

CARACTERIZACIÓN FONÉTICA DEL BILINGÜISMO, ANÁLISIS ACÚSTICO DEL HABLA ESPONTÁNEA Y EVALUACIÓN DE SISTEMAS DE SÍNTESIS DE HABLA

Joaquim Llisterri
Dolors Poch Olivé

Departamento de Filología Española - Laboratorio de Fonética
Facultad de Filosofía y Letras, Universidad Autónoma de Barcelona
08193 Bellaterra, Barcelona.

Las líneas de investigación en fonética experimental que actualmente se desarrollan en el Departamento de Filología Española de la Universidad Autónoma de Barcelona, en el marco del Laboratorio de Fonética de la misma universidad, pueden encuadrarse en los siguientes ámbitos : la caracterización fonética del bilingüismo, el análisis acústico del habla espontánea y la evaluación fonética de los sistemas de síntesis del habla.

1. CARACTERIZACION FONETICA DEL BILINGUISMO

La mayoría de estudios llevados a cabo sobre comunidades bilingües o plurilingües se sitúan en una perspectiva sociolingüística y tratan de establecer cuál es el estatus que tienen las lenguas en el grupo social examinado, cuántos hablantes hay de cada lengua, cuál es el nivel de conocimiento de ambas lenguas que poseen los diversos individuos, etc. Evidentemente, este tipo de trabajos es imprescindible para poder conocer las características lingüísticas de una comunidad pero es también muy importante determinar el comportamiento lingüístico de los hablantes, entendiéndolo por "comportamiento lingüístico" las modificaciones que llevan a cabo en sus mecanismos de producción de los sonidos del habla cuando hablan una u otra lengua. Estas modificaciones inciden en las características acústicas de los sonidos y se manifiestan de diversas formas: en el nivel segmental como variaciones en los valores de los parámetros acústicos de vocales y consonantes y en el nivel suprasegmental como diferencias en los correlatos acústicos del acento, en la realización de los patrones entonativos y en cambios en la calidad de la voz.

Una cantidad sustancial de trabajos sobre la interferencia lingüística se han realizado desde el punto de vista de la fonología y de la fonética pero, como ya hemos mencionado, son muy pocas las descripciones de los procesos de interferencia en hablantes bilingües llevadas a cabo con métodos experimentales que permitan llevar a cabo una caracterización fonética del bilingüismo¹. De ahí que una de las líneas de investigación que se desarrolla en el Departamento de Filología Española de la UAB, en el marco de las actividades del Laboratorio de Fonética, se centre en el estudio acústico de las realizaciones fonéticas de los hablantes bilingües catalán-español tanto cuando utilizan su segunda lengua (el español) como en el proceso de aprendizaje de una tercera (francés o inglés).

¹ Véase, por ejemplo, Caramazza *et al.* (1973), Mack (1982) o Williams (1982).

1.1. El comportamiento de los hablantes bilingües: nivel segmental

El primer proyecto emprendido en este campo fue el análisis acústico del vocalismo, uno de los rasgos que contribuye en mayor medida a caracterizar el "acento catalán" en las realizaciones españolas de los catalanohablantes bilingües (Balari *et al.* 1985). El corpus estudiado consistió en 33 palabras españolas contextualizadas, que fueron leídas por 15 hombres y 15 mujeres catalanohablantes cuya segunda lengua era el español. Para la determinación del timbre vocálico se llevó a cabo un análisis espectral utilizando la Transformada Rápida de Fourier (FFT) mediante el analizador de alta resolución Brüel & Kjaer 2033; por ese procedimiento se analizaron un total de 450 vocales.

Los resultados obtenidos mostraron que el español hablado por catalanohablantes bilingües constituye un sistema específico que no se caracteriza por una mayor abertura de las vocales medias [e] y [o], sino por una estructuración del triángulo vocálico que refleja la posición de los elementos vocálicos del catalán: [e] y [o] españolas se realizan como vocales medias cerradas y queda vacío el espacio correspondiente a las medias abiertas del catalán; en el español nativo, este espacio viene ocupado por las realizaciones de las vocales medias [e] y [o].

No ha escapado a nuestra atención el hecho de que los datos obtenidos pueden responder a la situación experimental creada: palabras insertadas en frases del tipo "ha dicho ... dos veces". La próxima etapa del trabajo consistirá en el estudio del vocalismo en el habla espontánea, tomando como punto de comparación los resultados hallados en el experimento que se acaba de describir.

1.2. El comportamiento de los hablantes bilingües: nivel suprasegmental (la calidad de voz)

Abercrombie (1967) define la calidad de voz como aquellas características que se encuentran de forma casi permanente en la actividad fonatoria de un individuo considerándola "*a quasi-permanent quality running through all the sound that issues from his mouth*" (p. 91). Diversos estudios experimentales han mostrado la influencia de la lengua utilizada en las propiedades globales del habla, usando técnicas de análisis acústico como el espectro medio (*long term average spectrum - LTAS*), que permite objetivar la calidad de la voz. Sin embargo, aún se conoce poco la incidencia en el espectro medio de la lengua del hablante y su posible utilidad para caracterizar objetivamente diferencias acústicas inter-lingüísticas. Harmegnies y Landercy (1985) desarrollan una metodología cuantitativa/cualitativa que aplican al neerlandés y al francés y que pone de manifiesto la existencia de una importante variabilidad en el LTAS ligada a la lengua pero que muestra también que los efectos de la lengua son menores que los ligados a los individuos (Harmegnies *et al.* 1987).

Con el fin de estudiar la variabilidad del LTAS en hablantes bilingües catalán-español se ha realizado un primer experimento (Harmegnies *et al.* 1989) consistente en la preparación de un corpus fonéticamente equilibrado en español y en catalán leído diez veces en catalán y en español por diez informantes bilingües catalanohablantes y en español por diez monolingües castellano hablantes. Para cada lectura de cada hablante en cada lengua se ha calculado el espectro medio utilizando la posibilidad de promediar espectros que ofrece el analizador Brüel & Kjaer 2033. Se han realizado a continuación

comparaciones entre locutores bilingües y monolingües y entre las lecturas en catalán y en español de los informantes bilingües mediante el índice de disimilaridad SDDD (*Standard Deviation of the Differences Distribution*) (Harmegnies *et al.* 1987) que calcula la diferencia entre los espectros que se comparan. Así, un valor alto en el índice de disimilaridad indica una considerable diferencia entre los espectros mientras que un valor bajo indica, por el contrario, una similitud. Los resultados de este trabajo muestran la sensibilidad del LTAS a los cambios de lengua realizados por el mismo hablante, es decir, los valores del índice SDDD son sensiblemente más altos en las comparaciones interlingüísticas (español - catalán) que en las comparaciones intralingüísticas (español - español y catalán - catalán) para todos los informantes.

Con el fin de determinar si el grado de bilingüismo influye en las modificaciones sufridas por el LTAS se ha llevado a cabo una segunda experiencia (Bruyninckx *et al.* 1991) seleccionando los informantes en función de su dominancia lingüística (español o catalán) que tiene como objetivo realizar una primera aproximación a la cuantificación de esta noción sobre unas bases acústicas objetivas.

