

**PARÀMETRES ACÚSTICS PER A LA SÍNTESI
DE CONSONANTS FRICATIVES CATALANES**

Per Josep Martí i Roca

CAP DE DEPARTAMENT D'ACÚSTICA DE L'ESCOLA
UNIVERSITÀRIA DE TELECOMUNICACIÓ
LA SALLE BONANOVA

Barcelona, febrer de 1986

PARÀMETRES ACÚSTICS PER A LA SÍNTESE DE CONSONANTS FRICATIVES CATALANES

1. OBJECTIU EXPERIMENTAL

Aquest treball s'emmarca dins d'un estudi experimental més ampli de tots els sons consonàntics del català en vistes a la seva síntesi automàtica a partir de text ortogràfic i que s'està portant a terme en els Laboratoris d'Acústica de l'Escola Universitària de Telecomunicació La Salle Bonanova. No pretenem, ni molt menys, fer un treball exhaustiu en l'anàlisi dels sons fricatius del català. Centrem la nostra atenció preferentment en la banda freqüencial compresa entre 0 i 4.000 Hz. És evident que els sons fricatius tenen components freqüencials que ultrapassen aquesta banda, però no són indispensables per a una bona intel·ligibilitat, com ho demostra el fet que la banda telefònica, limitada també a 4.000 Hz, no impedeix la correcta identificació dels fonemes fricatius. En principi, farem l'anàlisi fins els 10.000 Hz per estudiar el límit superior de l'espectre, però fixarem l'atenció particularment en la informació compresa per sota dels 4 KHz.

Les unitats bàsiques per a la síntesi posterior seran les síl·labes consonant-vocal [CV] obtingudes a partir de les tres fricatives sordes [s, ʃ, f] i les dues sonores [z, ʒ] combinades amb totes les 8 vocals [i, e, ε, a, o, u, ə]. No hem inclòs la versió sonora [v] ja que ens hem limitat a informants del català central, poc habituats a la pronunciació d'aquest fonema. Els trets estacionaris de les vocals catalanes havien estat ja objecte d'un estudi nostre anterior (MARTÍ, J., 1984)

del que farem referència quan sigui necessari. L'opció per aquest tipus d'unitats bàsiques es fonamenta en la conveniència d'agilitzar el procés informàtic a l'hora de generar les paraules, amb l'estalvi del temps que caldria dedicar a la generació de les transicions entre fonemes, si treballéssim en síntesi per regles a partir dels fonemes com a unitats bàsiques.

D'acord amb els paràmetres obtinguts reconstruirem les transicions entre les consonants fricatives i totes les vocals mitjançant un sintetitzador de formants, concretament el model PHILIPS MEA 8000. Posteriorment submetrem els sons sintetitzats a la percepció auditiva humana que ens permetrà d'introduir noves modificacions de les unitats sintetitzades i ens quantificarà fins a quin punt s'han aconseguit els objectius proposats.

Sortosament disposem de molta informació referent a l'estudi de consonants fricatives en altres llengües, que poden ser un bon punt de referència en l'estudi de les fricatives catalanes. Són ja clàssics els estudis de DELATTRE, LIBERMAN, COOPER (1962) sobre les transicions i el «locus» de les fricatives de l'anglès americà, i els d'STEVENS, K. N. (1971). Com també l'anàlisi dels efectes coarticulatoris entre fricatives i vocals de YENI-KOMSHIAN, G. H. & SOLI, S. D. (1981); de SOLI, S. D. (1981) i de WHALEN, D. H. (1981). També s'han estudiat els efectes coarticulatoris entre consonants fricatives i oclusives: REPP, B. H. & MANN, V. A. (1981 i 1982); o les diferències entre fricatives i africades: HOWELL, P. & ROSEN, S. (1983).

Quant a les fricatives del castellà sudamericà, tenim els treballs experimentals de GURLEKIAN, J. A. (1981) i de BORZONE, A. M. & MASSONE, M. I. (1981), que donen una visió més propera dels sons fricativus del català. Concretament, pel cas dels fricativus catalans, analitzats conjuntament amb altres sons, disposem dels estudis de RECASENS, D. (1984).

