

Efectos temporales de la integración fonética en sílabas FV

S. Feijóo, S. Fernández

Dpto. de Física Aplicada, Fac. de Física, Universidad de Santiago

15706 Santiago de Compostela

En este estudio investigamos la posible relación entre las características temporales de las sílabas fricativa-vocal (FV) y la percepción del lugar y modo de articulación. Para ello hemos realizado una serie de experimentos de percepción en los que se les presentaban a los oyentes porciones cada vez más largas de señal FV, comenzando desde un punto situado dentro de la vocal hasta un punto situado dentro del ruido fricativo. Los resultados muestran que los oyentes tienen dificultades en percibir el lugar de articulación de la fricativa cuando sólo disponen de la vocal. Para que el lugar de articulación pueda ser percibido, es necesario que la vocal vaya acompañada de una parte del ruido fricativo. Dependiendo de cada fricativa en particular, la percepción pasa de un modo plosivo cuando el ruido fricativo es de una duración entre 20-30 ms, hasta un modo fricativo para duraciones superiores. Sin embargo, para las fricativas sibilantes existe un estadio intermedio en el que predomina un modo africado. Estos resultados cuestionan la validez de ciertos modelos de integración fonética, y muestran la importancia de ciertos factores temporales en la percepción fonética.

1. Introducción

Uno de los fenómenos más importantes en la percepción del habla es la capacidad del sistema auditivo humano de integrar diversas fuentes de información para obtener una determinada percepción fonética. En este trabajo nos centramos en el estudio de un tipo de integración que tiene lugar en las sílabas formadas por consonante y vocal (CV), en concreto la interacción que tiene lugar entre el ruido fricativo (F) y la vocal que lo acompaña (V) en sílabas FV. Es bien sabido que el reconocimiento auditivo de las fricativas mejora cuando los oyentes escuchan la sílaba FV en relación al reconocimiento cuando sólo escuchan el ruido fricativo (Harris, 1958). Es decir, se produce una interacción fonética entre el ruido fricativo y la vocal. Esta interacción es particularmente importante para las fricativas no sibilantes (/T,ϕ/) (Feijóo *et al.*, 1999). En esos casos, el ruido fricativo no contiene toda la información necesaria para la identificación auditiva de la fricativa.

Una de las posibles hipótesis que podrían explicar dicha interacción se basa en la influencia coarticulatoria existente en las sílabas FV. La integración fonética podría ser el resultado de :

La influencia coarticulatoria de la vocal sobre la consonante: la vocal ejerce una influencia anticipatoria sobre la fricativa, de tal forma que las características de la fricativa están condicionadas por la “calidad” de la vocal que la sigue. Un caso típico es el redondeamiento de los labios que tiene lugar en la vocal /u/, que influye en las características espectrales de las fricativas /s/ y /S/, haciendo que su contenido espectral esté desplazado hacia las bajas frecuencias, en relación con las características de las mismas fricativas en los contextos de las vocales /a/ o /i/ (Johnson, 1991).

La influencia coarticulatoria de la consonante sobre la vocal: El lugar donde se produce la constricción fricativa (lugar de articulación) varía con las diferentes fricativas, y condiciona la transición de la vocal que la acompaña. Por ejemplo, la fricativa /s/ tiene un lugar de articulación alveolar, mientras que /f/ es labial o labiodental. Los movimientos articulatorios que se producen para una misma vocal en ambos contextos fricativos son diferentes (Whalen, 1981).

