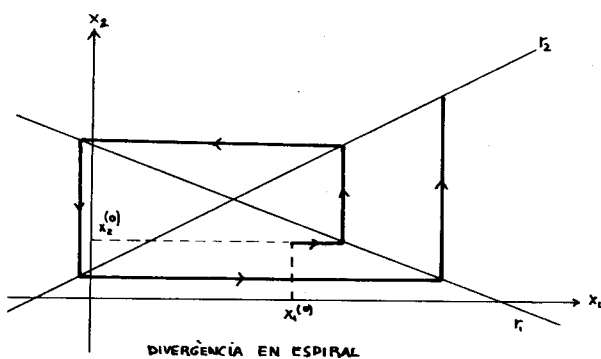
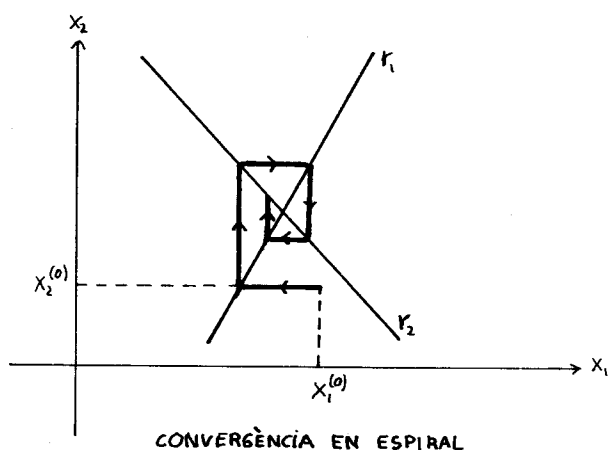
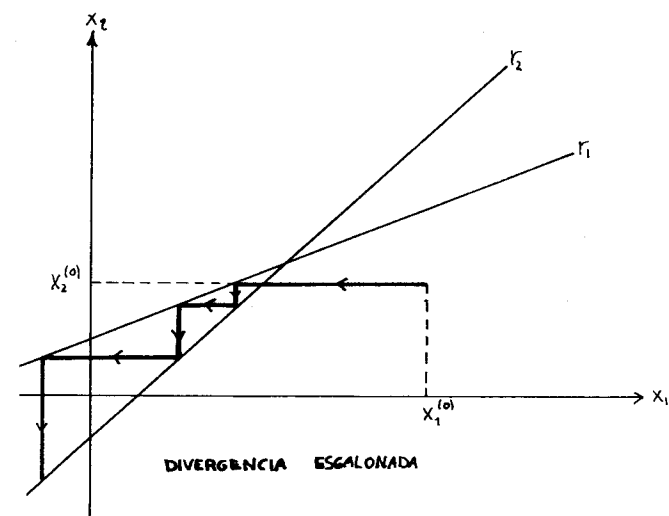


Sobre màquines parlants



Avantatges del mètode de Gauss-Seidel

El mètode d'eliminació de Gauss té l'avantatge de ser finit i aplicable a qualsevol sistema d'equacions lineals.

Tanmateix, quan els mètodes iteratius són aplicables, llavors és preferible la seva aplicació, atès que:

– El temps de càlcul és proporcional a n^2 , mentre que en els mètodes d'eliminació és proporcional a n^3 (n – nombre d'equacions i d'incògnites).

– L'error d'arrodoniment és, en general, més petit.

Ambdós aspectes fan interessant l'aplicació d'aquest mètode quan el nombre d'equacions és petit.

