

Llisterri, J. (1990). Algunes reflexions sobre el paper de la lingüística en la tecnologia de la veu. *Límits. Revista d'assaig i d'informació sobre les Ciències del Llenguatge*, 8, 19-32.

[http://liceu.uab.cat/~joaquim/publicacions/
Llisterri_90_Fonetica_TecnolHabla.pdf](http://liceu.uab.cat/~joaquim/publicacions/Llisterri_90_Fonetica_TecnolHabla.pdf)

ALGUNES REFLEXIONS SOBRE EL PAPER DE LA LINGÜÍSTICA EN LA TECNOLOGIA DE LA VEU

JOAQUIM LLISTERRI
Laboratori de Fonètica,
Universitat Autònoma de Barcelona
*Povera donna, sola, abbandonata
In questo popoloso deserto.*

Durant la lenta incorporació de les ciències del llenguatge al món de la tecnologia de la veu, més d'un lingüista treballant en aquest camp s'hauria pogut sentir com Violeta Valery en el París romàntic. I de fet, cal pensar si ara mateix no es dona encara una situació similar d'aïllament entre els professionals de les llengües i els experts en aplicacions informàtiques. Els sistemes de comunicació oral entre l'home i l'ordinador es troben, sens dubte, en un moment de gran desenvolupament, degut, entre d'altres factors, a l'interès que tots els mètodes d'interacció amb les computadores desperten en la indústria. Paral·lelament, tant entre els científics i tecnòlegs com entre el gran públic, s'han creat una sèrie d'expectatives sobre les immenses possibilitats que ofereix el fet de disposar de sistemes computacionals que sintetitzin i reconeguin la veu.

Davant d'aquest *boom*, que ha portat a autors com Bristow (1984) a parlar d'una revolució de la veu similar a l'explosió de la informàtica, cal preguntar-se si les noves formes de comunicació parlada amb l'ordinador deriven dels assoliments espectaculars de la tecnologia en els darrers anys o del coneixement aprofundit dels processos de producció i percepció de la parla en els humans a què han arribat els lingüistes.

Aquesta és justament la pregunta que es fan Lindblom,

Perkell i Klatt (1986) en el pròleg a un recull de treballs al voltant de la variació i la invariància en la parla; en llur opinió, l'aparició de les aplicacions industrials de la tecnologia de la veu no respon als progressos de les ciències del llenguatge, punt amb el qual coincideixen amb Fant (1983) quan, en obrir el desè Congrés Internacional de Ciències Fonètiques, feia veure la necessitat d'una cinquena generació d'investigadors de la parla més que no pas de la tan anunciada cinquena generació d'ordinadors.¹ De fet, cada vegada sembla més clar que si moltes de les aplicacions de la síntesi i el reconeixement de la veu a la vida diària no arriben a un nivell de qualitat suficient, pot arribar a produir-se un estancament en aquest camp.

A més de Fant, molts especialistes han coincidit a assenyalar que, en tecnologia de la veu, les eines fornides pels enginyers tenen un nivell de desenvolupament molt superior al dels models oferts des de la lingüística. Els mateixos autors esmentats abans parlen de la manca d'una teoria global que expliqui la producció i la percepció de la parla, reconeixent que aquest pot ésser un obstacle fonamental en l'aparició d'aplicacions comercials ben acceptades pels usuaris.²

No penso que sigui exagerat de dir que el naixement de la tecnologia de la veu ha posat un cop més de manifest la necessitat d'interacció entre les disciplines anomenades «humanístiques» i les que solen agrupar-se sota el rètol de «tècniques», alhora que ha demostrat els lligams indissolubles entre teoria i aplicació en la recerca científica. De fet, tant la síntesi com el reconeixement plantegen als especialistes en fonètica diversos problemes de tipus pràctic als quals cal donar una solució immediata. Però alhora —i aquest aspecte em sembla molt més interessant—, suggereixen reflexions que afecten els fonaments teòrics de la disciplina. La situació actual ha donat una marcada orientació tecnològica a les ciències fonètiques,

1. *We need a fifth generation of speech scientists rather than a fifth generation of computers* (Fant, 1983, p. 16).

