

HUMANITAS

ANUARIO DEL CENTRO DE ESTUDIOS HUMANÍSTICOS

13



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

1972

LETRAS
Sección Segunda

LOS SINTETIZADORES Y LA SÍNTESIS DEL LENGUAJE *

LIC. EDUARDO GUERRA CASTELLANOS
Centro de Estudios Humanísticos de la
Universidad de Nuevo León Escuela de
Letras del I.T.E.S.M.

DURANTE MUCHOS AÑOS, desde los inicios de la ciencia fonética, los lingüistas se dedicaron, principalmente, a la observación de la formación de los sonidos, pero atendiendo a sus características más generales. Estas características estaban basadas exclusivamente en descripciones fisiológicas o articulatorias. Es decir, el punto de articulación, el modo de articulación, la sonoridad, etc. Gracias a estas observaciones, el campo de la Fonética Genética se desarrolló de una manera considerable. Sin embargo, junto a esta Fonética descriptiva y fisiológica, hay otra rama, un tanto menos desarrollada, pero que tiene una primordial importancia en la lingüística contemporánea: la Fonética Acústica. Dice el Dr. D. Antonio Quilis: "La fonética acústica no fue olvidada por los antiguos fonetistas, pero tuvo un desarrollo menor que la fisiológica. Esto fue debido a que los estudios fisiológicos avanzaron mucho más rápidamente que los de la acústica".¹

Hay que subrayar un hecho que reviste una especial importancia. La Fonética no sólo debe contentarse con una descripción articulatoria de los sonidos, sino que además debe tomar en consideración su aspecto acústico.

Todas las actuales investigaciones en el campo de la Fonética se hacen ya desde los dos puntos de vista: articulatorio y acústico. "Como dice Bertil Malmberg, era errónea la opinión de toda la fonética clásica de que cada posición diferente de la lengua daba lugar a un sonido nuevo. Hoy podemos comprobar que en la caja de resonancia bucal se compensa la variación de

* El autor agradece la valiosa colaboración de la Biblioteca del Instituto Miguel de Cervantes del Consejo Superior de Investigaciones Científicas de Madrid.

¹ ANTONIO QUILIS MORALES, *El Método Espectográfico* (Notas de Fonética Experimental). *RFE*, XLIII, 1960, p. 415.

un órgano 'A' que producía antes de su movimiento un sonido 'X', con la modificación de otro órgano 'B' para dar un resultado acústico análogo al primitivo 'X'." ²

Los primeros intentos dentro de este campo de la fonética experimental se deben a Helmholtz, Rousselot, Gemeli y Pastori y otros. El primero —Helmholtz—, experimenta con sus resonadores. Rousselot trabaja con análisis quimográficos. Gemeli y Pastori, con sus métodos electroacústicos. Ellos, prácticamente, inician el movimiento que tan grandes avances habría de traer a esta importante parcela del conocimiento lingüístico.³ Sin embargo, no será sino hasta los finales de la segunda guerra mundial, precisamente en el año de 1947, cuando se dé el mayor paso dentro de este conocimiento. Es en este año de 1947 cuando aparece el libro de Potter, Kopp y Green, llamado *Visible Speech*, en donde se dan los primeros pasos en el análisis espectrográfico.

El espectrógrafo, cuya finalidad es descomponer el lenguaje automáticamente en sus componentes, se ideó con la intención de que los sordos pudieran leer, sobre una pantalla móvil, una conversación telefónica. Esta conversación se descomponía por medio de filtros en los formantes característicos de cada uno de los fonemas. Fue la Compañía Telefónica Bell, en sus laboratorios, quien hizo los primeros intentos. "Al poco tiempo de ver que las pruebas realizadas con el espectrógrafo daban resultados satisfactorios, la Kay Electric Company comenzó a construir este aparato.

Los espectrógrafos contruidos por esta casa antes de 1948, alcanzaban una frecuencia de sólo 3500 cps, lo que hacía difícil el estudio completo de muchas consonantes, cuyos rasgos espectrográficos comienzan a esa frecuencia. Más tarde, después del referido año, la escala de frecuencias espectrográficas fue aumentada hasta 8000 cps".⁴ El perfeccionamiento de este aparato ha dado una amplia gama de posibilidades en la experimentación.

El espectrógrafo de la Kay Electric Company está constituido por tres partes esenciales: *el sistema de filtros* (debemos recordar que el sistema de filtros es el elemento más importante del aparato ya que el espectro no es otra cosa que el resultado de la dispersión de un conjunto de radiaciones o de ondas. Los filtros tienen por objeto dispersar, conforme a módulos constantes, la energía acústica que llega a ellos), *el disco magnético y el cilindro reproductor*. Es evidente que el espectrógrafo lleva, aparte de estos instrumentos esenciales, otra serie de mecanismos que intervienen en la experimentación.⁵

² QUILIS MORALES, ANTONIO, *op. cit.*, p. 415.

³ Cfr. *Ibid.*, p. 415.

⁴ QUILIS MORALES, ANTONIO, *op. cit.*, p. 416. Cfr. DELATTRE, PIERRE, *Studies in French and Comparative Phonetics*. Edit. Mouton and Co. The Hague, 1966, pp. 250-51.

⁵ Cfr. QUILIS MORALES, ANTONIO, *op. cit.* et DELATTRE, PIERRE, *op. cit.*

Observa el Dr. Antonio Quilis que el tiempo máximo que puede analizarse con el espectrógrafo es de 2,4 segundos. Sin embargo, en investigaciones recientes realizadas en Estocolmo por H. Sund, se ha ideado un nuevo procedimiento que, aunque basado en la misma técnica espectrográfica, tiene la posibilidad de analizar contenidos mayores de sonidos. Este método de Sund consiste en reflejar el sonido descompuesto por medio de filtros en una pantalla.

Con una cámara fotográfica, agregada al espectrógrafo, se recoge toda la imagen del sonido. Tiene además una ventaja sobre los espectrógrafos de la Kay Electric Company y es que puede recoger sonidos cuya frecuencia esté en los 10000 cps. Observa Gunnar Fant: "The spectrograph proper is composed of 48 high quality bandpass filters which are scanned by an electronic switch at a rate of 500 times per second. The intensity of rather signal amplitude in each filter is represented by the height of a corresponding vertical line on the screen of a cathode ray tube where the separate filters are ordered in a horizontal row. This is the intensity versus frequency display providing the 'sections'. On a separate cathode ray tube the filter channels are ordered as points in a vertical row, and the intensity of the light beam at each point is modulated by the signal amplitude of the corresponding filter. The time axis has to be supplied by the continuously moving film of a camera attached to the oscilloscope. A film speed of 5 cm/s is utilized. A special time mark signal recurring with intervals of 1/5 second is displayed on the top of the spectrographic picture. All filters up to a centre frequency of 4000 c/s except the first one have a bandwidth of 300 c/s, and successively broader filters are used in the higher frequency region. This filter No. 48 covers the frequency region of 9400 - 10 000 c/s, and filter No. 1 the interval 0-200 c/s. Up to 1000 c/s there is an overlap by a factor of 4, which means that there is a distance of 75 c/s between the centre frequencies of adjacent filters. Between 1000-2000 c/s the filters are arranged with 3 times overlap, i.e. Above 100 c/s frequency intervals. Up to 3600 c/s this interval is 150 c/s. Above 4000 c/s there is no overlap. By this arrangement the centre frequencies of the filters are approximately distributed on the mel scale simulating the frequency to place conversion of the auditory mechanism. There are more filters (and thus more space on the spectrograms) devoted to the low and medium frequency region than to the higher frequency region. This is an advantage compared to the Sonagraph which has a linear frequency display (...). With the present camera it is possible to expose 10 minutes of speech at a time, but other cameras with larger reels may take up to 30 minutes of speech at a time".⁶ Evidentemente con este método se logran ventajas que con el sim-

⁶ FANT, GUNNAR, "Modern Instruments and Methods for Acoustic Studies of Speech", *Acta Polytechnica Scandinavica*. Oslo Univ. Press. Oslo, 1958. PH 1 (246/1968), pp. 44-47.

ple espectrógrafo, pero tiene un ligero inconveniente que es el del tiempo perdido en el revelado de la película. Con el espectrógrafo se tiene el resultado de inmediato. No obstante muchas veces el tiempo no importa tanto si se logran beneficios tales como una mayor extensión de trozo fónico para ser analizado.

Volviendo al espectrógrafo de la Kay Electric Company, que de suyo ya representa un gran avance en el campo de la acústica experimental, hay que hacer notar que el espectrograma que de él se obtiene presenta tres representaciones: 1o. Sobre el eje de las abscisas, y de izquierda a derecha, se nos da la *duración* total del espectrograma y, por supuesto, de cada uno de los sonidos; 2o. En el eje de las ordenadas, *la escala de frecuencias* y 3o. En el mayor o menor negro de los formantes observamos *la intensidad* con que se pronuncia un sonido con relación a otro.

Tenemos que recordar que cada uno de los formantes no representa un tono único, sino una zona en la cual se ponen de relieve una serie o conjunto de armónicos. Estos armónicos están a una frecuencia determinada.⁷

Entre las ventajas que de primera vista se resaltan en el análisis espectrográfico hay que señalar la coincidencia de movimientos articulatorios con determinadas señales que aparecen en el espectrograma. Basta citar algunas observaciones que el Dr. D. Antonio Quilis hace en su trabajo *El Método Espectrográfico*. Nos afirma: "Se ve claramente la sonoridad o sordez de los sonidos consonánticos... por la presencia o ausencia de un formante bajo (...). Cuando hay una completa cerrazón de la región bucal, no aparecen formantes altos que son los característicos de las vocales. En estas condiciones sólo pasa a través de la boca un sonido pequeño que procede de las vibraciones de las cuerdas vocales. Este sonido aparece reflejado en el espectrograma en una región de frecuencia muy baja que no llega al punto cero".⁸ Los primeros en observar estos fenómenos fueron Pierre Delattre y luego Gunnar Fant. De dichas observaciones se pueden deducir los siguientes aspectos:

El primer formante guarda una relación estrecha con la abertura del canal bucal: cuando la abertura es máxima, esto es, cuando la lengua está más separada del velo del paladar, la frecuencia de dicho formante es la más elevada; por el contrario, si la lengua se acerca más al paladar, la abertura vocálica decrece, y la frecuencia del formante también disminuye. El segundo formante puede sufrir modificaciones ya por la posición de la lengua (más elevación y anterioridad en su posición, más alta su frecuencia; más elevación y posterioridad en su posición, menor su frecuencia), ya por los movimientos

⁷ Cfr. QUILIS MORALES, ANTONIO, *op. cit.*, p. 422.