Se ha seleccionado en este caso una muestra de 24 informantes bilingües cuya utilización del español y del catalán es ligeramente distinta: la mitad de ellos emplea el catalán en un 60 % y el español en un 40 % en, por lo menos, dos tercios de las situaciones de comunicación sobre las que fueron interrogados. La otra mitad presenta un porcentaje de utilización inverso, es decir, 60 % catalán y 40 % español para, por lo menos, dos tercios de las situaciones de comunicación. Con el objetivo de determinar si el sexo es una variable importante en la variación del LTAS (Harmegnies 1988) se ha grabado a 6 hombres y 6 mujeres para cada uno de los dos grupos de bilingües.

El corpus utilizado ha sido el mismo texto equilibrado de la experiencia anterior y las condiciones de grabación y de realización de los espectros medios han sido también similares. Para cada informante se han realizado también en este caso dos tipos de comparaciones entre los LTAS: una comparación inter-lingüística y una comparación intra-lingüística. En el caso de las comparaciones inter-lingüísticas (catalán vs. español) cada uno de los LTAS de cada informante catalán ha sido comparado con cada uno de los LTAS del español del mismo informante. Este procedimiento implica 600 comparaciones inter-lengua/intra-informante (25 comparaciones x 2 grupos de dominancia x 2 sexos x 6 informantes). En el caso de las comparaciones intra-lingüísticas (catalán vs. catalán y español vs. español) se han realizado un total de 480 comparaciones (2 lenguas x 2 sexos x 6 informantes x 2 grupos de dominancia x 10 pares). Para el total de la muestra se han realizado pues 1080 comparaciones.

Los valores del índice SDDD en este caso se sitúan alrededor de 2.5 para los hombres y 3.2 para las mujeres lo cual pone de manifiesto ciertas diferencias en el LTAS debidas al sexo. Se observan, además, diferencias importantes en la variabilidad inter-informante: el rango de los valores del índice SDDD varía desde 1.93 a 4.43 (catalán vs. catalán), 1.83 a 4.59 (castellano vs. castellano) hasta 2.41 a 5.05 (catalán vs. español), lo cual confirma nuestra hipótesis acerca de la variación en el LTAS ligada a la utilización de cada una de las lenguas, puesto que los valores superiores en el índice de disimilaridad corresponden a las comparaciones inter-lingüísticas. El examen de las tendencias centrales muestra también que los valores del índice SDDD obtenidos en las comparaciones inter-

lingüísticas tienden a ser mayores que los obtenidos en las comparaciones intralingüísticas.

El estudio del espectro medio ha permitido pues, comprobar que la calidad vocal varía en función de la lengua que utilicen los hablantes. Por otra parte, el diferente comportamiento observado en los informantes en función de su lengua dominante permite apuntar la necesidad de continuar estas investigaciones, considerando el índice SDDD como una herramienta de caracterización del grado de bilingüismo.

1.3. Las interferencias fonéticas en hablantes bilingües

A pesar del número creciente de estudios experimentales y no experimentales sobre el papel de las características fonéticas primera lengua en la adquisición de la segunda ², la influencia de la segunda lengua --o de cualquier otra lengua distinta de la materna-- en el proceso de adquisición de una tercera es aún poco conocida y el caso concreto de bilingües hablantes de dos lenguas muy próximas cuando aprenden una tercera lengua parece no haber sido estudiado hasta el momento³. El interés que plantea esta investigación --además de sus aplicaciones inmediatas a la enseñanza de lenguas en nuestra comunidad-- reside también en la posibilidad de llegar a construir un modelo de adquisición de terceras lenguas en situaciones de bilingüismo, que podría ser aplicado a otras situaciones sociolingüísticas con características semejantes .

Hasta el momento se ha llevado a cabo el estudio de los sistemas vocálicos del inglés y del francés y de las fricativas del francés (Llisterri y Poch 1987a, 1987b). En lo que se refiere al estudio del vocalismo del francés (Llisterri y Poch 1986) se analizaron acústicamente 240 vocales pronunciadas por 10 informantes bilingües, observándose las mayores desviaciones respecto de las vocales francesas nativas en la serie de las vocales centrales labializadas y en las vocales medias; quizás la aportación más significativa de este trabajo es la verificación de que el área de dispersión de [e] y [o] francesas coinciden con las correspondientes catalanas y también con las de las realizaciones en español de los hablantes bilingües.

Un análisis similar del vocalismo inglés (Llisterri y Poch 1987c) realizado con cinco informantes bilingües y cuatro castellanohablantes monolingües permitió observar en ambos casos una gran superposición en las áreas de dispersión de las vocales centrales, y una distribución de las vocales medias abiertas y cerradas del inglés en los bilingües muy cercana a la de su L1 (catalán) y distinta de la que se aprecia en los castellanohablantes nativos. La misma influencia de la primera lengua, con independencia

² Son paradigmáticos desde el punto de vista experimental en este campo los diversos trabajos de Flege (1987a, b). En Flege y Bohn (1989) y en Flege y Eefting (1987) se estudia, por ejemplo, la influencia del español en la adquisición del inglés.

³ Como señala Ringbom (1986), la mayoría de los trabajos sobre la influencia de una segunda lengua en la adquisición de una tercera se refieren a aspectos léxicos y gramaticales, De hecho, estos trabajos no abordan tampoco el tema desde el punto de vista del bilingüismo. Sobre la adquisición de una tercera lengua véase, por ejemplo, Chamot (1973), Bentahila (1975), Mägiste (1984) y Ringbom (1987).

de la segunda, se observa en las características acústicas de las realizaciones inglesas del "schwa", que muestran una gran similitud con las catalanas.

El estudio de las fricativas del francés (Balari *et al.* 1988), en el que se analizaron cuatro informantes bilingües y cuatro monolingües, no mostró en cambio diferencias significativas en cuanto al comportamiento de hablantes bilingües y monolingües, dado que, desde el punto de vista acústico, las consonantes fricativas presentes en los sistemas del catalán y del español no presentan diferencias muy acusadas. Se hizo patente en cambio un fenómeno de ultracorrección en la producción de la fricativa alveolar francesa realizada, tanto por bilingües como por monolingües, con una concentración de la energía acústica a frecuencias superiores a las de las realizaciones francesas nativas.

Desde el punto de vista del análisis segmental se ha llevado a cabo también un estudio de la consonante alveolar lateral del francés producida por bilingües que aprenden dicha lengua (Llisterra y Martínez Dauden 1990, 1991). Este sonido es especialmente interesante dado que el español y el francés presentan la variedad "clara" (alveolar) y el catalán la variedad "oscura" (velarizada) de la misma. Se han utilizado informantes bilingües, con distinto grado de bilingüismo, que han leído un texto de ficción en las tres lenguas. El análisis acústico de las realizaciones de [l] ha puesto de manifiesto que el grado de velarización de la consonante lateral en los bilingües español - catalán está influenciado por su dominancia lingüística, de forma que la noción misma de bilingüismo incluye diferentes grados de competencia lingüística por parte de los hablantes.

