

LLISTERRI, J. (1984) "Aproximació a la síntesi de les vocals del català", *Folia Phonetica* 1: 45-78. DL: L-448-1984

[http://liceu.uab.es/~joaquim/publicacions/
Llisterri_84_vocals_catala.pdf](http://liceu.uab.es/~joaquim/publicacions/Llisterri_84_vocals_catala.pdf)

Aproximació a la síntesi de les vocals del català.

Joaquim Llisterrí.

Departament de Filologia Hispànica

Facultat de Lletres

Universitat Autònoma de Barcelona (Bellaterra)

Resum.

Aquest treball presenta els resultats d'un primer intent de sintetitzar vocals catalanes emprant paràmetres freqüencials derivats de produccions naturals. Es comparen les mesures preses amb un Sona-Graph i un analitzador de banda fina Bruel & Kjaer amb les obtingudes mitjançant un procés de tractament del senyal i una anàlisi per predicció lineal, alhora que s'avalua la precisió assolida de cara a la síntesi. Una sèrie d'estímulos reproduint les vuit vocals del català han estat sintetitzats emprant mitjanes lleugerament modificades de valors obtinguts de quatre parlants. Els resultats d'un test de percepció d'aquests estímulos donen una mitjana global del 88.87 % d'identificacions correctes, validant així els paràmetres utilitzats.

Finalment, es proposa una modificació del trapezi vocàlic de Cerdà (1972) i es suggereix que les dades obtingudes poden constituir el primer pas per a un sistema complet de síntesi per regles.

1. Introducció i objectius.

El 1981 apareixia a *Els Marges* un article de D. Recasens titulat "Futur dels estudis de fonètica experimental del català" (Recasens, 1981), en el qual l'autor donava una sèrie d'indicacions metodològiques per a l'anàlisi del timbre vocàlic en català. A més d'una revisió del llibre de R. Cerdà (1972), del qual assenyala algunes insuficiències, indica la necessitat de "determinar amb major precisió el triangle acústic de vocals catalanes fixat per Cerdà" (p. 55), i d'emprar tests de percepció d'estímulos sintetitzats per a una delimitació acurada del camp de dispersió (p. 59).

El present treball s'emmarca en aquesta línia de recerca, sense pretendre, però, aportar dades definitives, sinó resultats determinats per una situació experimental concreta. Tal com indica el títol, es tracta d'un primer experiment de síntesi de vocals catalanes encaminat a determinar la freqüència dels tres primers formants d'un corpus de vocals en context, i a verificar-la mitjançant una prova de percepció de reproduccions sintetitzades dels sons vocàlics estudiats¹.

En primer lloc s'obtenen dades sobre la validesa dels diversos mètodes d'anàlisi, que es contrasten amb les publicades per altres investigadors, i en segon, podem disposar també d'informacions sobre l'estructuració a nivell perceptiu del sistema vocàlic català. Finalment, s'arriba a possibles valors dels paràmetres per a la síntesi de les vocals catalanes, primer pas per a diverses aplicacions tecnològiques que esmentarem més endavant.

2. Corpus, informants i gravació.

L'esquema general del procediment experimental es presenta a la fig. 1. En conjunt, pot descompondre's en quatre etapes: (a) constitució del corpus i gravació; (b) anàlisi del corpus; (c) síntesi a partir de les dades obtingudes en l'anàlisi i (d) validació de la síntesi per mitjà d'un test de percepció. Ens referirem aquí a la primera etapa i exposarem detalladament les altres en els següents apartats.

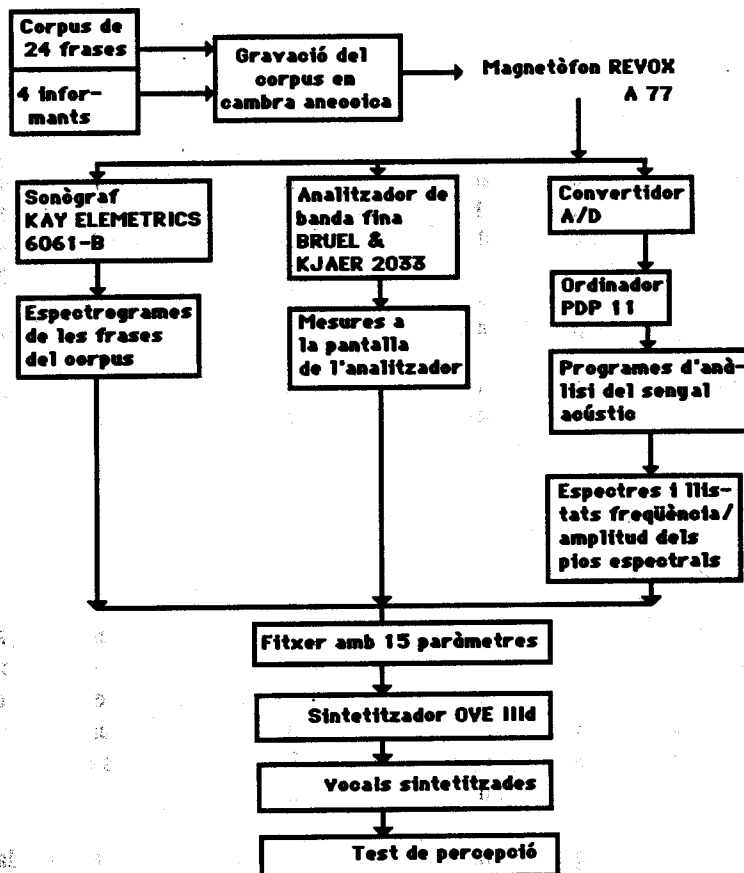


Fig. 1. Protocol experimental del treball.

2.1. Corpus.

El corpus sobre el que s'ha treballat consta de 24 frases que contenen els set sons vocàlics tòncics del català central a més de la vocal neutra. Reservo l'estudi de les vocals àtones per a un treball posterior. Els sons estudiats apareixen entre dos oclusius sords del mateix punt d'articulació en mots bisil·làbics o monosil·làbics contextualitzats en frases-marc del tipus "Va dir-li [CVC(V)] / [CV'CV] i va marxar/penjar/sortir". Tenim així dues variables: l'element vocàlic i l'entorn consonàntic, limitat però a tres possibilitats: [p_p], [t_t] i [k_k]. En aquest entorn la vocal queda ben delimitada entre les dues oclusions com pot apreciar-se a l'oscillograma de la fig. 2. Es representen així tres tipus d'entorn: [p] greu, [t] agut i [k] compacte (Jakobson-Waugh, 1979: 109).

La resta dels elements es manté constant: (a) posició de la vocal en el volum total de la frase que assegura una certa homogeneïtat pel que fa a la durada i a la distribució de l'energia global de l'enunciat -Cf. Murillo, 1981: IV-54-; (b) situació de la vocal en la corba entonativa; i (c) accent.

2.2. Informants.

El corpus fou enregistrat per quatre informants -J.C., M.O., E.B. i R.M.B.- de llengua materna catalana, parlants de la varietat central, estudiants universitaris en el moment de fer les gravacions, i sense cap defecte articulatori que afectés la qualitat de les vocals. J.C. i M.O. són informants masculins amb una freqüència del fonamental que a les vocals estudiades oscil·la entre 137.5 Hz i 162.5 Hz -mitjana = 155 Hz- per a J.C., i entre 87.5 Hz i 112.5 Hz -mitjana = 100 Hz- per a M.O. Pel que fa als locutors femenins, R.M.B. presenta un fonamental entre 212.5 Hz i 187.5 Hz -mitjana = 200 Hz- i E.B. varia entre 225 Hz i 275 Hz -mitjana = 245 Hz-.

2.3. Gravació.

La gravació del corpus va realitzar-se a la cabina insonoritzada del Laboratori de Fonètica de la Universitat Autònoma de Barcelona (Bellaterra), amb un magnetòfon Revox A77 i un micròfon Sennheiser cardioide MD44 sobre cinta professional Scotch Master.

Va demanar-se als locutors que llegissin les frases d'una manera natural, i se'ls va orientar sobre les possibles situacions en què podrien haver estat pronunciades. També es van donar precisions sobre casos potencialment conflictius com "còctel" i "poble", suggerint que es llegissin seguint els propis hàbits de pronúncia.

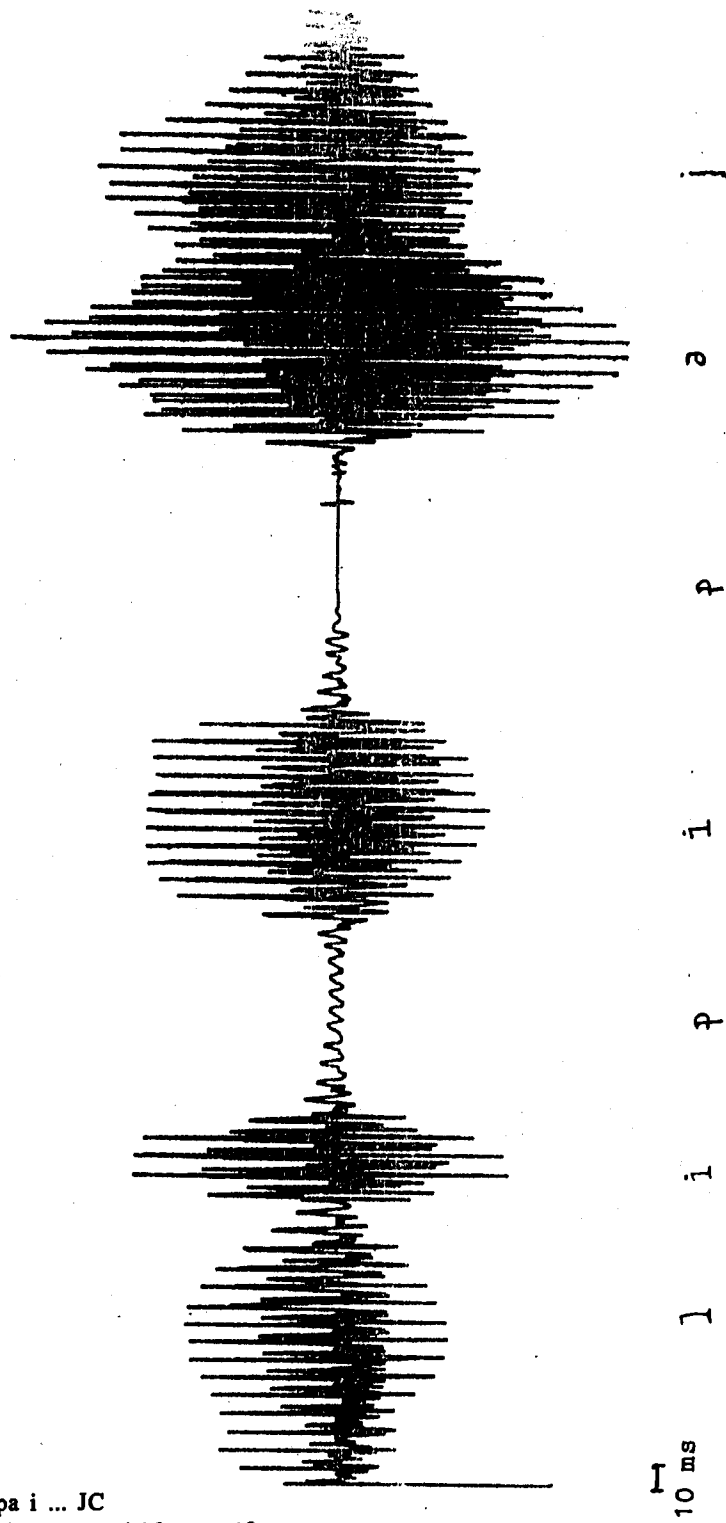


Fig. 2 ...dir-li pipa i ... JC
Escala de mesura: 4.16 mm= 10 ms

3. Mètodes i resultats de l'anàlisi.

Un cop enregistrat el corpus, va procedir-se a l'anàlisi acústica de les vocals a fi d'obtenir les mesures de la freqüència dels tres primers formants que més endavant constituïran els paràmetres de la síntesi. En primer lloc es va recórrer a l'espectrògraf, l'aparell més convencional amb el qual foren realitzats els clàssics estudis sobre vocals de l'anglès americà durant la dècada dels 50 -p.ex. Peterson-Barney, 1952-. La mesura dels espectrogrames fou complementada per la utilització d'un analitzador espectral de banda fina amb el qual també es va mesurar la freqüència del fonamental. Finalment va aplicar-se un programa de tractament del senyal acústic detallat a l'apartat 3.3.