⁸ *Ibid.*, p. 423.

de los labios (más redondamiento y abocinamiento de los labios, menor la frecuencia). El tercer formante varía con los movimientos velares (cuando el velo del paladar desciende, tomando una posición análoga a la adaptada para la producción de vocales nasales, la frecuencia se eleva y viceversa).⁹

El espectrógrafo ha servido también para clarificar algunos conceptos sobre la clasificación de los sonidos producidos por el hombre en función de la comunicación. Tal es el caso de los conceptos de Densidad y Difusión de los sonidos vocálicos y consonánticos. Las nuevas tendencias del estructuralismo lingüístico utilizan ya estos conceptos con toda propiedad. Tal es el caso de lo que hacen Román Jakobson y André Martinet. Dice el Dr. Quilis "Basado en el grado de separación o acercamiento de los formantes vocálicos es como se ha llegado a un rasgo distintivo de las vocales al considerarlas en oposición según su grado de densidad o difusión; llamando a las que tienen un mayor acercamiento entre sus dos primeros formantes, densas o compactas, por oposición a las difusas en las que la separación de las zonas de formantes es mayor".¹⁰ También las consonantes han sido clasificadas de igual forma, según su mayor o menor densidad de los formantes que las constituyen. Se dice que este grado de densidad o difusión depende de la cavidad resonadora.

Pero junto a este maravilloso aparato que tanto puede allegar a la experimentación en el laboratorio hay que mencionar otro, que de manera complementaria con el primero, o bien de manera autónoma puede dar nuevas luces en la investigación de la fonética acústica: *el sintetizador de lenguaje*.

Nuestro trabajo, bien humilde ciertamente, está orientado a presentar de una manera general y no, evidentemente, total, algunas de las investigaciones que con este último aparato, en todas sus versiones, se han venido realizando en los presentes años. Hemos de estudiar, en primer término, los tipos de sintetizadores que se han fabricado. En algunos casos entraremos a la especificación técnica de los sintetizadores. También creemos interesante el presentar, de una manera somera, algunos resultados de investigaciones y las consecuencias que de dichas observaciones han resultado como beneficio para la Ciencia Fonética Acústica. No es posible, y esto surge de manera necesaria, abarcar todo el estadio de conocimientos que implican dichos aparatos. Nuestro objeto es más simple.

⁹ Cfr. DELATTRE, PIERRE, *op. cit.* (passim) et QUILIS M., ANTONIO, *op. cit.*, pp. 425-26.

¹⁰ QUILIS MORALES, ANTONIO, *op. cit.*, p. 424.

Los sintetizadores del lenguaje, junto con los espectrógrafos y la cine-radiografía, son en la actualidad los más modernos auxiliares en la experimentación de la Fonética Acústica. Gracias a ellos, un sinnúmero de problemas que hasta hace muy pocos años aparentemente carecían de solución ahora han sido resueltos, aportando con ello no sólo un gran desarrollo a la propia ciencia de la fonética, sino también a la ciencia lingüística en general.

El propósito del presente trabajo, como ya lo hemos dicho anteriormente, es el presentar la descripción del primero de estos aparatos en sus diferentes versiones.

En términos muy generales se puede hablar de los tres sintetizadores de los laboratorios Haskins: los llamados SP, SV y SO. El sintetizador PAT del laboratorio de fonética de la Universidad de Edimburgo. Los dos sintetizadores del Royal Institute of Technology de Estocolmo: OVE I y OVE II. El sintetizador LEA, descrito por Gunnar Fant. Y, por último, el sintetizador DANA del M.I.T. el cual, en su versión actual, está controlado por un computador TX-O.

Evidentemente existen otros muchos aparatos cuya función sea la síntesis del lenguaje, sin embargo creemos que con penetrar en algunos de los ya mencionados se puede tener una idea de las posibilidades experimentales de los mismos.

Los sintetizadores de los Laboratorios Haskins:

Observa Pierre Delattre, en su obra *Studies in French and Comparative Phonetics*, que los laboratorios Haskins han construido tres sintetizadores de la palabra: el SP, el SV y el SO.

Los dos primeros, SP y SV, fueron construidos para convertir en sonidos los espectrogramas "naturales" de voz humana y los espectrogramas realizados a mano.

El sintetizador SO sólo puede articular sílabas aisladas.

Afirma Delattre: "Les trois (SP, SV et SO) sont faits, non pour produire la parole la plus réaliste possible (le phonographe et le magnétophone se chargent de cela), mais pour fournir de bons instruments de recherche. Les qualités qu'on exige d'eux sont la flexibilité et la versatilité: ils doivent permettre d'isoler, puis de faire varier dans toutes leurs dimensions, les nombreux

éléments acoustiques de la parole; et au fur et à mesure, de faire entendre les résultats de ces manipulations".¹¹

El sintetizador SP posee un juego de cincuenta tonos puros, es decir, ondas sinusoidales, que van desde una frecuencia de 120 ciclos por segundo hasta los 6000 cps. Estos tonos puros son accionados por 50 fotoceldas de una dimensión individual de 1/10 de pulgada. Las cincuenta fotoceldas tienen, en total, una dimensión de cinco pulgadas, las mismas que tiene el espectrograma, y están dispuestas para corresponder a las frecuencias de los cincuenta armónicos de un espectrograma natural de la misma dimensión. Dice Delattre: "Lors qu'un spectrogramme passe, à une vitesse constante donnée, sous les rayons de lumière, tous les tons purs qui correspondent aux formants se mettent à jouer. Chaque formant fait jouer, en moyenne, trois tons purs contigus celui du centre étant typiquement plus intense que les deux autres. Sauf tout au bas de l'échelle des fréquences, où les intervalles sont grands, un formant de trois tons contigus, joué seul, forme pour l'oreille une violente dissonance. Mais deux formants, donc six tons en deux groupes de trois tons contigus, s'entendant comme une belle voyelle —et non plus comme une dissonance— à condition seulement que les fréquences des formants correspondent à celles de voyelles connues du sujet entendant".¹²

En el sintetizador SP se pueden hacer pasar espectrogramas realizados a mano, es decir artificiales. Estos espectrogramas se pintan con pincel, teniendo en cuenta que para cada armónico debe corresponder una línea en el sentido del tiempo. La frecuencia será tanto más alta cuanto la línea sea más elevada. La duración más grande, cuanto la línea sea más larga (a razón de 7,2 pulgadas por segundo). Y, por último, la intensidad será más fuerte en la medida en que la línea sea más larga o más reflejante (longitud máxima: 1/10 de pulgada por armónico).

En la práctica se pinta de un solo trazo lo que cubriría un canal enteramente y la mitad de los dos canales contiguos al primero. Toda línea unida produce sonidos periódicos. Para producir sonidos no periódicos debe puntearse lo más irregularmente que sea posible. Los ruidos de fricción producidos de esta manera, tales como podrían ser una "s" o una "f", son menos naturales que los sonidos de una vocal, pero sin embargo son satisfactorios. Los ruidos de explosión, como lo podrían ser una "p" o una "t", se caracterizan con rasgos breves, más o menos verticales, con una longitud de frecuencia de aproximadamente 600 cps (cinco canales contiguos), siendo los mismos muy satisfactorios. En cambio, las variaciones de frecuencia de un tono la-

¹¹ DELATTRE, PIERRE, *Studies in French and Comparative Phonetics*. Edit. Mouton and Co. The Hague, 1966, p. 254.

¹² DELATTRE, PIERRE, *op. cit.*, p. 254

ríngeo, son imposibles de imitar por el sintetizador SP ya que las fundamentales están fijadas a 120 cps.¹³ En resumen, si se puede admitir, dice Delattre, que la palabra sea inteligible sin entonación, el sintetizador SP tiene un gran mérito.

El sintetizador SV, de los mismos Laboratorios Haskins, está más perfeccionado que el anterior. Pero este perfeccionamiento sólo es en un sentido, pues es menos flexible que el sintetizador SP.

La diferencia esencial del sintetizador SV con el SP, es que con el primero basta la pintura de una sola línea para producir automáticamente un formante completo de intensidad variable según la longitud de la línea. Esto hace que en última instancia los formantes sean más próximos a los formantes de la palabra natural tanto en el amortizamiento de las ondas, como en la relación de fase de los diversos armónicos que entran dentro de un formante. Sin embargo, dice Delattre: "Les formants de SV apportent donc une amélioration du point de vue du 'naturel'. D'autre part, ils enlevant un peu de flexibilité puisqu'on ne peut plus contrôler les harmoniques individuellement, come sur SP".¹⁴

Por otra parte, SV produce un verdadero ruido. El mismo trazo de pintura puede ser entendido como un formante de sonido periódico o como formante de sonido turbulento. Con el sintetizador SV se puede variar la frecuencia fundamental, haciendo con esto posible el estudio de la entonación y el acento.¹⁵

El sintetizador SO, por último, es en cambio de un tipo totalmente diverso a los dos anteriores. Este último aparato de los Laboratorios Haskins ha sido construido fundamentalmente para estudiar las transiciones. Este sintetizador se controla por una serie de botones de regulación que permiten construir una sílaba con entonación.

Con el sintetizador SO se puede controlar la frecuencia y tiempos de tres formantes automáticos (tal y como se puede hacer con el sintetizador SV, pero con la enorme ventaja en este de que son más naturales) los cuales pueden ser periódicos o no. Se puede, también, controlar la velocidad de los cambios en las curvas de transición y las intensidades de cada uno de los formantes, separadamente, así como la duración de cada segmento de sonido. Dice Delattre: "Quand tous les boutons de réglage sont au point désiré, on déclanche la production sonore de la syllabe entière".¹⁶

Analizados pues, en general, los sintetizadores de los Laboratorios Haskins, nos dan una idea de su uso y proyección en la experimentación de Fonética

¹³ Cfr. DELATTRE, PIERRE, *op. cit.*, p. 254.

¹⁴ DELATTRE, PIERRE, *op. cit.*, pp. 254, 255.

¹⁵ *Ibidem*. Cfr.

¹⁶ DELATTRE, PIERRE, *op. cit.*, p. 255.

Acústica. Cada uno de ellos supone un avance con respecto a los otros y, sin embargo, en alguna medida, el adelanto supone a la vez limitaciones.

