



ACÚSTICA DE AULAS: PERCEPCIÓN FONÉTICA EN PRESENCIA DE RUIDO

PACS: 43.71.Gv

S. Feijóo, J.M. Alvarez
Dpto. Física Aplicada, Facultad de Física, Universidad de Santiago de Compostela
15782 Santiago de Compostela. España
Tel: 981 563 100 ext. 14044
Fax: 981 520 676
E-mail: fasergio@usc.es

ABSTRACT

The results of a closed format phonetic perception test performed in the presence of pink noise with different Signal-to-Noise ratio are presented. Stimuli consisted in 90 CV syllables uttered by 2 men and 2 women. Results show similar error rates throughout SNR conditions. Nevertheless, the degree of consistency among results is low, particularly due to the inter-speaker variability, questioning the assumptions underlying this type of test: that the relation between phonetic identification and the acoustic characteristics of the stimuli in certain conditions determine the outcome of the test.

RESUMEN

En este trabajo presentamos los resultados de un test cerrado de percepción fonética realizado en presencia de ruido de fondo rosa con diferentes Relaciones Señal-Ruido. Los estímulos consistieron en 90 sílabas CV pronunciadas por 2 hombres y dos mujeres. Los resultados muestran porcentajes de error similares en todas las condiciones. Sin embargo, el grado de consistencia es bajo, debido principalmente al factor hablante, lo que cuestiona las suposiciones sobre las que se basa el test: la relación entre la identificación fonética y las características acústicas de los estímulos en determinadas condiciones.

INTRODUCCION

Durante los últimos años se ha venido incrementando la concienciación acerca de la calidad acústica de las aulas de centros de educación, tanto en la enseñanza primaria como en la secundaria o en la Universidad [1-4]. Dado que la transmisión oral de un mensaje en un aula sigue siendo el medio habitual de impartir conocimiento, puede existir una interacción importante entre las condiciones acústicas del aula y las características acústicas del mensaje. Esta interacción puede influir notablemente en las características acústicas de la señal que percibe el alumno, provocando una pérdida de inteligibilidad en el mensaje [5-7].

Existen circunstancias especiales que pueden agravar esta pérdida. Por ejemplo, en ciertas edades escolares los mecanismos auditivos de comprensión del lenguaje aún no están suficientemente desarrollados en comparación con la edad adulta, lo que implica que los alumnos no son capaces de usar con la máxima eficiencia todas las fuentes de información disponibles en el lenguaje hablado. Si las condiciones de transmisión del mensaje no son óptimas desde el punto de vista acústico (alto nivel de ruido ambiente, reverberación excesiva), es fácil que se produzca una mala comprensión del mensaje, con el inevitable retraso en la formación del alumno. Existen circunstancias adicionales que pueden agravar este hecho, como por ejemplo la alta incidencia durante los primeros años de edad escolar de enfermedades del oído, como la otitis media, que afectan seriamente las capacidades auditivas, y que en muchas ocasiones pasan desapercibidas durante algún tiempo. Una nueva circunstancia es el auge de la inmigración, que ha hecho que durante los últimos años haya aumentado la asistencia a nuestras aulas de alumnos cuya lengua materna es diferente de la lengua en la que se imparte la docencia. Estos alumnos tienen una dificultad mayor que el resto para comprender el mensaje del profesor, debido a su falta de eficiencia en el idioma oficial y a su mayor sensibilidad frente a la pérdida de calidad acústica en el mensaje.

Todos estos factores han llevado al desarrollo de una serie de técnicas de evaluación de la calidad acústica de las aulas en función de su efecto sobre la inteligibilidad del habla. Las medidas acústicas más habituales relacionadas con la inteligibilidad son índices como el *Speech Transmission Index (STI)*, o el *Articulation Index (AI)* [2]. Dichos índices están supuestamente relacionados con la inteligibilidad del habla, entendida ésta como el porcentaje de reconocimiento que demostrarían una serie de oyentes ante una serie de estímulos del habla. En estos estudios se han usado diferentes tipos de test de percepción, muchos de los cuales parten de consideraciones que no tienen por qué aplicarse a la acústica de salas, ni están relacionadas con el proceso de aprendizaje. Existe una gran controversia hoy en día sobre qué tipo de materiales deben ser usados, su validez, consistencia, adecuación al propósito del test, etc. (ver [8] para una revisión sobre el tema). La ventaja del uso de índices como el STI o el AI es que pueden ser obtenidos de forma sistemática mediante mediciones acústicas. Su desventaja es que la relación entre esos índices y la inteligibilidad aún no está suficientemente demostrada.