La imitació de la veu humana per mitjans artificials ha estat, i encara és avui en dia, un dels reptes –potser també un dels mites– al nostre desenvolupament tecnològic. Ja els grecs i els romans havien descobert maneres de «fer parlar» les estàtues dels déus, i convertiren les escultures en quelcom de sobrenatural. En el segle XVIII trobem un autèntic interès per les llavors anomenades «màquines parlants», entre les quals destaca la del baró Wolfgang von Kempelen (1791), que podia pronunciar frases en francès, llatí, italià i alemany. Amb l'arribada del segle XX, els sistemes mecànics van cedir el pas als elèctrics, els quals imitaven per mitjà de circuits el mode de producció dels sons de la parla, simulant la funció de la laringe i de la cavitat bucal. Un exemple clàssic fou el Voder (Voice Operation DEMonstrator), desenvolupat per H. Dudley, R.R. Riesz i S.S.A. Watkins i presentat al públic en la fira de Nova York l'any 1939. Avui, els circuits han deixat pas als ordinadors i als processadors, mini i micro, que compleixen la mateixa tasca: generar artificialment seqüències de sons que s'interpreten com a missatges en una llengua donada.¹

El mateix auge de la informàtica ha fet sorgir una altra necessitat: la comunicació oral entre l'home i la màquina. De fet, una de les fites que hom vol assolir amb la cinquena generació d'ordinadors és

el processament de les llengües naturals i la utilització interactiva per mitjà de la veu. Això implica, però, dos problemes diferents, per bé que relacionats. D'una banda, la interpretació del missatge oral per l'ordinador, que suposa el reconeixement i la segmentació dels elements fònics –segmentals i suprasegmentals– seguits d'una anàlisi sintàctica, semàntica i potser fins i tot pragmàtica a fi d'assignar un sentit a una sèrie de sons; d'altra, vol dir trobar la manera que l'ordinador respongui, també mitjançant la veu, al nostre missatge. Aquesta segona faceta es coneix com a síntesi de la parla –sovint hom es refereix a la parla sintetitzada o sintètica per oposició a la natural– i és aquella a la qual em referiré en aquest treball, deixant de banda el reconeixement de la veu, camp en el qual tot just ara es comencen a obtenir resultats encoratjadors.²

Cal començar preguntant-nos per la utilitat de comunicar-nos oralment amb els ordinadors i especialment de disposar de sistemes de resposta vocal en aparells que fem servir habitualment, per veure després quins mètodes s'empren en l'actualitat.

El primer avantatge de disposar d'ordinadors amb una sortida oral és que això deixa les mans lliures per fer qualsevol altra feina. No cal que l'atenció estigui concentrada en la pantalla d'una terminal, ni tampoc no ho ha d'estar la vista. Quan realitzem una tasca, normalment no ens cal interrompre-la si rebem un missatge parlat, però en canvi sí que l'hem de parar si ens obliguen a fixar-nos en allò que apareix en una pantalla. D'altra banda, si un ordinador disposa de sortida oral, ens hi podem comunicar per telèfon des de qualsevol lloc; gairebé sempre es pot utilitzar un telèfon, i en canvi a vegades és difícil tenir a mà una terminal: l'accés a la informació és, per tant, més immediat. Els components d'un sistema de producció de veu sintetitzada són –excepte els altaveus– completament electrònics, i per tant molt barats. Finalment, sembla que la llengua oral és un mitjà de comunicació molt més natural que l'escripta, i que el fet de disposar d'ordinadors que presentin un comportament fins a cert punt «humà» pot ajudar a facilitar les relacions entre l'home i la màquina.

Les aplicacions d'un sintetitzador que permeti d'emetre missatges parlats són moltes, i tenen tots els avantatges que acabem d'enumerar. En primer lloc, cal pensar en les perspectives que això representa per als disminuïts sensorials: un sistema de síntesi que faci possible produir frases que s'escriuin en el teclat d'un microprocessador portàtil, o fins i tot que imitin aquest teclat en una pantalla, obre infinites possibilitats de comunicació a persones que d'altra manera s'han de veure reduïdes al gest –i encara quan aquest és possible– com a únic mitjà de relació amb el seu entorn. Un invident que treballi amb una calculadora o una màquina d'escriure parlant, que representi en veu alta el caràcter corresponent a la tecla que ha pitjat o el resultat del càlcul o la frase que ha escrit, pot participar amb molta més eficàcia en la vida professional. També en aquest camp, la conjunció de diverses tecnologies ha fet possible l'aparició de màquines lectores –KRM («Kurzweil Reading Machine»)– que llegeixen en veu alta qualsevol material imprès, combinant una lectora òptica i un sintetitzador que trans-