El sintetizador "Pat" de la Universidad de Edimburgo:

El sintetizador PAT de los laboratorios de fonética de la Universidad de Edimburgo, produce cuatro formantes (automáticos en la misma medida en que los producen los sintetizadores SV y SO de los Laboratorios Haskins), los sonidos periódicos y los sonidos turbulentos así como los cambios de frecuencia en sus armónicos para dar la entonación. En términos generales el sintetizador PAT es del mismo tipo que los sintetizadores SP, SV y SO de Haskins, pero menos flexible que ellos en el sentido de que no puede manipular aisladamente las variables acústicas de la palabra, sea por los botones de control, sea por los diseños en perfil sobre la placa de proyección. Delattre afirma: "De plus PAT ressemble à SP et SV en ce qu'il suit les changements des formants dans le temps, et produit donc des phrases; les Analogues ne produisent encore que des sons isolés soutenus".¹⁷ Elizabeth T. Undall, de la Universidad de Edimburgo dice: "...this a resonance analogue synthesizer with the following parameters:

- Larynx Frequency
- Larynx Amplitude
- Frequency of Formant 1
- Frequency of Formant 2
- Frequency of Formant 3
- Amplitude of a 'hiss' with a single peak
- Centre frequency of this peak
- Amplitude of a 'hiss' with formant structure".¹⁸

Los sintetizadores de lenguaje del Royal Institute of Technology de Estocolmo:

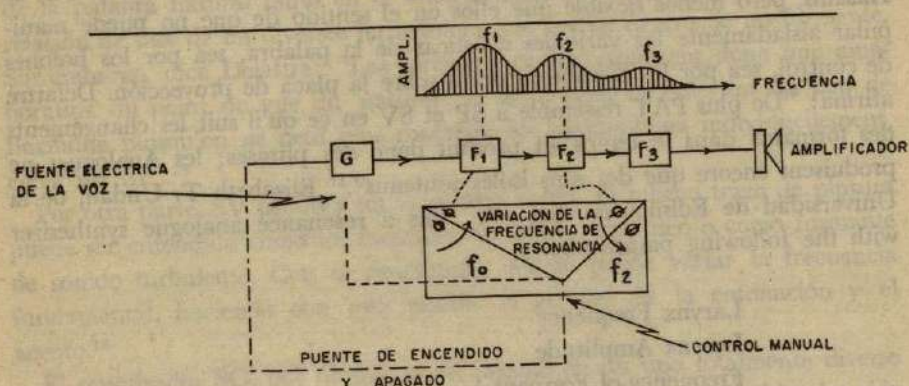
Los dos sintetizadores del R.I.T. de Estocolmo, OVE I y OVE II, pertenecen a lo que Gunnar Fant llama "Formant Circuit Synthesizers". El sintetizador OVE I pertenece a este tipo de aparatos en los cuales se comprende un cierto número de circuitos de formantes uno para cada formante que

¹⁷ DELATTRE, PIERRE, *op. cit.*, p. 255.

¹⁸ ULDALL, ELIZABETH, "The synthesis of some Sounds made on other than Pulmonic Air-Stream Mechanisms", *Phonetica*, XIII, p. 105 (1965).

va a ser representado. Los circuitos de formantes, dice Fant, pueden estar dispuestos en serie, tal y como se ha hecho en OVE y en otros sintetizadores (p. ej. DOVO), o bien en paralelo, como se ha hecho en algunas máquinas inglesas. Este último sistema es más flexible cuando vocales y consonantes tienen que ser producidos con un mínimo de circuitos, pero tiene la desventaja de no dar naturalidad.¹⁹

En el sintetizador OVE I, que como hemos visto pertenece a los aparatos cuyos circuitos se encuentran en serie, las frecuencias de formantes, la frecuencia de la voz fundamental y el inicio de la voz, son controlados manualmente. Esto se puede ver en la gráfica simplificada del aparato:



El desplazamiento angular de las dos varillas, cada una de las cuales está conectada a un potenciómetro, determina el tono de frecuencia de F_1 y F_2 . Dice Fant: "The position within the plane of the manocuvre board where the two rods meet in a joint constitutes a reference for the calibration. The fundamental pitch F_0 as well as the on-off switching of the voice source and these two formant frequencies are all varied by means of a single one-hand control. F_3 can be controled separately, and a special press button pitch box can be used for the F_0 -control to give a stepwise variation of the pitch as in song".²⁰ Afirma Fant que el aparato ha sido especialmente construido para demostrar la dependencia del color vocálico con las frecuencias de los formantes. Tanto las frases breves como los diptongos son fáciles de realizar en este sintetizador. Dentro de las frases breves, nuestro autor hace hincapié en

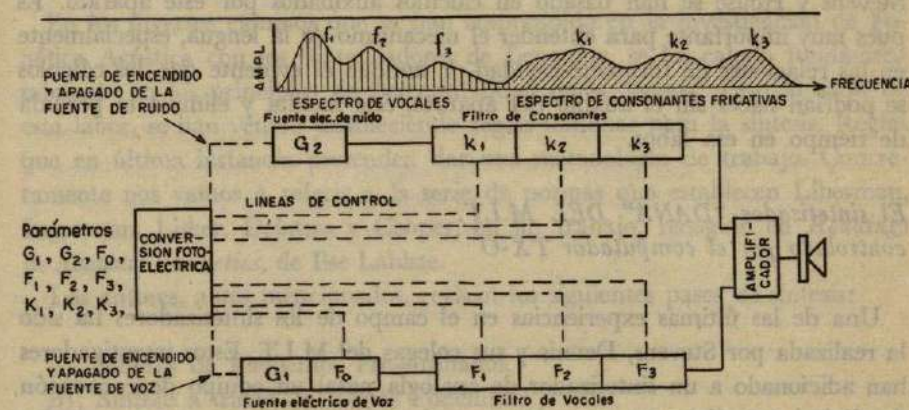
¹⁹ FANT, GUNNAR, "Modern Instruments and Methods for Acoustic Studies of Speech", *Acta Polytechnica Scandinavica*. Ph 1 (246/1958), Oslo Univ. Press. Oslo, 1958 (*passim*).

²⁰ FANT, GUNNAR, *op. cit.*, pp. 72-73.

aquellas que contienen las consonantes /w, v, j, r, l/. Se subraya que la lengua inglesa puede ser muy bien simulada con este aparato, al grado de que no es necesario, para reconocer muchas de las frases inglesas, que el oyente tenga especiales condiciones.

En resumen, han sido tan buenas las condiciones que ofrece el sintetizador OVE I que puede utilizarse para estandarizar las cualidades vocálicas y los símbolos fonéticos. Aparte del uso que de él se puede hacer como instrumento de demostración en las clases de fonética.

El sintetizador OVE II es más complicado. Se controla con base a un generador fotoeléctrico que convierte las curvas previamente dibujadas tanto de variaciones temporales de frecuencias de formantes, el tono fundamental y el carácter originario así como la intensidad en instrucciones continuas a las unidades particulares de la máquina parlante. Se puede decir que su operación es muy similar al sintetizador inglés diseñado por Lawrence.²¹ El esquema simplificado de OVE II puede darnos una idea general de su construcción y avance con respecto a OVE I.



El sintetizador OVE II se ha construido para investigar la importancia diferencial de las grandes variables del habla. Es decir, no sólo distinciones de carácter fonemático, sino problemas de los correlatos acústicos de la naturaleza del habla. Sin embargo, como dice Gunnar Fant: "The machine will not be used for any larger project of this type before it has been developed further. A more stable function generator is also needed".²²

²¹ Cfr. FANT, GUNNAR, *op. cit.*, pp. 73-74.

²² FANT, GUNNAR, *op. cit.*, p. 74.

El sintetizador LEA:

Este aparato sintetizador, descrito por Gunnar Fant en su artículo "Modern Instruments and Methods for Acoustic Studies of Speech", es un poco diferente a los otros modelos mencionados anteriormente. Pertenece a la clasificación de "Configurative Analogs". La fuente eléctrica simula, en este sintetizador, la función de las cuerdas vocales. Los resultados a los que lleva son muy similares a los logrados por OVE I y OVE II. El aparato está compuesto por un gran número de bobinas y condensadores que están relacionados a la cámara de aire cercana al tracto vocálico. Tiene cuarenta y cinco secciones de filtrado, cada una de las cuales está compuesta por bobinas y condensadores de eliminación. Cada una de estas secciones representa 0.5 cm. del corte perpendicular de las cavidades en dirección a donde el aire sale. Cada uno de estos cortes, en su área de cruce seccional, describe las variaciones del área respectiva desde las cuerdas vocales hasta los labios.²³

Es un hecho muy significativo que todas las investigaciones sobre las relaciones entre la articulación y los patrones de formantes que han realizado Stevens y House se han basado en cálculos auxiliados por este aparato. Es pues muy importante para entender el mecanismo de la lengua, especialmente en las relaciones de formante-cavidad. Y aunque es evidente que los cálculos se podrían hacer sin el auxilio del aparato, éste facilita y elimina la pérdida de tiempo en esa labor.

El sintetizador "DANA" DEL M.I.T. controlado por el computador TX-O

Una de las últimas experiencias en el campo de los sintetizadores ha sido la realizada por Stevens, Dennis y sus colegas del M.I.T. Estos investigadores han adicionado a un sintetizador de analogía nasal un equipo de conversión de digital a análogo que permite controlar el aparato con un computador TX-O. Dice Franklin S. Cooper en su artículo "Instrumental Methods for Research in Phonetics": "This step was aimed at the greatly increased flexibility of control by computer as compared with that provided by the original circuitry".²⁴

En resumen, podemos ver cómo se han desarrollado diferentes tipos de

²³ FANT, GUNNAR, *op. cit.* (*passim*). Cfr.

²⁴ COOPER, FRANKLIN S., "Instrumental Methods for Research in Phonetics", *Proceedings of the Fifth International Congress of Phonetic Sciences*, Edit. S. Karger, Basel, Switzerland, 1965, p. 152.

sintetizadores, cada uno de los cuales, no sólo representa un avance con respecto a los demás, sino que, cada uno de ellos, en lo particular, tienden a investigar un determinado aspecto del lenguaje del hombre. En el Congreso de Helsinki se discutió muy ampliamente sobre esta problemática y se puso un especial énfasis en la efectividad con que la síntesis pudiera ser controlada. Lo que se buscaría, en todo caso, sería la simplicidad mecánica que los aparatos pudieran ofrecer. Pero aparte de ello, un mayor realismo en la imitación de la voz humana. Todas las investigaciones que en este orden se han venido haciendo tienden, ciertamente, hacia esa finalidad última: Simplicidad y Realismo.

Una seguridad que tenemos es que, gracias a los grandes pasos que continuamente se dan en la ciencia moderna, pronto se podrá contar con aparatos más perfectos que coadyuven a la labor de investigación en este campo, prácticamente virgen, de la Fonética Acústica.

DE LA SÍNTESIS DEL LENGUAJE

En los diversos caminos que se han emprendido en la investigación de Fonética Acústica con los Sintetizadores de Lenguaje, se ha calado fundamentalmente en los principios de perceptibilidad de la lengua. Pero al lado de esta labor, se han venido estableciendo reglas mínimas para la síntesis. Reglas que en última instancia pretenden dar una metodología de trabajo. Concretamente nos vamos a referir a la serie de normas que establecen Liberman, Ingemann, Lisker, Delattre y Cooper, en un trabajo, recogido en *Readings in Acoustic Phonetics*, de Ilse Lahiste.

Los autores, antes mencionados, revisan los siguientes pasos de síntesis:

- a). Síntesis de Elementos Pre-Grabados.
- b). Síntesis a través de Reglas Fonémicas.
- c). Síntesis por Reglas Sub-Fonémicas.
- d). Reglas adicionales para la Posición.
- e). Reglas para el acento y silabación.

A través de todos estos pasos se van destacando determinados elementos que deben ser reconocidos como reglas de ese método propuesto. Analicemos cada uno de estos pasos propuestos por separado.