Existen diferentes tipos de tests auditivos, diseñados principalmente para el idioma inglés, en función de objetivos concretos. Por ejemplo, un test diseñado para comprobar la calidad de transmisión de un sistema de comunicación no tiene por qué ser exactamente igual a un test diseñado para realizar una audiometría vocal, cuya característica principal es mostrar capacidad de diagnóstico. Por otra parte, los estímulos del test pueden haber sido producidos por un *único hablante*, con lo que evitamos la variabilidad entre individuos, pero limitamos su reproducibilidad, o pueden haber sido producidos por un *grupo de hablantes*, con lo que sus resultados son más generalizables. Los estímulos producidos por un *hablante* pueden ser mensaje *grabados* o de *viva voz*, mientras que si intervienen varios hablantes probablemente tendrán que ser *grabados*. Otro factor importante es el tipo de estímulos, que pueden ser *sílabas* o *palabras sin sentido*, o bien *palabras del lenguaje natural*, que a su vez se pueden presentar al oyente de forma *aislada* o incluidas *dentro de frases*. En cuanto a las respuestas del test, este puede tener un formato *abierto*, en el que el oyente debe reproducir de forma oral o escrita el estímulo percibido, o bien puede ser *cerrado*, en el que se le presentan al oyente un grupo de respuestas, y éste debe escoger entre una de ellas. En cuanto al tipo de resultado, podemos contar los fallos que se producen en cada *fonema* las *palabras* o *frases* falladas, etc. Todos estos factores y algún otro adicional condicionan los resultados del test.

Los parámetros acústicos que determinan la calidad sonora de un recinto dedicado a la enseñanza son, básicamente, el tiempo de reverberación y la relación señal-ruido [1]. Una forma de enfocar el problema es la realización de test auditivos en una serie de aulas reales, entre las que habrá algunas diferencias en ambos parámetros, y efectuar análisis de regresión del porcentaje de inteligibilidad en función de la relación señal-ruido y tiempo de reverberación. Otro posible enfoque es hacer análisis de regresión entre índices como el STI o el AI y el porcentaje de inteligibilidad. En

cualquiera de los dos casos, el porcentaje de inteligibilidad es usado como parámetro de referencia para evaluar la influencia de las condiciones acústicas sobre la inteligibilidad. Sin embargo, los resultados de un trabajo previo nos muestran que esa puede ser una opción inadecuada [10]. En dicho trabajo empleamos palabras bisílabas sin sentido que debían ser reconocidas por una serie de alumnos de enseñanza primaria. Aunque los resultados obtenidos contando los fallos producidos en *palabras* podían ser considerados consistentes si se consideraba el *porcentaje total de palabras con fallos*, un análisis detallado de los resultados nos mostró que existía una enorme variabilidad en los fallos cometidos, de tal manera que cuando se administró el mismo test en una segunda ocasión, los resultados diferían considerablemente de los obtenidos en la primera ocasión, aun cuando el porcentaje global de aciertos era similar. Por otra parte, cuando la población de alumnos a los que se administró el mismo test se dividió en dos partes iguales, se comprobó que las respuestas de los dos grupos tenían un nivel bajo de consistencia. Eso demuestra que el test administrado no permite establecer una relación entre características acústicas del estímulo e inteligibilidad, ya que ante el mismo estímulo dos grupos de oyentes reaccionan de manera diferente, y por tanto habrá otros factores desconocidos que jueguen un papel. Una de las causas a las que atribuimos esa falta de consistencia es el hecho de que el test fuera abierto y las respuestas se escribieran sobre una plantilla, con los consiguientes problemas de interpretación.

De acuerdo con esos resultados pensamos en el diseño una serie de experimentos de percepción que nos permitan predecir la inteligibilidad en un aula a partir del tiempo de reverberación y la relación señal-ruído. Para ello debemos disponer de tests auditivos que sean fáciles de administrar, de duración razonable y que muestren: 1) relación con las características acústicas de la sala; 2) relación con la inteligibilidad esperada de un mensaje de lenguaje natural. Estas dos características no son fáciles de combinar en un mismo test, ya que en la comprensión del lenguaje natural intervienen factores extra-acústicos. Por tanto nos decidimos a diseñar : a) Un test de sílabas CV formadas por las posibles consonantes en posición inicial, acompañadas por las cinco vocales; b) Un test con frases del lenguaje natural. La identificación auditiva en el test a) estaría basada exclusivamente en las características acústicas de los estímulos. Además nos proporcionaría una información muy interesante sobre qué tipo de fonemas sufren errores de reconocimiento para diferentes condiciones, lo que nos permitiría afinar el diseño del test b) escogiendo palabras que incluyan los fonemas que proporcionen mayor información, y por tanto reducir el tamaño del mismo.