Es ésta una breve presentación de las investigaciones llevadas a cabo hasta ahora en el campo de la caracterización fonética del bilingüismo. Actualmente, se está profundizando en el estudio de la modificación de la calidad de la voz provocada por los cambios de lengua y, en concreto, sobre la cuestión de si el inventario fonético de las lenguas estudiadas (español y catalán en este caso) influye en dichos cambios o si se trata de un factor totalmente independiente.

2. ANALISIS ACUSTICO DEL HABLA ESPONTANEA

El análisis y descripción del habla continua no es una línea de investigación nueva dentro de la fonética, pues hace ya algunos años se vienen publicando trabajos que estudian aspectos tales como la influencia de la velocidad de elocución sobre la duración de los sonidos, la naturaleza y función de las pausas en el habla, lo que se podría denominar las "unidades de procesamiento" de la información, etc. Lo que sí supone una novedad y una aportación es el estudio sistemático de los fenómenos que se producen en el habla espontánea y su reducción a una serie de reglas que permitan predecir la variabilidad.

Ha sido desde el punto de vista de las aplicaciones, y en concreto de la síntesis y el reconocimiento del habla, desde donde se ha planteado la necesidad de un nuevo enfoque descriptivo: un sistema de reconocimiento, por ejemplo, debiera ser capaz de identificar como pertenecientes a la misma categoría las diversas ondas sonoras que corresponden a distintas realizaciones de un mismo sonido y que sabemos son diferentes. En muchos casos, los sistemas de reconocimiento no consiguen establecer esta identidad con lo que su rendimiento se ve disminuido. El problema de base es cómo relacionar una serie de entidades o unidades fonológicas (que utilizamos para caracterizar los sonidos del habla

y su estructura) con los datos que nos proporcionan los análisis de los movimientos articulatorios, las ondas sonoras o el comportamiento de la percepción. Las entidades fonológicas suponen la existencia de elementos invariantes y en cambio los datos "físicos" nos indican la existencia de un alto grado de variación.

Ahora bien, el hecho de que la comunicación se produzca nos indica que debe existir una estrecha relación entre ambos aspectos. En un intento de determinar por qué no se ha resuelto hasta ahora el problema de la invariación, Lindblom (1987) se pregunta dónde se ha buscado hasta ahora, es decir, qué tipo de material se ha estudiado para determinar los elementos invariables del acto de comunicación. A partir del vaciado de los artículos contenidos en una serie de las más importantes revistas de fonética experimental (como *JASA* o *Journal of Phonetics*) se pone de manifiesto que para la descripción del habla se ha utilizado únicamente material "de laboratorio", es decir, sílabas, palabras aisladas o palabras contenidas en "frases marco". Este tipo de habla permite controlar buena parte de los factores que intervienen en un experimento, mientras que establecer un protocolo experimental que permita controlar el habla real presenta mayores dificultades; por otra parte, algunos fonetistas defienden que el estudio del material "de laboratorio" es el primer paso en la solución del problema de la invariación siendo el caso del habla espontánea tan sólo un tipo más complejo de habla.

El estudio del acto de comunicación pone claramente de manifiesto que en el habla se dan diversos estilos y que todos los hablantes somos capaces de cambiar nuestro "modo de hablar" y de flexibilizar al máximo nuestras realizaciones en función del contexto, de los interlocutores, de las reglas sociales, etc. Estas "modificaciones" se efectúan, por supuesto, sobre la señal sonora y ejercen fuertes influencias sobre el modo de considerar la cuestión de la invariación. Lindblom (1984) formula teóricamente estos problemas al demostrar que el proceso del habla es de naturaleza teleológica: tiene una finalidad (que la comunicación efectivamente se produzca), es plástico (las realizaciones de los distintos sonidos se adaptan a las diversas situaciones de comunicación) y es económico (se tiende a realizar siempre el mínimo esfuerzo posible). Para este autor, la finalidad, entendida como el hecho de que la comunicación se produzca o no, es lo que determina las características de las ondas sonoras realizadas (con sus reducciones, modificaciones, adaptaciones, etc.), así como los movimientos articulatorios necesarios para su producción. De esto se desprende, por tanto, que los factores "independientes" de la señal (sociales, situacionales, etc.) tienen tanta importancia en el estudio del habla como los "dependientes" de la señal (los movimientos articulatorios o las ondas sonoras). Lo más importante, pues, en todo acto de comunicación es la interacción entre hablante y oyente, pues el hablante es capaz de variar su *output* en función de distintos factores (Lindblom 1990).

El habla "de laboratorio" encaja difícilmente en este esquema, pues en ella los factores "independientes de la señal" no son suficientemente tomados en consideración, es decir, se los toma en cuenta para aislarlos pero no como un elemento que intervenga decisivamente en el proceso de comunicación. Por ello, los resultados obtenidos de su análisis no corresponden totalmente al análisis del habla espontánea y las aplicaciones tecnológicas sobre habla real plantean problemas a la hora de utilizarlas con bases de datos proporcionados por habla "de laboratorio".

Si se acepta la hipótesis de Lindblom como "estrategia" de comunicación, es posible resolver el problema de la invariación: no existe una invariación absoluta sino que se trata de un problema ecológico: lo invariante es distinto cada vez y se adapta (= ecología) a la situación (= entorno). Lindblom (1987) propone el concepto de variabilidad adaptativa para referirse a esta cuestión. Esta concepción explica, además, por qué hasta ahora no se ha llegado a resultados definitivos en la búsqueda de la invariación: tomada en sentido estricto la concepción tradicional de invariación contiene una idea del habla como no-adaptativa a las diferentes situaciones.

Se puede así hablar de diversos "estilos de habla" capaces de hacer variar las características de los sonidos cuyos valores de frecuencias y duraciones se ven modificados según el tipo de corpus utilizado para el análisis. La descripción del habla, vista desde esta perspectiva, se plantea, por tanto, la descripción de las transformaciones fonéticas y fonológicas que se producen sistemáticamente en función de los diversos "estilos" de modo que se pueda llegar a la formulación de un algoritmo que permita aplicar, a partir de la especificación acústica de una palabra de referencia, las transformaciones necesarias requeridas por cada contexto concreto. Los datos obtenidos en la descripción servirán, como se desprende de lo expuesto hasta ahora, para mejorar las aplicaciones tecnológicas.

De entre los trabajos realizados en el Departamento de Filología Española de la UAB en esta perspectiva de análisis, cabe destacar el estudio de la coarticulación en las consonantes oclusivas comparando habla "de laboratorio" y habla espontánea (Poch *et al.* 1989). El corpus de habla espontánea procede de una conversación informal y el de habla "de laboratorio" está constituido por las mismas palabras analizadas en habla espontánea pero leídas, esta vez, en forma de lista por el mismo informante. Como procedimiento de comparación se ha utilizado la "ecuación del locus" (Krull 1987) que permite relacionar los valores de frecuencias del inicio y el centro de los formantes, en este caso del F2 de las vocales situadas junto a consonantes oclusivas. Los resultados ponen de manifiesto que el grado de coarticulación es mucho mayor en habla espontánea que en habla de laboratorio, es decir, que la diferencia entre el inicio y el centro del segundo formante es considerablemente menor en habla espontánea que en habla de laboratorio.