A). Síntesis de elementos pre-grabados:

Dicen Liberman y los otros autores: "In this system the incoming phonemes

simply Key the prerecorded sounds. If we instrument such an arrangement, we will almost surely find it quite unsatisfactory. Of all the various difficulties that one will ultimately experience with this system, the most immediately obvious will be a noticeable bumpiness and roughness in the output. One thinks, then, of setting up various smoothing operations, and, indeed, it is surely possible to improve the output by such means. But no amount of smoothing will solve what is here a very fundamental, and, by now, familiar problem. One has only to look at spectrograms to see that speech tends to vary more or less continuously over stretches of greater than phonemic length. The patterns rarely break at what might be considered to be phoneme boundaries, and those who have tried to find acoustic limits of the phoneme have come to know this as the problem of segmentation. Now none of this should be taken to deny the existence of the phoneme, either as a convenient linguistic abstraction or as a perceptual unit.²⁵ Todo lo anterior no quiere indicar otra cosa que, perceptual y lingüísticamente hablando, los fonemas discretos están siempre combinados y, también, en muchos casos encadenados en unidades que son más que un fonema en longitud. Esta, creemos, sea la dificultad que nuestros autores encuentran cuando se trata de tomar o aislar un fonema de la cinta grabada.

Se señala de que si, sin embargo, se trata de producir lenguaje de fonemas pre-grabados se pueden tomar dos vías (las cuales, señalan nuestros autores, son indeseables): Una de estas vías es emplear diferentes grabaciones, o variantes alofónicas, de la mayor parte de los fonemas en la mayor parte de las combinaciones en que éstos participen. Esto, obviamente, requiere de un inventario fabuloso de elementos pre-grabados. Ahora bien, el número de elementos se puede reducir creando clases de variantes en donde cada una de las clases esté representada por una sola y típica forma. Sin embargo esta reducción en el número de variantes sólo se puede hacer comprometiendo muy severamente el punto de partida de la investigación. Otra vía o alternativa es tratar de grabar, o restablecer, los sonidos del lenguaje en una forma muy breve como, por ejemplo, en una rápida recitación del alfabeto. Sin embargo hay una dificultad y es que "the phonemes have become a kind of 'spelling bee'. Nor is this difficulty avoidable: a shift from spelling to phonetic pronunciation only shortens and centralizes the vowel that clings to almost every consonant; indeed, it is difficult to imagine how a voiced stop, for example, could possibly be produced or heard without some vowel-like sound

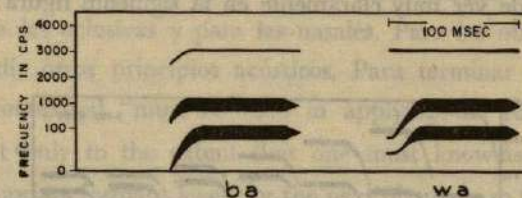
²⁵ LAHISTE, ILSE, *Readings in Acoustic Phonetics*. The M.I.T. press. M.I.T., Cambridge, Mass., 1967, p. 334.

preceding or following it".²⁶ Esta alternativa pues, pone en violencia el proceso de la lengua.

Todo lo anterior no quiere decir que no se pueda trabajar en la síntesis con elementos pre-grabados. Los trabajos de Petersen, Wang y Sievertsen, en relación a la síntesis con elementos discretos, se ha realizado con base a lo que ellos han llamado "diadas", las cuales no son otra cosa que "partes de dos fonos con su influencia mutua en la mitad del segmento". Estos autores estiman que se necesitarían ocho mil diadas para producir un idiolecto.²⁷

B). *Síntesis a través de reglas fonémicas:*

A pesar de que es muy difícil producir lenguaje de segmentos fonémicos pre-grabados, sin embargo es posible, en cambio, generar lenguaje a partir de instrucciones fonémicas discretas. En otras palabras, las reglas para la síntesis pueden ser escritas cuando se es posible ir de unidades fonémicas al lenguaje y luego reducir, por un largo factor, el total de reglas que se necesitan. Esto se hace posible tomando en consideración lo que se sabe acerca de los principios de percepción del lenguaje.



Patrón espectrográfico pintado a mano que ilustra algunos principios de la consonante oclusiva /b/ y de la semivocal /w/.

La figura anterior nos ilustra, dicen los autores de este importante artículo, algunos de los principios y nos apunta una de las razones por las cuales es muy difícil cortar y ensamblar segmentos fonémicos.

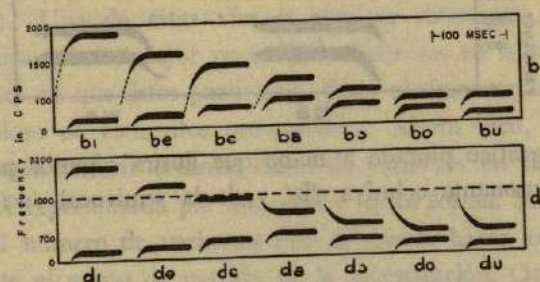
Al investigar con patrones como el anteriormente citado se puede ver que un principio primario para la percepción de esta consonante o de otras es un rápido cambio o desviación en la frecuencia de los formantes y que se

²⁶ LAHISTE, ILSE, *op. cit.*, p. 334.

²⁷ Cfr. LAHISTE, ILSE, *op. cit.*, p. 334.

puede ver a la izquierda de cada patrón. Esta desviación se llama transición (nombre, por cierto infortunado porque da la idea de un mero incidente en el proceso de formación de fonema a fonema). Sin embargo esta transición es uno de los principios fundamentales en la percepción de las consonantes. Dicen nuestros autores: "For the consonant phonemes of Fig. 1, and for others too, it is unqualifiedly true that there is no position in the pattern that will be perceived as the intended consonant, or, indeed, as any consonant when it is in steady state. Sounding the initial steady-state portion of /w/ will cause the listener to hear the vowel /u/. Every point of the transition leading into the steady-state vowel will, if prolonged, produce a vowel-like sound. The listener will perceive /w/ only if he is given information about where the formant begins, where it ends, and how long it takes to move from the one frequency to the other. Normally, this information is conveyed continuously by the transitions. It is always possible, of course, to degrade the patterns to some degree, as for example by erasing parts of the transitions, without utterly destroying the phoneme as perceived".²⁸

Para llegar a las reglas fonémicas para la generación de sílabas como las que aparecen en la figura anterior, se tiene que empezar por tomar en cuenta que todas las transiciones para dar consonantes tienen una característica común. Esta se puede ver muy claramente en la siguiente figura:



Transición del segundo formante apropiado para /b/ y /d/ colocadas ante vocal.

Podemos observar que a pesar de que la extensión y dirección de las transiciones son diferentes para /d/ antes de vocales diferentes, es, sin embargo

²⁸ LAHISTE, ILSE, *op. cit.*, p. 336.

claro que las transiciones han sido originadas aproximadamente en un mismo lugar. Dicho lugar se le ha llamado "Locus".

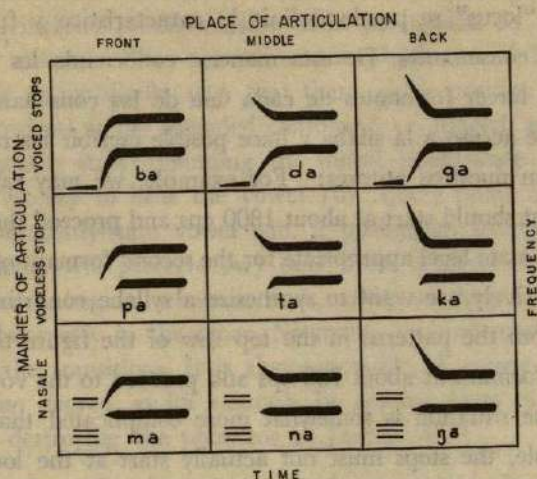
A través del "locus" se puede definir lo característico y fundamental de cada una de las consonantes. De esta manera, conociendo los "loci" de primero, segundo y tercer formantes de cada una de las consonantes se tiene la llave que da libre acceso a la sílaba y hace posible escribir las reglas del nivel fonémico. Dicen nuestros autores: "For example, we may say of /d/ that its second formant should start at about 1800 cps and proceed then at a certain rate to the steady-state level appropriate for the second formant of the following vowel. If, alternatively, we want to synthesize a syllable consisting of /b/ plus vowel, we see from the patterns in the top row of the figure that we should start the second formant at about 700 cps and proceed to the vowel level from there. In fact, the situation is somewhat more complicated than this several ways. For example, the stops must not actually start at the loci-rather, they should only 'point' to them. In the patterns of the figure the dashed lines represent non explicit portions of the complete transition specified by the locus hypothesis".²⁹ Se puede decir que la característica del "locus" es un marcador para las oclusivas y para las nasales. Para las otras consonantes es necesario añadir otros principios acústicos. Para terminar se afirma que la información contextual "must be used in applying the rules for successive phonemes, but only to the extent that one must know-as he must in any case- the appropriate formant levels for the next phoneme so that the transitions may be properly connected".³⁰ Ante una situación como la anterior se puede observar cómo, en principio, el número de reglas se aproxima al número de fonemas.

c). *Síntesis por reglas sub-fonémicas:*

Sin embargo se puede buscar todavía una mayor economía en el número de reglas. Esto se puede hacer mediante su establecimiento a través de una dimensión sub-fonémica. La figura que sigue puede ilustrar esto:

²⁹ LAHISTE, ILSE, *op. cit.*, p. 336.

³⁰ LAHISTE, ILSE, *op. cit.*, p. 336.



Patrón ilustrando algunos principios acústicos para consonantes oclusivas y nasales.

En la figura se puede ver un espectrograma realizado a mano que produce una aproximación entre las sílabas señaladas. Todos los sonidos que tienen la misma transición en el segundo formante se agrupan en una misma columna. Lo mismo sucede para aquellos sonidos que teniendo la misma manera de articulación y la misma transición en el primer formante y, en algunos casos, marcas adicionales como la nasalidad. Con esto es posible establecer una regla para el lugar frontal de articulación, otra, para el lugar medio y, otra más, para el lugar posterior de la articulación. Asimismo, una regla para las oclusivas sonoras, otra para las sordas y una más para las nasales. En esta forma se pueden obtener nueve fonemas con sólo seis reglas.

Sin embargo "it should be noted here that when the rules are written at subphonemic level, arrangements must be made for simultaneous (as well as sequential) combination".³¹

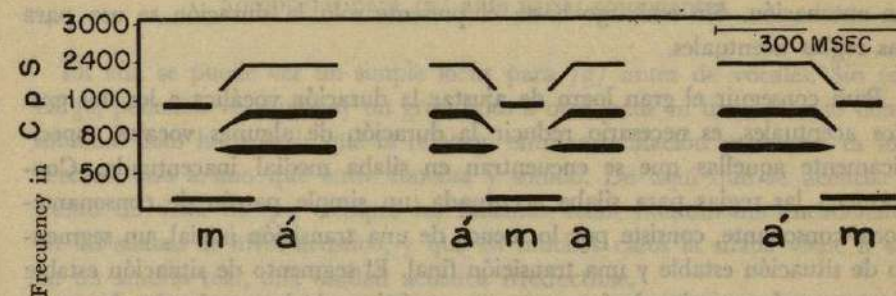
³¹ LAHISTE, ILSE, *op. cit.*, p. 337.