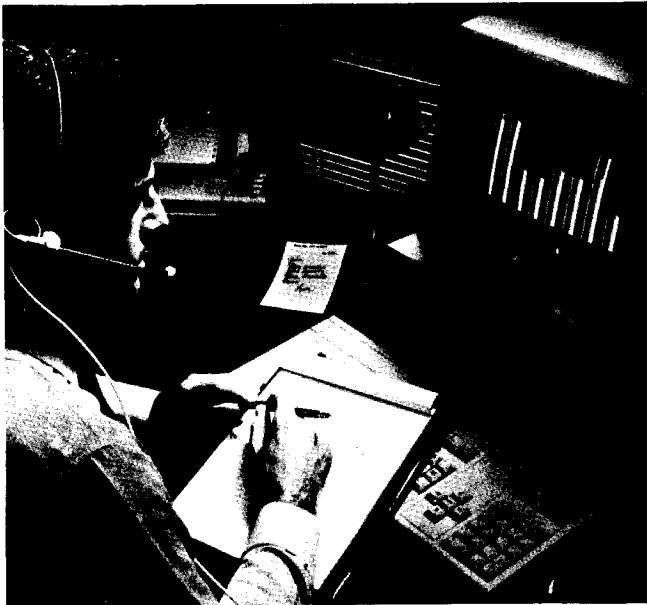
forma els caràcters llegits en parla. Aquestes màquines, tot i que encara requereixen un cert període d'adaptació al tipus de veu i a les imperfeccions de lectura –penseu en els problemes de conversió automàtica d'un text en ortografia convencional a una transcripció fonètica que pugui ser «llegida» per un sintetitzador en una llengua com l'anglès, amb un grau molt baix de regularitat en les relacions entre el so i la grafia– per part de l'usuari, ja es troben en funcionament a diverses biblioteques públiques d'Amèrica i Europa i són regularment emprades per estudiants invidents.

La parla sintetitzada pot també fer-se servir en tots aquells serveis d'informació als quals es pot accedir telefònicament: des del simple rellotge parlant fins a les informacions meteorològiques, de carreteres, esportives, etc., si se selecciona el mode d'operació per mitjà de les tecles o les xifres del marcadors del telèfon. Els ordinadors també poden ajudar a donar instruccions orals a un operari que executi tasques complexes, en les quals tingui les mans ocupades i necessiti consultar constantment documentació: instal·lació de circuits, cables telefònics, reparacions, etc. Ja en un altre pla podem pensar també en totes les indicacions sobre el seu funcionament que ens poden donar els diversos electrodomèstics habitualment usats en una casa: forns, rentadores i –¿per què no?– potser seria fins i tot agradable que la torradora o la cafetera ens diguessin «bon dia» cada matí. Algunes marques de cotxes han incorporat ja indicacions verbals sobre l'estat de manteniment i les incidències de la conducció.

Per acabar, entrant en camps una mica menys frívols, els missatges orals serien un bon complement als programes d'ensenyament per ordinador, que ja comencen a introduir-se a les escoles de casa nostra. En l'estat actual dels sistemes de síntesi, encara és prematur d'utilitzar-los en l'ensenyament de llengües estrangeres vista la qualitat de la veu produïda, però el fet de rebre una indicació verbal seria potser molt més motivador que una sessió de treball davant d'una pantalla muda. Quan els sistemes de traducció automàtica estiguin més avançats, podrà emprar-se igualment un sintetitzador de parla per aconseguir una sortida oral.³

En aquest punt, cal adonar-se que cada aplicació requereix un sistema de síntesi diferent, que ha de ser més sofisticat com més complexa sigui la prestació que li requerim. En referir-nos a la síntesi de la parla, podem distingir inicialment tres modalitats: les tècniques de codificació i reconstrucció de l'ona sonora, la síntesi que globalment anomenarem analògica i la síntesi per regles.