2. *The lack of a comprehensive theory of speech production and perception will still turn out to be a major stumbling block in current engineering approaches to man-machine interfaces* (Lindblom, Parkell i Klatt, 1986, p. i).

mentre que, per altra banda, potser encara no ha arrelat prou entre els experts que dissenyen aplicacions la necessitat de col·laboració continuada amb els qui, per definició, coneixen el funcionament de la parla en tant que sistema de comunicació entre humans.

1. LA PARLA COM A CODI

Des del punt de vista del lingüista, la parla constitueix un procediment codificat de manera eficient per a l'intercanvi ràpid d'informació. Aquesta és una idea que es troba ja en la teoria motriu de la percepció de la parla propugnada des dels laboratoris Haskins a finals dels anys seixanta (Liberman, Cooper, Shankweiler i Studdert-Kennedy, 1967), de la qual ens interessa destacar aquí el paper central que adquireix la noció de solapament.

En el moment de la producció dels enunciat, les unitats lingüístiques —posem per cas els fonemes— no es manifesten de manera seqüencial en la cadena sonora. Seguint el clàssic exemple emprat per Liberman (1970), en una paraula com *bag*, els indicis que ens permeten identificar /b/ se superposen —és a dir, apareixen en el mateix punt en el temps— amb els emprats en la percepció de /æ/, mentre que aquests es donen simultàniament amb els corresponents a /g/: els segments discrets i successius de la representació fonològica s'entrellacen de manera que en la part central del mot trobem una zona en la qual se solapa informació corresponent als tres segments que el componen.

L'explicació d'aquest fenomen rau, en opinió dels proponents de la teoria motriu, en la realització simultània dels moviments dels articuladors que corresponen a segments diferents; per exemple, el tancament labial requerit per /b/ pot portar-se a terme alhora que la llengua adopta la posició pròpia de /æ/. Les conseqüències són ben conegudes pels fonetistes, i solen agrupar-se sota la denominació de coarticulació.

Un mode de funcionament com el que acabem de descriure és extraordinàriament útil per a la transmissió de grans quan-

titats d'informació en poc temps, operació difícil d'acomplir donada la lentitud de moviments de l'aparell fonatori. Observen Liberman i els seus col·laboradors que una velocitat d'elocució de 10 a 15 fonemes per segon no seria possible sense la codificació simultània d'informació sobre més d'un segment en paral·lel. Aquesta estratègia no es troba en cap altre sistema de comunicació del món animal, i no sembla desencertat de pensar que ha pogut constituir un avantatge evolutiu important.

Tornant al nostre tema, caldria fer veure ara de quina manera la síntesi i el reconeixement poden fer ús de les nocions que acabem d'exposar. Per una banda, la divisió en un nivell fonològic d'unitats abstractes i un nivell fonètic, en el qual es materialitzen aquestes unitats mitjançant el procés de codificació en paral·lel esmentat, ha de permetre orientar el disseny de sistemes de comunicació home-màquina que vagin més enllà de les simples tècniques de gravació i reproducció de la veu.

Seguint les orientacions d'Allen (1985), tant un sistema de síntesi com un de reconeixement es poden concebre com un conjunt de transformacions entre diversos nivells de representació.³ El model de codificació esquematitzat més amunt ha de servir de base per a aquestes operacions, i, de fet, aquesta és la filosofia de les regles que en un conversor de text a veu fan el pas de la transcripció fonètica al conjunt de valors de paràmetres acústics adequats per a controlar el sintetitzador que genera l'ona sonora final. Els diversos tipus de transicions de consonant a vocal emprats en un sistema com el MITalk (Allen, Hunnicutt i Klatt, 1987) constitueixen una mostra dels intents per arribar a modelitzar de manera completament automàtica el resultat dels mecanismes de coarticulació en la parla natural.