En un trabajo anterior exploramos ambas posibilidades usando “características conflictivas” (Fernández y Feijóo, 2000). Los resultados indicaron que la hipótesis b) podría jugar un cierto papel en la integración FV, aunque también aparecieron algunos efectos asociados al contexto vocálico. Sin embargo, ninguna variable acústica parecía ser exclusivamente responsable de dicha influencia coarticulatoria de modo claro, lo que nos llevó a desarrollar un modelo probabilístico en el que fricativa y vocal son evaluadas por separado, y el resultado final depende de la combinación de las probabilidades de pertenencia a una determinada clase fricativa asociadas a ambos segmentos. Para evaluar el modelo realizamos una serie de experimentos de percepción en los que intentamos usar el sistema auditivo humano como una especie de “análizador” que nos permitiera evaluar la influencia del lugar de articulación de la fricativa sobre la vocal (Fernández *et al.*, 2000). Los resultados no fueron concluyentes debido principalmente a que no existía una relación clara entre el lugar de articulación reflejado en la vocal y la percepción de los oyentes, cuando estos disponían sólo de la vocal para la audición. En este trabajo, ampliamos el campo de estudio sobre diferentes duraciones de la sílaba FV para comprobar los efectos temporales de la integración fonética en dichas sílabas sobre la percepción del lugar y modo de articulación.

2. Experimentos de percepción

2.1 Estímulos

Las señales usadas en los experimentos de percepción estaban formadas por la combinación de las fricativas /ʈ,ɸ,σ,Σ/ con las vocales /a,e,i,o,u/, pronunciadas por un hombre y una mujer gallegos. Para construir los estímulos que fueron usados en los experimentos de percepción, se procedió, en primer lugar, a determinar el lugar de unión entre la fricativa y la vocal. A continuación seleccionamos segmentos cada vez mayores de señal a partir de un punto situado en la vocal a 40 ms de la unión entre fricativa y vocal. La duración del primer segmento era de 60 ms, y los segmentos siguientes eran de duraciones que se incrementaban en pasos de 10 ms, cada vez incluyendo más porciones de señal hacia la unión FV, primero, y después incluyendo segmentos subsiguientes de fricativa a partir de dicha unión. La figura 1 muestra los diversos pasos en la construcción de los estímulos. La condición de offset 0 ms se corresponde con el estímulo formado por 100 ms de vocal; la condición de offset 40 ms se corresponde con el estímulo formado por 60 ms de la vocal a partir de un punto situado a 40 ms de la unión FV; la condición de offset -50 se corresponde con el estímulo formado por 150 ms de la señal a partir de un punto situado a 50 ms de la unión FV dentro de la fricativa, etc. Los extremos de los estímulos se suavizaron con una ventana semisenoidal de 10 ms de duración, para evitar la aparición de "clics" que afectaran a la percepción de los segmentos. Por tanto el número total de estímulos usados en los experimentos fue de 2 hablantes x 4 fricativas x 5 vocales x 15 condiciones = 600.

2.2 Oyentes

Cincuenta estudiantes de la Escuela Universitaria de Óptica, con edades comprendidas entre 20 y 30 años, participaron como oyentes en el experimento como parte de sus créditos de prácticas de la asignatura de Acústica. Ninguno de ellos presentaba problemas de audición, y todos hablaban y comprendían el gallego.

2.3 Presentación de los estímulos y respuestas al experimento

Los estímulos se presentaron directamente a través de auriculares, y a un nivel de intensidad confortable (alrededor de 70 dB SPL). Dada la naturaleza del experimento, como opciones de respuesta se consideraron las siguientes : vocal, diptongo, p, b, t, d, ch, z, f, s, x. Las 6 primeras respuestas se pensaron especialmente para las condiciones en las que no hubiera ningún segmento de fricativa en los estímulos: si los oyentes no percibían ninguna consonante inicial deberían responder “vocal” o “diptongo”, este último caso especialmente asociado a la transición de la fricativa Σ /, que suele dar lugar a la aparición de un diptongo que comienza por /i/; si los oyentes perciben alguna consonante inicial en los estímulos, probablemente serán plosivas, sonoras o sordas, asociadas al lugar de articulación correspondiente a la fricativa; así tendremos, por ejemplo, que las fricativas dentales y alveolares (/T/ y /σ/) podrán dar lugar a la percepción de una /t/ o /d/, mientras que /f/ puede dar lugar a una percepción de plosiva labial /p/ o /b/. Las otras respuestas se corresponden con las fricativas z - /T/, f - /φ/, s - /σ/, x - /Σ/. La respuesta ch se corresponde con la africada /τΣ/, y se consideró necesaria especialmente para /Σ/, y en menor grado para /σ/, debido a que en algunas condiciones el acortamiento de la duración natural de la fricativa y la subida brusca de amplitud podían hacer que se percibiera una africada en lugar de una fricativa.