Actualment podem fer ús de diverses tècniques que permeten de «comprimir» l'ona sonora i d'extreure del seu desenvolupament temporal tota la informació que cal per a una síntesi d'alta qualitat. El concepte fonamental per entendre el seu funcionament és la digitalització. L'ona –variacions d'amplitud al llarg del temps que se succeeixen a una freqüència determinada– ha de transformar-se en una representació que pugui conservar-se en la memòria d'un ordinador. El problema rau en el fet que els ordinadors no accepten dades analògiques, sinó només digitals. Per tant, hem de començar per buscar



un mètode que converteixi l'ona sonora en una representació digital. Per tal d'arribar-hi, cal passar per un mostratge de l'ona. La idea del mostratge s'entén fàcilment si pensem en la imatge que tindrem d'una autopista si la mirem fixament durant una estona o si la miren només a intervals: com menys seguits siguin els intervals tindrem una idea més imprecisa de la circulació de l'autopista, però alhora hauré reduït el nombre de dades; com més freqüent és el mostratge, més ens acostem a la realitat. Per aquest procediment convertim un senyal analògic com és l'ona sonora en una sèrie de polsos. Això té lloc un cop hem introduït el resultat del mostratge en un convertidor analògic/digital que ens transforma les mostres de l'ona en elements discrets que s'emmagatzemen en un memòria i que després, amb la intervenció d'un convertidor invers —és a dir, digital/analògic—, es transformen en una reconstrucció de l'ona original, que l'oïda humana interpreta com un missatge. Així es pot conservar la parla en un ordinador i fer que emeti les seqüències que prèviament hem enregistrat, mostrejat i digitalitzat en el moment necessari: quan s'acaba la benzina del cotxe, després d'una determinada instrucció, etc.

El problema d'aquesta tècnica és que per obtenir una reproducció acurada de l'ona sonora hem de guardar un nombre molt elevat de dades. Una manera de resoldre'l és codificant, en compte de les mostres, les diferències entre mostres consecutives, és a dir, en lloc de guardar l'amplitud del senyal, conservar només els canvis d'amplitud de les mostres. Amb aquest procediment —anomenat Modulació Delta— es redueix l'espai necessari en la memòria de l'ordinador.

Fins aquí només hem vist modificacions de l'ona sonora en el domini del temps, però encara podem reduir més la quantitat de memòria necessària si la codifiquem —és a dir, la «comprimim», tornant a la metàfora que feïem servir al principi— en el domini de la freqüència. Potser la tècnica més emprada en l'actualitat és la codificació per predicció lineal (LPC

«Linear Predictive Coding»), desenvolupada a mitjan anys 60, que perfecciona un mètode de càlcul que permet la predicció dels esdeveniments futurs en un sistema lineal —és a dir, sense ramificacions— fent servir informació derivada dels que ja han passat. Això s'aconsegueix gràcies a una propietat de l'ona sonora: la periodicitat. Si l'ona és periòdica —com en el cas dels sons—, una mostra d'aquesta ona serà previsible a partir de mostres anteriors. Llavors, és lògic pensar que, en comptes de codificar totes les mostres de l'ona, podem representar-la en termes de la diferència entre dues mostres consecutives, és a dir, mitjançant l'error de predicció. Aquesta mena d'anàlisi podem ampliar-la, de manera que es considerin totes les mostres precedents. Les característiques acústiques de l'ona sonora depenen de la configuració del tracte vocal, i per això podem definir les seves variacions en termes de coeficients de predicció, considerats —d'una manera molt simplificada— com un resum del mode en què les cavitats bucals han modificat els sons produïts per la font d'energia —és a dir, les cordes vocals. Alhora ofereixen una representació paramètrica de l'ona, altament econòmica. Això es deu, com hem dit, al fet que els coeficients es relacionen amb les ressonàncies del tracte vocal, que normalment es poden especificar en prou detall si s'empren entre 10 i 12 coeficients.