El análisis de vocales iguales en contacto (Benese y Machuca 1989) extraídas de una conversación espontánea ha permitido poner de manifiesto tres tipos básicos de comportamiento de estos sonidos: la reducción a una sola vocal cuya duración es similar a la de las vocales que no han sufrido este proceso; la reducción "parcial", es decir, la presencia de una vocal alargada y el mantenimiento de dos vocales iguales. La aparición de cada uno de estos comportamientos está condicionada por las necesidades comunicativas y se han establecido una serie de reglas que predicen el grado de reducción.

Igualmente se ha estudiado el comportamiento de vocales diferentes en contacto (Aguilar 1991a) comparando esta vez las extraídas de una conversación informal con las obtenidas a partir de la lectura de un texto. También en este caso se detectan distintos grados de reducción y mantenimiento de diferencias entre las vocales en contacto y también un comportamiento diferente según el tipo de corpus del que se han extraído.

La comparación entre las consonantes nasales en habla espontánea y en habla "de laboratorio" (Machuca 1991), de las vibrantes (de la Mota 1991) y de las aproximantes (Aguilar y Andreu 1991) en las mismas condiciones ha puesto asimismo de manifiesto las modificaciones sufridas por estas clases de sonidos en función del estilo de habla.

Finalmente, el estudio del sistema vocálico del español también en habla espontánea y en habla "de laboratorio" (Harmegnies y Poch 1991) a partir del análisis acústico de quinientas vocales de cada uno de los dos estilos, muestra una importante "desorganización" del espacio vocálico en habla espontánea. La modificación afecta los valores de frecuencia de los formantes que son más centralizados en el habla espontánea, así como más variables, de tal forma que las áreas de dispersión son mucho menos claras en habla espontánea que en habla de laboratorio. Estas modificaciones en la organización del sistema vocálico pueden ser consideradas como diferentes grados de entropía que presenta dicho sistema en función del estilo de habla considerado conservando, no obstante, los rasgos imprescindibles para facilitar la identificación de las distintas unidades pues si la entropía llegara a un punto en que se confundieran los elementos el sistema como tal dejaría de existir.

3. LA EVALUACION DE LOS SISTEMAS DE SINTESIS DEL HABLA

3.1. El problema de la inteligibilidad y la naturalidad en los sistemas de conversión de texto a voz

Los avances experimentados en las últimas décadas en el campo de la tecnología del habla permiten disponer actualmente de sistemas de conversión de texto a voz que alcanzan un elevado grado de inteligibilidad y que, como ya apuntaban Pisoni *et al.* (1985) hace seis años, pronto llegarán a los niveles propios del habla natural. Sin embargo, tal como afirmaba Allen (1985) por las mismas fechas - y su opinión sigue siendo válida hoy en día -, aún no disponemos de sistemas de síntesis que reproduzcan adecuadamente la variabilidad fonética observada en el habla natural.

Vemos así cuáles son los dos grandes ejes alrededor de los cuales gira el problema de la utilización del habla sintetizada en aplicaciones reales: la inteligibilidad por un lado y la naturalidad por otro. Si bien se han realizado notables avances en la primera, no se ha conseguido aún alcanzar el grado de naturalidad deseable para una difusión masiva de los sistemas de conversión de texto a voz. La modelización de las variaciones fonéticas propias de un mismo locutor - cambios de timbre de la voz, de velocidad de elocución, de estilo de habla, de estado emotivo - o de las diferencias que se dan entre hablantes distintos - diferencias en el habla y en la voz de hombres y mujeres, marcas de la procedencia social o geográfica - son objeto de las investigaciones más recientes en el terreno de los conversores de texto a voz (Granström y Nord 1991).

La relación entre naturalidad e inteligibilidad es, en cierto modo, complementaria y depende en parte de la aplicación del sistema de síntesis. Bristow (1984) menciona, por ejemplo, que en el caso de una alarma industrial, es esencial mantener la inteligibilidad ; sin embargo, en el complemento sonoro de un video juego la naturalidad es primordial; otras aplicaciones requieren un grado intermedio de ambas categorías, aunque una puede primar sobre otra: la inteligibilidad en las telecomunicaciones, cuando no hay un soporte visual complementario, y la naturalidad - junto con el tipo de voz - en las aplicaciones

para usuarios no especialmente motivados que van acompañadas de un soporte visual, como podría ser el caso de la voz sintetizada en los automóviles.

Tal como ha ido avanzando el desarrollo de los sistemas de conversión de texto a voz, haciendo posible su utilización práctica, se ha ido imponiendo la necesidad de disponer de herramientas que permitan comparar las prestaciones de diversos sistemas de síntesis. Al mismo tiempo, se ha visto también el interés en estudiar la reacción al uso de la voz sintetizada y, muy especialmente, las diferencias en los mecanismos de percepción entre el habla sintetizada y el habla natural.

En concreto, siguiendo a Pisoni *et al.* (1985) pueden plantearse las cuestiones siguientes:

- (1) ¿ Con qué grado de precisión de reconocen los sonidos y las palabras sintetizadas ?
- (2) ¿ Con qué precisión se entiende el sentido de una frase en habla sintetizada ?
- (3) ¿ Qué dificultades plantea la percepción y la comprensión del habla sintetizada ?

El desarrollo de estos temas de investigación ha llevado a la creación de un campo de trabajo conocido como *Speech Assessment* que se ocupa, en conjunto, del estudio de la posibilidad de utilización de la síntesis (*output*) y el reconocimiento (*input*) del habla en situaciones reales. Una muestra del interés que ha despertado entre la comunidad científica puede verse en las publicaciones monográficas que recientemente han aparecido sobre el tema (ESCA 1989; Pols (Ed) 1990)

Entre los proyectos europeos encuadrados en el programa ESPRIT figura el conocido como SAM - Harland *et al.* (Comps) (1989) -, dentro del cual se incluye un apartado dedicado a la evaluación de la inteligibilidad y la calidad de los sistemas de conversión de texto a voz. En este marco se han desarrollado una serie de herramientas estandarizadas en inglés, holandés, alemán, francés, italiano y sueco, sin que el español haya tenido de momento cabida en el proyecto.

Actualmente contamos con más de un sistema de conversión de texto a voz para el castellano⁴, y parece por tanto que ha llegado el momento de dotarse de mecanismos que permitan evaluar tanto su inteligibilidad como su naturalidad y que favorezcan además la comparación objetiva entre sistemas distintos. Tal como se acaba de mencionar, estos instrumentos se han desarrollado ya para otras lenguas - especialmente para el inglés - y

⁴ Entre los que han llegado a nuestro conocimiento cabe citar el sistema desarrollado en la ETS de Ingenieros de Telecomunicación de Madrid en colaboración con el Departamento de Lengua Española de la UNED (Olabe *et al.* 1984), el desarrollado en la ET de Ingenieros de Telecomunicación de la Salle de Barcelona (Martí y Niñerola, 1987) , la versión española del sistema Infovox basado en los trabajos del KTH de Estocolmo (Carlson y Granström 1990), la versión española del sistema Multivoix del Instituto de Lingüística de la Academia Húngara de Ciencias (Olaszy, 1980) y los recientemente desarrollados en el Departamento de Lenguajes y Sistemas Informáticos de la Universidad Politécnica de Valencia (Pérez y Vidal, 1991) y por la empresa belga Lernout & Hauspie (Conejo y van Coile, 1991); Vieregge *et al.* (1987) y Gerwen *et al.* (1989) mencionan un sistema para el español creado en la Universidad Católica de Nimega.

están siendo aplicados con éxito a un gran número de sistemas de síntesis existentes en la actualidad, tanto comercializados como a prototipos de laboratorio ⁵.