Los estímulos se presentaron en bloques correspondientes a cada hablante. El orden de presentación de los estímulos fue aleatorio y diferente para cada oyente. Se explicó la naturaleza del experimento a los oyentes indicando que debían seleccionar tan sólo una de las respuestas. Estos disponían, además, de un guión explicativo que podían consultar siempre que lo considerasen oportuno. Se permitió que los oyentes repitieran la audición de cada estímulo una sola vez, después de la cual deberían seleccionar obligatoriamente una respuesta. Los propios oyentes realizaron el experimento al ritmo que consideraron oportuno, monitorizando su tarea a través de la pantalla de un ordenador, en el cual “pinchaban” para señalar sus respuestas. El programa de percepción utilizado controlaba todo el proceso, tanto la presentación de los estímulos como el almacenamiento de las respuestas y el cálculo de los perfiles de respuesta individuales y totales de cada experimento.

3. Resultados

Debido a la gran cantidad de datos y respuestas posibles, la información más relevante del resultado de los experimentos se ha condensado en las figuras 2- 9. Las respuestas correspondientes a las plosivas sonoras y sordas correspondientes al mismo lugar de articulación se juntaron bajo una denominación común : labial (/p/ o /b/), dental (/t/ o /d/).

Las figuras 2 y 3 muestran los resultados para la fricativa / ϕ /, para los estímulos del hombre y de la mujer. La percepción de la plosiva labial es predominante entre los 10 y -30 ms. La percepción de la fricativa parece ocurrir entre -40 y -60 ms, dependiendo del hablante particular. La percepción de sólo una vocal se produce sorprendentemente rápidamente, alrededor de los 10 ms para el hombre y en la condición de offset 0 para la mujer. En conjunto, los resultados indican que el lugar de articulación que se refleja en la

vocal es ambiguo, incluso en un caso claro como el de la fricativa / ϕ /, en el cual sería de esperar que, por ejemplo, en la condición 0, se produjera una percepción clara de una plosiva labial. Esa percepción parece estar asociada a la inclusión de una pequeña parte del ruido fricativo de duración aproximadamente igual a la del ruido plosivo de una plosiva labial sorda.

Las figuras 4 y 5 muestran los resultados para / π / . El primer resultado sorprendente (aunque no totalmente inesperado) es que cuando los oyentes escuchan la vocal sin ningún ruido fricativo, la percepción predominante parece ser la de un lugar de articulación labial. El lugar de articulación dental se percibe cuando la vocal se escucha acompañada de un segmento de ruido de aproximadamente 30 ms de duración, que corresponde también de manera aproximada a la duración del ruido plosivo de una plosiva alveolar sorda. Este resultado podría interpretarse como un efecto de sesgo en la percepción de estímulos procedentes de sílabas CV naturales en los que el ruido de la consonante está ausente: la ausencia de ruido parece favorecer la percepción de una plosiva labial, independientemente del lugar de articulación original (Smits, 1996). A pesar de eso, la percepción predominante a partir de la condición de offset 0 es la de una vocal. La excepción a este caso se ve reflejada en la curva sólida, que corresponde a la percepción de la fricativa en los estímulos de la mujer: entre 0 y 10-20 ms a partir del comienzo de la vocal parece existir una fuente adicional de información en esos estímulos, que podría corresponderse a la presencia de parte del flujo de aire correspondiente al ruido fricativo en la vocal. Al igual que en el caso anterior, la percepción de la fricativa parece estabilizarse a partir de las condiciones de -40 ó -60 ms.