En este trabajo presentamos los resultados obtenidos con un test del tipo a), para diferentes condiciones de relación señal-ruído. El objetivo es determinar la consistencia del test, su validez y utilidad, y determinar los fonemas consonánticos en posición inicial que son afectados por la pérdida de relación señal-ruído, así como establecer la mejor estrategia posible para evaluar estos resultados.

METODO

Los estímulos utilizados estaban formados por sílabas CV en los que la consonante inicial era uno de los fonemas /p,t,k,b,d,g,ʔ,f,s, ʃ, x, ʧ, m,n,ʎ, l, ʎ, r /, en combinación con las cinco vocales /a,e,i,o,u/. Las 90 sílabas fueron pronunciadas en una cabina acústica absorbente por 2 hombres y dos mujeres de edades comprendidas entre 23-30 años, y se muestrearon a una frecuencia de 32 kHz. A continuación, procedimos a superponer un ruido rosa sobre las señales así obtenidas, con diferentes valores de Relación Señal-Ruido (RSR) : 24, 12, 6 y 0 dBs. Decidimos emplear ruido rosa en vez de ruido blanco por que el ruido rosa tiene una mayor concentración de energía en las bandas de baja frecuencia, de manera similar a lo que le sucede al ruido de fondo examinado por nosotros en aulas reales. Por lo tanto, el número total de estímulos empleados en los experimentos es de 4 hablantes x 90 sílabas x 4 condiciones de ruido = 1440.

Como oyentes han participado en los experimentos un total de 49 estudiantes (22 mujeres y 27 hombres), de la Asignatura de Acústica, de edades comprendidas entre 19-30 años, que han

realizado los tests como parte de los créditos de prácticas para la asignatura. El desarrollo del test se llevó a cabo en una serie de sesiones en días diferentes, para evitar en lo posible efectos de aprendizaje. Los oyentes realizaron el test en una cabina acústica, en la que se les presentaban los estímulos vía auriculares, en bloques correspondientes a un hablante con una determinada RSR. Los estímulos se presentaron de forma aleatoria, y con diferente orden para cada bloque de experimentos. El experimento se desarrolló al ritmo de cada oyente, que utilizaba el ratón para dar paso a cada estímulo, que sólo se escuchaba una vez. Una vez escuchado el estímulo, el oyente seleccionaba en la pantalla del ordenador una de las cinco posibles respuestas (test cerrado). Las posibles respuestas se escogieron en función de los resultados obtenidos en los tests mencionados anteriormente (10). Para comprobar la adecuación de las respuestas elegidas, incluimos la opción de una respuesta escogida por el oyente en el caso de que el fonema que percibe no estuviera incluido entre una de las cinco respuestas obligatorias. Dado el alto número de estímulos, la muestra total se dividió en dos partes, de tal manera que la mitad de los oyentes realizó los experimentos correspondientes a un hombre y una mujer, y la otra mitad hizo los experimentos correspondientes al otro hombre y la otra mujer.

RESULTADOS

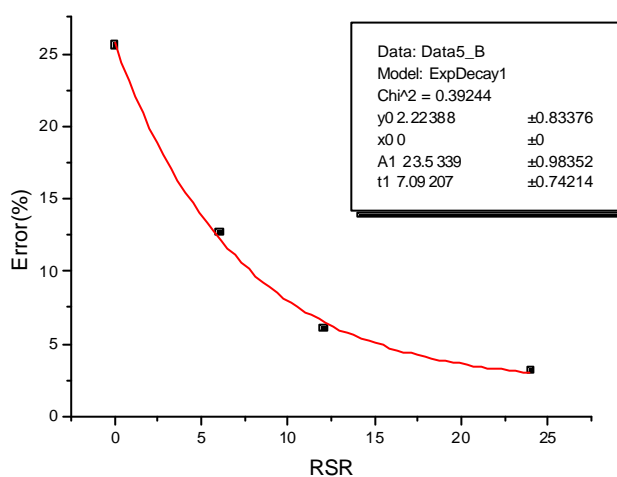


Fig. 1 Porcentaje de error total en la identificación de los fonemas promediado entre todos los hablantes, y separado por condición de ruido.