A continuación se revisan algunas de las técnicas de evaluación más comúnmente utilizadas y se presentan algunas sugerencias para su adaptación al castellano, extraídas de trabajos en curso realizados en el Departamento de Filología Española de la UAB en el marco del Laboratorio de Fonética.

3.2. La evaluación de la inteligibilidad de los elementos segmentales aislados

Existen actualmente diversas propuestas de técnicas para llevar a cabo la evaluación de la inteligibilidad de los elementos segmentales aislados en un conversor de texto a voz. Todas ellas se basan en pruebas en las que un grupo de oyentes debe responder a una serie de estímulos sintetizados, identificando los elementos segmentales en los que se centra el estudio. Algunas derivan de pruebas de inteligibilidad diseñadas para el habla natural ya a finales de los años 50, mientras que otras se han concebido específicamente para la evaluación del habla sintetizada.

Uno de los tests más habituales lo constituye el llamado Test de Rimas (RT, *Rhyme Test*) que en su versión más actual y utilizada se conoce como Test de Rimas Modificado (MRT, *Modified Rhyme Test*); fue diseñado por Fairbanks en 1958 y modificado por House *et al.* en 1965. Se trata de un test formado por estímulos consistentes en palabras monosilábicas con la estructura consonante-vocal-consonante, en el que los oyentes deben elegir una palabra entre seis alternativas. Las palabras difieren en un único segmento, que se encuentra o en posición inicial o en posición final; una de las ventajas de este test es que ha sido utilizado en el habla natural, y existen por lo tanto medidas estandarizadas que pueden utilizarse como punto de referencia.

La adaptación al castellano ha sido llevada a cabo por Aguilar (1991 b), manteniendo las características del test original inglés: estímulos monosilábicos y estructura CVC (aceptando CV o VC en algunos casos). El requisito de la monosilabidad plantea problemas importantes, dado que en ciertos casos ha debido recurrirse a palabras poco familiares o reducirse el número de alternativas ante la imposibilidad de encontrar seis palabras que sólo difieran en la consonante inicial o en la final. Por este motivo, el número de alternativas en la respuesta se ha reducido a 4. En el momento de seleccionar los monosílabos, se ha tenido en cuenta la mayor o menor frecuencia de aparición de la consonante, tendiendo a una aparición proporcional a la que se encuentra en la lengua; aún así, el equilibrio fonético no es un requisito del Test de Rimas. En la figura 1 se presentan algunos de los estímulos del test, tal como figuran en la hoja de respuesta; para realizarlo, el oyente debe señalar únicamente cuál es la palabra que oye entre todas las de la serie.

⁵ Véase para una panorámica general van Bezooijen y Pols (1990), Pisoni *et al.* (1985) y Pols (1989)

	A	B	C	D
1	van	vas	bah	bar
2	ved	ven	ves	ver
3	dad	dan	dar	das
4	sol	son	sor	sos
1	sal	cal	chal	tal
2	gol	col	sol	rol
3	can	tan	pan	dan
4	hez	pez	tez	vez

Figura 1: Algunos estímulos correspondientes a la adaptación castellana del test de rimas modificado. En los cuatro primeros se evalúa la inteligibilidad de la consonante inicial, y en los cuatro últimos la de la consonante final (Aguilar, 1991 b)

Existen además otras herramientas, entre las que citaremos el *Diagnostic Rhyme Test* (DRT) de Voiers (1984), adaptado al castellano por Nadeu (1987) y actualmente en curso de revisión. También se dispone de una versión castellana del *Fast Diagnostic Test* (FDT), originalmente concebido por Loman y van Beezoujen (1988), en el que se contempla la inteligibilidad de todas las combinaciones posibles de consonante vocal en palabras de estructura CVC y VCV. La adaptación castellana realizada por Carbonell (1991), cubre de forma exhaustiva el inventario de fonemas del castellano y permite especialmente obtener matrices de confusiones consonánticas en distintas posiciones silábicas a partir de las cuales basar una evaluación sistemática de un sistema de conversión de texto a voz.

3.3. La evaluación de la inteligibilidad de palabras en contexto

Evidentemente, las pruebas anteriores nos ofrecen sólo una indicación de la inteligibilidad de elementos segmentales en palabras aisladas, con lo cual obtenemos con resultados que no reflejan la utilización de real del contexto que los hablantes realizan a la hora de percibir el habla, tanto natural como sintetizada. Por ello se han puesto a punto diversas pruebas en las que se evalúa la inteligibilidad de elementos segmentales en palabras situadas en el interior de una frase.

Como en el caso anterior, las primeras pruebas se desarrollaron para el estudio de la comprensión del habla natural y fueron también utilizadas para el estudio de los trastornos de la producción y de la percepción del habla en personas con dificultades en este terreno.

Nos referiremos en primer lugar al corpus de enunciados conocido como las Frases Psicoacústicas de Harvard (*Harvard Psychoacoustic Sentences*), desarrolladas por Egan en 1948 para aplicaciones audiométricas. Su característica principal reside en el hecho de que se trata de frases con sentido, sintácticamente bien formadas y fonéticamente equilibradas. En nuestro Departamento este corpus ha sido adaptado al

castellano por Valero (1991), siguiendo, en lo que se refiere al equilibrio fonético, la frecuencia de distribución de fonemas del español propuesta por Navarro Tomás (1946). En su versión actual se dispone de 10 grupos de 10 frases, manteniendo el equilibrio fonético en el interior de cada grupo. La figura 2 ilustra algunas de las frases propuestas.

1. Puedes fumar, pero vete al balcón
2. La tía Carmen quería guisar el conejo
3. No des patadas a los rosales
4. Necesito un centímetro para medir las dos telas
6. Con un quilo de boniatos ya tengo de sobras
7. Si quieres rezar, te dejo solo
8. Nunca debí dinero ni lo tomé prestado
9. Sus cacerías acabaron con la fauna de la zona
10. Me diste la paga de este mes

Figura 2: Grupo de diez frases fonéticamente equilibradas correspondientes a la adaptación al castellano de las frases psicoacústicas de Harvard. La prueba de χ^2 entre la distribución teórica y la real da un valor de 2.61 (Valero, 1991)