Tal como se puede ver, con este método el número de reglas es considerablemente reducido. Sin embargo, este ideal no puede ser absolutamente cumplido por las reglas adicionales de posición a las que se va a hacer alusión ahora.

d). Reglas adicionales para la posición:

Se puede decir que una de las complicaciones con que se tropieza tanto el método de reglas fonémicas como el de reglas sub-fonémicas es, en muchas ocasiones, la especial provisión de variaciones posicionales.

Sin embargo, en muchos casos es posible producir patrones con posibilidad de que sirvan, a partir de las mismas reglas básicas, para otras posiciones. Es posible pues hacer reglas básicas para la dimensión fonémica o sub-fonémica y de allí derivar los patrones peculiares para cada una de las distintas posiciones. Como ejemplo, se podría tomar el de los patrones para la consonante labial nasal /m/ en posición inicial, intervocálica y final. La siguiente figura nos muestra, en espectrograma realizado a mano, las mencionadas posiciones.



Las reglas básicas de /m/ requieren un estado estable de formantes con duración, intensidades y frecuencias específicas. Además se requiere que todo formante adyacente tenga transiciones de duración específica, que sea discontinuo en relación a los formantes nasales y que apunte a ciertas frecuencias en el "locus".

Se puede ver en la figura de arriba las diferencias entre los patrones de inicial, intervocálica y final de /m/ que implica la presencia o ausencia de transiciones en uno u otro lado de los formantes nasales.

Si se debe o no pintar la transición depende de si los formantes adyacentes están especificados y depende, por supuesto, de las reglas apropiadas para los vecinos inmediatos a /m/ en la secuencia de los fonemas insertos. En otras

palabras, se puede decir que antes de tener una transición se necesitan dos contiguos cuyas reglas nos digan acerca de su condición acústica.

En ciertos casos, sin embargo, no es posible derivar determinados patrones de las reglas básicas y entonces, es necesario escribir enteramente una regla nueva para estas ocasiones. Se puede encontrar que un patrón apropiado puede ser producido simplemente por la cualificación o modificación de la posición a una regla básica.³²

e). Reglas para el acento y silabación:

Quedan aún dos tipos de problemas que hay que tomar en consideración. El primero de ellos se refiere a los hechos prosódicos y en particular al acento. La calidad del lenguaje sintético es considerablemente mejorada tanto en inteligibilidad como en naturalidad si se toman en consideración en las reglas los dos grados de acentuación. Las diferencias acentuales se pueden especificar por una o más cualidades acústicas tales como podrían ser la frecuencia fundamental, la intensidad o la duración. Hay que considerar, por supuesto, que la frecuencia fundamental es la base para las variaciones en entonación. Sin embargo hasta el presente sólo la duración se usa para las reglas acentuales.

Para conseguir el gran logro de ajustar la duración vocálica a los dos grados acentuales, es necesario reducir la duración de algunas vocales, específicamente aquellas que se encuentran en sílaba medial inacentuada. Conforme a las reglas para sílaba acentuada, un simple patrón de consonante-vocal-consonante, consiste por lo menos de una transición inicial, un segmento de situación estable y una transición final. El segmento de situación estable tiene unas frecuencias de formante características de la vocal sola; las transiciones tienen duraciones y puntos de final fijados de acuerdo de lugar y manera para las consonantes. Para convertir cualquier sílaba en una forma apropiada para una condición inacentuada, se debe omitir el segmento de situación estable. Esto significa que tanto el segundo como el tercer formantes son de hecho dibujados como líneas rectas que se conectan con las frecuencias del punto final dadas por las reglas de lugar para consonantes adyacentes. Es necesario que el segundo formante, por otra parte, pase a la región de los 1000 a 2000 cps. Donde la regla de la línea recta puede violar esta restricción —por ejemplo en el caso de vocal entre dos labiales el segundo formante— el segundo formante debe curvarse para encontrarse arriba o

³² Cfr. LAHISTE, ILSE, *op. cit.*, p. 337.

abajo dentro del rango de frecuencia requerido. La configuración del primer formante dependerá de la situación de que sean o no sonoras las consonantes adyacentes. Si son sonoras, el primer formante debe moverse de su posición inicial a la frecuencia de 500 cps. para luego llegar a su frecuencia final; si es sorda, debe permanecer en una frecuencia de condición estable de 500 cps.

La otra complicación de la que hablábamos no es tan frecuente y no entraña una grave consecuencia, pero si reviste de cierto interés. Esta dificultad surge porque ocasionalmente hay una relación más bien compleja entre el fonema como unidad perceptual y el sonido que él representa. Por ejemplo, en la figura de abajo:



Patrón ilustrando las transiciones del segundo formante apropiadas para /g/ ante varias consonantes.

En ella se puede ver un simple locus para /g/ antes de vocales. Sin embargo, podemos ver que hay un gran salto a otro locus en un momento determinado. Esto hace decir que la relación entre articulación y fonema es más cercana uno a uno que entre fonema y sonido. De aquí que se acentúe el punto de vista de que siempre los fonemas están literalmente encadenados en las sílabas al nivel acústico, y que en muchos casos la sílaba viene a ser, en un sentido real, una unidad acústica irreductible.

Veamos por último un ejemplo, que nuestros autores dan, de la aplicación de las reglas que arriba se sostienen. El término a analizar y sintetizar es "LABS" que en secuencia fonemática sería /Laebz/. /l/: El primer fonema pertenece a las consonantes resonantes /l, w, r, j/. La regla de la manera de estas consonantes habla de tres formantes que deben ser mantenidos (con intensidades específicas) en una frecuencia de locus por 30 msecs. La misma regla especifica que los formantes adyacentes deben tener transiciones de 75 msecs., dibujados para ser continuados con las frecuencias del locus. También se fija que el locus del primer formante debe estar a 360 cps. Por último, fija que debe existir un sonido armónico. La regla de lugar para /l/ especifica frecuencias de locus de 360, 1260 y 2880 cps.

/æ/: El siguiente fonema del contexto pertenece a las vocales largas. La

regla de la manera de esta clase habla de tres formantes con la variedad de armónicos, teniendo una duración de 150 msecs. La regla de lugar para /æ/ fija frecuencias de formante a 750, 1650 y 2460 cps. y también específica intensidades para los formantes.

/b/: El siguiente fonema conecta su regla de manera al de las oclusivas /b, d, g, pt, k/. La regla habla de un intervalo de silencio seguido de una explosión y después especifica que los formantes adyacentes tienen un punto de transición de 50 msecs. hacia las frecuencias del locus dadas por la regla de lugar apropiada para esta particular oclusiva. La regla de manera fija, también el locus del primer formante en la frecuencia de la barra de sonoridad. La regla de las labiales que sirve igualmente para /b, p, m, f, v/ especifica que las transiciones adyacentes de los segundo y tercer formantes apunten a las frecuencias de 720 y 2100 cps. respectivamente. La regla de sonoridad para oclusivas (aplicable igualmente para /b, d, g/) requiere que el intervalo de silencio tenga una duración de 70 msecs., y que este intervalo esté cubierto por la barra de sonoridad.

/z/: Para el fonema final la regla de manera que es apropiada para las fricativas /f, v, s, z et al/ habla de un intervalo de ruido de banda limitada. La regla para las fricativas establece que los formantes adyacentes tengan transiciones de 50 msecs. apuntando hacia los loci virtuales dados por la regla de lugar de esta particular fricativa. También, por la regla de manera el locus del primer formante se encuentra a 240 cps. La regla de las alveolares /z/ o /s/ especifica que el ruido (requerido por las fricativas) un bajo corte de frecuencia a 3600 cps. y que las transiciones de los segundo y tercer formantes apunten a una frecuencia de 1800 y 2700 cps. respectivamente. La regla de sonoridad establece que el ruido debe estar en baja intensidad teniendo una duración de 100 msecs. y acompañado por la barra de sonoridad.

Finalmente se aplica el modificador de posición para las sílabas inmediatamente antes del silencio o juntura el cual dobla la duración de la vocal marcando una duración de 300 msecs. para /æ/. Véase el cuadro de la página siguiente:

SYNTHESIS BY RULES: /æbz/. (33)

Manner	<i>Resonants /wrlz/:</i> Periodic sound (buzz); formant intensities and durations are specified. F1 locus is high. Formants have explicit loci	<i>Long Vowels /æææ/:</i> Periodic sound (buzz); formant intensities and durations are specified.	<i>Stops /pbrdkg/:</i> No sound at formant frequencies, i. e. "silence" Burst of specified frequency so band width follows "silence" F1 locus is low F2 and F3 have virtual loci	<i>Fricatives /fvθʃzʒs/:</i> Aperiodic sound (hiss); intensity and band width are specified F1 locus is intermediate. F2 and F3 have virtual loci
Place	// F2 and F3 loci are specified	/l/ Formants frequencies specified	<i>Labials /pbfvm/:</i> F2 and F3 loci are specified Frequencies of buzz and hiss are specified	<i>Alveolars /tdzʒ/:</i> F2 and F3 loci are specified. Frequencies of buzz and hiss are specified
Voicing	(The voicing rules are only applied to those phonemes for which the condition of voicing has differential value. For the resonants and vowels, which are invariably voiced, the acoustic features correlated with voicing are specified under Manner)		<i>Voiced /bdgz/:</i> Voice bar Duration of "silence" is specified F1 onset is not delayed.	<i>Voiced /vzʒ/:</i> Voice bar Duration of hiss is specified. F1 onset is not delayed.
Position		Vowels in final syllable: Duration is double that specified under Manner.		

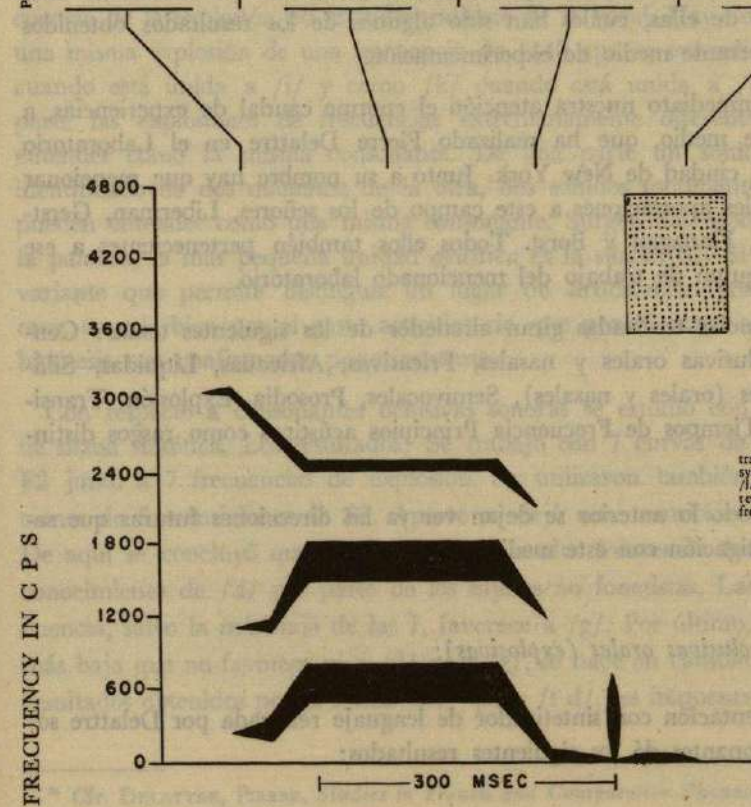


Fig. 8. Table illustrating the rules for synthesizing the word /æbz/, and the pattern derived there from.