3.1. Anàlisi espectrogràfica.

El corpus sencer fou espectrografiat emprant un aparell Sona-Graph Kay Elemetrics model 7029A. Per arribar a una millor definició dels formants en les produccions dels informants femenins i augmentar alhora la resolució en el domini temporal es va fer servir el filtre de banda ampla (300 Hz) i una velocitat de reproducció la meitat de la normal (3 3/4 ips) (Fant, 1962: 23; Fant, 1968: 181). Tot i aplicant aquesta tècnica que millora notablement els resultats resten encara una sèrie de problemes en la interpretació i mesura dels sonagrames, molts dels quals han estat ja assenyalats a la bibliografia (Lindblom, 1962; Ladefoged, 1967: 81-83):

(a)- L'escala de mesura. Treballant, com en el cas d'aquest estudi, amb una escala de freqüència entre 40 i 4.000 Hz i a una velocitat de reproducció la meitat de la normal, 1 mm sobre l'espectrograma equival a 86 Hz en l'escala de freqüència i a 8.5 ms en l'escala temporal. Si tractem de sintetitzar vocals, un error de medició d'un mil·límetre en un primer formant baix pot superar fàcilment el valor del llindar diferencial de percepció de la freqüència -el 3% de la freqüència del formant segons Flanagan, 1957- i per tant introduir canvis en el timbre que poden afectar la qualitat de la vocal sintetitzada.

(b)- Resolució dels formants. La definició dels formants en un sonagrama no sempre és clara; sorgeixen problemes especialment quan: (a) F1 és baix i tendeix a barrejar-se amb el fonamental, fenomen que potser explicaria el sorprenent resultat de 195 Hz per a la freqüència de F1 a [i] i [u] en l'estudi de R. Cerdà (op. cit.: 162-3); (b) F1 i F2 són molt propers, com és el cas de les vocals posteriors, i és difícil de destriar l'un de l'altre; (c) F2 i F3 tenen una intensitat molt baixa -p.ex. en [u]- i apareixen molt feblement registrats en el sonagrama. Aquests problemes es tracten amb detall a Ladefoged (1967).

Considerant, doncs, les limitacions del nostre instrument, foren mesurades només les produccions de l'informant M.O. en el context [pVp], amb els resultats que es presenten a la taula I. A la figura 3 es comparen aquestes mesures amb el trapezi proposat per R. Cerdà (loc. cit.); pot observar-se que si bé es mantenen les proporcions, el desplaçament és considerable. Cal remarcar, però, el fet ja ben conegut que un entorn amb oclusius bilabials tendeix a fer més baixes les freqüències dels dos primers formants; en el treball de Cerdà les vocals apareixen en tots els contextos consonàntics.

3.2. Anàlisi espectral.

El corpus fou posteriorment estudiat amb un analitzador de banda fina Bruel & Kjaer 2033. Les mesures dels espectres es van prendre directament sobre la pantalla, localitzant primer el segment més estacionari de la vocal, i avaluant després visualment el centre de cada formant a partir dels pics de l'envolupant espectral. La resolució de cada línia de la pantalla -que representa un harmònic- és de 12.5 Hz, amb la qual cosa l'error de mesura es redueix sensiblement respecte al Sona-Graph.

Vocal	MO			Cerdà (1972)	
	F1	F2	F3	F1	F2
[i]	405	2340	2925	195	2050
[e]	495	2160	2745	330	1775
[ɛ]	630	1890	4655	415	1625
[ɐ]	585	1350	2250	370	1225
[a]	810	1260	2700	540	1125
[ɔ]	720	1080	2340	400	890
[o]	495	900	1890	300	670
[u]	405	855	2430	195	480

Taula I. Freqüència dels tres primers formants mesurada sobre espectrogrames de l'informant MO en el context [pVp] comparades amb les dades de Cerdà (1972).

Aquest procediment no evita, però, alguns problemes als quals s'ha fet referència en l'apartat anterior, especialment pel que fa als informants femenins: la major separació dels harmònics dificulta considerablement la detecció dels pics corresponents al centre dels formants amb un mínim de precisió. També en els casos de F1 baix aquest tendeix a confondre's amb un Fo resultant en un únic pic en el qual és difícil d'avaluar correctament el centre del primer formant. El mateix succeeix quan F1 i F2 estan molt junts i es confonen també en una sola ressonància.

A la taula II es donen els resultats corresponents als informants masculins obtinguts pel procediment que acabem d'explicar i es comparen amb els obtinguts per J. Martí (1983) en la seva anàlisi de les vocals del català emprant el mateix instrument Bruel & Kjaer 2033². Pel que fa al primer formant, s'observa que la diferència més gran rau en les mitjanes corresponents a la vocal neutra -prop de 46.5 Hz-, però que en els altres casos no arriben mai a 35 Hz, mantenint-se entre un 1% i un 10% del valor del formant. A F2 les diferències no són tan elevades i en cap cas, llevat de [u] que Martí discuteix a part -(1983): 123-25-, no superen el 7% del valor del formant i no excedeixen el 5% si eliminem [ə] que presenta habitualment un camp de dispersió molt ampli. Quant al tercer formant, les diferències poden arribar al 10% del valor del formant en el cas de [i] o a un 3% en el cas de [ə], [o] i [u].

La comparació entre els triangles delimitats pels dos grups de mitjanes pot veure's a la figura 4, en la qual cal destacar la superposició obtinguda -compareu-la amb la figura 3-, llevat del cas de [i] i [e] i també de [u] que, com hem dit, ja es comenta en el mateix treball de Martí.

Primer Formant F1 (Hz)						
Vocal	MO i JC			J. Martí (1983:79) 105 m./voc.		
	Mitjana	D. tít.	D. rel. %	Mitjana	D. tít.	D. rel. %
[i]	287.5	27	9.39	258.3	29.9	11.6
[e]	372.9	20.93	5.6	404.8	54.7	13.5
[ɛ]	547.91	38.47	7.02	581.0	35.3	6.1
[ə]	458.33	18.63	4.06	504.8	49.8	9.9
[a]	685.41	48.09	7.01	694.5	68.2	9.8
[ɔ]	556.25	58.51	10.52	586.9	39.2	6.7
[o]	441.66	42.49	9.6	471.4	42.8	9.1
[u]	318.75	17.30	5.42	329.8	37.6	11.4

Segon Formant F2 (Hz)						
Vocal	MO i JC			J. Martí (1983:80) 105 m./voc.		
	Mitjana	D. tít.	D. rel. %	Mitjana	D. tít.	D. rel. %
[i]	2274.5	129.8	5.71	2179.8	7.3	
[e]	2025	31.45	1.55	1929.8	138.4	7.2
[ɛ]	1743.5	44.34	2.54	1713.1	129.1	7.5
[ə]	1383.33	166.24	12.01	1287.5	213.7	16.6
[a]	1347.91	138.52	10.27	1365.5	109.1	8.0
[ɔ]	1054.16	95.10	9.02	1014.3	64.5	6.4
[o]	928.75	83.21	8.95	947.6	79.4	8.4
[u]	795.83	119.38	15	1097.6	122.5	11.2

Tercer Formant F3 (Hz)

Vocal	MO i JC			J. Martí (1983:81) 105 m./voc.		
	Mitjana	D. tít.	D. rel. %	Mitjana	D. tít.	D. rel. %
[i]	3106.25	365.7	11.77	2791.7	239.8	8.6
[e]	2533.33	62.77	2.47	2619.8	155.5	5.9
[ɛ]	2431.25	117.20	4.82	2545.2	144.4	5.7
[ə]	2241.66	159.86	7.13	2435.7	219.5	9.8
[a]	2272.5	199.74	8.78	2406.0	224.4	9.3
[ɔ]	2293.75	188.98	8.23	2376.2	254.4	10.7
[o]	2385	295.45	12.38	2461.9	209.0	8.5
[u]	2379.91	65.21	2.7	2310.7	244.3	10.6

Taula II. Comparació entre les mitjanes dels dos informants masculins i els resultats de Martí (1983) obtinguts amb l'analitzador de banda fina. S'hi afegeix la desviació relativa que apareix en el treball de Martí com a mesura de la dispersió.

3.3. Anàlisi per predicció lineal.

Finalment, va aplicar-se un programa concret de tractament del senyal acústic, esquematitzat i explicat a la figura 5. La part essencial de tot el procés és l'anàlisi per predicció lineal³.

Segons Monsen (1981) amb una freqüència del fonamental entre 100 Hz i 350 Hz els tres primers formants poden mesurar-se amb una precisió de ± 60 Hz, amb l'avantatge que no disminueix quan augmenta el fonamental, mentre no vagi més enllà dels 350 Hz.

La idea bàsica del mètode és que una mostra d'un enunciat pot ésser representada com una combinació d'un nombre de mostres anteriors del mateix enunciat. A partir de la diferència entre les mostres reals i les mostres predites, es determinen els coeficients de predicció -normalment entre 10 i 15- que constitueixen una mena de model o de representació paramètrica de l'ona (Schäfer-Rabiner, 1975: 673). A partir d'aquests coeficients pot reconstruir-se l'envolupant i obtenir les freqüències dels formants amb un programa de detecció automàtica de pics. A la taula III es presenten els resultats corresponents als informants masculins i femenins d'aquest treball obtinguts pel procediment que s'acaba d'exposar⁴.

Finalment, la freqüència del fonamental amb què es pronuncià cada vocal fou mesurada a la pantalla de l'analitzador de banda fina Bruel & Kjaer 2033, pel mètode de trobar la freqüència del primer harmònic de l'espectre. Els resultats, igual que en altres llengües, confirmen la correlació entre F_0 i $F1$ ⁵.