Sin embargo, pronto surgió la necesidad de desarrollar un nuevo corpus de frases en el que la redundancia contextual jugara un papel más reducido, y para ello se elaboraron las que se conocen como Frases Semánticamente Anómalas de Haskins, creadas por Nye y Gaitenby en 1974. Se trata de un conjunto de enunciados fonéticamente equilibrados, sujetos a restricciones en cuanto a su estructura sintáctica y que cumplen la condición de no tener un significado predecible. En el momento de responder al test, el oyente debe escribir la palabra que falta en la hoja de respuesta. En la adaptación al castellano realizada por Serra (1991), se ha elegido la estructura sintáctica artículo + nombre + adjetivo + verbo + artículo + adjetivo + nombre, siguiendo muy de cerca el modelo inglés; se dispone actualmente de 50 frases fonéticamente equilibradas siguiendo los criterios expuestos en el párrafo anterior. La comparación con la distribución teórica de Navarro Tomás (1946) arroja un valor de χ^2 de 2.02. En la figura 3 se ofrece una muestra de estas frases:

1. El niño sedoso tuerce el piso
2. La acera fina cuenta la muesca
3. El océano generoso pinta la sobriedad
4. La butaca fatídica amasa la tristeza
5. La cámara divina mina los nidos

Figura 3: Muestra de frases correspondientes a la adaptación al castellano de las frases semánticamente anómalas de Haskins (Serra, 1991)

Cabe mencionar también en este apartado el corpus de frases que ha sido desarrollado para varias lenguas europeas en el marco del proyecto SAM por Benoit *et al.* (1989), para el cual no se dispone aún de una versión castellana.

3.4. La evaluación de la comprensión

Naturalmente, el próximo paso una vez estudiado el nivel segmental y el nivel de la frase, es la evaluación de la comprensión de textos completos. Para ello, se han utilizado para el inglés pruebas de comprensión lectora similares a las usadas en el habla natural. Algunas de ellas se comentan en Pisoni *et al.* (1985) y en Pisoni (1987). Alberte (1991) ha preparado una serie de 15 textos en castellano seguido cada uno de ellos de un conjunto de preguntas en el formato de elección múltiple siguiendo el modelo de Pisoni (1987). Con ellos se evalúa la comprensión, por parte del oyente, de los contenidos de un fragmento coherente sintetizado; los textos cubren una amplia gama temática, oscilando entre las 147 y las 371 palabras con distintos grados de dificultad en función del tema.

Tal como se ha visto en los párrafos anteriores, empezamos a disponer de métodos que nos permiten plantearnos la evaluación de diversos sistemas de conversión de texto a voz existentes para el castellano, especialmente en el ámbito de la inteligibilidad. Motivos de espacio nos impiden tratar con cierto detalle la cuestión, más compleja y menos estudiada hasta el momento, de la evaluación de la naturalidad y la aceptabilidad por parte del usuario de los sistemas de síntesis. Remitimos al lector a los trabajos de Pavlovic *et al.* (1989 a, b) y Robert *et al.* (1989) en los que se describen algunas de las técnicas más habituales. Su desarrollo para el castellano está aún pendiente, aunque contamos con poder iniciarlo en un breve plazo.

Finalmente, cabe mencionar que el campo de la evaluación del habla sintetizada constituye un terreno ideal para la colaboración entre los técnicos que se ocupan del diseño de los sistemas y los lingüistas, que tienen en su haber la experiencia - y en algunos casos las herramientas - necesaria para observar las diferencias entre el producto sintetizado y el habla natural.

Referencias bibliográficas

ABERCROMBIE, D. (1967) *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

AGUILAR, L. (1991a) *Algunas cuestiones en torno a la reducción fonética: secuencias de vocales en contacto*. Universidad Autónoma de Barcelona, Departamento de Filología Española. Proyecto de investigación.

AGUILAR, L. (1991b) *Propuesta de un test de evaluación segmental del habla para el castellano: el Test de Rimas Modificado*. Universidad Autónoma de Barcelona, Departamento de Filología Española. ms no publicado.

AGUILAR, L.- ANDREU, M. (1991) "Acoustic description of Spanish approximants in laboratory speech and in continuous speech", *Actes du XIIème Congrès International des Sciences Phonétiques*. 19-24 août 1991, Aix-en-Provence, France. Aix-en-Provence: Université de Provence, Service des Publications. Vol 3 pp. 362-365.

ALBERTE, M. (1991) *Evaluación del habla sintetizada: test de comprensión auditiva*. Universidad Autónoma de Barcelona, Departamento de Filología Española. ms no publicado.

ALLEN, J. (1985) "A Perspective on Man-Machine Communication by Speech", *Proceedings of the IEEE* 73,11 : 1541-1550.

BALARI, S.- LLISTERRI, J.- POCH, D. (1985) "Estructuración fonética de la segunda lengua en hablantes bilingües", XV Simposio de la Sociedad Española de Lingüística, Córdoba, diciembre de 1985. Publicado en *Estudios de Lingüística* (Alicante), 5 (1988/1989) : 93-100.

BALARI, S.- LLISTERRI, J.- POCH, D. (1988) " Structuration de la troisième langue chez les locuteurs bilingües", *Actas de las Xas Jornadas Pedagógicas sobre la enseñanza del francés en España*. Barcelona : Institut de Ciències de l'Educació, Universitat Autònoma de Barcelona, pp. 89-98.

BENESEY, C.- MACHUCA, M.J. (1989) "Analysis of vowel coarticulation in continuous speech", in SZENDE, T. (Ed) *Proceedings of the Speech Research '89 (Magyar Fonetikai Füzetek / Hungarian Papers in Phonetics 21)*. Budapest: Linguistics Institute of the Hungarian Academy of Sciences. Vol. 2. pp. 9-13.

BENOIT, C.- ERP, A. VAN - GRICE, M.- HAZAN, V.- JEKOSCH, U. (1989) " Multilingual synthesizer assessment using semantically unpredictable sentences " in TUBACH, J.P.- MARIANI, J.J. (Eds) *Eurospeech 89. European Conference on Speech Communication and Technology*. Paris, September 1989. Edinburgh: CEP Consultants Ltd. vol 2. pp. 633-636

BENTAHILA, A. (1975) *The Influence of L2 on the Learning of L3*. MA Dissertation. Bangor: University College of North Wales, Department of Linguistics

BEZOOIJEN, R. van - POLS, L.C.W. (1990) " Evaluating text-to-speech systems: some methodological aspects", *Speech Communication* 9,4: 263-270

BRISTOW, G. (1984) "Designing a Speech System" in BRISTOW, G. (Ed) *Electronic Speech Synthesis. Techniques, Technology and Applications*. London: Granada pp. 229-240.

BRUYNINCKX, M.- HARMEGNIES, B.- LLISTERRI, J.- POCH, D. (1991) "Effects of language change on voice quality. An experimental study of Catalan - Castilian bilinguals", *Actes du XIIIème Congrès International des Sciences Phonétiques*. 19-24 août 1991, Aix-en-Provence, France. Aix-en-Provence: Université de Provence, Service des Publications. Vol 2. pp. 398-402.

CARAMAZZA, A.- YENI-KOMSHIAN, G.- ZURIFF, E.- CARBONE, E. (1973) "The Acquisition of a New Phonological Contrast: The Case of Stop Consonants in French-English Bilinguals", *Journal of the Acoustical Society of America* 54, 2: 421-428.

CARBONELL, J.F. (1991) *Adaptación castellana del Fast Diagnostic Test (FDT)* . Universidad Autónoma de Barcelona, Departamento de Filología Española. ms no publicado.