Los resultados globales promediados entre todos los hablantes pueden verse en la fig. 1. Sobre los resultados hemos ajustado una curva de decaimiento exponencial. Tal como se ve, parece existir una relación de decaimiento exponencial del porcentaje total de error a medida que aumenta la RSR. Esta relación tendría la forma

$$\% \text{ Error} = y + A e^{-\text{RSR} / t}$$

donde y sería el valor de offset para el porcentaje de error; A sería la amplitud; y t sería la constante de caída.

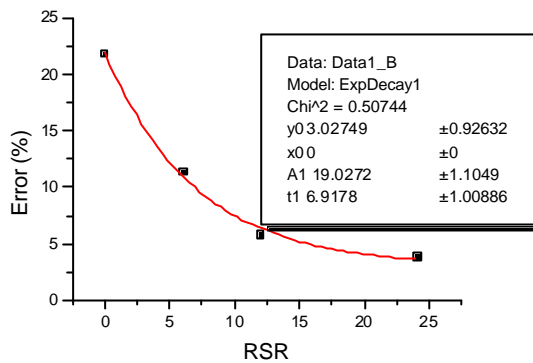


Fig. 2 Porcentaje total de error para los estímulos del hablante 1 para cada condición de ruido

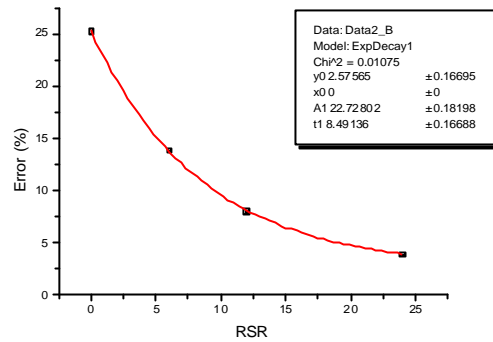


Fig. 3 Porcentaje total de error para los estímulos del hablante 2 para cada condición de ruido

Los resultados para cada hablante pueden verse en las fig. 2-5. En todos los casos se obtiene un buen ajuste entre los resultados y una curva exponencial decreciente.

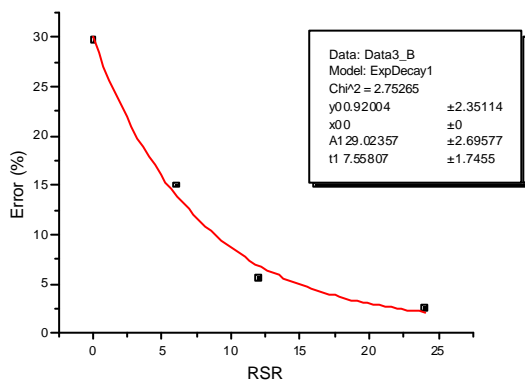


Fig. 4 Porcentaje total de error para los estímulos del hablante 3 para cada condición de ruido

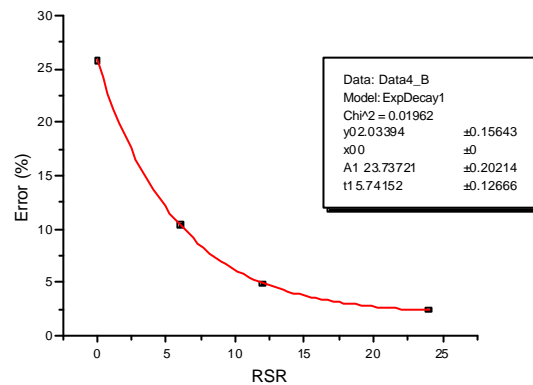


Fig. 5 Porcentaje total de error para los estímulos del hablante 4 para cada condición de ruido

Se realizó un test ANOVA (one way), con el porcentaje de error como variable dependiente, y el hablante como variable independiente, obteniéndose que no había diferencias significativas en cuanto a los porcentajes globales de error obtenidos para las diversas condiciones entre los diferentes hablantes ($\alpha=0.05$). Por tanto, parece que el test diseñado muestra una cierta consistencia, al menos en los porcentajes globales de errores obtenidos entre hablantes. A continuación decidimos investigar esta cuestión de manera más detallada.

Una inspección cuidadosa de los resultados reveló numerosas inconsistencias en las respuestas obtenidas. Dado que usamos estímulos sin sentido, esperamos que las características acústicas de los fonemas que forman cada sílaba y su interacción con el ruido enmascarante determinen de forma apreciable la identificación auditiva. Ello también implica que el factor hablante tenga una importancia relativamente menor en comparación. Sin embargo, los fonemas fallados masivamente para uno de los hablantes no eran los mismos que para otro hablante. Por ejemplo, el fonema /f/ de la sílaba /fa/ obtuvo los siguientes resultados para las condiciones (24,12,6,0): para el hablante 1 (0,0,0,0); para el hablante 2 (0,12,12,9); para el hablante 3 (3,2,3,0); para el hablante 4 (0,0,0,4).