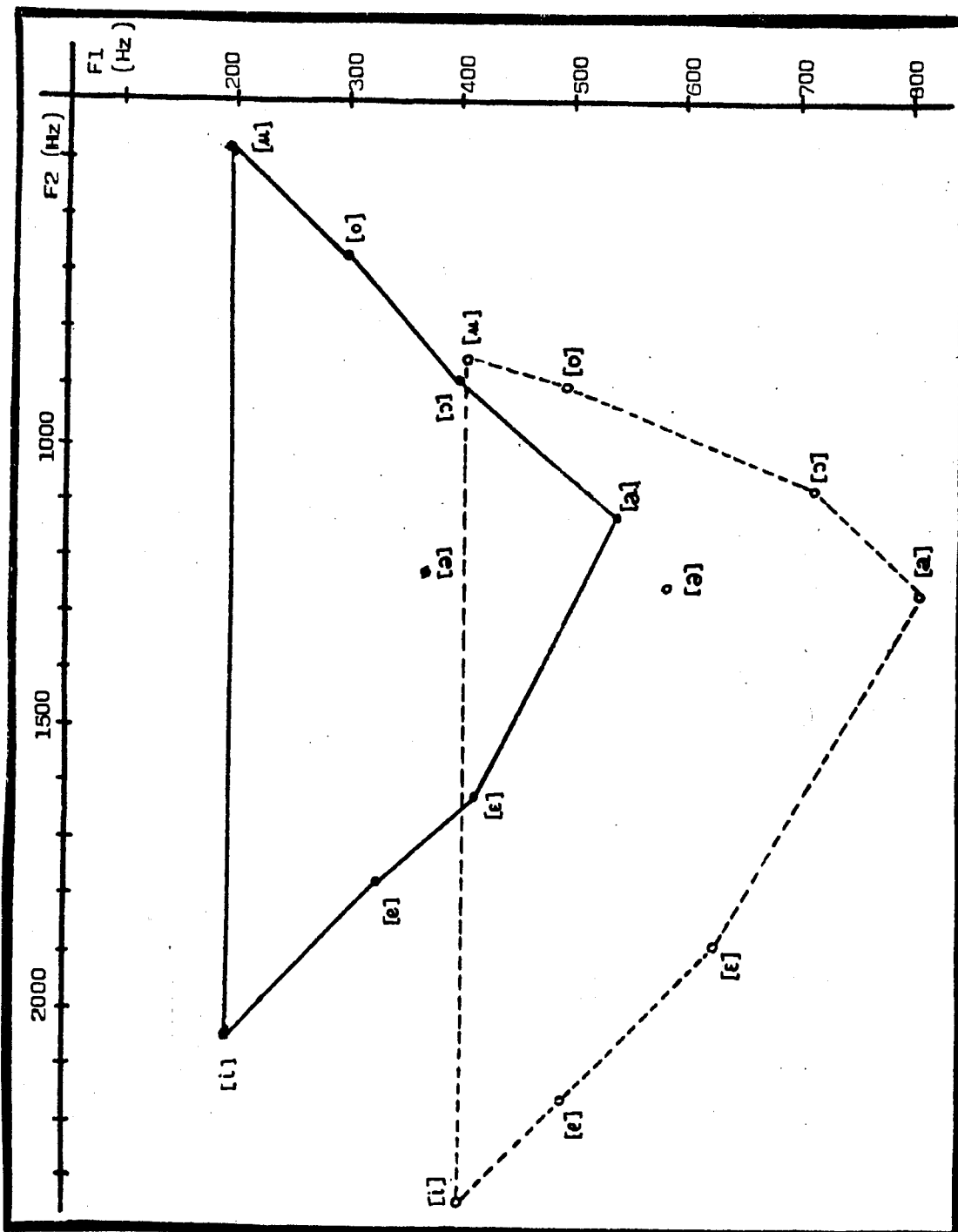


Fig. 3. Triangles vocàlics obtinguts a partir de mesures d'espectrogrames.
 — Cerdà (1972)
 - - - Informant MO

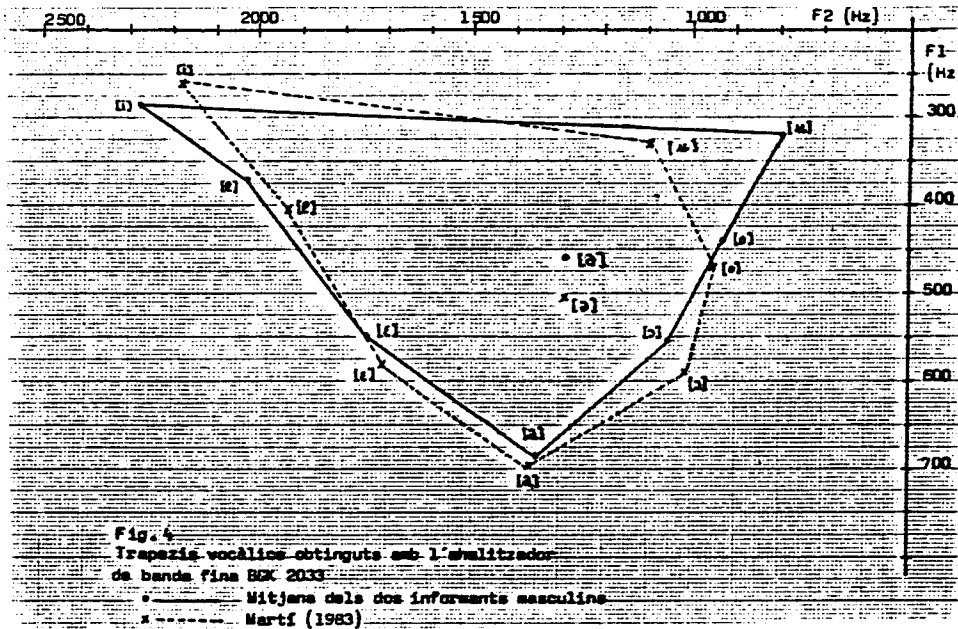
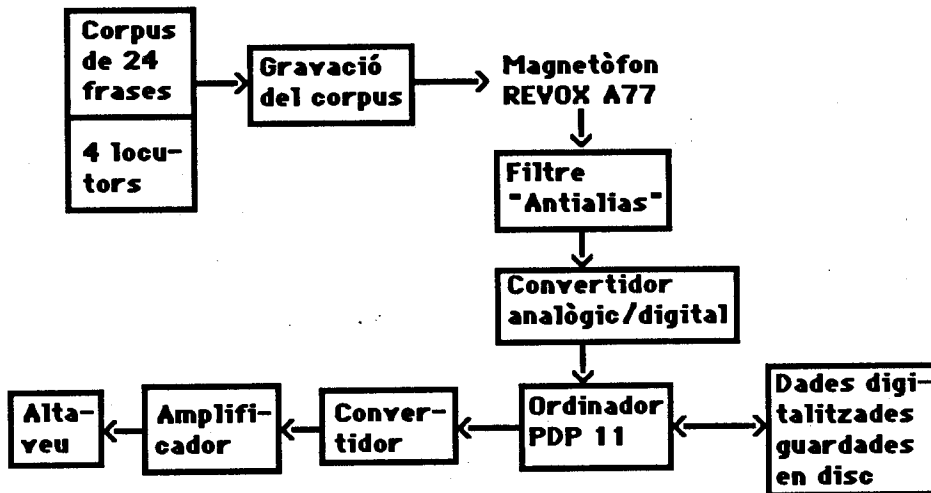


Fig. 4. Trapezis vocàlics obtinguts amb l'anàlitzador de banda fina B&K 2033
— Mitjana dels informants masculins
- - - Martí (1983)

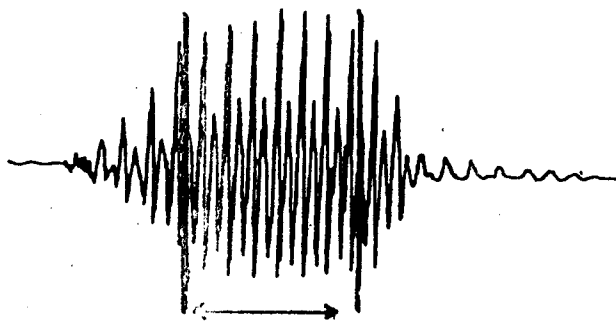
Fig. 5. Procediment de tractament del senyal acústic.



Programes

1. "DISK": localitza al disc el segment que es vol analitzar (un bloc de 0.6 segons), reproduït a l'altaveu.

2. "SIGPRA": delimita mitjançant dos cursors la part del segment que s'estudia.
Representació oscil·logràfica a la pantalla de la terminal:

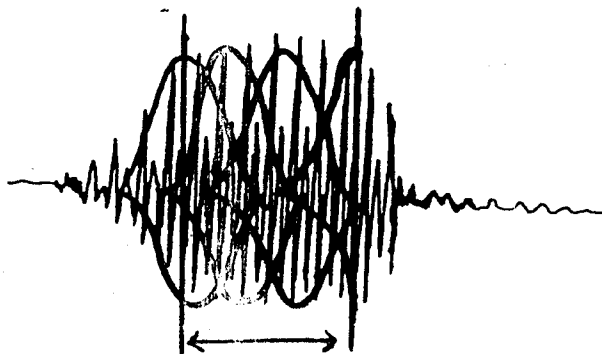


Cursors delimitant la part del segment que s'estudia.
Les dades dels cursors es guarden en un fitxer.

3. "SIGPRB": fa l'anàlisi per predicció lineal.

El programa parteix de les dades dels cursors de "SIGPRA", analitzant la part de segment seleccionada.

El senyal és multiplicat per una sèrie de finestres "Hanning" de 12.8 ms amb un 50% de superposició:



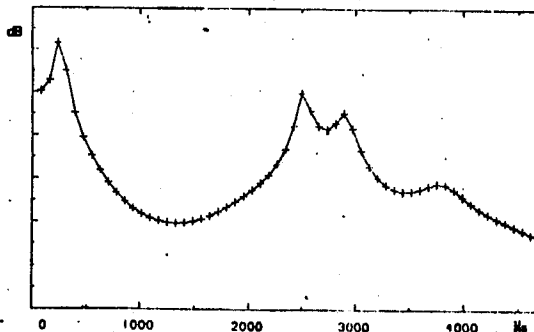
Cada finestra constitueix un bloc.

A continuació es porta a terme una anàlisi per predicció lineal a cada bloc amb un nombre de coeficients definit prèviament¹².

Les dades de tota la sèrie de blocs es guarden en un fitxer.

Dades del fitxer amb els coeficients de predicció lineal enviades a l'ordinador per a la resta del tractament

4. "SIGPRØ-2H": Reconstitueix l'espectre de cada bloc o finestra a partir dels coeficients de l'anàlisi per predicció lineal:



Les dades es guarden en un fitxer.

5. "PPMWH": Detecta els pics a partir de les dades de l'espectre obtingudes amb el programa anterior. El programa detecta automàticament els pics dels espectres i dóna:

(a) Un llistat de freqüències i amplituds de la banda dels formants presents en cada bloc

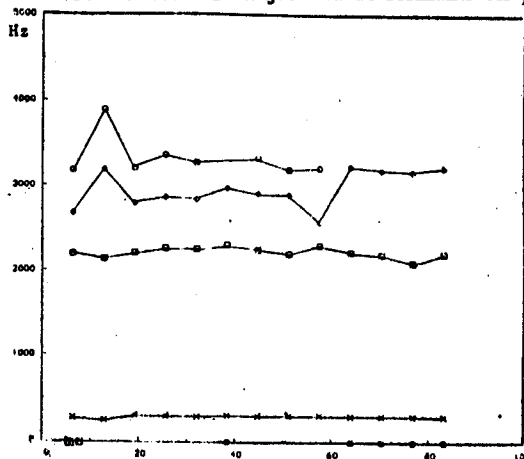
Bloc 1

274.0	118.4
2191.8	156.2
2675.3	765.5
3177.3	346.2
4053.1	237.3

Bloc 2

250.8	139.2	
2132.1	109.4	
3184.0	128.6	
3884.2	528.7	Hz

(b) Un "pseudo-espectrograma" reconstituït en forma de trajectòria de formants els pics detectats:



Taula III. Freqüències dels tres primers formants obtingudes per predicció lineal.