CARLSON, R.- GRANSTRÖM, B. (1990) " An environment for multilingual text-to-speech development ", *Proceedings of the ESCA Tutorial Day and Research Workshop on Speech Synthesis*. Autrans, France, 25-28 September 1990. pp. 73-82

CONEJO, J.M.- VAN COILE, B. (1991) : Desarrollo de un conversor de texto a voz en español dentro de una arquitectura multilingüe", Comunicación presentada en el VII Congreso de la SEPLN, Valencia, 3-4 de octubre de 1991.

CHAMOT, A.U. (1973) "Phonological Problems in Learning English as a Third Language" , *International Review of Applied Linguistics IRAL* 11, 3: 243-250.

DE LA MOTA, C. (1991) "A study of [r] and [] in spontaneous speech", *Actes du XIIIème Congrès International des Sciences Phonétiques*. 19-24 août 1991, Aix-en-Provence, France. Aix-en-Provence: Université de Provence, Service des Publications. Vol 4. pp. 386-389.

EGAN, J.P. (1948) "Articulation Testing Methods", *Laryngoscope* 58: 955-991

ESCA (1989) *Proceedings of the ESCA Tutorial Day and Workshop on Speech Input/Output Assessment and Speech Databases*. Noordwijkerhout, the Netherlands, 20-23 September 1989.

FAIRBANKS, G. (1958) " Test of Phonemic Differentiation: The Rhyme Test ", *Journal of the Acoustical Society of America* 30,7: 596-600

FLEGE, J.E. (1987 a) " The Instrumental Study of L2 Speech Production : Some Methodological Considerations", *Language Learning* 37,2: 285-296

FLEGE, J.E. (1987 b) "The production of 'new' and 'similar' phones in a foreign language: evidence for the effect of equivalence classification", *Journal of Phonetics* 15: 47-65.

FLEGE, J.E.- BOHN, O, (1989) " An Instrumental Study of Vowel Reduction and Stress Placement in Spanish-Accented English ", *Studies in Second Language Acquisition* 11,1: 35-62.

FLEGE, J.E.- EEFTING, W. (1987) "Production and perception of English stops by native Spanish speakers", *Journal of Phonetics* 15: 67-83.

GERWEN, R.P.M.W. van - VIEREGGE, W.H.- KERKHOF, M.P.A.M. (1989) " Evaluation of an automatic text-to-speech conversion system for Spanish ", *Proceedings of the ESCA Tutorial Day and Workshop on Speech Input/Output Assessment and Speech Databases*. Noordwijkerhout, the Netherlands, 20-23 September 1989. pp. 3.5.1.- 3.5.4.

GRANSTRÖM, B. - NORD, L. (1991) " Ways of exploring speaker characteristics and speaking styles ", *Actes du XIIème Congrès International des Sciences Phonétiques*. 19-24 août 1991, Aix-en-Provence, France. Aix-en-Provence: Université de Provence, Service des Publications. Vol 4. pp. 278-281

HARLAND,G. - FOURCIN, A.- BARRY, W.J. - GRICE, M. (Comps) (1989) *Esprit Project 1541 (SAM) Multilingual Speech Input/Output: Assessment, Methodology and Standardisation. Extension Phase. Final Report*. London: University College London.

HARMEGNIES, B. (1988) *Contribution à la caractérisation de la qualité vocale: analyses plurielles de spectres moyens à long terme de parole*. Thèse de doctorat. Mons: Université de Mons.

HARMEGNIES, B.- BRUYNINCKX, M.- LLISTERRI, J.;- POCH, D. (1989) "Effects of language change on voice quality. An experimental contribution to the study of the Catalan-Castilian case", in TUBACH, J.P.- MARIANI, J.J. (Eds) *Eurospeech 89. European Conference on Speech Communication and Technology*, Paris, September 1989. Edinburgh: CEP Consultants Ltd., 1989. vol 2 pp. 489-492.

HARMEGNIES, B.- LANDERCY, A. (1985) "Language features in the long-term average spectrum", *Revue de Phonétique Appliquée* 73-75:69-80.

HARMEGNIES, B.- LANDERCY, A.- BRUYNINCKX, M. (1987) "An experiment in inter-languages speaker recognition using the SDDD index ", *Proceedings of the XIth ICPhS. The Eleventh International Congress of Phonetic Sciences* . Estonia: Academy of Sciences of the Estonian SSR, 1987. Vol 2. pp. 241-244.

HARMEGNIES, B. - POCH, D. (1991) "Some aspects of vowel reduction in Spanish spontaneous speech", *Proceedings of the ESCA Workshop on ' Phonetics and Phonology of Speaking Styles'* Barcelona, 30 September - 2 October 1991.

HOUSE, A.S.- WILLIAMS, C.E.- HECKER, M.H.L.- KRYTER, K.D. (1965) "Articulation-Testing Methods: Consonantal Differentiation with a Closed-Response Set", *Journal of the Acoustical Society of America* 37,1: 158-166.

KRULL, D. (1987) "Second formant locus patterns as a measure of consonant-vowel coarticulation", *PERILUS* 5:43-61.

LINDBLOM, B. (1984) "On the teleological nature of speech processes", *Speech Communication* 2, 2/3: 155-158.

LINDBLOM, B. (1987) "Phonetic invariance an the adaptive nature of speech", Conferencia presentada en el simposio 'Working Models of Human Perception', Instituut voor Perceptie Onderzoek, Eindhoven, August 26-28, 1987. ms no publicado

LINDBLOM, B. (1990) "Explaining phonetic variation: A sketch of the H and H theory ", in HARDCASTLE, W.J.- MARCHAL, A. (Eds) *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers (NATO ASI Series, Behavioural and Social Sciences, 55) pp. 403-440.

LOMAN, H.- VAN BEZOOIJEN, R. (1988) " Evaluation of Dutch rule synthesis consonants by means of a fast diagnostic test ", *AFN - Proceedings, University of Nijmegen* 12 : 57-72

LLISTERRI, J.- MARTINEZ DAUDEN, G. (1990) " Phonetic interference in bilingual speakers learning a third language: the production of lateral consonants " Comunicació presentada en el 9th World Congress of Applied Linguistics, Tesalónica, Grecia, 12-21 de abril de 1990.

LLISTERRI, J.- MARTINEZ DAUDEN, G. (1991) "Lateral consonant production in bilingual speakers learning a third language", *Actes du XIIème Congrès International des Sciences Phonétiques*. 19-24 août 1991, Aix-en-Provence, France. Aix-en-Provence: Université de Provence, Service des Publications. Vol 2 . pp. 318-321.

LLISTERRI, J.- POCH, D. (1986) "Influence de la L1 (catalan) et de la L2 (castillan) sur l'apprentissage du système phonologique d'une troisième langue (français)", *Actas de las IX Jornadas Pedagógicas sobre la enseñanza del francés en España*. Barcelona: Institut de Ciències de l'Educació, Universitat Autònoma de Barcelona. pp. 153-167.

LLISTERRI, J.- POCH, D. (1987a) "Les réalisations sonores dans l'apprentissage d'une troisième langue par des bilingues", *Revue de Phonétique Appliquée* 82-83-84:209-219.