Las figuras 6 y 7 muestran los resultados para / Σ / . La percepción de un diptongo con /i/ como vocal inicial predomina cuando los oyentes disponen de tan sólo la vocal como estímulo. La presencia de esa /i/ inicial parece ocupar los primeros 20ms de la vocal, predominando la percepción de la vocal V correspondiente a partir de ese punto. Sin

embargo, y a diferencia de lo que ocurre con los estímulos contruidos a partir de /T/, en este caso la plosiva labial es percibida con mucha menor claridad (no se presentan datos por ser de valor muy pequeño). Igual que en el caso de /T/, cuando la vocal está acompañada de un ruido de unos 10-30 ms de duración, predomina la percepción de una plosiva dental/alveolar. A medida que se añade más ruido, la percepción predominante es la de una africana / $\tau\Sigma$ /, siendo necesarios unos 80-90 ms de ruido para que comience a ser predominante la percepción de la fricativa Σ /. La percepción de la africana podría deberse al acortamiento del ruido fricativo, ya que las características espectrales de ambos fonemas son muy similares. Otra hipótesis es la ya mencionada de que el aumento brusco de amplitud debido al corte de la señal no pueda ser compensado por la ventana de suavizado, y al coincidir el punto de corte con una zona del ruido de amplitud relativamente alta, se acorte el “tiempo de subida” del ruido fricativo confundiendo con el de una africana (Howel y Rosen,1983). Es difícil establecer la validez de ambas hipótesis, ya que no parece existir una explicación satisfactoria hasta el momento acerca de las diferencias acústicas entre / Σ / y / $\tau\Sigma$ /.

Las figuras 8 y 9 muestran los resultados para / σ /. Cuando los oyentes sólo disponen de la vocal, predomina la percepción de una vocal. En el caso de los estímulos de la mujer, y al igual que sucede con /T/, parte del flujo de aire del ruido fricativo parece haber pasado a los primeros 30 ms de la vocal, dando lugar a la percepción de una /T/. La percepción de una plosiva dental/alveolar domina fuertemente las condiciones entre 0 y -40 ms, con mayor fuerza que en /T/, quizá debido a que el lugar de articulación de / σ / es más parecido al de /t/ o /d/ que el de /T/. Un tanto sorprendente es la aparición de la percepción de una africana, ya que, aunque las características espectrales de / σ / y Σ / tienen algunos puntos en común, los ruidos de ambas fricativas son perfectamente distinguibles por parte de los oyentes. La percepción de la fricativa comienza a predominar a partir de que se añaden unos 60 ms de ruido, que es un período más corto que en el caso de Σ /, probablemente debido a la diferencia entre las características espectrales de / σ / y / Σ / en condiciones naturales.

4. Discusión

Los resultados obtenidos muestran claramente que la percepción del lugar y modo de articulación en sílabas FV depende de numerosos factores, entre los cuales el factor temporal tiene una importancia particular. En primer lugar, y a pesar de la existencia de ciertas diferencias individuales entre los dos hablantes, los resultados parecen ser bastante consistentes, en el sentido de que los mismos efectos principales aparecen en los dos hablantes. En segundo lugar, existe una clara incapacidad por parte de los oyentes de extraer información relativa al lugar de articulación de la parte vocálica de la sílaba. Este hecho cuestiona las teorías de integración fonética basadas en la evaluación separada de fricativa y vocal. Incluso para condiciones de duración intermedia, en las que predominan percepciones de modo plosivo o africada, parece necesario que la morfología de la señal sea comparable a la de las señales naturales para que la percepción sea más o menos clara. De alguna manera podríamos decir que el modo de articulación resultante de las señales construidas a partir de las originales afecta al lugar de articulación percibido.

Los valores de duración necesarios para la correcta percepción del modo de articulación fricativo no coinciden con los obtenidos por (Jongman, 1989). Los métodos utilizados en ambos trabajos son diferentes, ya que en el nuestro hemos comenzado a añadir porciones de señal desde el punto de unión entre fricativa y vocal, y no desde el inicio del ruido fricativo.