Resumiendo lo anteriormente expuesto, podríamos decir que la aplicación de las reglas establecidas por Liberman, Ingemann, Lisker, Delattre y Cooper, permiten un método objetivo y científico de trabajo en la síntesis del lenguaje.

DE LOS RESULTADOS EN ALGUNOS EXPERIMENTOS CON LA SÍNTESIS DEL LENGUAJE.

Entre las muy diversas experiencias que se han realizado con los sintetizadores del lenguaje, se ha llegado a establecer una corriente de estudios, en su mayor parte comprobatorios, de diversos aspectos acústicos del habla. Los sintetizadores han dado resultados muy importantes en el establecimiento de determinados principios, aplicables ahora al estudio sistemático de la Fonética Acústica.

Nuestro propósito, al llegar a esta parte final de nuestro trabajo, es entrar en algunas de estas múltiples experiencias, desarrolladas en los últimos años y ver, dentro de ellas, cuáles han sido algunos de los resultados obtenidos por este importante medio de experimentación.

Llama de inmediato nuestra atención el enorme caudal de experiencias, a través de este medio, que ha realizado Pierre Delattre en el Laboratorio Haskins de la ciudad de New York. Junto a su nombre hay que mencionar las considerables aportaciones a este campo de los señores, Liberman, Gerstman, Cooper, Flanagan y Borst. Todos ellos también pertenecientes a ese maravilloso equipo de trabajo del mencionado laboratorio.

Las experiencias realizadas giran alrededor de los siguientes temas: Consonantes (Oclusivas orales y nasales, Fricativas, Africadas, Líquidas, Silábicas), Vocales (orales y nasales), Semivocales, Prosodia, Explosión, Transición, Locus, Tiempos de Frecuencia Principios acústicos como rasgos distintivos etc.

Aparte de todo lo anterior se dejan ver ya las direcciones futuras que seguirá la investigación con este medio.

Consonantes oclusivas orales (explosivas):

La experimentación con sintetizador de lenguaje realizada por Delattre sobre estas consonantes dá los siguientes resultados:

³³ Cfr. LAHISTE, ILSE, *op. cit.*, pp. 340-41.

1) Las oclusivas orales se distinguen sobre todo por su grado de interrupción del sonido bucal (tenue), por la brevedad del sonido turbulento intenso (explosión) y por la rapidez de la transición que lleva hacia la vocal siguiente o que viene de la vocal precedente.³⁴

2) Las explosiones se han estudiado con la ayuda del sintetizador SP. La experiencia se hace sobre oclusivas iniciales sordas. Las sílabas a identificar están formadas por una explosión sintética seguida de una vocal sintética con dos formantes derechos constituidos por tres armónicos. Las explosiones presentan la forma de óvalos verticales de 600 cps. y a 15 mseg. Teniendo 12 frecuencias diferentes que van de 360 a 4320 cps., éstas se combinan con siete vocales cardinales [i e ε a o u] dando 84 sílabas sintéticas. Las sílabas se graban magnetofónicamente y se dan a identificar a 30 sujetos no fonetistas los cuales las identifican como /p/, /t/ o /k/. Los resultados son claros: Las explosiones altas en los 3000 cps. son identificadas como /t/; las otras sobre los 3000 cps. como /k/ o /p/ según estén situadas por arriba del inicio de F2 (/k/) o en otra parte (/p/). Se revela que el efecto de la frecuencia de la explosión no es independiente de la vocal. Hay una excepción: una misma explosión de una frecuencia de 1440 cps. es entendida como /p/ cuando está unida a /i/ y como /k/ cuando está unida a /a/. Por otra parte las explosiones de frecuencias extremadamente diferentes se pueden entender como la misma consonante. De una parte un sonido puede ser identificado de dos maneras; de la otra, dos sonidos totalmente distintos se pueden entender como una misma consonante. Surgen dos hipótesis: a).—En la palabra, la más pequeña unidad acústica es la sílaba. b).—Si existe un invariante que permite distinguir un lugar de articulación consonántico de otro, es más bien por el gesto articulatorio que por el rasgo acústico. Estas hipótesis son confirmadas posteriormente.

Con respecto a consonantes oclusivas sonoras se estudió con 294 modelos de sílaba sintética. Los resultados: Se trabajó con 7 curvas de transición de F2 junto a 7 frecuencias de explosión. Se utilizaron también 35 combinaciones de 5 transiciones de F3. Aparte se dejó una transición fija para F1. De aquí se concluyó que las explosiones de alta frecuencia favorecen el reconocimiento de /d/ por parte de los sujetos no fonetistas. Las de baja frecuencia, salvo la más baja de las 7, favorece a /g/. Por último, la frecuencia más baja que no favorece ni a /d/ ni a /g/, lo hace en cambio con /b/. Los resultados obtenidos por la síntesis son "pour /t d/, les fréquences sont hautes;

³⁴ Cfr. DELATTRE, PIERRE, *Studies in French and Comparative Phonetics*. Ed. Mouton and Co. The Hague, 1966, p. 256.

pour /p b/ elles sont basses; et pour /k g/, elles sont intermédiaires mais dans une très grande marge de fréquences parce qu'elles suivent les transitions de F 2, qui varient d'environ 3000 cps. à 600 cps."³⁵

3) La transición de las oclusivas: Se conviene en llamar a las transiciones correspondientes a los tres formantes: T1, T2 y T3. Los índices que se obtienen con T1 son de modo de articulación. Los índices de T2 y T3 son de lugar de articulación. En T1 se da la distinción entre clases de consonantes y distinción entre sordas y sonoras.³⁶

Para F1 se observa que será más alta su frecuencia cuanto más abiertas sean las vías bucales. De aquí que aplicado este concepto a las consonantes, se puede decir que la consonante sonora y abierta implica una T1 que comience en frecuencias altas. El estudio del locus de F1 muestra que dentro de él, las variaciones de frecuencia de F1 derecho combinadas a F2 curvo, indica que el punto de partida más bajo para T1 es el mejor para las explosivas y que, a medida que el punto se acopla en frecuencia, esto hace que se aproxime a la percepción de las consonantes más abiertas. En el caso de las consonantes fricativas, el punto de partida de T1 es menos bajo que para las explosivas. En cuanto a las líquidas y semivocales iniciales T1 debe partir de más alto —400 cps. en su medio— para evitar toda percepción de oclusión.

Se observa que la velocidad de transición y la duración de T1 contribuyen a las distinciones de clase. Estos dos factores —velocidad y duración— variando a la vez para T1 y T2 permiten distinguir entre: vocales, semivocales y explosivas sonoras. T2: En una duración muy corta (o una velocidad muy rápida), las T2 dan los mejores índices de distinción de los lugares de articulación. La duración de T2 en las explosivas está entre los 50 msecs., sin embargo tiende a ser más corta en las labiales y más larga en las dentales ante vocales posteriores.

Las dimensiones de T2 que contribuyen a la identificación del lugar de articulación son: Su dirección, llamada positiva cuando es más alta que la F2 de la vocal; y negativa, si ella es más baja que la de la vocal. b).—La diferencia de frecuencia entre el momento de su inicio y el momento en que alcanza a F2 de la vocal. "Cette dimension est généralement donnée, dans les travaux de Haskins, par un multiple de 120 cps.—ainsi une transition de 3 atteint une fréquence de 360 cps. en dessous du formant correspondant de la voyelle"³⁷

³⁵ DELATTRE, PIERRE, *op. cit.*, p. 259.

³⁶ Cfr. *Ibid.*, p. 259.

³⁷ DELATTRE, PIERRE, *op. cit.*, p. 261.

4) Locus: el locus correlativo a cualquier lugar de articulación está determinado no por la extrapolación de curvas de T2, sino por la variación de formantes derechos, evitando así el error que pudieran ocasionar las curvas. Dice John M. Borst: "One of the rules which is most helpful in painting or synthesizing speech is the concept of the locus; this gives a simple way to account for the many different transitions shown in Fig. 8.

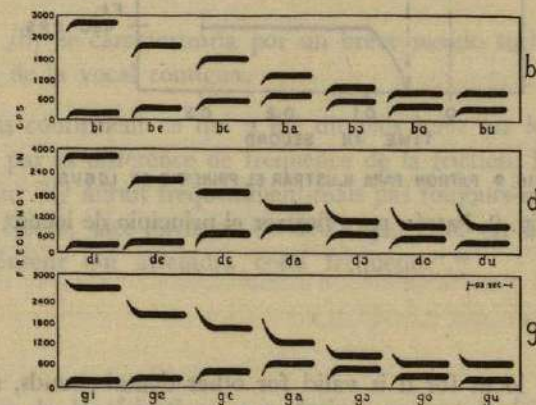


Fig. 8. Patrón espectrográfico simplificado de las consonantes /b/, /d/ y /g/.

Thus for a d preceding any vowel, we can see from the Fig. 9 that the second formant transition appears to have started from a virtual point, or locus, at about 1800 cps.; similarly, there is a locus for b and another for g.

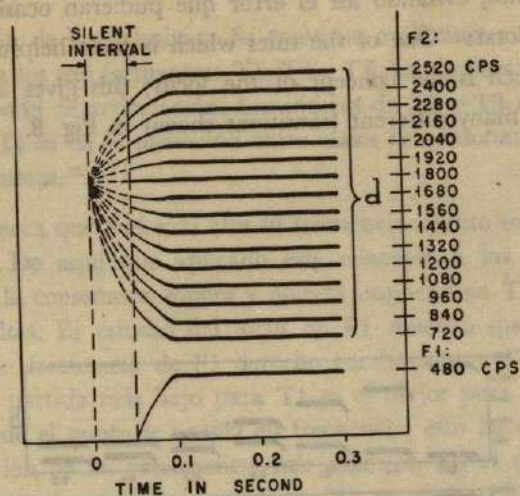


FIG. 9 PATRON PARA ILUSTRAR EL PRINCIPIO DE LOCUS

Fig. 9. Patrón para ilustrar el principio de locus.