El reconeixement pot plantejar-se també des d'aquesta pers-

3. *The task converting a symbolic input (e.g. text) to an output speech waveform or the need to convert a speech waveform into a symbolic representation (such as a message designator or a speaker's identity) can be seen as a transformation between two distinct representations* (Allen, 1985, p. 1542).

pectiva. El fet que sigui relativament més senzill de dissenyar un sistema de reconeixement de text escrit —en el qual la informació és codificada seqüencialment— que no pas un mètode que impliqui la llengua oral, ens mostra ja la importància del procés de solapament entre unitats. El reconeixedor hauria de funcionar doncs de manera inversa al conversor de text a veu, operant, en algun dels seus nivells, el pas de cadenes d'elements superposats a sèries d'elements discrets.

Tot i així, cal no oblidar un advertiment de Fant sobre la tasca que encara queda per realitzar abans de poder aplicar a sistemes reals de síntesi o de reconeixement els models de producció i percepció de la parla desenvolupats des de la lingüística:

even if we possessed perfect general models of production and perception and a maximally effective linguistic framework, we would still have to derive an immense amount of rules and reference data relating message units and speaker categories to observed phonetic sound shapes (Fant, 1983, p. 30).

Aquest comentari ens introdueix alhora el tema central del proper apartat, que dedicarem a la variabilitat en la realització fonètica de les unitats lingüístiques.

2. LA VARIACIÓ EN LA PARLA

Mentre que en lingüística sol assumir-se el caràcter discret i invariant d'unitats com ara el fonema, és habitual per al fonetista de contemplar com aquest pren diferents característiques articulatòries i acústiques en funció del context en el qual es produeix. Aquesta variabilitat obeeix, naturalment, al mateix procés de codificació que esmentàvem al principi: tornant a l'exemple de *bag*, és clar que les propietats de [æ] vindran condicionades per les de [b] i [g], i viceversa. Justament per això parla Fant en la cita anterior de la necessitat de relacionar sistemàticament les diverses realitzacions contextuals de [æ] amb la unitat subjacent /æ/.

En una primera aproximació, podem adonar-nos que el grau de variabilitat motivada per l'entorn no és el mateix en tots els segments: els al·lòfons vénen més condicionats contextualment que les síl·labes, les quals a llur torn canvien més que els mots. En el moment de decidir les unitats per a un sistema de síntesi o de reconeixement, val la pena de considerar com a criteri no tan sols les possibles modificacions que pot experimentar cadascuna, sinó també el punt on solen donar-se: en el mot, per exemple, la variabilitat es troba més en els elements inicials i finals que en el centre,⁴ per aquest motiu un sistema de concatenació de mots haurà de preveure regles d'ajustament a les fronteres entre mots, mentre que un procediment d'enllaç d'al·lòfons considerarà ben probablement molts més canvis.

Aquesta és la raó per la qual s'han desenvolupat unitats com el difonema —segment format per la part estacionària del primer element, la transició vers el segon i la part estacionària d'aquest darrer— o la semisíl·laba —fragment entre l'inici de la síl·laba i el centre de la vocal que actua de nucli—, sense cap mena de tradició en la lingüística, però que en canvi responen a la necessitat de resoldre els problemes derivats de la variació contextual.

Tot i així, la qüestió no es redueix a trobar la unitat més adient. Des del punt de vista de la generació automàtica de les manifestacions fonètiques de les unitats triades en un sistema de síntesi determinat, ens cal disposar de dues informacions complementàries: per una banda, l'inventari dels factors que incideixen en la variabilitat de cadascuna de les unitats i, per una altra, les modificacions en els valors dels paràmetres acústics produïdes per aquests factors.