Primer Formant				
Vocal	EB, RMB 6 m./voc.		MO, JC 6 m./voc.	
	Mitjana	D. tít.	Mitjana	D. tít.
[i]	298	48.53	307	4.84
[e]	433.5	41.72	390.5	48.54
[ɛ]	638	98.81	555.16	23.26
[ə]	493.5	39.22	491.5	29.38
[a]	701.5	106.60	684	33.57
[ɔ]	582.5	117.82	581.33	49.23
[o]	454.66	31.72	448.16	28.55
[u]	341	52.09	317.83	15.83

Segon Formant				
Vocal	EB, RMB 6 m./voc.		MO, JC 6 m./voc.	
	Mitjana	D. tít.	Mitjana	D. tít.
[i]	2512.5	146.51	2272.33	91.94
[e]	2293.5	199.30	2137.5	188.23
[ɛ]	2155.5	214.73	1736.66	45.95
[ə]	1651	174.76	1853.16	109.66
[a]	1522.5	226.32	1366.16	107.06
[ɔ]	1244	96.54	1082.16	85.91
[o]	941.66	115.72	930	153.51
[u]	867.2	119.32	778.5	120.27

Tercer Formant				
Vocal	EB, RMB 6 m./voc.		MO, JC 6 m./voc.	
	Mitjana	D. tít.	Mitjana	D. tít.
[i]	3077.33	296.26	2878.5	318.63
[e]	2934.5	181.34	2627.5	217.87
[ɛ]	2912.5	101.71	2559.83	462.55
[ə]	2592.16	216.22	2311	195.36
[a]	2622.33	523.04	2481.16	256.25
[ɔ]	2673.66	192.01	2226.5	110.94
[o]	2679.16	307.76	2538.5	330.67
[u]	2585.2	189.62	2394.5	64.63

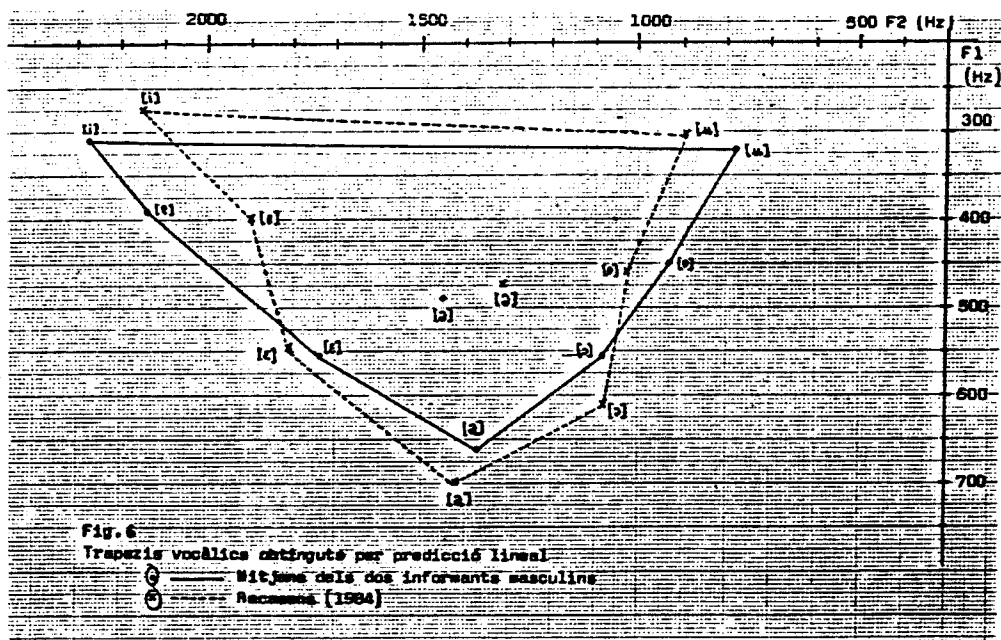


Fig. 6. Trapezis vocàlics obtinguts per predicció lineal.
 — Mitjana dels dos informants masculins
 - - - - Recasens (1984)

4. Síntesi de les vocals.

4.1. Procediment.

4.1.1. Descripció del sintetitzador.

El model de sintetitzador emprat en el present treball és l'OVE IIIId⁶ manufacturat per la casa sueca Fonema. Es tracta d'un sintetitzador de formants en sèrie, que tal com s'observa a l'esquema de la fig. 7 consta de tres branques: una per als sons vocàlics, una altra per als nasals, i una tercera per als fricatius. Els oclusius s'obtenen introduint un soroll aleatori d'una durada molt curta simulant l'explosió, un espai de silenci si cal, corresponent al VOT-MEL en la terminologia de Julià (1981)-, i reproduint les transicions amb variacions de la freqüència dels formants de la branca vocàlica (Llisterra-West, 1983).

Les dades arriben al sintetitzador a través d'un ordinador -en el present estudi fou emprat un PDP 11- per mitjà d'un fitxer de paràmetres, procediment que permet de sintetitzar des de sons aïllats fins a síl·labes o frases senceres. Un cop arriben al sintetitzador, els senyals digitals enviats per l'ordinador es guarden en una memòria intermèdia des d'on passen a controlar els circuits analògics que generen els impulsos necessaris perquè un amplificador i un altaveu reproduïxin el so. A la fig. 8 s'esquematitza aquest procés.

L'OVE III d opera amb quinze paràmetres, relacionats principalment amb la freqüència i l'amplitud. Cada paràmetre té un valor màxim i un mínim, que poden alterar-se en increments del 3 % en els de freqüència i de 2 en 2 dB en els d'intensitat. Per a la síntesi de vocals s'han utilitzat els següents:

(a) amplitud de la sonoritat: controla el nivell de la font d'excitació, que determina la intensitat de la vocal.

(b) freqüència del fonamental (F_0)

(c) freqüència dels tres primers formants vocàlics (F_1 , F_2 i F_3)

(d) increment de l'amplitud de banda dels tres primers formants (B_1 , B_2 i B_3).

F_4 i F_5 venen ja fixos a 3500 Hz i 4000 Hz respectivament; a causa de la disposició en sèrie, el nivell dels formants s'ajusta automàticament en funció de la seva freqüència.

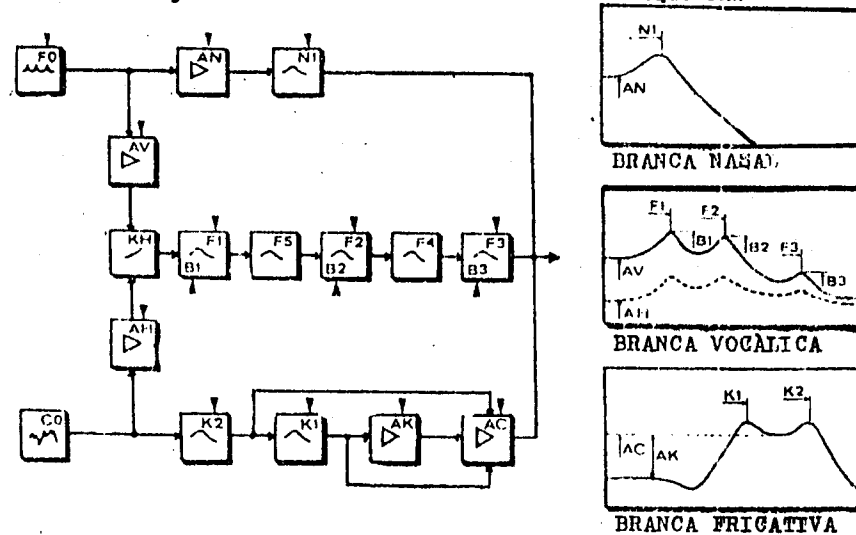


Fig. 7. Esquema del sintetitzador OVE III d emprat en aquest treball. Extret del Manual d'utilització.

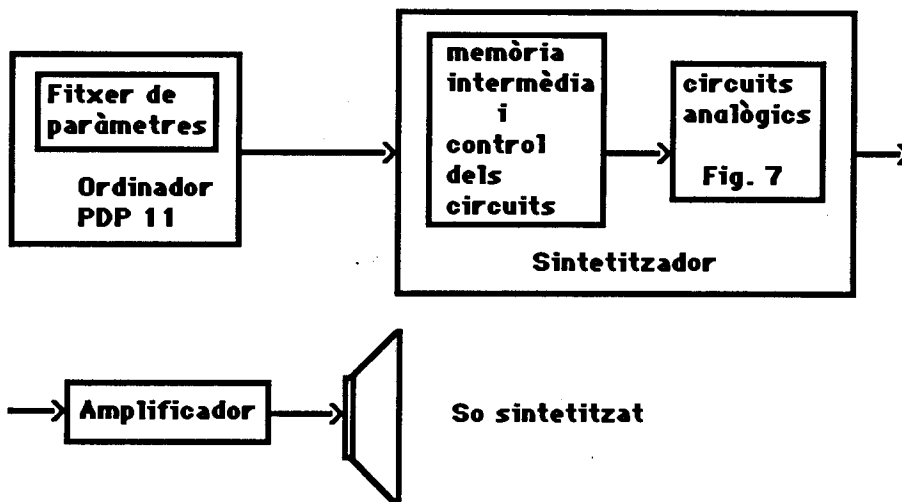


Fig. 8. Diagrama simplificat del procés de control de la síntesi en l'OVE III d.

4.1.2. Les dades

Tal com s'ha vist, les dades corresponents als paràmetres de control s'introdueixen al sintetitzador per mitjà d'un fitxer de 15 paràmetres creat a la memòria de l'ordinador. El seu format correspon al de la fig. 9: a la columna de l'esquerra apareix el temps en relació al principi del so, i a la dreta els valors dels paràmetres -freqüències o amplituds- en un moment determinat. Els valors de l'exemple són en mm, ja que en principi el programa va dissenyar-se per acceptar dades extretes directament d'espectrogrames; l'equivalència en Hz i en ms es troba multiplicant per 44 i per 8.5 respectivament.

Per evitar la repetició d'informació constant, va crear-se un fitxer permanent amb una sèrie de paràmetres invariables, en el qual s'anava modificant els valors de F1, F2 i F3 a fi de sintetitzar les diverses vocals. Van mantenir-se fixos:

(a) l'amplitud de la sonoritat, creixent de 5 a 20 dB des del principi fins al centre de la vocal, i després decreixent, imitant la distribució de l'energia en una vocal natural.

(b) la freqüència del fonamental a 100 Hz, o a la freqüència mitjana de cada locutor.

(c) la freqüència del formant nasal a 250 Hz, que no altera la qualitat oral de la vocal.

(d) els increments de les amplituds de banda dels formants: 60 Hz per a F1, 100 Hz per a F2 i 120 Hz per a F3.

	temps	freqüència en mm x 10	
AV	1	0	
	3	5	Amplitud de la
	5	10	sonoritat
	7	15	
	9	20	
	11	25	
	245	25	
	247	20	
	249	15	
	251	10	
253	5		
255	0		
999	9999		
AH	1	0	Amplitud de
	255	0	l'aspiració
999	9999		
AN	1	0	Amplitud de la
	255	0	nasalitat
999	9999		
Fo	1	22	Freqüència del (fonamental)
	256	22	
999	9999		
F1	1	44	Freqüència del
	256	44	primer formant
999	9999		
F2	1	132	Freqüència del
	255	459	segon formant
999	9999		
F3	1	264	Freqüència del
	255	694	tercer formant
999	9999		
N1	1	56	Freqüència del
	255	56	formant nasal
999	9999		
B1	1	13	Amplitud de banda del
	255	12	primer formant
999	9999		
B2	1	22	Amplitud de banda del
	225	22	segon formant
999	9999		
B3	1	28	Amplitud de banda
	255	24	del tercer formant
999	9999		
AG	1	0	Amplitud de l'energia
	255	0	fricativa a les freqüències altes
999	9999		
K1	1	0	Freqüència de l'energia fricativa
	255	0	a les freqüències baixes: primer formant
999	9999		
K2	1	0	Freqüència de l'energia fricativa
	255	0	a les freqüències altes: segon formant
999	9999		
AC	1	0	Ampli-
	255	0	formant fricatiu
999	9999		

Fig. 9. Exemple de paràmetres i fitxer de dades emprats per a la síntesi de vocals.

