchos casos, con lo cual nos acercamos a la idea según la cual los índices invariantes hay que buscarlos en realidad en el seno de la sílaba.

En las vibrantes se ha demostrado que las breves oclusiones, o si se desea los cambios bruscos de la intensidad, son índices suficientes para su identificación: una para la vibrante simple y dos o más para la múltiple.

Se puede decir que hemos partido, siempre que ha sido posible, de la replicación de los experimentos del equipo formado por S. Blumstein, K. Stevens y otros; y para todo aquello que no ha podido ser validado para el español hemos buscado nuevas hipótesis y métodos que nos proporcionasen un buen resultado final, como, por ejemplo, ha sucedido en el punto de articulación de las oclusivas, que ha sido la dificultad mayor con la que nos hemos tropezado.

A continuación se exponen todos estos experimentos que nos han permitido llegar a formarnos una idea más precisa de los parámetros invariantes existentes en la señal acústica.

3. REFERENCIAS

- BEHRENS, S.J. y BLUMSTEIN, S.E., (1988), "Acoustic characteristics of English voiceless fricatives: a descriptive analysis", *Journal of Phonetics*, 16, 295-298.
- (1988), "On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants", *JASA*, 84(3), 861-867.
- BLUMSTEIN, S. E. (1986), "On acoustic invariance in speech" en PERKELL, J.S. y KLATT, D.H. (Eds), *Invariance and Variability in Speech Processes*, Hillsdale NJ, Erlbaum.
- BLUMSTEIN, S.E. y STEVENS, K.N. (1979), "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants", *JASA*, 66(4), 1001-1017.
- (1980), "Perceptual invariance and onset spectra for stop consonants in different vowel environments", *JASA*, 67(2), 648-662.
- (1981), "Phonetic features and acoustic invariance in speech", *Cognition*, 10, 25-32.
- (1985), "On some issues in the pursuit of acoustic invariance in speech: a reply to Lisker", *JASA*, 77(3), 1203-1204.

- BURTON, M.W.; BLUMSTEIN, S.E. y STEVENS, K.N. (1992): "A phonetic analysis of prenasalized stops in Moru", *Journal of Phonetics*, 20(1), 127-142.
- DELGUTTE, B. (1986), "Analysis of french stop consonants using a model of the peripheral auditory system", en J. S. PERKELL y D. H. KLATT (Eds.), *Invariance and variability of Speech processes*, Hillsdale, N.J., Lawrence Erlbaum Associates, Inc.
- DELGUTTE, B. y KIANG, N. Y. S. (1984) "Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics", *JASA*, 75, 897-907.
- HENTON, C.; LADEFOGED P. y MADDIESON, Y. (1992), "Stops in the World's languages", *Phonetica*, 49, 2, p.95.
- JONGMAN, A., BLUMSTEIN, S.E. y LAHIRI, A. (1985), "Acoustic properties for dental and alveolar stop consonants: a cross-language study", *Journal of Phonetics*, 13, 235-251.
- KUROWSKI, K. y BLUMSTEIN, S.E (1984), "Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants", *JASA*, 76(2), 383-390.
- LAHIRI, A. y BLUMSTEIN, S.E. (1981), "A reconsideration of acoustic invariance in stop consonants: evidence from cross-language studies", *JASA*, suppl 1, 70, S39.
- LAHIRI, A., GEWIRTH, L. y BLUMSTEIN, S.E., (1984), "A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: evidence from a cross-language study", *JASA*, 76(2), 391-404.
- LIBERMAN, A.M.; COOPER, F.S.; SHANKWEILER, D.P. y STUDDERT-KENNEDY, M. (1967): "Perception of speech code", *Psychol. Rev.*, 74, 431-461.
- MACK, M. y BLUMSTEIN, S.E. (1983): "Further evidence of acoustic invariance in speech production: the stop-glide contrast", *JASA*, 73(5), 1739-1750.
- MARTÍNEZ CELDRÁN, E. (1993): "La percepción categorial de /b-p/ en español basada en las diferencias de duración", *Estudios de Fonética Experimental*, V, Barcelona, PPU, 223-239.

- REPP, B. H. (1984): "Categorical perception: Issues, methods, findings", en LASS, N. J. (Ed.), *Speech and Language: Advances in Theory and Practice*, vol. 10, New York, Academic Press, 243-335.
- STEVENS, K.N. (1972): "The quantal nature of speech: evidence from articulatory -acoustic data", en DENES, P. B. y DAVIS, E. E. (Eds.), *Human Communication, An Unified View*, New York, McGraw-Hill, 51-66
- STEVENS, K.N. (1989): "On the quantal nature of speech", *Journal of Phonetics*, 17, 3-45.
- STEVENS, K.N. y BLUMSTEIN, S. E. (1975): "Quantal aspects of consonant production and perception: a study of retroflex consonants", *Journal of Phonetics*, 3, 215-234.
- (1978), "Invariant cues for place of articulation in stop consonants", *JASA*, 64, 1358-1368.
- (1981), "The search for invariant acoustic correlates of phonetic features", en EIMAS, P.D. y MILLER, J.L. (Eds), *Perspectives on the Study of Speech*, Hillsdale NJ, Erlbaum.
- STEVENS, K.N, BLUMSTEIN, S.E., GLICKSMAN, L., BURTON, M. y KUROWSKI, K. (1992), "Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters", *JASA*, 91(5), 2979-3000.
- STEVENS, K.N.; KEYSER, S.J. y KAWASAKI, H. (1986), "Toward a phonetic and phonological theory of redundant features", en J. S. PERKELL y D.H. KLATT (Eds.), *Invariance and variability of Speech processes*, Hillsdale, N. J., Lawrence Erlbaum Associates, Inc.
- SUSSMAN, H.M.; MCCAFFREY, M.A. y MATTHEWS, S.A. (1991), "An investigation of locus equations as a source of relational invariance for stop place categorization", *JASA*, 90, 1309-1325.

JASA = Journal of Acoustical Society of America