El principal avantatge de la predicció lineal amb vista a la síntesi és que la codificació de l'ona en coeficients de predicció ocupa molt poc espai en la memòria d'un microprocessador. Per exemple, un microprocessador de 48 K de memòria podria retenir fins a 5 minuts de parla codificada per predicció lineal, mentre que si s'empren altres mètodes el total pot baixar a 24 segons o fins i tot a 3.2 segons, segons la qualitat que es busqui.

La primera etapa del procés consisteix a enregistrar les frases que es volen conservar en la memòria i digitalitzar-les amb un convertidor analògic/digital. Allò que es fa en aquest punt és una codificació de l'ona tal com hem explicat abans. Però si en compte de guardar-la així per reconstruir-la posteriorment, la codifiquem mitjançant el mètode de la predicció lineal, obtindrem un conjunt de paràmetres que caracteritzen la seqüència sonora que hem enregistrat i que hem anomenat coeficients de predicció. Guardats en una memòria —habitualment una memòria ROM («Read Only Memory»)— poden després, seguint el procediment invers, reconvertir-se en una ona equivalent a la inicial. Els sintetitzadors per predicció lineal que avui es troben al mercat tenen una memòria com la que hem descrit, en la qual s'han emmagatzemat els coeficients de predicció corresponents als missatges que el sintetitzador pot emetre; l'única tasca que porten a terme és la reconstrucció de l'ona partint d'aquests coeficients. Per aquesta raó tenen una capacitat molt limitada i un vocabulari molt reduït. El primer «chip» de síntesi per predicció lineal que va introduir-se massivament fou desenvolupat per Texas Instruments i es comercialitzà en forma de joguina: el «Speak'n Spell», que fins i tot s'ha fet famós en una pel·lícula recent, utilitzat per ensenyar a escriure anglès a una mexicana fronterera.⁴

A hores d'ara, ja haurà quedat clar que els siste-

mes de síntesi que acabem de presentar tenen un inconvenient greu: la limitació del seu vocabulari. Algunes de les aplicacions a què hem fet referència no poden funcionar amb un lèxic reduït; per exemple, una màquina que llegeixi en veu alta no pot emmagatzemar tot el vocabulari d'una llengua, i el mateix s'aplica a qualsevol sistema que necessiti convertir en veu seqüències que no hagin estat previstes. Per això ens calen uns models de sintetitzador que treballin amb una estratègia diferent. Ens podem apropar a la síntesi des d'una altra perspectiva, i en comptes de treballar sobre la mateixa ona sonora, intentar de fer una imitació —és a dir, un model— de la manera com els humans produïm la parla. Tant podem prendre de punt de referència els mecanismes articuladoris com els mecanismes acústics de la producció de la parla. Si agafem els primers, la síntesi s'haurà de fer a partir de paràmetres articuladoris, és a dir, haurem d'especificar al programa del nostre sintetitzador quina és la configuració del tracte vocal en la producció d'un determinat so. Li haurem de donar informació sobre l'avançament i el tancament dels llavis, l'elevació de les diverses parts de la llengua, etc. S'aconsegueix així una simulació de la cavitat bucal per mitjà d'un programa, i cada so de la llengua en la qual treballem quedarà emmagatzemat com un conjunt de paràmetres articuladoris. Després caldrà preveure les influències entre els sons adjacents, ja que aquestes poden modificar les seves característiques —es tracta del fenomen conegut per coarticulació—, i així arribem a sintetitzar frases senceres amb una certa naturalitat. Tot i així, ja fa temps que s'ha constatat que cada parlant pot adoptar les seves estratègies particulars en pronunciar un so determinat, i que el mateix resultat acústic —és a dir, el so que sentim i que interpretem— pot ser produït per mitjans articuladoris diferents.⁵