2. CONDICIONS EXPERIMENTALS

Les consonants a estudiar són de dues categories: les sordes [s, f, ʃ] y les sonores [z, ʒ]. No estudiarem aquí la fricativa sonora [v],

ja que transcriurem sempre la grafia «v» per [b] tal com s'acostuma a fer en la parla normal del català central. L'estructura de les primeres està configurada per unes determinades bandes de soroll aleatori amb uns pics de ressonància característics que anomenarem formants fricativus o simplement formants. Tant les bandes com els pics són elements característics de cada una de les consonants. Les consonants fricatives sonores afegeixen, a més a més, uns components periòdics que ocupen la part baixa de l'espectre amb una distribució harmònica. En la generació d'aquest segon grup de consonants intervenen simultàniament una font sonora (les cordes vocals) i una font sorda (la constricció del conducte bucal que genera la turbulència de l'aire).

Les paraules seleccionades per a fer l'estudi d'aquests sons foren les següents:

Per a l'estudi de [si]: *sidra*
 » [se]: *sempre*
 » [sɛ]: *sèptim*
 » [sa]: *savi*
 » [sɔ]: *sol*
 » [so]: *sorra*
 » [su]: *suma*
 » [sə]: *sardana*

Per a l'estudi de [fi]: *xifra*
 » [fe]: *aixecat*
 » [fɛ]: *marxem*
 » [fa]: *marxar*
 » [fɔ]: *això*
 » [fo]: *moixó*
 » [fu]: *eixut*
 » [fə]: *marxen*

Per a l'estudi de [fi]: *físic*
 » [fe]: *feſta*
 » [fɛ]: *fèrtil*
 » [fa]: *fals*

- » [fɔ]: *foli*
- » [fɔ̃]: *fonda*
- » [fu]: *fusta*
- » [fə]: *ferida*

- Per a l'estudi de [zi]: *alzina*
- » [ze]: *desembre*
 - » [zɛ]: *zero*
 - » [za]: *resar*
 - » [zɔ]: *resol*
 - » [zo]: *divisor*
 - » [zu]: *Zúric*
 - » [zə]: *ase*

- Per a l'estudi de [ʒi]: *llegir*
- » [ʒe]: *indigest*
 - » [ʒɛ]: *la gerra*
 - » [ʒa]: *rajar*
 - » [ʒɔ]: *la jota*
 - » [ʒo]: *la jove*
 - » [ʒu]: *ajut*
 - » [ʒə]: *ajegut*

L'article «la» s'utilitza en algunes paraules per tal d'assegurar una pronunciació correcta del fonema que es pretén; com per exemple en el cas de [ʒ] al principi de paraula que tindria el perill de perdre la sonoritat i convertir-se en la corresponent versió sorda [ʃ]. Davant de «xifra» s'utilitza per evitar el so africacat [ç].

Totes les paraules foren pronunciades per 7 locutors de ràdio masculins en una sala tractada acústicament, i enregistrades sobre cinta magnètica. La tasca professional dels informants ens va assegurar una elocució molt acurada del «corpus» d'anàlisi. Tots ells van ser preparats amb una lectura prèvia del text i una invitació a que pronunciessin les paraules pausadament i amb un màxim de naturalitat.

Sobre els enregistraments magnetofònics s'han analitzat els següents paràmetres:

- Durada de la consonant des del moment que apareix la fricció fins a l'inici de la vocal.
- Valor freqüencial al qual s'inicia la transició dels formants F1, F2, F3 i F4 de la vocal següent.
- Durada de la transició del segon formant.
- Ressonàncies més importants de la part fricativa.
- Límits inferior i superior de la banda de soroll.
- Valor final de les transicions de primer i segon formant.

Hem pogut observar la gran importància que té delimitar degudament les resonàncies de la part fricativa per a una identificació correcta de la consonant. De fet hem experimentat la identificació dels diferents fonemes treballant fins a 4 KHz, per tant ens hem preocupat fonamentalment dels pics espectrals situats dins d'aquesta banda.

Resulta també de molta importància la transició del segon formant per a una correcta identificació de les fricatives en contacte amb les vocals. La transició del primer formant és sempre creixent, per tant contribueix poc a diferenciar els diferents fonemes fricatius. Quant al tercer formant, quasi mai presenta transicions importants. Els increments absoluts de freqüències en aquesta tercera transició són d'un ordre semblant als de la primera, però relativament representen variacions molt menys importants (fig. 9 i 10).