Per aquest procediment foren sintetitzades 48 vocals, corresponents a les dades extretes d'espectrogrames i a les obtingudes amitjanant els resultats de l'anàlisi per predicció lineal de les tres produccions de la mateixa vocal en els tres contextos diferents per a cada locutor.

Cal aclarir que sobre les dades inicials s'han operat dos reajustaments. El primer ve donat per la pròpia dinàmica del sintetitzador, que només accepta increments freqüencials del 3 %, i això implica que s'ha agafat el valor més proper al real dins de les possibilitats de l'aparell. El segon és que en fer la mitjana resultant dels tres contextos diferents en que fou pronunciada cada vocal s'han eliminat els valors molt desviats, clarament producte d'un error de mesura.

Finalment, també es sintetitzaren estímuls a partir dels valors amitjanats per als quatre locutors, descartant de nou aquells que presenten una desviació anormal respecte del conjunt. El total de vocals sintetitzades s'eleva així a 56.

4.2. Resultats.

4.2.1. Resultats obtinguts amb les dades inicials.

Començarem aquest apartat exposant els resultats de la síntesi a partir dels valors mesurats en els espectrogrames. Van donar bons resultats -és a dir, reproduccions acurades de la vocal original- les dades emprades per a [e] en els dos locutors, i les mesurades a les produccions de [a] de MO, i de [ɔ] i [u] de JC. Altres vocals van obtenir-se amb un timbre aproximat però amb algunes distorsions, com ara [ɛ] i [ɔ] dels dos locutors, mentre que d'altres foren clarament diferents de les inicials.

En conjunt, pot apreciar-se la dificultat de sintetitzar partint de mesures obtingudes amb el Sona-Graph, fet que confirma les limitacions de l'aparell a les quals ja s'ha al·ludit.

Pel que fa als resultats obtinguts emprant com a paràmetres els valors trobats pel programa de detecció automàtica de formants, després de l'anàlisi per predicció lineal, cal destacar d'entrada que foren sensiblement millors que les anteriors.

En conjunt, les aproximacions més bones a [i] van obtenir-se amb les freqüències mesurades a les produccions de MO i JC. La síntesi de [e] i [ɛ] no plantejà problemes, però en canvi la síntesi de la vocal neutra fou ja més difícil. La millor aproximació fou deguda als valors mesurats a RMB i JC, mentre s'obtenia una vocal labialitzada amb les dades de MO i EB. Cap informant no va fornir dades que permetessin d'arribar a bones reproduccions de la vocal [a], i les millors s'obtingueren amb els valors mesurats per MO, l'informant que presenta valors relativament més baixos.

De nou, les dades per a [ɔ] que donaren una aproximació més identificable foren les de MO; en els altres casos es constatà un desplaçament cap a [a] probablement ocasionat per un F2 elevat. La síntesi de [o] no presentà dificultats especials i també s'aconseguien bones reproduccions sintetitzades de [u] amb les dades de MO, JC i EB.

Treballant amb les mitjanes dels quatre locutors la vocal [i] fou correctament reproduïda, i també ho fou [e], la qual ja hem vist que no presentava problemes. El mateix pot dir-se de [ɛ] i [ɐ], més difícils de sintetitzar amb els valors de cada informant. El timbre de [a] fou bo d'entrada, però s'observà una lleugera tendència cap a [ɔ], fenomen repetit a [ɔ], encara que en aquesta vocal el resultat fou sensiblement pitjor. [o], en canvi, proporcionà una reproducció acceptable. El mateix pot afirmar-se de [u], que no presentà especials dificultats.

Cal recordar, a l'hora d'avaluar aquests resultats, que en el càlcul de les mitjanes foren descartats els valors que presentaven desviacions prou altes com per pensar que eren producte d'una confusió entre formants.

4.2.2. Modificació dels paràmetres i valors finals.

L'objectiu principal del present treball és, com s'ha dit al primer apartat, arribar, partint de dades reals, a uns valors de la freqüència dels tres primers formants que es puguin emprar per a produir reproduccions sintetitzades de les vocals catalanes. Per tal d'assolir-lo, un cop mesurades les produccions dels locutors i comprovats els resultats que aquestes mesures donen en la síntesi, cal procedir a modificar els paràmetres que ho requireixin. Aquesta tècnica ha estat aplicada només als valors de les mitjanes.

En alguns casos -[i], [e], [ɛ], [ɐ], i [u]- no va caldre modificar les dades inicials perquè va arribar-se ja d'entrada a un resultat satisfactori. En d'altres, les variacions foren mínimes, com en el cas de [a] on la reducció del valor de F2 contribuï a una millora del timbre de la vocal. En canvi, per a la síntesi de [ɔ] i [o] va caldre més modificacions: per a [o] F1 s'augmentà molt lleugerament i F2 fou reduït a 845 Hz, allunyant així la vocal de la zona de les vocals centrals. Per a [ɔ], es va recórrer a les dades mesurades en el context [p_p] del locutor MO, que van oferir una reproducció perfectament identificable.

A la taula IV es donen els valors, considerats definitius, dels paràmetres que permeten d'obtenir les millors aproximacions sintetitzades a les produccions dels informants, corregides en funció dels valors reals del sintetitzador.

4.2.3. Discussió.

Un cop obtinguts aquests resultats, resten encara per comentar alguns aspectes del procés de la síntesi que fins ara no hem tractat. Ens referirem en primer lloc a les diferències entre les dades obtingudes de cadascun dels quatre locutors.

D'entrada, els millors resultats s'aconsegueixen amb dades de locutors masculins amb un fonamental baix. En termes quantitius, sobre el total d'estímuls sintetitzats per cada locutor, poden considerar-se aproximacions raonablement bones a les originals el 62.5 % de les vocals obtingudes a partir de les dades de MO, el 50 % de les de JC, i el 12.5 % de les de RMB i EB. Es desprèn d'aquí la necessitat de cara a la síntesi de disposar de locutors amb un fonamental de freqüència baixa, fet que no s'ha d'atribuir al procediment de síntesi mateix, sinó a la insuficiència dels instruments d'anàlisi per al tractament de la veu femenina.

Taula IV. Valors amitjanats (4 locutors) i corregits de F1, F2 i F3 que van donar les millors reproduccions sintetitzades de les vocals estudiades.

Vocal	F1(Hz)	F2(Hz)	F3(Hz)
[i]	291	2397	2934
[e]	412	2075	2540
[ɛ]	550	1745	2329
[ɔ]	490	1510	2468
[a]	692	1307	2468
[ɔ]	617	924	2136
[o]	462	848	2397
[u]	300	800	2468

Hem de preguntar-nos, doncs, per la validesa de les mitjanes trobades analitzant veu femenina i veu masculina, i pensar si no són valors en certa manera ficticis, ja que en principi no corresponen a la realitat de cap de les vocals pronunciades. De fet, corresponen a realitzacions vocàliques que entren en el camp de dispersió delimitat a partir de tots els valors mesurats en el corpus -fig. 10-, i que posteriorment han estat correctament identificades. Notem també a la fig. 10 que els valors situats a la zona més extrema del camp de dispersió són els de la sèrie posterior, caracteritzada com la més problemàtica.

En relació amb la qüestió de la freqüència fonamental, van fer-se diverses proves alterant aquest paràmetre. Finalment, va emprar-se un Fo estàndard de 100 Hz per a tots els estímuls, tot i que hem comprovat posteriorment que un valor de 120 o 130 Hz contribueix a augmentar la seva naturalitat.

Cal assenyalar en aquest punt que l'avaluació dels resultats de la síntesi exposats fins aquí ha estat feta pel mateix experimentador. C.G. Fant ha escrit a propòsit dels perills d'aquesta situació:

"An important conditioning factor when the subject is the designer of the experiment is that he hears what he expects to hear and not much else" (Fant, 1968: 254)

Per tant, s'imposà la realització d'un test de percepció amb altres subjectes validant els estímuls sintetitzats considerats provisionalment com les millors aproximacions a les vocals dels locutors.

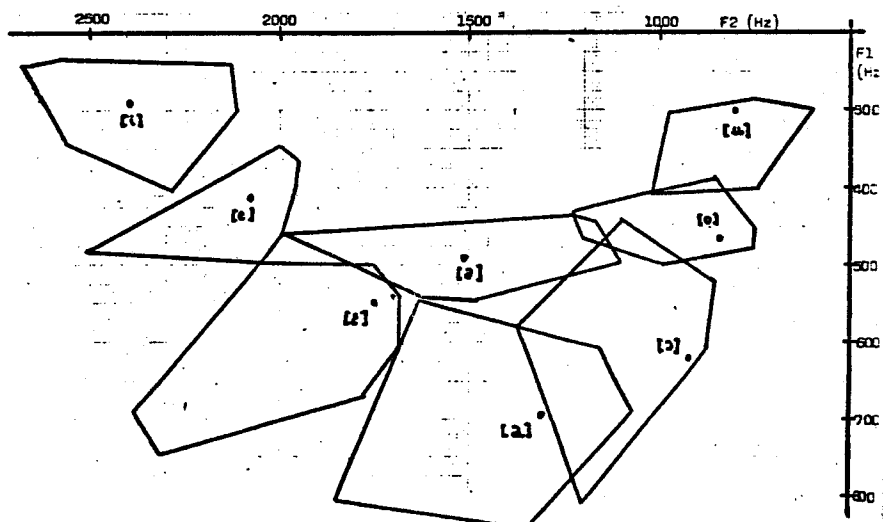


Fig. 10. Situació en el pla F1- F2 dels valors amittjanats seleccionats com les millors aproximacions a les vocals estudiades en relació amb el camp de dispersió delimitat a l'anàlisi.

5. Percepció de les vocals sintetitzades.

Cal començar indicant una limitació, i és que no s'intenta aquí de fer un estudi complet de la percepció de les vocals del català variant sistemàticament la freqüència dels formants tal com el van portar a terme per a l'anglès l'equip dels Haskins als anys 50 (Delattre, 1950; Delattre & *al.*, 1952) -o com més recentment ho han fet per al francès a partir de filtratges A. Landercy i R. Renard (1974, 1975)-. L'objectiu present és senzillament de validar els estímuls sintetitzats, per no caure en el parany anunciat per Fant.

5.1. Preparació i realització del test de percepció.

Per preparar el test va emprar-se una gravació de les vocals sintetitzades amb les freqüències de la taula IV i normalitzades pel que fa a la durada i al fonamental. L'ordre de presentació dels estímuls fou el següent:

- una sèrie de 16 estímuls on cadascun es repeteix dues vegades, d'anterior a posterior.
- una sèrie de 7 estímuls en ordre aleatori, i
- una sèrie de 8 estímuls en ordre també aleatori.