Aquest fet ens porta a pensar que potser és més encertat de considerar un model del procés de producció de la parla basat en paràmetres acústics. Fent-ho així, cada so quedaria especificat a la memòria de l'ordinador en termes de les seves característiques acústiques: freqüència, amplitud i durada. Els valors d'aquests paràmetres s'obtenen de mesurar produccions de parlants de la llengua en la qual treballem i comprovar que, un cop reproduïts per la màquina, arribem a seqüències de sons que reben una interpretació lingüística correcta per part dels altres parlants. Cada so pot definir-se només amb 10 o 20 paràmetres, i això fa que hi hagi també un estalvi considerable de memòria. Els paràmetres que es fan servir habitualment són la freqüència del que és fonamental —responsable, per exemple, dels canvis melòdics en les frases—, les freqüències i amplituds dels formants (zones d'harmònics reforçats especialment significatives en la descripció i molt possiblement també en la percepció de la parla), l'amplitud de l'energia, etc.⁶

Aquests darrers sistemes de síntesi a partir dels valors dels paràmetres, tant si són articuladoris com acústics, ens porten a una altra concepció de la síntesi, que es coneix com a síntesi per regles. Essencialment, es tracta d'obtenir els mateixos resultats que amb un model, però partint d'una cadena de símbols discrets —una transcripció fonètica, un text

escrit— que es converteixen en sons mitjançant un sistema de regles prèviament definides. La síntesi per regles s'aplica típicament als sistemes de conversió de text a parla, que transformen un text escrit en ortografia convencional en una seqüència de sons. L'operador comença entrant a l'ordinador un text per mitjà del teclat. Les primeres regles que el modifiquen assignen els accents, hi afegeixen les fronteres morfològiques que calen per a la presentació fonètica posterior i el text modificat es transforma, gràcies a una segona sèrie de regles de conversió grafia-so en una cadena fonètica —és a dir, en una mena de transcripció en alfabet fonètic. La tercera classe de regles altera també aquesta cadena, ja que les característiques dels sons es modifiquen segons el context i cal operar reajustaments deguts als fets de contacte entre sons. Finalment, un darrer conjunt de regles assigna a cada so els valors dels paràmetres —habitualment acústics— que han estat recollits en el diccionari de sons que l'ordinador guarda a la memòria. Agafant aquests valors, el sintetitzador pot generar la seqüència sonora corresponent al text que hem introduït. Alguns sistemes incorporen també un «diccionari de pronunciació», en el qual es recullen els casos que no es poden predir per regles.⁷

Malgrat llur capacitat infinita, aquests sistemes presenten en l'actualitat diversos problemes. En primer lloc, la concatenació entre els segments no sempre és perfecta, ja que, com hem dit anteriorment, el context modifica les característiques acústiques dels sons de la parla que, per altra banda, no es produeixen mai aïlladament. Cal disposar també d'una certa informació sobre l'estructura sintàctica de l'enunciat, per tal d'establir regles que permetin de donar una entonació adequada a qualsevol frase. Encara hi ha un altre factor: la qualitat de la parla sintetitzada per mitjà de regles és molt més baixa que l'obtinguda amb sistemes de reconstrucció i codificació de l'ona. Aquests darrers ofereixen certament una major qualitat, però tenen molta menys versatilitat i requereixen més memòria. Els sistemes més flexibles ofereixen, en canvi, una intel·ligibilitat molt baixa —com és el cas de la KRM— i els usuaris han de passar per un període d'adaptació a la qualitat —«l'accent»— de la veu sintetitzada. En conjunt, es tracta de veure quin és el sistema més convenient en funció de l'aplicació que li volem donar i de conèixer les limitacions i els avantatges de cadascun. Tot i així, un dels principals experts actuals en síntesi de la parla, el britànic John Holmes, en el darrer Congrés Internacional de Ciències Fonètiques celebrat a Utrecht l'estiu de 1983, anunciava que d'aquí a pocs anys els sistemes de síntesi per regles substituiran tots els mètodes basats en l'emmagatzemament de parla codificada. Evidentment això obre moltes perspectives, ja que els sistemes de conversió de text a parla requereixen la intervenció d'un equip d'especialistes. L'elaboració dels programes correspon a un expert en informàtica, el disseny del sintetitzador que ha de convertir els valors dels paràmetres en ones sonores requereix la intervenció d'especialistes en acústica i en tècniques de tractament del senyal, mentre que les regles de transcripció del text ortogràfic en una representació fonètica són responsabilitat del lingüista. I en el mo-