Moreover, the locus for *d* is valid for other dental sounds, such as *t* and *n*, is approximately correct for *s* and *z*.³⁸

Fricativas:

Las fricativas se distinguen de las africadas y explosivas por la duración del ruido (sonido turbulento) y por la rapidez por la cual la intensidad del ruido crece.³⁹

Las fricativas se pueden dividir en tres clases que se distinguen entre ellas por la intensidad de la fricción, por la extensión de la frecuencia de la fricción y por sus transiciones.

³⁸ BORST, JOHN M., "The use of Spectrograms for Speech Analysis and Synthesis", *Journal of the Audio Engineering Society*. Vol. 4, 1956, p. 18.

³⁹ DELATTRE, PIERRE, *op. cit.*, p. 264. Cfr.

/s S/ presentan una fuerte intensidad y una extensión mediana; */O F/*, una intensidad menor y una extensión grande y */ç X/* una intensidad media y una extensión estrecha.

Una misma fricción, en la síntesis, ambigua a los 3500 cps. se entiende como */s/* o */ç/* según se una a la vocal por una T3 (positiva-dental) o por una T2 (positiva-palatovelar).

En el interior de estas tres clases, las distinciones son simples: */s/* se distingue de */S/* principalmente por la frecuencia de la fricción (*/s/* descende a 3500 cps., en tanto que */S/* a los 2000 cps.); */O/* se distingue de */f/* principalmente por las transiciones (*/O/* tiene un locus dental, */f/* un locus labial); */ç/* se distingue de */X/* por su frecuencia de fricción y por las transiciones.

La fricativa */h/* se caracterizaría por un breve sonido turbulento a la frecuencia de F2 de la vocal contigua.

"Les résultats confirment ce qui a été dit plus haut sur la distinction entre */s/* et */S/* par la différence de fréquence de la friction. Ils apportent un facteur nouveau: */f/* aurait fréquemment-mais pas toujours-une concentration d'énergie très élevée, aux environs de 8000 cps. Ceci est à vérifier par le premier synthétiseur qui atteindra cette fréquence".⁴⁰

Las africadas:

Se distinguen de las fricativas y explosivas en que la fricción tiene una interrupción completa. "En gros, comparé au bruit des fricatives, le bruit des affriquées est (après interruption) plus court en durée totale et plus court en durée de croissance. Comparé aux explosives, le bruit des affriquées est plus long en durée totale. Pour des valeurs moyennes de durée de croissance, on perçoit des fricatives sourdes quand le bruit total dure au moins 110 ms, des affriquées sourdes quand le bruit total dure au moins 50 ms, et des explosives sourdes quand le bruit total dure au plus 30 ms".⁴¹

Las oclusivas nasales:

Las oclusivas nasales se distinguen de las orales por: a).-La forma de T1 y por b).-Los formantes nasales de la tenue. "Le premier formant nasal, à

⁴⁰ DELATTRE, PIERRE, *op. cit.*, pp. 264-265.

⁴¹ DELATTRE, PIERRE, *op. cit.*, p. 265.

240 cps., était seulement un peu plus faible que celui d'une voyelle normale (environ 6 db de moins) et avait un puissant effet perceptuel de nasalisation. Il apparaît donc que le mode nasal des consonnes dépend de la forme de T1, d'un formant de tenue dand les 250 cps., et des formes de T2 et T3 semblables à celles des explosives de même lieu d'articulation".⁴²

En cuanto a su lugar de articulación se puede decir que su rol de transiciones, definido por el locus correlativo a los lugares de articulación es el mismo que para las explosivas. Es importante el hecho de que el análisis espectrográfico producido por un Análogo eléctrico de la boca, confirma la importancia de F2 para distinguir /m/ de las otras dos nasales /n η/ y nos indica la posibilidad de distinguir /n/ de /η/ por un formante a los 3000 cps.

Consonantes líquidas y semivocales:

Las cuatro consonantes /wjr/ tienen acústicamente algo en común "a).- Pendant la tenue, un F1 de fréquence relativemente haute (près de 400 cps. de moyenne) . . . b).- Pendant la tenue, des formants supérieurs à F1, d'intensité plus grande que ceux de la tenue nasale mais plus faibles que ceux des voyelles. c).- Des transitions en continuité avec les formants de tenue. d).- Un lenteur relative des transitions".⁴³

En cuanto a su lugar de articulación las cuatro consonantes se distinguen las unas de las otras por la frecuencia de sus formantes de tenue superiores a F1 y las transiciones entre la tenue consonántica y la tenue de la vocal contigua.

Las consonantes silábicas:

Apenas se empiezan a hacer estudios en este sentido. Dice Delattre: "Des travaux en cours, par la synthèse, étudient la réduction d'intensité de F2, et la form implosive des transitions qui précèdent, comme indices des modes d'articulation qui distingueront entre les consonnes /l r m n η/ en position syllabique et les voyelles qui ont leurs formants presque aux mêmes fréquences".⁴⁴

⁴² *Ibid.*, p. 266.

⁴³ DELATTRE, PIERRE, *op. cit.*, p. 267.

⁴⁴ DELATTRE, PIERRE, *op. cit.*, pp. 267-268.

Prosodia:

Los elementos prosódicos de la palabra tales como el acento, el ritmo y entonación se han comenzado a estudiar por medio de la síntesis del lenguaje. En dichos estudios resaltan los valores objetivos de la duración, cantidad y frecuencia pudiendo ser éstos de hechura independiente. Entre los resultados obtenidos en investigaciones realizadas sobre el acento inglés se ha observado que más la duración que la intensidad, sin que se excluyan, contribuyen a la percepción del lugar del acento.

De los principios acústicos a los rasgos pertinentes:

Muy importante punto de vista tiene el trabajo de Pierre Delattre: "From Acoustic Cues to Distinctive Features". En él se dan, en un sistema completo, los principios acústicos que son necesarios y suficientes para sintetizar todas las consonantes del francés. La manipulación experimental, en este orden, de las variables controladas por medio del lenguaje sintético ha dado once principios de lugar de articulación y 17 de modo de articulación. Estos principios, los agrupa el autor en tres categorías de lugar y seis de modo de articulación. Dice el autor: "The 11 place cues fall into three categories: A) The direction of the second-formant transitions, as defined by an implicit locus. B) The direction of the third-formant transitions, as defined by a locus. C) The frequency of the turbulence (only for plosives and fricatives), wheather it be a burst noise or a friction noise. In French the band-width and the intensity of the turbulence are not relevant.

The 17 manner cues fall into six categories:

- 1).-The shape and tempo of all transitions.
- 2).-The locus of the first-formant transition.
- 3).-The presence of a high turbulence vs. the presence of a low link (mutually exclusive).
- 4).-The lenght of the turbulence, if any. (This takes care of the conventional opposition: interrumped vs. not -interrupted).
- 5).-Continuity vs. discontinuity of the links (if any—).
- 6).-The time and intensity factors, known as 'voicing'.⁴⁵

⁴⁵ DELATTRE, PIERRE, "From Acoustic Cues to Distinctive Features", *Phonetica*, 18, 1968, p. 216.

Asimismo se subraya que las consonantes francesas no se distinguen las unas de las otras por menos de dos principios o por más de 11. La contribución perceptual de cada uno de los principios (los once de lugar y los 17 de modo de articulación) se demuestran en el trabajo por el contraste de parejas sintéticas de formas espectrográficas transformadas por un sintetizador del tipo de "playback", siendo todos los principios neutralizados en cada pareja contrastiva.

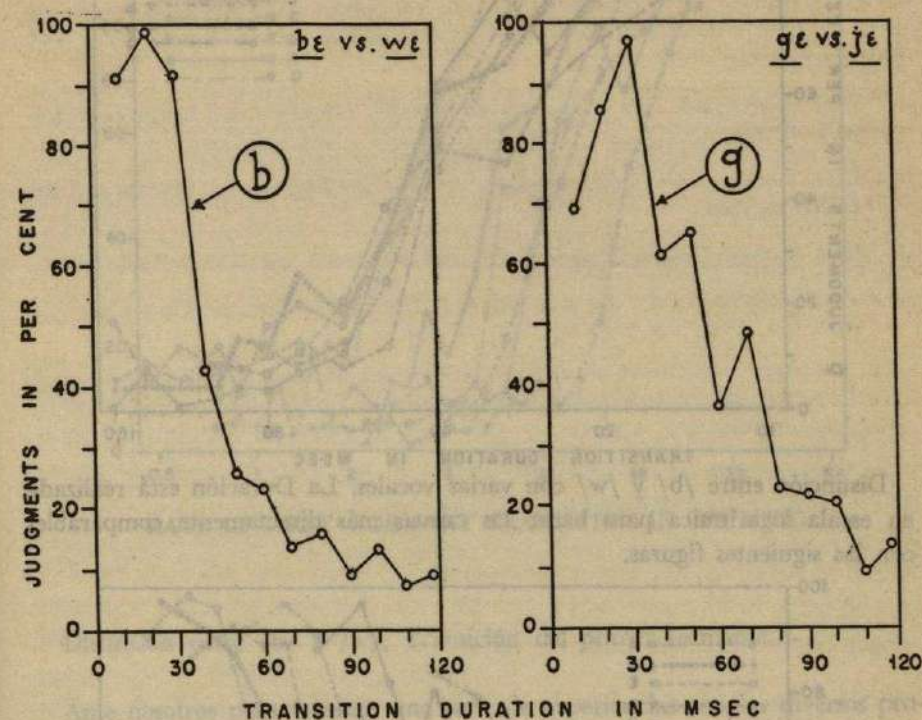
Los cambios de "tempo" de frecuencia como principios acústicos para la distinción de sonidos del lenguaje

En un reporte aparecido en *Journal of Experimental Psychology*, Liberman, Delattre, Gerstman y Cooper, realizan una serie de experiencias con lenguaje sintético asilando una serie de principios acústicos para distinguir, mediante ellos, los sonidos del lenguaje.

El primer experimento consiste en determinar si el cambio de "tempo" en las transiciones del primero y segundo formantes es suficiente para convertir las sílabas /b E/ y /g E/ (consonante oclusiva más vocal) en /w E/ y /j E/ (semivocal más vocal) y, en el extremo /u E/ e /i E/. Los resultados son los siguientes: las transiciones de primero y segundo formantes en cambio de "tempo" son principios para distinguir entre consonante oclusiva y semivocal. El cambio ocurrió cuando la duración de la transición alcanzó 40 msecs., para b>w. El cambio de g>j en las vecindades de 50 a 60 msecs. El segundo experimento consiste en encontrar cómo el escucha responde a las variaciones en el "tempo" de las transiciones cuando hay otras vocales que no sean /E/ en la segunda parte de la sílaba.