Cada estímül es repeteix, doncs, 4 vegades, excepte [a] que no fou introduït en la segona sèrie. La gravació va fer-se directament del sintetitzador a un Revox A77.

Fou preparat un full de respostes amb instruccions i espai per a les dades dels informants: llengua materna, lloc de naixement, lloc de residència i edat. A les instruccions va especificar-se de quina mena d'estímul es tractava, i com és habitual, es demanava de no deixar cap resposta en blanc. En les respostes, els informants havien d'encerclear el símbol de l'AFI corresponent a la vocal que sentien, símbol que podia comprovar-se en una llista de mots-clau que encapçalava el full.

El test s'administrà a un grup d'estudiants de 1er curs de l'Escola Universitària de Traductors i Intèrprets de la U.A.B. (Bellaterra) a l'aula on normalment es donen les classes. S'emprà un magnetòfon Tandberg i un altaveu. En total van contestar 33 informants de llengua materna catalana, nascuts i residents dins del domini del català central. La mitjana d'edat es situà entre els 18 i 19 anys.

5.2. Resultats.

Com pot observar-se a la taula V i a la fig. 11, el percentatge total d'identificacions correctes per a cada vocal oscil·la entre el 60.6 % per a [i] i el 97.7 % per a [u], amb una mitjana del 88.87 % de transcripcions encertades. Les vocals que ofereixen més dificultats d'identificació són [i] i la vocal neutra, mentre que la més ben percebuda ha estat [u]. La resta de vocals es mou entre el 92.4 % i el 96.6 % d'identificacions correctes.

Taula V. Resultats totals per a cada vocal.

Vocal	nº repeticions	identificacions correctes	
		total	percentatge
[i]	4	80	80.6 %
[e]	4	122	92.4 %
[ɛ]	4	124	93.9 %
[ɐ]	4	112	84.8 %
[a]	3	97	94.6 %
[ɔ]	4	125	94.6 %
[o]	4	122	92.4 %
[u]	4	129	97.7 %
Mitjana d'identificacions correctes			88.87 %

L'ordre d'aparició no sembla haver afectat els resultats, si no és en una lleugera disminució dels encerts en la tercera aparició de la vocal, primera de les sèries en ordre aleatori.

La taula VI presenta una visió global en forma de matriu de confusió dels problemes d'identificació de cada vocal. La primera columna correspon a les vocals que apareixen en el test, i la segona a les transcrites pels informants, quantificades en tants per cent.

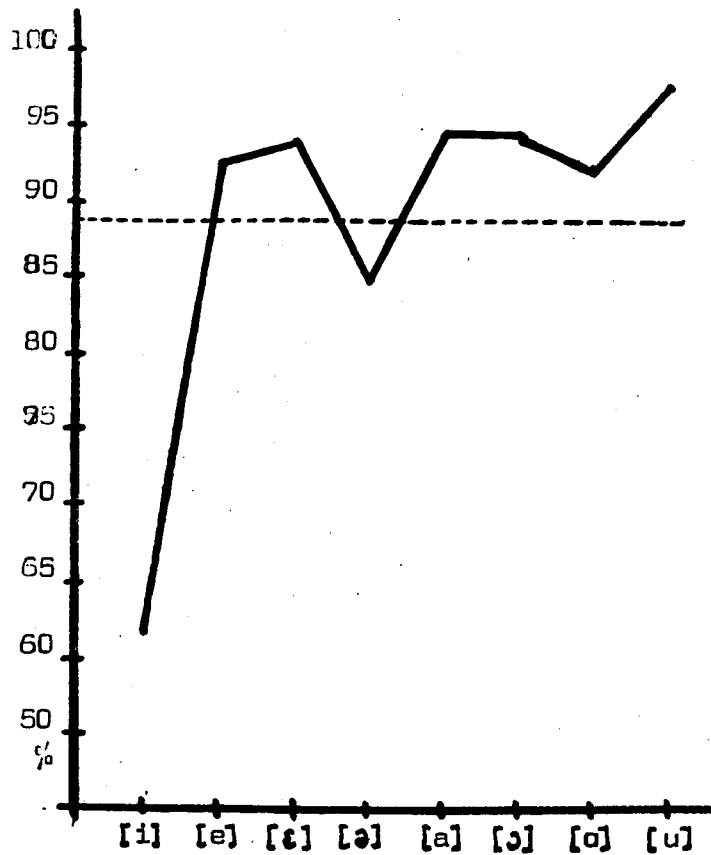


Fig. 11. Percentatge total d'identificacions correctes per a cada vocal sintetitzada.

Vocals del test	Respostes (%)							
	[i]	[e]	[ɛ]	[ə]	[a]	[ɔ]	[o]	[u]
[i]	60.6	34	4.5	0.75				
[e]	1.5	92.4	3.7	0.75	0.75			0.75
[ɛ]		4.5	93.9	0.75				
[ə]		1.5	0.75	84.8	0.75	2.2	7.5	2.2
[a]	1				94.6	1		
[ɔ]						94.6	3.7	0.75
[o]						2.2	92.4	5.3
[u]		0.75	0.75				0.75	97.7

Taula VI. Matriu de confusions per a cada vocal (resultats en percentatges).

Les confusions més importants es donen entre [i] i [e], que arriben a un 34 %, xifra no assolida per cap altra vocal. També hi ha vacil·lacions en la identificació del timbre de [e]-[e] i [o]-[ɔ], encara que molt menys importants. Cal remarcar l'elevat grau de diversitat de respostes pel que fa a la vocal neutra, confosa en major o menor grau amb totes les vocals del test excepte [i]. Exactament el cas contrari es dona amb [a], amb un 5 % només dels casos interpretada com a [ɔ], molt menys del que era de preveure donada la superposició dels camps de dispersió en l'anàlisi.

5.3. Discussió dels resultats.

5.3.1. Resultats generals.

En el moment d'avaluar els resultats generals del test, cal pensar que un 88.7 % d'identificacions correctes és una mitjana acceptable, considerant que es tracta d'estímuls sintetitzats. En una experiència semblant, Lehiste i Meltzer (1973) van obtenir un 75.23 % d'identificacions encertades amb vocals sintetitzades a partir de les mitjanes de Peterson-Barney (1952), i un 50.87 % emprant mesures derivades d'espectrogrames, percentatge que concorda amb les nostres proves en aquest sentit. Treballant amb vocals holandeses igualment sintetitzades, Cohen, Slist & 't Hart (1967) van arribar a un 87.3 % d'identificacions en un test de 12 vocals que es repetien 10 vegades cadascuna, passat a 167 subjectes. Val a dir que en el seu estudi la vocal amb un percentatge més baix de reconeixement fou [e] (58.7 %) seguida de [i] (75 %) i de [ɛ] (78.3 %).

En relació a la identificació de vocals aïllades no sintetitzades, podem comparar els nostres resultats amb els de Borzone (1979) sobre el castellà, que va obtenir un 97 % d'identificacions correctes; en els seus resultats de nou fou [i] la vocal amb un percentatge més alt d'error (5.6 %). Segons els treballs de Landercy-Renard ja citats, la mitjana de reconeixements de vocals franceses aïllades sense cap mena de filtratge no supera el 80 %.

5.3.2. Factors en la identificació de les vocals.

Els resultats globals als quals s'ha arribat s'expliquen, però, per una combinació de factors que incideixen sobre les diverses etapes de l'experiència, i que han estat ja àmpliament tractades a la bibliografia sobre el tema.⁷

En primer lloc, cal referir-se a una sèrie de factors que podem anomenar intrínsecs a l'estímul. En aquest aspecte, s'ha limitat el test a vocals aïllades, sense introduir-les en un context. En principi, sembla que aquest fet ha de contribuir a uns resultats més baixos, ja que es desprèn de diversos estudis que la identificació de vocals en contextos consonàntics millora respecte a la de les vocals aïllades (Ainsworth & Millar, 1972; Strange, Verbrugge & Shankweiler, 1974; Strange, Edman & Jenkins, 1979; Strange & Gottfried, 1980; Howell, 1981).

La causa del fenomen sembla raure en la informació aportada per les transicions (Shankweiler, Verbrugge & Studdert-Kennedy, 1978) tant pel que fa a la direcció (Lindblom & Studdert-Kennedy, 1967) com per la informació que aporten sobre les relacions temporals en el marc de la síl·laba, element bàsic en la identificació de la vocal segons Strange & Verbrugge (1976); Strange, Jenkins & Edman (1978); Gottfried & Strange (1978) i Strange, Jenkins & Johnson (1983)⁸.

Un segon factor que cal considerar és la freqüència fonamental dels estímuls sintetitzats, constant a 100 Hz en el nostre test, però que pot influir en concomitància amb altres elements (Miller, 1953) encara que no d'una manera decisiva (Ainsworth, 1971).

La durada s'ha mantingut constant a 250 ms, per la qual cosa aquest paràmetre tampoc no ha tingut un pes important en els resultats, i més considerant que el català no és una llengua que presenti oposicions de durada en el seu sistema vocàlic, i que per tant, la configuració espectral és l'índici predominant (Bennet, 1965; Ainsworth, 1972).

A més d'aquests elements, relacionats directament amb la forma de l'estímul, existeixen una sèrie de característiques del disseny del test que intervenen en els resultats obtinguts. El primer, i potser un dels més importants i més discutits, és la presència d'un marc de referència, sigui la veu del locutor que enuncia les vocals del test (Ladefoged, 1967; Verbrugge & al., 1976), una frase (Ainsworth, 1974), o bé la resta d'estímuls (Fry & al., 1962; Thompson & Hollien, 1970; i Ainsworth & Millar, 1971). Aquests últims autors conclouen que existeix un factor d'aprenentatge i de familiarització amb la parla sintetitzada, però que, superada una primera etapa, l'actuació dels subjectes es manté constant. En les nostres dades, pot observar-se una lleugera disminució d'encerts en la primera sèrie en ordre aleatori, que en molts casos no arriba a ésser significativa (fig. 12). Donat que la presència d'una frase sembla aportar informació de tipus temporal, però no freqüencial, no hem considerat aquesta variable en el test.

L'ordre d'aparició dels estímuls pot ésser rellevant tal com acabem de veure, i també arriba a tenir una influència sobre els resultats l'interval que els separa (Healy, Repp & Crowder, 1978). Aquest factor es relaciona amb la memòria auditiva i la capacitat d'establir comparacions entre les vocals anteriors i la que s'elicita en un moment del test.

Finalment, cal prendre en consideració la tasca de resposta que es demana dels subjectes. En el nostre cas va emprar-se la transcripció fonètica, donat que els informants eren estudiants de fonètica familiaritzats amb l'AFI; s'eviten així les interferències amb les representacions ortogràfiques, a les que s'ha reconegut un paper important a l'hora de computar els resultats (Assman, Nearey & Hogan, 1982).

Ja només queda referir-se als factors intrínsecs als informants. En una experiència de comparació d'estímuls vocàlics, Fox (1982) trobà diferències individuals a nivell de percepció que poden explicar-se en funció de diferències en les estructures perceptuals dels subjectes i que confirmen l'existència d'una relació entre el procés de producció i el de percepció. Els factors fonològics juguen també un paper en la identificació i discriminació dels estímuls (Repp, Healy & Crowden, 1979). En relació amb aquest aspecte, és un factor de primer ordre la llengua o la variant dialectal de l'informant, que pot donar lloc a fenòmens d'interferència fonètica constatats des de Polivanov (1931) i base dels treballs de Scholes (1967) (1968), Willis (1971) i Stevens & al. (1969), entre altres. Vam intentar mantenir el màxim d'homogeneïtat quant al lloc de naixement i residència dels nostres informants, parlants de la varietat central a la qual corresponen les produccions dels locutors que van gravar el corpus.