LLISTERRI, J.- POCH, D. (1987b) "Phonetic Interference in Bilingual's Learning of a Third Language", *Proceedings of the XIth ICPhS. The Eleventh International Congress of Phonetic Sciences* . Estonia: Academy of Sciences of the Estonian SSR, 1987. Volume 5. pp.134-137.

LLISTERRI, J.- POCH, D. (1987c) "Phonetic Aspects of Third Language Learning in Bilingual Students", *Sisenes Jornades Pedagògiques per a l' Ensenyament de l' Anglès* , Barcelona, 4-7 de marzo de 1987, Instituto de Ciencias de la Educación de la Universidad Autònoma de Barcelona

MACK, M. (1982) "Voicing-dependent vowel duration in English and French: Monolingual and Bilingual Production", *Journal of the Acoustical Society of America* 71,1: 173-178.

MACHUCA, M.J. (1991) *Análisis acústico de las consonantes nasales del español en habla espontánea y en habla de laboratorio*. Universidad Autònoma de Barcelona, Departamento de Filología Española. Proyecto de investigación.

MÄGISTE, E. (1984) " Learning a Third Language ", *Journal of Multilingual and Multicultural Development* 5,5: 415-421

MARTI, J.- NIÑEROLA, D. (1987) "SINCAS: un conversor texto-voz en castellano", *Procesamiento del lenguaje natural, Boletín nº 5*: 111-122.

NADEU, C. (1987) *Evaluación subjetiva de la calidad del habla*. Universitat Politècnica de Catalunya, Departament de Teoria del Senyal i Comunicacions. ms no publicado

NAVARRO TOMAS, T. (1946) " Escala de frecuencia de fonemas españoles " in *Estudios de fonología española*. New York: Las Américas Publishing Company, 1966 2a ed. pp. 15-30

NYE, P.W.- GAITENBY, J. (1974) "The Intelligibility of Synthetic Monosyllable Words in Short, Syntactically Normal Sentences", *Haskins Laboratories Status Report on Speech Research SR-37/38*: 169-190

OLABE, J.C.- SANTOS, A.- MARTINEZ, E.- MUÑOZ, E.- MARTINEZ, M.- QUILIS, A.- BERSTEIN, J. (1984) "Real Time Text-to-Speech System for Spanish", *Proceedings of the IEEE-ICASSP-84*, San Diego, 2-10 pp. 1-3.

OLASZY, G. (1980) " MULTIVOX - A flexible text-to-speech synthesis for Hungarian, Finnish, German, Esperanto, Italian and other languages for IBM - PC " in TUBACH, J.P.- MARIANI, J.J. (Eds)

(1989) *Eurospeech 89. European Conference on Speech Communication and Technology*. Paris, September 1989. Edinburgh: CEP Consultants Ltd. vol 2 pp. 525-29

PAVLOVIC, C.- ROSSI, M.- ESPESSER, R. (1989 a) " Direct scaling of the performance of text-to-speech synthesis systems " in TUBACH, J.P.- MARIANI, J.J. (Eds) *Eurospeech 89. European Conference on Speech Communication and Technology*. Paris, September 1989. Edinburgh: CEP Consultants Ltd. vol 2. pp. 644-647

PAVLOVIC, C.- ROSSI, M.- ESPESSER, R. (1989 b) " Subjective assessment of acceptability, intelligibility and naturalness of text-to-speech synthesis ", *Proceedings of the ESCA Tutorial Day and Workshop on Speech Input/Output Assessment and Speech Databases*. Noordwijkerhout, the Netherlands, 20-23 September 1989. pp. 3.2.1.- 3.2.5

PÉREZ, J.C.- VIDAL, E. (1991) " Un sistema de conversión de texto a voz para el castellano", Comunicación presentada en el VII Congreso de la SEPLN, Valencia, 3-4 de octubre de 1991.

PISONI, D. (1987) " Some measures of intelligibility and comprehension " in ALLEN, J.- HUNNICUTT, M.S. - KLATT, D.H. *From Text to Speech. The MITalk System*. Cambridge: Cambridge University Press. pp. 151-171.

PISONI, D. B.- NUSBAUM, H. C.- GREENE, B. G. (1985) "Perception of Synthetic Speech Generated by Rule", *Proceedings of the IEEE* 73,11: 1665-1676.

POCH, D. - FERNANDEZ, N.- MARTINEZ, G. (1989) "Some problems of coarticulation in CV stop syllables in Spanish and Catalan spontaneous speech", in SZENDE, T. (Ed) *Proceedings of the Speech Research '89 (Magyar Fonetikai Füzetek / Hungarian Papers in Phonetics 21)*. Budapest: Linguistics Institute of the Hungarian Academy of Sciences. Vol. 2. pp. 111-115.

POLS, L.C.W. (1989) " Improving synthetic speech quality by systematic evaluation " , *Proceedings of the ESCA Tutorial Day and Workshop on Speech Input/Output Assessment and Speech Databases*. Noordwijkerhout, the Netherlands, 20-23 September 1989. pp. 3-11

POLS, L.C.W. (Ed) (1990) *Speech Input / Output Assessment and Speech Databases*, Special Issue, *Speech Communication* 9,4.

RINGBOM, H. (1986) " Crosslinguistic influence and the foreign language learning process " in KELLERMAN, E.- SHARWOOD-SMITH, M. (Eds) *Crosslinguistic Influence in Second Language Acquisition*. New York: Pergamon Press. pp. 150-162

RINGBOM, H. (1987) *The Role of the First Language in Foreign Language Learning*. Clevedon, Avon: Multilingual Matters.

ROBERT, J.M.- CHOINIERE, A.- DESCOUT, R. (1989) " Subjective evaluation of the naturalness and acceptability of three text-to-speech systems in French " in TUBACH, J.P.- MARIANI, J.J. (Eds) *Eurospeech 89. European Conference on Speech Communication and Technology*. Paris, September 1989. Edinburgh: CEP Consultants Ltd. vol 2. pp. 640-643

SERRA, A. (1991) *Un test de evaluación de habla sintetizada para el castellano: las frases semánticamente anómalas de Haskins*. Universidad Autónoma de Barcelona, Departamento de Filología Española. ms no publicado.

VALERO, A. (1991) *El corpus de las frases psicoacústicas de Harvard: una adaptación al castellano*. Universidad Autónoma de Barcelona, Departamento de Filología Española. ms no publicado.

VIEREGGE, W.H.- KERKOF, P.A.M.- BOVES, L.- VAN GERWEN, R. (1987) " Automatic Text-to-speech Conversion for Spanish " *Proceedings of the Institute of Phonetics, Catholic University of Nijmegen* 11: 29-30.

VOIERS, W D. (1984) "Evaluating Processed Speech using the Diagnostic Rhyme Test", *Speech Technology* 1,4: 30-39.

WILLIAMS, L. (1982) "Phonetic Variation as a Function of Second Language Learning" in YENI-KOMSHIAN, G. - KAVANAGH, J.F.- FERGUSON, C.A. (Eds.) *Child Phonology. Vol 2: Perception*. New York: Academic Press. pp. 185-215.