ment de provar la intel·ligibilitat del resultat, pot intervenir el psicòleg en la confecció d'una bateria de tests per tal de verificar l'eficàcia de la parla sintetitzada des del punt de vista de la comunicació. Potser en una etapa no gaire llunyana, la intel·ligència artificial confluirà amb la síntesi de la parla i farà possible noves recerques i aplicacions.

El meu objectiu en aquest treball era de presentar d'una manera simplificada i esquemàtica alguns dels procediments de què disposem actualment per obtenir una resposta verbal de les màquines, i d'aquí el títol de l'article. Espero haver fet veure que el camp de la síntesi de la parla és un camp ric en aplicacions futures i que, sobretot, permet i demana la presència conjunta de professionals de disciplines que fins ara havien restat, potser, incomunicades entre elles i que d'ara endavant poden col·laborar en aquesta aventura apassionant que és donar veu a les màquines.

JOAQUIM LLISTERRI I BOIX

Professor del departament de Filologia Hispànica de la Universitat Autònoma de Barcelona

1. Per a una visió general del desenvolupament històric, vegeu DAVID, E. (1973) «Artificial Speech» a Miller, G. A. (Ed) *Communication, Language and Meaning*. New York: Basic Books Inc. pp. 141-8; FLANAGAN, J. L. (1972) «Voices of men and machines», *Journal of the Acoustical Society of America*, 51, 5: 1375-86 i a Flanagan, J. L.-Rabiner, L. R. (Eds) (1973) *Speech Synthesis*. Stroudsburg, Penn.: Dowden, Hutchinson & Ross, Inc. pp. 9-21; METTAS, O. (1965) «Aperçut historique sur les appareils de synthèse de la parole», *Travaux de Linguistique et Littérature* (Strasbourg), III, 1: 185-200. Sobre la màquina de von Kempelen: DUDLEY, H.-TARNOCZY, T. H. (1950) «The Speaking Machine of Wolfgang von Kempelen», *Journal of the Acoustical Society of America*, 22: 151-166, i VAN DEN BROECKE, M. P. R. (1983) «Wolfgang von Kempelen's Speaking Machine as a Performer» a Van den Broecke, M. P. R.-Van Heuven, V.-Zonneveld, W. (Eds) *Sound Structures. Studies for Antonie Cohen*. Dordrecht: Foris Publications, pp. 9-20. Pel que fa al Voder, cf. DUDLEY, H. RIESZ, R.R.-WATKINS, S. S. A. (1939) «A Synthetic Speaker», *Journal of the Franklin Institute* (Philadelphia), 227: 739-64 i a Flanagan-Rabiner (Eds), *op. cit.*, pp. 190-215.

2. Sobre reconeixement de parla vegeu, per exemple: CATER, J. P. (1984) *Electronically Hearing: Computer Speech Recognition*. Indianapolis, Ind.: Howard W. Sams & Co.; Lea W. A. (Ed) (1980) *Trends in Speech Recognition*. Englewoods Cliffs, N.J.: Prentice Hall.; LEVINSON, S. E.-LIBERMAN, M. Y. (1981) «Speech Recognition by Computer», *Scientific American* 244: 64-76; i NOLAN, F. (1983) *The Phonetic Basis of Speaker Recognition*. Cambridge: Cambridge University Press.