Si bé els fonetistes coneixen quins són els elements que poden modificar les propietats articulatòries i acústiques dels segments, hi ha poques propostes globals per a integrar-les en un model coherent.⁵ Solen distingir-se els factors individuals

4. Per a més detalls sobre els criteris de selecció de les unitats en síntesi, vegeu Allen (1984).

5. Potser la proposta de Nolan (1983) és una de les més ben estruc-

—entre els quals destaquen els derivats de l'anatomia de l'aparell fonatori de cada parlant—, els que es relacionen amb la pròpia producció de l'enunciat com la qualitat i la dinàmica de la veu, el mode de fonació, o la velocitat d'elocució, i els factors contextuals que responen a la posició de cada al·lòfon en la síl·laba, el mot i l'enunciat, o a l'accentuació. Cal afegir encara les variacions de tipus geogràfic, les que vénen donades per l'origen social i les que deriven de la utilització de registres diferents.

Malgrat que el repertori sembla relativament ben establert, els estudis que relacionen els factors esmentats amb les característiques fonètiques dels sons són encara insuficients. Podem consultar treballs que revisen globalment, per exemple, quina és la incidència dels factors anatòmics (Stevens, 1972) o de la velocitat d'elocució (Miller, 1981), i les aportacions de sociolingüistes com Labov han estat importants per a conèixer els condicionants socials de la variació al·lofònica;⁶ en canvi, estem lluny encara de disposar tant de les dades de les regles que, seguint les indicacions de Fant, relacionin les propietats fonètiques observables en els segments amb les unitats lingüístiques abstractes de les quals es deriven.

Tant de cara a la síntesi com al reconeixement, la possibilitat d'incorporar aquesta variació obre un nou ventall de possibilitats. En síntesi, s'ha de tendir a sistemes que ofereixin la màxima naturalitat —fins ara els esforços s'han centrat més aviat en la qüestió de la intel·ligibilitat—, reproduint de manera fidel les diverses manifestacions fonètiques de cadascuna de les unitats. També, de cara als diversos entorns en què es pot emprar la síntesi, cal pensar en la incorporació dels factors de variació lligats a l'ús social de la llengua esmentats abans, i fins i tot en la possibilitat de modelitzar tipus i quali-

turades. Consisteix, a grans trets, en un model d'actuació que permet de caracteritzar els nivells de representació en els quals els parlants tenen la possibilitat de seleccionar aquells trets que els poden distingir dels altres parlants.

6. Vegeu, per exemple, Labov (1986).

tats diferents de veu si els sistemes de síntesi han de servir com a pròtesi per a persones privades de la parla.⁷

Certes aplicacions del reconeixement han de poder ser emprades per un gran nombre d'usuaris —és a dir, cal concebre-les com a independents del locutor—, i això requereix també una base de dades molt àmplia a fi de recollir les variants que parlants diferents puguin presentar al sistema.

La consideració de les diverses manifestacions d'una mateixa unitat en l'ona sonora porta a analitzar dos temes centrals tant en la teoria fonètica com en les aplicacions a la tecnologia de la veu: es tracta dels problemes de la invariància i de la segmentació, que es deriven també del procediment específic de codificació de la informació lingüística en el senyal sonor propi de la parla.

3. LA INVARIÀNCIA I LA SEGMENTACIÓ

Entre les moltes formulacions que s'han fet d'aquestes qüestions, sembla particularment clara la que presenta Lindblom:

For a given language there seems to be no unique set of acoustic properties that will always be present in the production of a given unit (feature, phoneme, syllable), and that will reliably be found in all conceivable contexts (Lindblom, 1986, p. 495).

Plantejat així, és fàcil de veure que la manca d'un conjunt de propietats acústiques invariants presents en l'ona sonora cada vegada que apareix una determinada unitat lingüística —i cal fer notar que aquí no es redueix la qüestió als al·lòfons, sinó que es fa també referència a la síl·laba i al tret distintiu— constitueix un entrebanc greu tant per als sistemes de síntesi com per als de reconeixement.