3. EL MÈTODE DE MESURA

El mètode concret d'estudi i de mesura s'ha bassat fonamentalment en una anàlisi espectral per transformada ràpida de Fourier (FFT) de banda fina i la corresponent evolució espectral en el decurs del temps. D'aquesta forma podem apreciar simultàniament els paràmetres temporals de la veu i els paràmetres espectrals; o més concretament: l'evolució temporal dels paràmetres espectrals. Donat que l'analitzador utilitzat (Brüel & Kjaer model 2033) opera sempre sobre un total de 1.024 mostres separades temporalment segons el fons d'escala freqüencial seleccionat, ens hem vist en la necessitat d'optar per un com-

promís entre la resolució temporal i la resolució espectral: en augmentar el fons d'escala, la resolució espectral es veu mimbrada mentre que la resolució temporal es veu afavorida. Aquí hem optat concretament per un fons d'escala de 10 KHz que comporta una resolució espectral de 25 Hz per línia i una finestra temporal d'anàlisi de 40 ms. Aquestes finestres de senyal temporal es veuen afectades per una ponderació «Hanning» i una juxtaposició del 75 % entre finestres successives, la qual cosa fa que la resolució temporal es pugui considerar pràcticament de 10 ms. La representació gràfica del mòdul de la transformada de Fourier s'ha fet en forma comprimida i juxtaposada cada 10 ms amb una impressora gràfica, la qual cosa ens dóna una representació (com la de la fig. 1) que ens recorda la mateixa tècnica de la representació espectrogràfica tradicional. En ordenades tenim l'escala temporal i en abscises l'escala freqüencial, mentre que el nivell d'ennergiment representa l'energia.

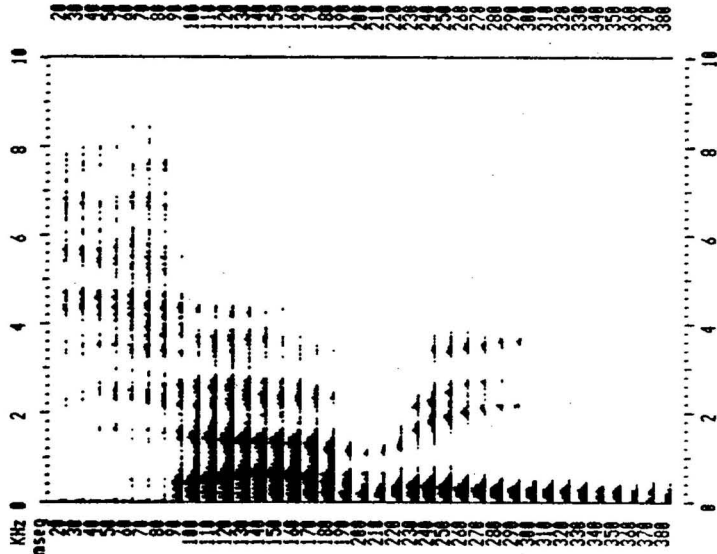


Fig. 1 Representació espectrogràfica de la paraula "savi".

4. CONSONANTS FRICATIVES SORDES

4.1. Paràmetres del fonema [f]

La consonant [f] correspon, sens dubte, al fonema fricatiu català menys definit. Precisament la seva característica més específica radica en la dispersió de l'espectre. Quasi bé no presenta ressonàncies particulars i el seu nivell global resulta molt més baix que el de la resta de consonants fricatives. És per això que, en l'estudi d'aquest so, ens hem vist impossibilitats de puntualitzar quines són les principals ressonàncies i la seva evolució al fer el contacte amb la vocal. Els formants de la vocal apareixen bruscament al seu valor estacionari just després de la zona fricativa (fig. 2), sense solució de continuïtat.

En tot cas l'únic que hem pogut observar és una lleugera tendència a preparar els formants F2 i F3 de la vocal següent amb una petita concentració d'energia al voltant dels seus valors respectius durant la part fricativa. De totes formes, el fenomen no es dona d'una

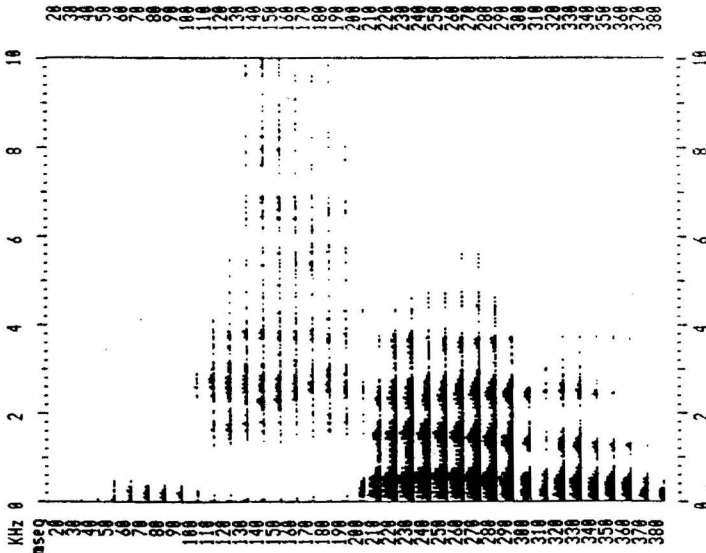


Fig. 2. Evolució espectral del difonema. [fɛ].