Obviamente, el factor hablante es más decisivo que el factor fonema. Además, ni siquiera parece existir la tendencia esperada: a menor Relación Señal-Ruido, mayor número de fallos.

	RSR 24 dBs	RSR 12 dBS	RSR 6 dBs	RSR 0 dBs
Hablante 1-2	0.17	0.42	0.29	0.06
Hablante 1-3	0.48	0.15	0.21	0.18
Hablante 1-4	-0.20	0.75	0.47	0.34
Hablante 2-3	-0.25	-0.14	0.04	0.16
Hablante 2-4	0.36	0.01	-0.07	0.22
Hablante 3-4	-0.33	0.21	0.22	0.22

Tabla 1. Correlaciones entre las respuestas dadas a los estímulos de cada hablante.

Como consecuencia, nos decidimos a investigar la consistencia de las respuestas en función del hablante. Procedimos a calcular los coeficientes de correlación entre las respuestas dadas a los estímulos de cada hablante, para cada condición de ruido. Los resultados pueden verse en la matriz de correlación de la Tabla 1. Como puede observarse, los coeficientes de correlación son muy bajos, y además sus valores tienden lógicamente a disminuir a medida que la RSR disminuye, ya que en este caso el factor ruido va ganando importancia frente a otros factores. Incluso en el caso más favorable, para la condición de RSR de 24 dBs, el porcentaje promedio de varianza explicado sería inferior al 10%, sobre el conjunto de los cuatro hablantes.

CONCLUSIONES

El test de percepción de auditiva se propuso en base a las siguientes suposiciones: la identificación fonética se basa en las características acústicas de los estímulos y en las características acústicas ambientales (ruido de fondo, reverberación), con una contribución relativamente menor debida a otros factores (hablante, particularidades auditivas del oyente, variabilidad intrafonética, etc.). Los resultados nos muestran un panorama bien diferente. Sorprendentemente, los porcentajes globales de error son muy similares entre hablantes para las diferentes condiciones de RSR. Sin embargo, no podemos atribuir esa similitud a una consistencia de nuestras suposiciones, ya que los fallos ocurren de manera casi impredecible, variando los fonemas percibidos erróneamente de hablante a hablante. ¿Cómo se combina esa aleatoriedad para dar unos resultados globales similares? Esta cuestión no puede responderse de manera definitiva. Lo cierto es que lo observado detalladamente no nos muestra un patrón coherente, como cabría esperarse, mientras que la mezcla de resultados no coherentes, nos proporciona patrones aparentemente coherentes.

REFERENCIAS

1. Speech intelligibility studies in classrooms, J.S. Bradley, J. Acoust. Soc. Amer., 80(3):846-854 (1986)
2. Predictors of speech intelligibility in rooms, J.S. Bradley, J. Acoust. Soc. Amer., 80(3):837-845 (1986)
3. Measurement and prediction of typical background-noise levels in university classrooms during lectures, M. Hodgson, R. Rempel, S. Kennedy, J. Acoust. Soc. Amer., 105(1):2276-233 (1999)
4. Experimental investigation of the acoustical characteristics of university classrooms, M. Hodgson, J. Acoust. Soc. Amer., 106(4):1810-1819 (1999)
5. Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics, S. Bistafa, J.S. Bradley, J. Acoust. Soc. Amer., 107(2):861-875 (2000)
6. Empirical prediction of speech levels and reverberation in classrooms, M. Hodgson, Journal of building acoustics, 8(1):1-13 (2001)

7. Revisiting speech interference in classrooms, M. Picard, J.S. Bradley, *Audiology*, 40:221-244 (2001)
8. *Speech perception assessment*, L.L. Mendel, J.L. Danhauer, Ed. Singular Publishing Group, San Diego (1997)
9. Context effects and acoustic cues for the auditory identification of spanish fricatives /f/ and /ʃ/, S. Feijóo, S. Fernández, N. Barros, R. Balsa, *Acustica-Acta Acustica*, 88:113-126 (2002)
10. A speech perception test for children in classrooms, S. Feijóo, S. Fernández, J.M. Alvarez, *J. Acoust. Soc. Amer. (A)*, 112(5):2 (2002)