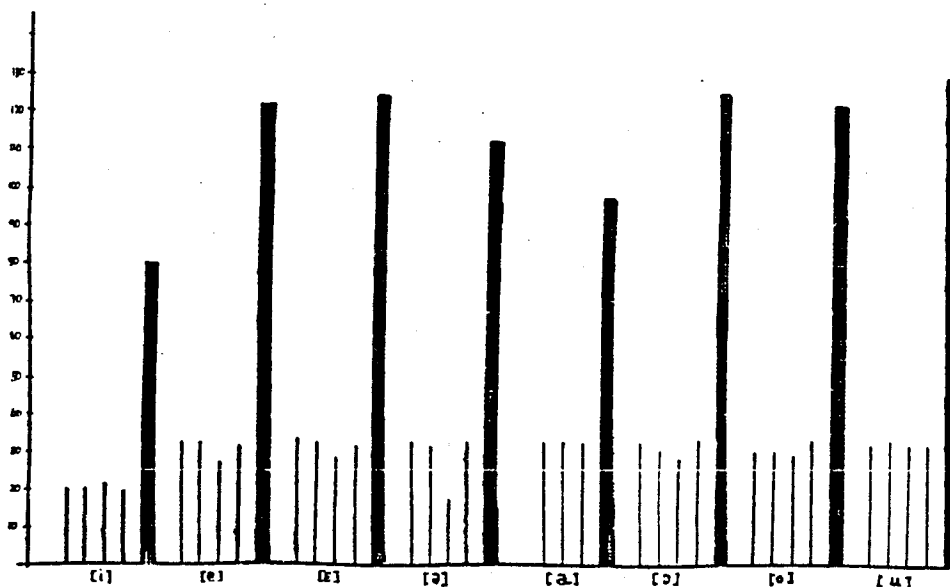


Fig. 12. Resultats totals i resultats per a cada aparició de la vocal.

Les línies es presenten per ordre d'aparició al test; la més gruixuda representa el total.

6. Conclusions i perspectives.

S'ha encetat aquest treball afirmant que el seu objectiu és d'arribar a tenir dades extretes de produccions reals que permetessin de sintetitzar adequadament les vocals del català. Els resultats del test de percepció efectuat amb estímuls sintetitzats emprant com a paràmetres les freqüències de la taula IV permeten de pensar que l'objectiu ha estat assolit, tot i que són necessàries modificacions en algunes de les dades, principalment en les de les vocals [i] i [e], que també hem vist que són les que més s'aparten dels resultats obtinguts per d'altres investigadors.

Cal insistir, però, que aquestes no són "les" freqüències de les vocals catalanes, sinó un conjunt de valors que han proporcionat bons resultats en un context ben precís. Per tal de delimitar el triangle vocàlic del català caldria treballar amb molts més informants i aplicar procediments de normalització. Un cert grau de coincidència entre les dades del present treball i les de Martí (1983) i Recasens (1984), enfront de les de Cerdà (1972) -Cf. fig. 13- fan pensar en la necessitat d'una revisió d'aquestes últimes, tal com ja s'ha assenyalat en l'article que citàvem al principi. Sembla, però, indefugible que les futures propostes en aquest sentit s'hauran de validar amb experiències sobre percepció emprant estímuls sintetitzats, que confirmen els resultats de l'anàlisi.

A més de la determinació precisa a nivell perceptual del camp de dispersió de les vocals catalanes, les dades sobre freqüència -completades amb mesures de durada, intensitat i amplitud de banda- tenen una utilitat immediata en un sistema de síntesi al·lofònica⁹ incorporable a un ordinador o un microprocessador, que proporciona una capacitat de resposta vocal il·limitada, alhora que constitueix la base d'un treball posterior en reconeixement de parla.

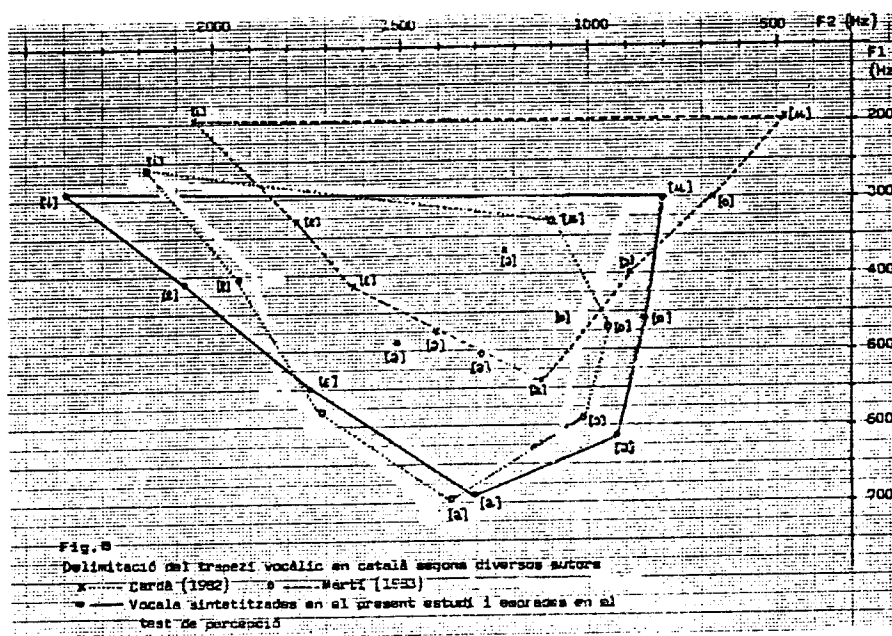


Fig. 13. Delimitació del trapezi vocàlic en català segons diversos autors.
 Cerdà (1982)
 - - - - Martí (1983)
 ——— Vocals sintetitzades en el present estudi i emprades en el test de percepció.

Notes.

- (1) Em baso principalment en l'últim capítol de Llisterra (1984). Haig de fer constar el meu agraïment al Dr. Julio Murillo (Dep. de Filologia Romànica, UAB), al Dr. Martin West (Dpt. of Applied Acoustics, Univ. of Salford, U.K.), que va posar a la meva disposició tot l'equip del seu laboratori i que és autor dels programes d'anàlisi i de síntesi utilitzats en el treball i al Dr. José-Manuel Bleuca, que em va permetre de fer els enregistraments al Laboratori de Fonètica de la UAB. Vull agrair igualment la col.laboració de la Dra. Dolors Poch (Escola Universitària de Traductors i Intèrprets, UAB), que es va encarregar de passar el test, i d'en Joan Hostau que va contribuir a la seva preparació i que supervisà les gravacions del corpus. Agraïxo també els comentaris del Dr. Daniel Recasens a una primera versió d'aquest article. Com sol dir-se, els errors que resten són del tot meus.
- (2) Un treball d'anàlisi d'un corpus de vocals catalanes emprant aquest aparell i aplicant el mètode de suavitzat de l'envolupant espectral per facilitar la detecció dels pics ha estat realitzat per J. Martí (1983). Part dels resultats es publica en aquest mateix volum (Martí, 1984). Haig d'agrair al professor Martí la comunicació personal de les dades d'aquest estudi amb antelació a la publicació.
- (3) Sobre la predicció lineal cf. Atal-Hanauer (1971) i també el llibre clàssic de Markel-Gray (1976) per a un tractament matemàtic detallat. Algunes aplicacions en síntesi es presenten a Cater (1983) i Poulton (1983).
- (4) En la seva tesi doctoral, D. Recasens (1984) presenta dades sobre la freqüència dels quatre primers formants de l'espectre de vocals tòniques i àtones del català en diferents contextos simètrics CVC i aïllades. L'autor (comunicació personal) considera que aquests valors no són encara prou definitius per a ésser reproduïts en el present treball.
- (5) El tema de la freqüència del fonamental intrínseca de les vocals ha generat una gran quantitat de literatura revisada a Petersen (1978), Ewan (1979), Gandour-Weinberg (1980) i Rossi & Autesserre (1981). Les explicacions es divideixen entre: (a) les de tipus acústic basades en un acoblament entre el tracte vocal i la font glotal (Flanagan & Landgraf (1968)); (b) les de base aerodinàmica que proposen un augment de la impedància del tracte vocal i un augment de la pressió subglòtica (Gandour & Weinberg (1980), Martí (1983); i (c) les articulatòries, fonamentades en l'increment de tensió de les cordes vocals a causa de l'elevació de la llengua, amb intervenció de l'hioides i moviment de la laringe (Ohala & Eukel (1975) i Lehiste (1970) entre altres). Cal afegir que la correlació entre la freqüència del fonamental i la del primer formant es troba ja esmentada a Lehiste & Peterson (1961). Vegi's també Rossi & al. (1981).
- (6) Descrit tècnicament a Liljencrants (1968). Trec moltes de les informacions que segueixen del manual d'utilització.

- (7) Una visió global i una bibliografia molt àmplia dels treballs sobre síntesi i percepció es troba a Chafcouloff (1976).
- (8) Aquests estudis han estat recentment contestats per Macchi (1980) i Diehl, Mc Custer & Chapman (1981), que no troben en les seves experiències un augment substancial dels percentatges d'identificacions correctes de les vocals aïllades quan s'inclouen en un context consonàntic.
- (9) A més del clàssic treball de Liberman & al. (1959), poden veure's el panorama general presentat per Mattingly (1974) i els estudis recollits a Flanagan & Rabiner (eds.) (1973); com a treball recent és interessant Hertz (1982).

Bibliografia.

- AINSWORTH, W. A. (1971) "Perception of isolated synthesized vowels and h-d words as a function of fundamental frequency". *JASA* 49, 4: 1323-1324.
- AINSWORTH, W.A. (1972) "Duration as a cue in the recognition of synthetic vowels". *JASA* 50, 1: 648-651.
- AINSWORTH, W.A. (1974) "The influence of precursive sentences on the perception of synthesized vowels". *Language and Speech* 17, 2: 103-109.
- AINSWORTH, W.A. & MILLAR, J.B. (1971) "Methodology of experiments on the perception of synthesized vowels". *Language and Speech* 14, 3: 201-212.
- AINSWORTH, W.A. & MILLAR, J.B. (1971) "Identification of synthetic vowels and vowels in h-d context". *Acustica* 27: 278-282.
- ASSMAN, P.F.; NEAREY, T.N. & HOGAN, J.T. (1982) "Vowel identification: Orthographic, perceptual and acoustic aspects". *JASA* 71, 4: 975-989.
- ATAL, B.S. & HANAUER, S.L. (1971) "Speech analysis and synthesis by Lineal Prediction of the speech wave" *JASA* 50: 637-55; FLANAGAN, J.L. & RABINER, L.R. (eds.) (1973) : 270-288.
- BENNET, D.C. (1968) "Spectral form and duration as cues in the recognition of English and German Vowels". *Language and Speech* 11: 65-85.
- BORZONE, A.M. (1979) "On the recognition of isolated Spanish Vowels". HOLLIEN, H. & P. (eds.) *Current Issues in Phonetic Sciences*. Amsterdam: John Benjamins. Vol. II : 677-681.
- CATER, J.P. (1933) *Electronically speaking: Computer Speech Generation*. Indianapolis, Indiana: Howard W. Sams & Co.