3. Vegeu, entre altres: CATER, J. P. (1983) *Electronically Speaking: Computer Speech Generation*. Indianapolis: Howard W. Sams & Co.; FLANAGAN, J. L. (1972) «The Synthesis of Speech», *Scientific American*, 226, 2: 48-58; FLANAGAN, J. L. (1976) «Computers that Talk and Listen. Man Machine Communication by Voice», *Proceedings IEEE*, 64,4: 405-415; FLANAGAN, J. L. (1982) «Talking with Computers: Synthesis and Recognition of Speech by machines» *IEEE Transactions on Biomedical Engineering* BM-29, 4: 223-32; FLANAGAN, J. L. et al. (1970) «Synthetic Voices for Computers», *IEEE Spectrum*, 7, 10: 22-45; MORGAN, N. (1984) *Talking Chips. IC Speech Synthesis*. New York: McGraw-Hill Book Company (National's Semiconductor Technology Series). POULTON, A. S.

(1983) *Microcomputer Speech Synthesis and Recognition*. Wilmslow, Cheshire: Sigma Technical Press. SCLATER, N. (1983) *Introduction to Electronic Speech Synthesis*. Indianapolis, Indiana: Howard W. Sams & Co.; WITTEN, I. H. (1982) *Principles of Computer Speech*. New York: Academic Press (Computers and People).

4. Sobre la predicció lineal vegeu: ATAL, B. S. - HANAUER, S. L. (1971) «Speech Analysis and Synthesis by Linear Prediction of the Speech Wave», *Journal of the Acoustical Society of America*, 50: 637-75, i Flanagan-Rabiner (Eds.), *op. cit.*, pp. 270-288.; MARKEL, J. D.-GRAY, A. H. (1976) *Linear Prediction of Speech*. Berlin-Heidelberg-New York: Springer Verlag (Communications and Cybernetics, 12); WAKITA, H. (1980) «New Methods of Analysis in Speech Acoustics», *Phonetica*, 37, 1/2: 87-108.

Descripcions del «Speak'n Spell» poden trobar-se als treballs citats en la nota anterior, especialment a SCLATER (1983) i WITTEN (1982).

5. Un exemple de síntesi emprant paràmetres articuladoris el dona HAGGAR, M. P. (1979) «Experiences and Perspectives in Articulatory Synthesis» a Lindblom, B.-Ohman, S. (Eds.) *Frontiers of Speech Communication Research*. New York: Academic Press, pp. 259-74.

6. Vegeu, per exemple, HOLMES, J. N. (1979) «Synthesis of Natural-Sounding Speech Using a Formant Synthesizer» a Lindblom, B.-Ohman, S. (Eds.), *op. cit.*, pp. 275-85; KLATT, D. H. (1980) «Software for a cascade/parallel formant synthesizer», *Journal of the Acoustical Society of America*, 67: 971-995.

7. Els treballs clàssics sobre síntesi per regles són: HOLMES, J.-MATTINGLY, I.-SHEARME, J. N. (1964) «Speech Synthesis by Rule» *Language and Speech*, 7: 127-143 i a Flanagan, J. L.-Rabiner, J. L. (Eds.), *op. cit.*, pp. 351-367; i LIBERMAN, A. M.-INGEMAN, F.-LISKER, L.-DELATTRE, P. C.-COOPER, F. S. (1959) «Minimal rules for synthesizing speech» *Journal of the Acoustical Society of America* 31: 1490-1499, i a LEHISTE, I. (Ed.) (1967) *Readings in Acoustic Phonetics*. Cambridge, Mass: MIT Press, pp. 333-42 i a Fry, D. B. (Ed.) (1967) *Acoustic Phonetics*. Cambridge: Cambridge University Press, pp. 445-66. Un treball més actual pot trobar-se a HERTZ, S. (1982) «From text to speech with SRS», *Journal of the Acoustical Society of America* 72, 4: 1155-1170. En llengua castellana es poden veure els treballs del grup d'anàlisi, síntesi i reconeixement de la parla de l'ETSIT i el Laboratori de Fonètica del CSIC de Madrid, alguns dels quals es publiquen a *Mundo Electrónico* d'octubre de 1984.

8. HOLMES, J. N. (1984) «Speech Technology in the Next Decades» a Van den Broecke, M. P. R.-Cohen, A. (Eds.) *Proceedings of the Tenth International Congress of Phonetic Sciences*. Dordrecht: Foris Publications, pp. 125-139.