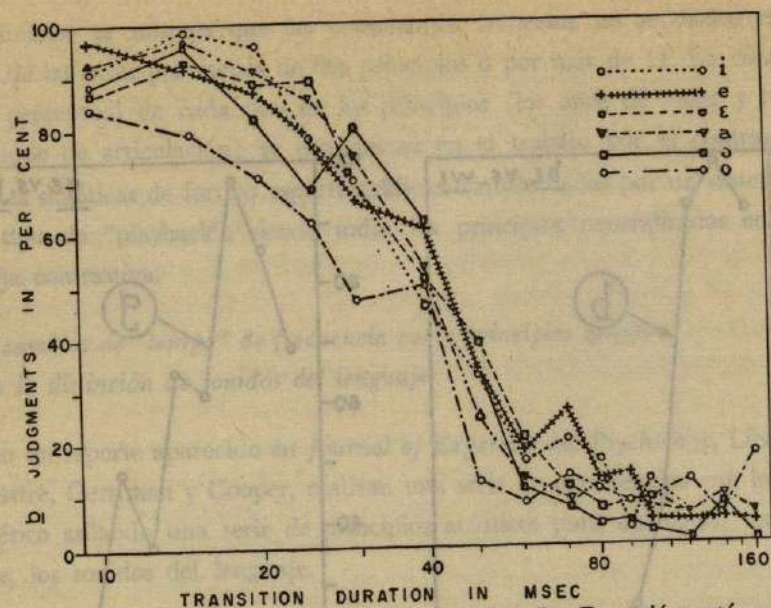
El experimento se realizó sólo con la distinción entre la oclusiva /b/ y la semivocal /w/. Los resultados fueron el "tempo" de la transición es suficiente para distinguir entre la oclusiva /b/ y la semivocal /w/ en la gama de variedad de vocales.⁴⁶

La figura primera representa el resultado del primer experimento. La segunda, los resultados del segundo experimento.

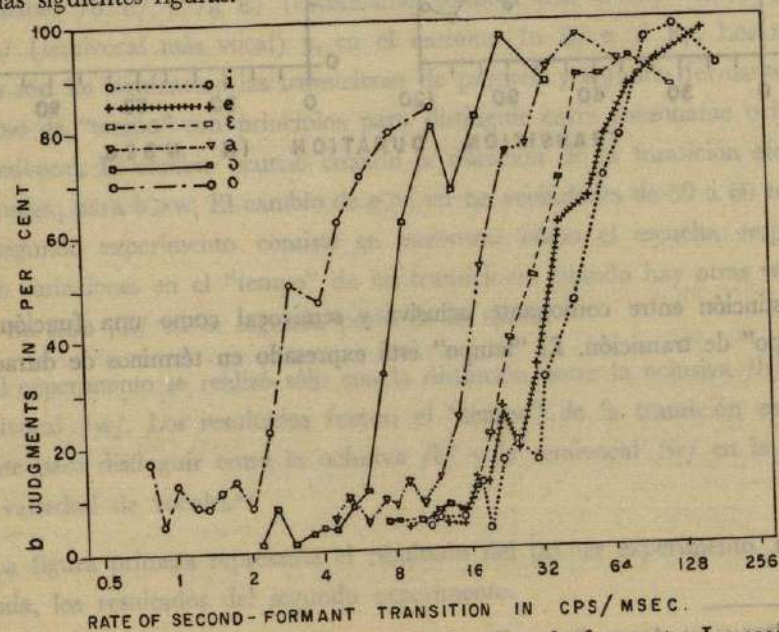


Distinción entre consonante oclusiva y semivocal como una función del "tempo" de transición. El "tempo" está expresado en términos de duración.

⁴⁶ LIBERMANN, ALVIN, PIERRE DELATTRE, et al. "Tempo of Frequency as a Cue for Distinguishing Classes of Speech Sounds", *Journal of Experimental Psychology*. Vol. 52, No. 2, August, 1956. Cfr. (*passim*).

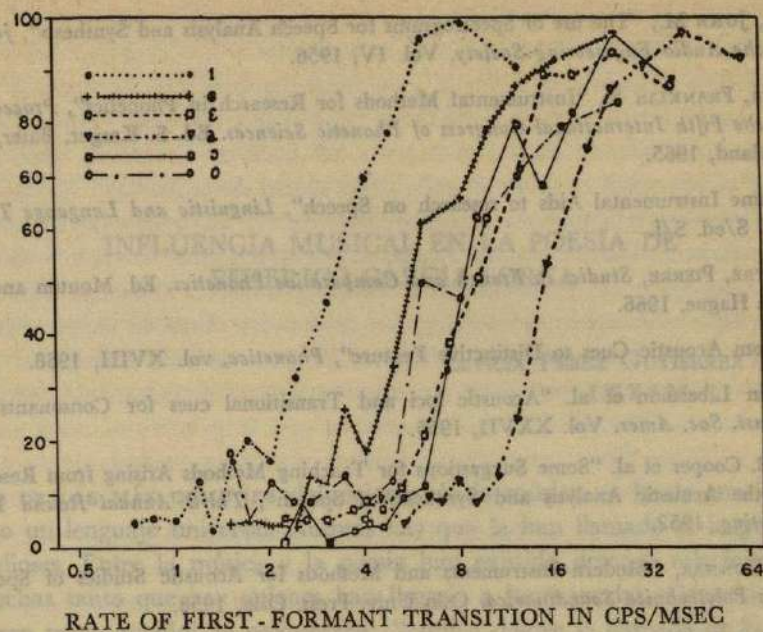


Distinción entre /b/ y /w/ con varias vocales. La Duración está realizada en escala logarítmica para hacer las curvas más directamente comparables con las siguientes figuras.



Distinción entre /b/ y /w/. Transición de segundo formante. La escala logarítmica sirve para igualar la distancia de la abscisa ocupada por la extensión de la proporción de cada una de las vocales.

b. JUDGMENTS IN PER CENT



Distinción entre /b/ y /w/. Transición del primer formante.

Ante nosotros pues tenemos una serie de experiencias con los diversos problemas y sus resultados.

Creemos que los sintetizadores de lenguaje como la síntesis misma han de seguir dando frutos y grandes enseñanzas para el desarrollo de la Fonética Experimental. Los resultados y las soluciones propuestas hasta ahora, ya se han incorporado en las más recientes investigaciones en el campo de la lingüística contemporánea.

Madrid, Enero-Junio de 1971.

- BORST, JOHN M., "The use of Spectrograms for Speech Analysis and Synthesis", *Journal of the Audio Engineering Society*. Vol. IV, 1956.
- COOPER, FRANKLIN S., "Instrumental Methods for Research in Phonetics", *Proceedings of the Fifth International Congress of Phonetic Sciences*. Ed. S. Karger. Baser, Switzerland, 1965.
- "Some Instrumental Aids to research on Speech", *Linguistic and Language Teaching*. S/ed. S/f.
- DELATTRE, PIERRE, *Studies in French and Comparative Phonetics*. Ed. Mouton and Co. The Hague, 1966.
- "From Acoustic Cues to Distinctive Feature", *Phonetica*, vol. XVIII, 1968.
- Alvin Liberman et al. "Acoustic loci and Transitional cues for Consonants", *J. acoust. Soc. Amer.* Vol. XXVII, 1955.
- F. S. Cooper et al. "Some Suggestions for Teaching Methods Arising from Research on the Acoustic Analysis and Synthesis of Speech", *Third Annual Round Table Meeting*, 1952.
- FANT, GUNNAR, "Modern Instruments and Methods for Acoustic Studies of Speech" *Acta Polytechnica Scandinavica*. Oslo Univ. Press, Oslo, 1958.
- FLANAGAN, JAMES, "Difference Limen for Formant Amplitude", *The Journal of Speech and Hearing Disorders*. Vol. XXII, 1957.
- HOWARD, C. R. et al., "Analysis and Synthesis of Formants and Moments of Speech Spectra", *J. acoust. Soc. Amer.* Vol. XXVIII, 1956.
- LAHISTE, ILSE, *Readings in Acoustic Phonetics*. The M.I.T. press, M.I.T. Cambridge, Mass, 1967.
- LIBERMAN, ALVIN, Pierre Delattre et al. "Tempo of Frequency as a Cue for Distinguishing Classes of Speech Sounds", *Journal of Experimental Psychology*, vol. LII, No. 2, August, 1956.
- PETERSON, GORDON E. y Harold L. Barney, "Control Methods used in the Study of the vowel", *J. Acoust. Soc. Amer.* Vol. XXIV, No. 2 March, 1952.
- PETTER, R. K., G. A. KOPP et al., *Visible Speech*. Ed. Van Nostrand, New York, 1947.
- QUILIS M., ANTONIO, "El Método Espectrográfico. (Notas de Fonética Experimental)". *R.F.E.*, vol. XLIII, 1960.
- *Curso de Fonética y Fonología Españolas*. 4a. edic. Edit. C/S/I/C. (Inst. Miguel de Cervantes), Madrid, 1969.
- SCHATZ, CAROL D., "The Role of Context in the Perception of Stops", *Language*, vol. 30, 1954.
- ULDALL, ELIZABETH, "The synthesis of Some Sounds made on other than Pulmonic Air-Stream Mechanisms", *Phonetica*. Vol. XIII, 1965.

INFLUENCIA MUSICAL EN LA POESÍA DE FEDERICO GARCÍA LORCA

LETICIA PÉREZ GUTIÉRREZ M.L.E.
I.T.E.S.M.

UNA DE LAS MÁS COMUNES definiciones de la música es la de considerarla como un lenguaje universal. Autores hay que la han llamado el lenguaje de los dioses. Entre la música y la poesía han existido siempre relaciones muy estrechas tanto que hay quienes han llegado a hermanarlas. Gerardo Diego, el gran poeta español, se preguntaba: "¿quién desoye la llamada de la poesía y de su hermana la música?" Muchos han sentido este "vocare" y la han estudiado como un lenguaje. Sin embargo pocos han podido conjugar las dos vocaciones, la de músico y la de poeta. Una de estas personalidades singulares lo fue Federico García Lorca.

En una nota autobiográfica cuenta el poeta su incipiente amor por la música. A los siete años fue al colegio de los Escolapios donde comenzó su estudio. Después pasó a Granada en donde tomó lecciones con don Antonio Segura, un viejo compositor discípulo de Verdi. Hasta 1917 toda su vida gira alrededor de la música. Cree que esa es su auténtica vocación. Da varios conciertos. Funda la Sociedad de Música de Cámara. Como su familia se opone a que continúe sus estudios musicales en París, es entonces cuando se refugia en el afán creacionista de la poesía.

Si hay un poeta en el que la música haya influido de una manera decisiva este es García Lorca. Jorge Guillén decía que su amigo "Habría podido ser un compositor si se lo hubiese propuesto. Se contentó con ser de verdad, un apasionado muy competente. En música fue tal vez donde el gusto de Federico se refinó con más pureza. De su piano surgían la interpretación fiel o estupendas imitaciones que implicaban conocimiento y crítica. A petición de alguno, que proponía un nombre, tocaba trozos no recordados, sino inventados, con el inconfundible estilo del modelo. ¡Qué inteligencia y qué gracia una vez más! El Lorca músico se sitúa así, bromeando y estudiando, entre