Els sistemes més sofisticats de conversió de text a veu solen

7. Més comentaris sobre la possibilitat de reproduir diferències entre parlants diferents i variacions en la veu del mateix parlant es troben a Nolan (1984).

tenir taules de valors de paràmetres acústics per a la síntesi de cadascun dels al·lòfons de la llengua en què es treballa; aquests valors —denominats molts cops *target values*— es refereixen a la part estacionària o central, sense considerar la incidència dels elements adjacents. En termes pràctics, aquest mode de funcionament implica assumir un cert grau d'invariància de les unitats. Les regles del component fonètic —entre les quals es compten les que fan referència a les transicions esmentades més amunt— s'encarreguen de modificar aquests valors en funció dels segments anteriors i posteriors, amb la qual cosa s'introdueixen els factors de variació contextual. Les limitacions a aquest conjunt de regles vénen donades, tal com es pot suposar, pel coneixement actual dels mecanismes que incideixen en les característiques fonètiques de la realització de les unitats lingüístiques.

La noció d'invariància és també central en el reconeixement, ja que de la definició d'un conjunt de propietats acústiques estables per a una determinada unitat depèn el bon funcionament dels sistemes que porten a terme una decodificació de les ones sonores en unitats abstractes; els sistemes basats en entitats fonèmiques o subfonèmiques assumeixen la invariància en aquest nivell, mentre que els fonamentats en la comparació de plantilles de mots consideren l'existència de propietats constants dins d'aquesta unitat.

Tal com assenyala Lea (1986), bona part de la recerca en reconeixement s'ha portat a terme basada en la hipòtesi de la invariància, combinada amb un principi similar al dels nivells de representació en lingüística: els trets fonètics que apareixen simultàniament —manifestats en forma d'un conjunt de valors de paràmetres acústics que es donen en un mateix punt de l'ona sonora— s'associen a un al·lòfon, interpretat com la realització d'un fonema; aquest es combina en seqüències per a formar paraules, les quals, al seu torn, s'integren en enunciats.⁸

8. Benveniste (1964) parla de relacions de constituència i d'integració per descriure les relacions entre unitats que pertanyen a diferents nivells de representació o, per emprar els seus propis termes, nivells d'anàlisi.

Intrínsecament lligat a la invariància trobem el problema de la segmentació, especialment rellevant en el terreny del reconeixement que ara ens ocupa. Llegim de nou com ho formula Lindblom:

The signal cannot be unambiguously segmented into temporally, non overlapping chunks corresponding to e.g. phones, syllables and words (Lindblom, 1986, p. 495).

El principi de la codificació en paral·lel amb el qual començàvem les nostres reflexions ens dóna la clau del perquè d'aquesta impossibilitat de segmentar els senyals sonors en unitats que no se superposin. Les dificultats en la segmentació, conjuntament amb la manca d'invariància acústica, sembla que hagin de fer impossible de concebre sistemes automàtics de reconeixement. Tanmateix, no solament es tracta d'una operació que qualsevol persona és capaç de realitzar en la seva pròpia llengua, sinó que també existeixen en el mercat diversos programes que porten a terme el reconeixement de seqüències orals amb un èxit notable.

Amb una certa ingenuïtat es podria pensar que els problemes de la invariància i de la segmentació desapareixeran quan haurem acumulat un conjunt suficient de coneixements sobre les propietats fonètiques dels segments: aquesta era la recomanació de Klatt en la seva avaluació del projecte ARPA, magna empresa americana que havia de donar com a resultat un sistema de reconeixement de parla:

The main objective of future research in this area should therefore be the accumulation of more detailed linguistic and acoustic-phonetic facts about English sentences (Klatt, 1977, p. 1358).

No hi ha cap mena de dubte que aquesta és una activitat útil; autors com Stevens i Blumstein (1981) han mostrat diverses manifestacions de la invariància acústica i han suggerit una certa tendència a subestimar-la, de manera que molts dels problemes sembla que quedarien resolts amb una major precisió

en l'anàlisi acústica i amb una selecció adequada dels paràmetres rellevants.