- CERDA, R. (1972) *El timbre vocálico en catalán*. Madrid: CSIC.
- COHEN, A.; SLIS, I.H. & HART, J. (1967) "On tolerance and intolerance in vowel perception". *Phonetica* 16, 2: 65-70.
- CHAFCOULOFF, M. (1976) *Vingt cinq années de recherches en synthèse de la parole*. Paris: Ed. du CNRS.
- DELATTRE, P.C. (1950) "The use of pattern play-back in studies of vowel colour by synthesis". *JASA* 22, 5: 678.
- DELATTRE, P.C. & al. (1952) "An experimental study of the acoustic determinants of vowel colour. Observations on one- and two-formant vowels synthesized from spectrographic patterns". *Word* 8: 195-210; FRY, D.B. (ed.) (1976) : 221-237.
- DIEHL, R.L.; McCUSKER, S.B. & CHAPMAN, L. (1981) "Perceiving vowels in isolation and in consonant context". *JASA* 69, 1: 239-248.
- EVAN, W.G. (1979) "Can intrinsic vowel Fo be explained by source-tract coupling?". *JASA* 66, 2: 358-362.
- FANT, C.G. (1962) "Sound Spectrography". SOVIJARVI, A. & AALTO, P. (eds.) *Proceedings of the Fourth International Congress of Phonetic Sciences*. The Hague: Mouton: 14-31.
- FANT, C.G. (1968) "Analysis and synthesis of the speech processes". MALMBERG, B. (ed.) *Manual of Phonetics*. Amsterdam: North Holland: 173-277.
- FLANAGAN, J.L. (1957) "Estimates of the maximum precision necessary in quantizing certain "dimensions" of vowel sounds". *JASA* 29, 4; LEHISTE, I. (ed.) (1967): 324-325.
- FLANAGAN, J.L. & LANDGRAF, F. (1968) "Self-Oscillating Source for Vocal tract Synthesizers". *IEEE Transaction in Audio and Electroacoustics* AU-16, 1: 57-64; FLANAGAN, J.L. & RABINER, L.R. (eds.) (1973) : 140-147.
- FLANAGAN, J.L. & RABINER, L.R. (eds.) (1973) *Speech Synthesis*. Stroudsburg, Pennsylvania: Dowden Hutchinson & Ross Inc.
- FOX, R.A. (1982) "Individual variations in the perception-production link". *Phonetica* 39, 1: 1-22.
- FRY, D.B. & al. (1962) "The identification and discrimination of synthetic vowels". *Language and Speech* 5: 171-189; FRY, D.B. (ed) (1976): 238-257.
- FRY, D.B. (ed) (1976) *Acoustic Phonetics. A course of basic readings*. Cambridge: Cambridge University Press.

- GANDOUR, J. & WEINBERG, B. (1980) "On the relationship between height and fundamental frequency: evidence from esophageal speech". *Phonetica* 37, 5/6: 344-54.
- GOTTFRIED, T.-STRANGE, W. (1980) "Identification of coarticulated vowels". *JASA* 68, 6: 1626-1635.
- HEALY, A.F.-REPP, B. & CROWDER, R.G. (1978) "Effect of interstimulus interval on identification and discrimination of isolated vowels". *JASA* 63, 2: 53.
- HERTZ, S. (1982) "From text to speech with SRS". *JASA* 72, 4: 1155-1170.
- HOWELL, P. (1981) "Identification of vowels in and out of context". *JASA* 70, 2: 350-355.
- JAKOBSON, R. & WAUGH, L. (1979) *The sound shape of language*. Brighton: Harvester Press.
- JULIA, J. (1981) "Estudi contrastiu dels oclusius de l'anglès i del català. Un experiment acústic". *Estudi General* 1, II: 75-85.
- LADEFOGED, P. (1967) "The nature of vowel quality". *Three areas of Experimental Phonetics: 50-142*. London: Oxford University Press.
- LANDERCY, A. & RENARD, R. (1974) "Perception des voyelles françaises filtrées". *Revue de Phonétique Appliquée* 32: 11-32.
- LANDERCY, A. & RENARD, R. (1975) "Zones fréquentielles et reconnaissance de voyelles françaises". *Revue de Phonétique Appliquée* 33/34: 51-79.
- LEHISTE, I. (1970) *Suprasegmentals*. Cambridge Mass: MIT Press.
- LEHISTE, I. (1967) (ed) *Readings in Acoustic Phonetics*. Cambridge Mass: MIT Press.
- LEHISTE, I. & MELITZER, D. (1973) "Some basic considerations in the analysis of intonation". *JASA* 33, 4: 419-425.
- LIBERMAN & al. (1959) "Minimal rules for synthesizing speech". *JASA* 31: 1490-1499; FRY, D.B. (ed) (1976): 445-65; LEHISTE, I. (ed) (1967): 333-342.
- LILJENCRANTS, J.C.W.A. (1968) "The OVE III Speech Synthesizer". *IEEE Transactions on Audio & Electroacoustics* AU-16, 1: 137-140.
- LINDBLOM, B. (1962) "Accuracy and limitations of Sonagraphic measurements". SOVIJARVI, A.-AALTO, P. (eds) *Proceedings of the Fourth International Congress of Phonetic Sciences: 138-202*. The Hague: Mouton.

- LINDBLOM, B. & STUDDERT-KENNEDY, M. (1967) "On the role of formant transitions in vowel recognition". *JASA* 42: 830-843.
- LLISTERRI, J. & WEST, M. (1983) "Analysis of stop-vowel transitions in Catalan". 11ème Congrès International d'Acoustique, *Revue d'Acoustique (hors série)* 4: 279-282.
- LLISTERRI, J. (1984) *Anàlisi i síntesi de vocals catalanes* (Determinació experimental de la freqüència dels tres primers formants en un corpus de vocals en context realitzades per 4 parlants de català central.) Memòria de llicenciatura inèdita. Facultat de Lletres, Universitat Autònoma de Barcelona, Bellaterra.
- MACCHI, M.J. (1980) "Identification of vowels spoken in isolation versus vowels spoken in consonantal context". *JASA* 68: 1636-1642.
- MARKEL, J.D. & GRAY, A.H. Jr. (1976) *Linear Prediction of Speech*. Berlin: Springer.
- MARTI, J. (1983) *Estudi acústic de les vocals catalanes*. Memòria de llicenciatura no publicada. Universitat de València, Facultat de Ciències Físiques.
- MARTI, J. (1984) "Paràmetres vocàlics del català". (*Folia Phonetica* 1).
- MATTINGLY, I.G. (1974) "Speech synthesis for phonetic and phonological models". SEBEOK, T.A. (ed) *Current Trends in Linguistics Vol. 12: "Linguistics and Adjacent Arts and Sciences, part IV"*. The Hague Mouton: 2451-2487.
- MILLER, R.L. (1953) "Auditory tests with synthetic vowels". *JASA* 25, 1: 114-121.
- MONSEN, R.B. (1981) "Accuracy of formant frequency estimation by spectrograms and by linear prediction analysis". *JASA* 69 suppl. 1: 917.
- MURILLO, J. (1981) *El umbral de la fonologización de los sonidos agudos turbulentos del habla en francés y en español*. Tesi doctoral no publicada. Universitat Autònoma de Bellaterra, Facultat de Lletres. 2 Vols.
- OHALA, J. & EUKEL, B.W. (1976) "Explaining the intrinsic pitch of vowels". *JASA* 60, suppl. 1: 944.
- PETERSEN, N.R. (1978) "Intrinsic fundamental frequency of Danish vowels". *Journal of Phonetics* 6: 177-189.
- PETERSON, G.E. & BARNEY, H.L. (1952) "Control methods in a study of the vowels" *JASA* 24, 2: 175-184; FRY, D.B. (ed) (1976): 104-122; LEHISTE (ed) (1976): 118-127.
- POLIANOV, E. (1931) "La perception des sons d'une langue étrangère" *Travaux du Cercle Linguistique de Prague* 4: 79-96; "Le Cercle de Prague" *Change* 3 (1969): 111-114.

- POULTON, A.S. (1983) *Microcomputer Speech Synthesis and Recognition*. Wilmslow, Cheshire: Sigma Technical Press.
- RECASENS, D. (1981) "Futur dels estudis de fonètica experimental del català". *Els Marges* 21: 47-64.
- RECASENS, D. (1984) *Producció i coarticulació de vocals i consonants del català en el decurs*. Tesis doctoral inèdita. Barcelona: Universitat de Barcelona.
- REPP, B.; HEALEY, A.F. & CROWDER, R.G. (1979) "Categories and context in the perception of isolated steady-state vowels". *Journal of Experimental Psychology, Human Perception and Performance* 5: 129-145.
- ROSSI, M. & AUTESSERRE, D. (1981) "Movements of the hyoid and the larynx and the intrinsic frequency of vowels". *Journal of Phonetics* 9: 233-249.
- ROSSI, M. & al. (1981) *L'intonation. De l'acoustique à la semantique*. Paris: Klincksieck.
- SCHAFFER, R. & RABINER, L.R. (1975) "Digital representation of Speech signals" *Proceedings of the IEEE* 63, 4: 662-677.
- SCHOLES, R.J. (1967) "Phoneme categorization of synthetic vocalic stimuli by speakers of Japanese, Spanish, Persian and American". *Language and Speech* 10.
- SCHOLES, R.J. (1968) "Phonetic interference as a perceptual phenomenon". *Language and Speech* 11: 86-103.
- SHANKWEILER, D.; VERBRUGGE, R.R. & STUDDERT-KENNEDY, M. (1978) "Insufficiency of the target for vowel perception". *JASA* 63, 2: 84.
- STEVENS, K.N. et al (1969) "Crosslanguage study of vowel perception". *Language and Speech* 12: 1-23.
- STRANGE, W.; EDMAN, T & JENKINS, J.J. (1979) "Acoustic and phonological factors in vowel identification". *Journal of Experimental Psychology, Human Perception and Performance* 5: 643-656.
- STRANGE, W. & GOTTFRIED, T. (1980) "Tasks variables in the study of vowel perception". *JASA* 68: 1622-1625.
- STRANGE, W.; JENKINS, J.J. & JOHNSON, T.L. (1983) "Dynamic specification of coarticulated vowels". *JASA* 74, 3: 595-705.
- STRANGE, W.; VERBRUGGE, R.R. & SHANKWEILER, D. (1974) "Consonantal environment specifies vowel identity". *JASA* 55 suppl. 1: 554.
- STRANGE, W. & al. (1975) "Consonantal environment specifies vowel identity". *JASA* 60, 1: 213-224.

- THOMPSON, C.L. & HOLLIEN, H. (1970) "Some contextual effects on the perception of synthetic vowels". *Language and Speech* 13: 1-13.
- VERBRUGGE, R.R. & al. (1975) "What information enables a listener to map a talker's vowel space?". *JASA* 60: 198-212.
- WILLIS, C. (1971) "Synthetic vowel categorization and dialectology" *Language and Speech* 14: 213-218.

Abstract.

Approach to the synthesis of Catalan vowels.

This paper outlines the results of a first attempt at synthesizing Catalan vowels using frequency parameters derived from natural speech. Measurements taken with a Sona-Graph and a Bruel & Kjaer 2033 narrow band analyzer are compared to those obtained applying signal processing techniques and linear predictive analysis, and their accuracy for speech synthesis purposes is evaluated.

Stimuli imitating the eight Catalan vowels synthesized using slightly modified averaged frequency data from four speakers are tested. The result shows an average of 88.87 % of correct identifications, validating the parameters used.

A modification of Cerdà's (1972) vocalic triangle is suggested, and it is also proposed that the data obtained will be a first step towards a complete system of synthesis by rule.