forma sistemàtica en totes les paraules analitzades. L'única transició que hem pogut observar és la que correspon al segon formant de [fe] i [fi] que és lleugerament creixent, degut al valor alt que de per si presenta el segon formant de les vocals [i, e]. Les altres transicions són quasi bé imperceptibles sobre els espectrogrames; però a partir de la percepció de la veu sintetitzada hem pogut observar com millora la naturalitat si es provoca una petita transició creixent a F1 i F2 durant uns 32 ms. El que sí hem pogut quantificar degudament és la durada de la zona fricativa amb els resultats amitjanats i les desviacions típiques (DT) de la taula I.

El valor amitjanat global de totes les durades és de 98 ms i no s'aprecien fluctuacions notables segons la vocal. En tot cas es podria insinuar una tendència a allargar el temps pel cas de les vocals més tancades. De totes formes, no es poden treure massa conclusions d'aquests resultats per l'excessiu grau de desviació que presenten. El que sí es pot afirmar és que la durada no és un tret gaire decisiu per a la definició d'aquest so.

Quant al nivell del soroll aleatori, hem observat que per a cada informant resulta molt constant i independent de la vocal següent. A partir de proves auditives de veu sintetitzada amb el MEA 8000 hem fixat un nivell de — 18 dB respecte de la part estacionària de la vocal [a]. Els nivells de les altres vocals s'han alterat d'acord amb uns criteris perceptius que asseguressin una sensació auditiva d'igual sonoritat, però el nivell corresponent a les respectives [f] s'ha mantingut sempre inalterat.

TAULA I. Estadística de la durada de la consonant [f] en contacte amb totes les vocals i DT.

	[i]	[e]	[ɛ]	[a]	[ɔ]	[o]	[u]	[ə]
Durada (ms)	101	100	106	95	90	97	103	95
DT	26	25	35	28	19	28	28	15

La distribució espectral de l'energia és molt uniforme i s'estén per igual entre uns 500 Hz i uns 8 o 10 KHz, sense concentracions apreciables d'energia en tota la banda.

TAULA II. Valors estacionaris dels dos primers formants de les vocals que segueixen a [f] i DT.

	[i]	[e]	[ɛ]	[a]	[ɔ]	[o]	[u]	[ə]
F1 (Hz)	300	433	617	775	633	520	340	575
DT	0	41	52	29	58	45	42	50
F2 (Hz)	2.233	2.067	1.767	1.250	1.090	1.075	750	1.575
DT	137	108	137	55	74	88	50	50

L'efecte coarticulatori d'aquesta consonant sobre les vocals següents és realment poc important com es pot veure a la taula II si la comparem amb els valors típics dels formants de les vocals (taula III), obtinguts en un estudi anterior (MARTÍ, J., 1984). La desviació més significativa és la de la vocal neutra amb un augment de 71 Hz per F1 i de 290 Hz per F2, i la del primer formant de [a] amb un augment de 80 Hz.

TAULA III. Valors típics dels dos primers formants de les vocals catalanes (en Hz) segons el nostre propi estudi (MARTÍ, J., 1984).

	[i]	[e]	[ɛ]	[a]	[ɔ]	[o]	[u]	[ə]
F1	258	405	581	695	587	471	338	505
F2	2.180	1.930	1.713	1.366	1.014	948	737	1.286

4.2. Paràmetres del fonema [s]

Aquest fonema, més energètic que [f] té un contingut espectral considerable fins a 10 KHz i està limitat per la part baixa al voltant dels 2 KHz. Presenta zones espectrals amb un major contingut energètic, particularment al voltant dels 4 KHz. A la taula IV es pot veure el valor amitjanat dels paràmetres que s'han quantificat i les desviacions típiques (DT) obtingudes a partir dels 7 informants.

TAULA IV. Paràmetres de la consonant [s] en contacte amb totes les vocals i DT.