Tot i així, cal també fer entrar en joc plantejaments com el de Lindblom (1987), que consideren la interacció entre emissor i receptor en els processos comunicatius i distingeixen la informació present en les ones sonores d'aquella que aprenem a derivar del context: moltes vegades interpretem correctament missatges en absència d'indicis reals en el senyal acústic —quan, per exemple, es troben emmascarats per soroll com el de la línia telefònica— utilitzant l'elevat grau de redundància present en els intercanvis verbals, o fent ús simplement del nostre coneixement de la situació. Des d'aquesta perspectiva, el comportament fonètic dels parlants es caracteritza per la seva plasticitat —la capacitat de variar en funció del marc comunicatiu— i pel seu caràcter teleològic, és a dir, l'orientació de tots els esforços vers el requisit principal, que és fer-se entendre (Lindblom, 1983).

Es ben probable que la resolució dels problemes de la invariància i de la segmentació hagi d'apropar-se si més no a aquests camins, que no estan lluny de la consideració de la parla com a procediment codificat per a l'intercanvi d'informació. L'adaptabilitat dels sistemes de síntesi i la introducció en els reconeixadors de coneixements lingüístics sembla conseqüència lògica d'una via com la que aquí se suggereix.

4. CONSIDERACIONS FINALS

Al llarg d'aquestes reflexions, potser una mica massa vagues, he provat de fer veure de quina manera, partint de la concepció de la parla com a vehicle codificat per a l'intercanvi d'informació, és possible de destriar uns camps de contacte entre la lingüística i la tecnologia de la veu. Nocions com la de solapament, variació i invariància, centrals en la teoria fonètica, haurien d'orientar el disseny dels programes de recerca sobre els quals es fonamenten les aplicacions reals de la sin-

tesi i el reconeixement. L'estudi de les propietats acústiques⁹ de les manifestacions de les diverses unitats lingüístiques i —sobretot— de les lleis que governen la seva variabilitat seria un dels objectius comuns, però tampoc no s'han de negligir els elements que no es troben directament en els senyals sonors però que contribueixen a la comunicació.

En realitat, ni l'estudi de la parla és ja un terreny exclusiu dels fonetistes ni les aplicacions han de quedar només en mans d'aquells que tradicionalment s'han dedicat a allò que s'anomenaven ciències. Un equip que no sigui autènticament interdisciplinari té poques probabilitats de sobreviure en aquest terreny; el marc administratiu en el qual se situï té ben poca importància, però en canvi hem d'esperar que l'aventura entre la lingüística i la tecnologia esdevingui, com per Violeta i Alfredo, *un serio amore*.

REFERÈNCIES BIBLIOGRÀFIQUES

- ALLEN, J. (1984): «Symposium 2: Units in Speech Synthesis», dins M. P. R. VAN DEN BROECKE i A. COHEN (eds.) (1984), *Proceedings of the Tenth International Congress of Phonetic Sciences*. Dordrecht, Foris Publications, pp. 151-155.
- ALLEN, J. (1985): «A Perspective on Man-Machine Communication by Speech», *Proceedings of the IEEE* 73.11, pp. 1541-1550.
- ALLEN, J., M. S. HUNNICUTT i D. H. KLATT (amb R. C. ARMSTRONG i D. PISONI) (1987): *From Text to Speech: The MITalk System*. Cambridge, Cambridge University Press.
- BENVENISTE, E. (1964): «Les niveaux de l'analyse linguistique», dins E. BENVENISTE (1965), *Problemes de linguistique générale I*. París, Gallimard, pp. 119-131; trad. cast. de J. Almela: «Los niveles del análisis lingüístico», dins *Problemas de lingüística general I*. México, Siglo XXI, 1974, pp. 118-130.
- BRISTOW, G. (1984): «The Speech Revolution», dins G. BRISTOW (ed.), *Electronic Speech Synthesis Techniques, Technology and Applications*. Londres, Granada, pp. 7-13.