Los resultados presentados, sin embargo, no nos permiten averiguar cuál es la posible causa de esa incapacidad del sistema auditivo para evaluar el contenido fonético de segmentos escindidos de la señal original. Parece claro que el diseño de ciertas condiciones en las que se suelen llevar a cabo algunos experimentos de percepción, en los que se “cortan” las señales, puede llevar a conclusiones precipitadas. Sin embargo, el

problema se acentúa cuando consideramos que tampoco las variables acústicas medidas sobre esos segmentos “cortados” tienen la información precisa que permita resolver el dilema (Fernández *et al.*, 1999).

REFERENCIAS

- Fernández, S. & S. Feijóo. A probabilistic model of integration of acoustic cues in FV syllables. En *Proceedings of ICSLP 2000* (CD-ROM), 2000.
- Fernández, S., S. Feijóo, R. Balsa & N. Barros. Perceptual effects of coarticulation in fricatives. *Proceedings of ICASSP 2000* (CD-ROM), 2000.
- Feijóo, S., S. Fernández, N. Barros & R. Balsa. Context effects and acoustic cues for the auditory identification of Spanish fricatives /f/ and /T/. (En proceso de revisión), 1999.
- Harris, K. S. Cues for the discrimination of American English fricatives in spoken syllables. En *Lang. Speech*, 1, 1-7, 1958.
- Howel, P. & S. Rosen. Production and perception of rise time in the voiceless affricate/fricative distinction. En *J. Acoust. Soc. Am.*, 73, 976-984, 1983.
- Jongman, A. Duration of frication noise required for identification of English fricatives. En *J. Acoust. Soc. Am.*, 85, 1718-1725, 1989.
- Johnson, K. Differential effects of speaker and vowel variability on fricative perception. En *Lang. Speech*, 34, 265-279, 1991.
- Smits, R., L. Ten Bosch & R. Collier. Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants.I. Perception experiment. En *J. Acoust. Soc. Am.*, 100, 3852-3864, 1996.
- Whalen, D. H. Effects of vocalic formant transitions and vowel quality on the English [s-Σ] boundary. En *J. Acoust. Soc. Am.*, 69, 275-282, 1981.

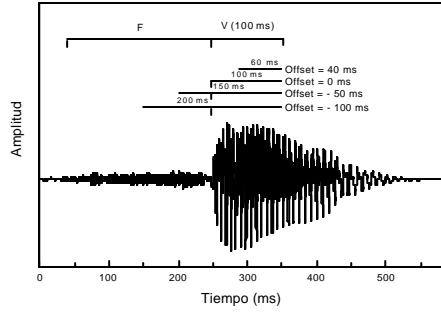


Figura 1: Construcción de los estímulos empleados en los experimentos de percepción.

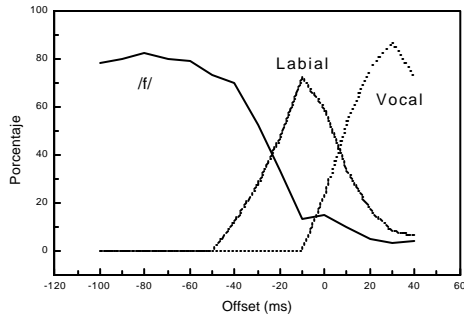


Figura 2: Resultados para la fricativa /f/. Estímulos pronunciados por un hombre.

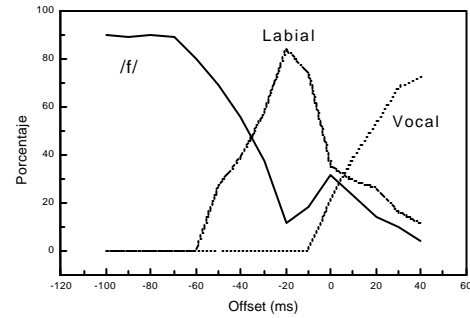


Figura 3: Resultados para la fricativa /f/. Estímulos pronunciados por una mujer.