	[i]	[e]	[ɛ]	[a]	[ɔ]	[o]	[u]	[ə]
Durada (ms)	115	112	114	118	111	114	143	108
DT	30	24	16	15	17	21	15	20
F2 de [s] (Hz)	1.842	1.742	1.808	1.785	1.664	1.578	1.502	1.645
DT	127	45	169	186	217	191	146	201
F3 de [s] (Hz)	2.628	2.728	2.625	2.476	2.514	2.554	2.500	2.592
DT	198	243	126	204	69	177	82	169
F4 de [s] (Hz)	3.728	3.833	3.775	3.606	3.508	3.371	3.483	3.650
DT	138	186	133	359	180	246	68	132
Transició F2 (ms)	90	85	88	69	68	62	76	65
DT	23	10	10	24	16	24	9	8

	[i]	[e]	[ɛ]	[a]	[ɔ]	[o]	[u]	[ə]
F1 vocal (Hz)	275	408	525	783	588	492	342	600
DT	27	49	106	82	25	20	49	115
F2 vocal (Hz)	2.200	1.917	1.750	1.400	1.042	1.083	858	1.350
DT	167	214	71	130	49	93	66	100

Cal advertir que la nomenclatura F2, F3, F4 que utilitzem aquí referida al soroll de les consonants fricatives és arbitrària. Els anomenem així perquè normalment marquen el «locus» dels corresponents formants F2, F3 i F4 de la vocal següent. No parlem en canvi del formant F1 ja que en la zona baixa de l'espectre no es detecta energia. D'altres autors prefereixen parlar directament de F1, F2 i F3 per referir-se al que nosaltres anomenem F2, F3 i F4.

Donem un valor del formant F2 d'[s], encara que el seu nivell resulta bastant baix; ho fem així ja que té particular importància quant a marcar l'inici de la transició del segon formant que acaba en el F2 de la vocal. Aquesta transició és important per millorar la qualitat dels difonemes (sV). Els formants F3 i F4 presenten uns nivells molt més alts però no aporten transicions significatives per a la identificació. Contribueixen més aviat a realçar la zona espectral més característica de la part estacionària d'[s]. A la figura 3 es pot veure un exemple característic amb la transició [su].

L'efecte coarticulatori de les vocals sobre la consonant s'accentua sobre tot a la pujada del F2 d'[s] seguida de les vocals anteriors, particularment pel cas de [si], [sɛ], i a la devallada del mateix formant per les vocals posteriors: [so], [sɔ], [su]. A l'inrevés: la consonant actua sobre les diferents vocals generalment augmentant els seus formants característics. Destaca el F1 de [a] amb un augment de 88 Hz i de [ə] amb un augment de 95 Hz, y també l'augment del F2 de [o] i [u] amb un increment de 135 Hz i 121 Hz respectivament.

La durada del soroll aleatori és molt uniforme si exceptuem el cas

de [su] amb una durada notablement més alta. El seu valor amitjanat total resulta de 117 ms., més alt que el de totes les altres consonants fricatives, tant sordes com sonores.

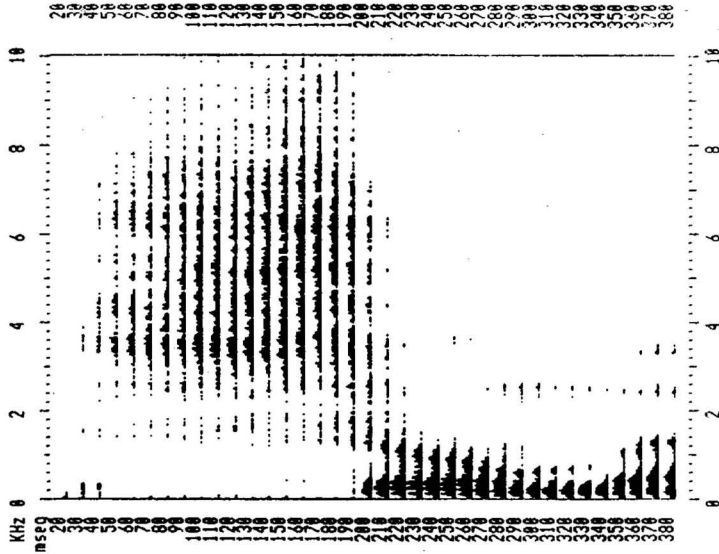


Fig. 3 Evolució espectral del difonema [su] .