9. He deixat pràcticament de banda tot el que fa referència als mecanismes auditius i perceptius; una mostra de la seva pertinència pot trobar-se a Klatt (1986).

- FANT, C. G. (1983): «Phonetics and Speech Technology», *Speech Transmission Laboratory - Quarterly Progress and Status Report* 2-3: 20-35, dins M. P. R. VAN DEN BROECK i A. COHEN (eds.) (1984), *Proceeding of the Tenth International Congress of Phonetic Sciences*. Dordrecht, Foris Publications, pp. 13-24.
- KLATT, D. H. (1977): «Review of the ARPA Speech Understanding Project», *Journal of the Acoustical Society of America* 62.2, pp. 1345-1366.
- (1986): «Problems of Variability in Speech Recognition and in Models of Speech Perception», dins J. S. PERKELL i D. H. KLATT (eds.), *Invariance and Variability in Speech Processes*. Hillsdale, Lawrence Erlbaum, pp. 300-319.
- LABOV, W. (1986): «Sources of Inherent Variation in the Speech Processes», dins J. S. PERKELL i D. H. KLATT (eds.), *Invariance and Variability in Speech Processes*. Hillsdale, Lawrence Erlbaum, pp. 402-422.
- LEA, W. A. (1986): «The Elements in Speech Recognition», dins G. BRISTOW (ed.), *Electronic Speech Recognition. Techniques, Technology and Applications*. Londres, Collins, pp. 49-129.
- LIBERMAN, A. M. (1970): «The Grammars of Speech and Language», *Cognitive Psychology* 1.4, pp. 301-323.
- LIBERMAN, A. M., F. S. COOPER, D. P. SHANKWEILER i M. STUDDERT-KENNEDY (1967): «Perception of the Speech Code», *Psychological Review* 74.6, pp. 431-461.
- LINDBLOM, B. (1983): «On the Teleological Nature of Speech Processes», *Speech Communication* 2, pp. 155-158.
- (1986): «On the Origin and Purpose of Discreteness and Invariance in Sound Patterns» dins J. S. PERKELL i D. H. KLATT (eds.), *Invariance and Variability in Speech Processes*. Hillsdale, Lawrence Erlbaum, pp. 496-510.
- (1987): «Adaptive Variability and Absolute Constancy in Speech Signals: Two Themes in the Quest for Phonetic Invariance», dins *Proceeding of the Eleventh International Congress of Phonetic Sciences*. Tallin, Academy of Sciences of the Estonian SSR, vol. 3, pp. 9-18.
- LINDBLOM, B., J. S. PERKELL i D. H. KLATT (1986): «Preface», dins J. S. PERKELL i D. H. KLATT (eds.), *Invariance and Variability in Speech Processes*. Hillsdale, Lawrence Erlbaum, pp. I-IV.
- MILLER, J. L. (1981): «Effects of Speaking Rate on Segmental Distinctions», dins P. D. EIMAS i J. L. MILLER (eds.), *Perspectives in the Study of Speech*. Hillsdale, Lawrence Erlbaum, pp. 39-74.

- NOLAN, F. (1983): *The Phonetic Basis of Speaker Recognition*. Cambridge, Cambridge University Press.
- (1984): «Applying Linguistics to Synthesis», dins G. BRISTOW (ed.), *Electronic Speech Synthesis Techniques, Technology and Applications*. Londres, Granada, pp. 320-335.
- STEVENS, K. N. (1972): «Sources of Inter and Intra-Speaker Variability in the Acoustic Properties of Speech Sounds», dins R. CHARBONNEAU i A. RIGAUULT (eds.), *Proceedings of the Seventh International Congress of Phonetic Sciences*. La Haia, Mouton, pp. 206-232.
- STEVENS, K. N. i S. BLUMSTEIN (1981): «The Search for Invariant Acoustic Correlates of Phonetic Features», dins P. D. EIMAS i J. L. MILLER (eds.), *Perspectives in the Study of Speech*. Hillsdale, Lawrence Erlbaum, pp. 1-38.