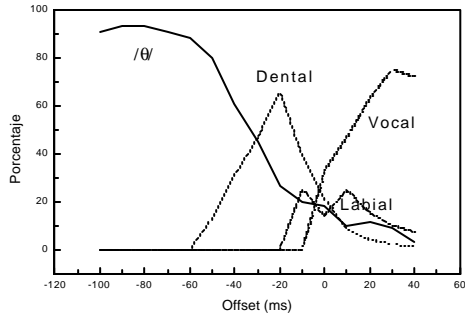


Figura 4: Resultados para la fricativa /θ/.
Estímulos pronunciados por un hombre.

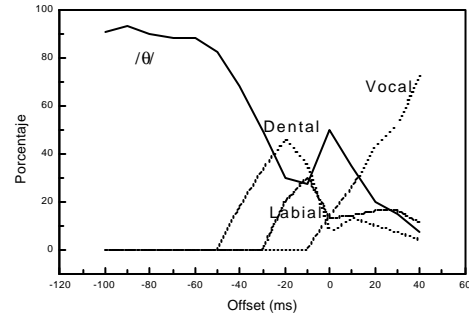


Figura 5: Resultados para la fricativa /θ/.
Estímulos pronunciados por una mujer.

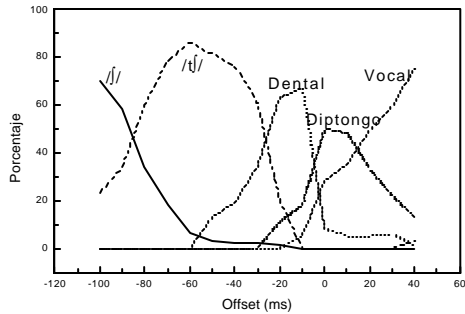


Figura 6: Resultados para la fricativa /ʃ/.
Estímulos pronunciados por un hombre.

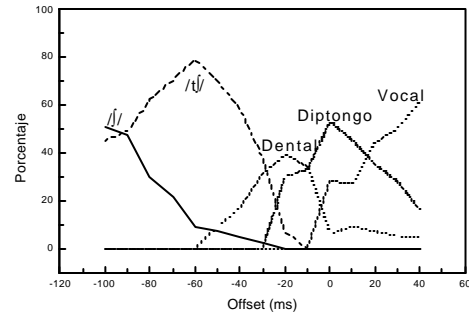


Figura 7: Resultados para la fricativa /ʃ/.
Estímulos pronunciados por una mujer.

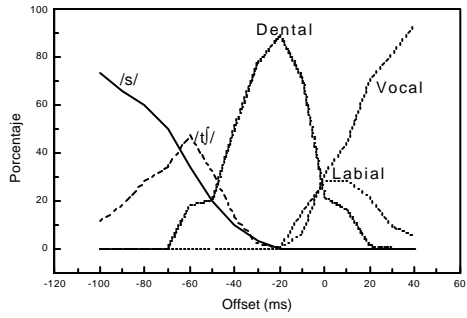


Figura 8: Resultados para la fricativa /s/.
Estímulos pronunciados por un hombre.

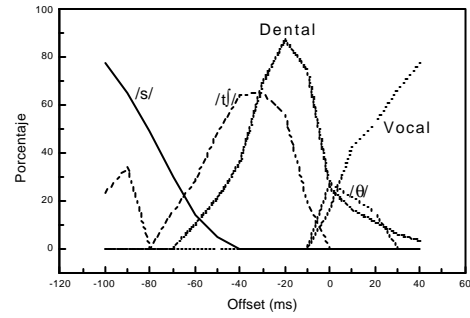


Figura 9: Resultados para la fricativa /s/.
Estímulos pronunciados por una mujer