El nivell global de la zona fricativa és bastant més alt que el corresponent a la consonant [f]. Sigui com sigui, en el cas d'una síntesi limitada a 4 KHz, com és el nostre cas, hem pogut observar que els nivells són molt semblants, ja que bona part de l'energia d'[s] queda més enllà dels 4 KHz, mentre que [f] estén el seu contingut energètic fins a les zones més baixes de l'espectre. A partir de proves de percepció hem pogut detectar nivells més alts (uns 3 dB) per la [s] davant de les vocals més obertes [a, ə] respecte de la resta de vocals. En general els nivells de la part fricativa s'han sintetitzat a 18 o 21 dB per sota del nivell corresponent a la part estacionària de les vocals.

En els registres espectrogràfics hem pogut observar una atenuació del nivell fricatiu abans d'entrar la vocal, durant uns 20 ms. A l'hora de fer la síntesi de cada difonema hem aplicat una atenuació de 9 dB en aquest punt, amb la qual cosa ha millorat la qualitat de l'emissió.

4.3. Paràmetres del fonema [ʃ]

Aquesta consonant presenta un espectre de soroll més desplaçat cap a les bandes baixes que la [s]. Les resonàncies més importants estan entre 2 KHz i 3 KHz, encara que l'espectre es perllonga aproximadament fins els 8 KHz. Dóna lloc a transicions molt marcades del segon formant, particularment per les vocals posteriors (Fig. 4).

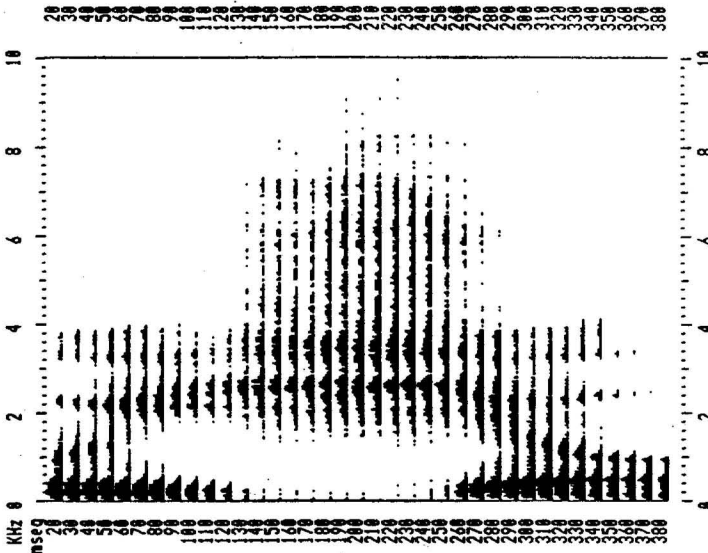


Fig. 4 Transició [ʃo] dins la paraula "moixó".

El nivell global és bastant semblant al d' [s]; ara bé, per una síntesi fins a 4 KHz el contingut energètic de [ʃ] s'ha de considerar més alt a causa de les ressonàncies de més baixa freqüència d'aquest últim fonema. La diferència de nivell s'ha estimat al voltant de 9 dB a partir de proves de percepció. No hem observat diferències de nivell segons la vocal següent, i per tant hem utilitzat sempre un mateix valor: exactament 9 dB per sota del nivell corresponent a la part estacionària d' [a]. També aquí es dóna una atenuació del soroll durant uns 20 ms abans de l'inici de la vocal, igual que en el cas d' [s]. A l'hora de fer la síntesi d'aquests difonemes hem aplicat una atenuació de 6 dB just en aquest punt.

TAULA V. *Paràmetres de [ʃ] en contacte amb totes les vocals i DT*

	[i]	[e]	[ɛ]	[a]	[ɔ]	[o]	[u]	[ə]
Durada (ms)	120	105	105	111	110	108	96	103
DT	18	23	13	12	18	15	23	19
F2 de [ʃ] (Hz)	2066	2092	1921	2025	1990	2055	2041	1941
DT	121	110	98	175	222	212	163	111
F3 de [ʃ] (Hz)	2810	2810	2683	2744	2600	2750	2435	2742
DT	162	179	133	180	89	225	63	215
F4 de [ʃ] (Hz)	3650	3800	3550	3650	3650	3620	3475	3550
DT	187	200	105	207	265	179	96	129
Transició F2 (ms)	78	74	91	105	97	107	86	78
DT	24	23	10	27	21	27	11	11
F1 vocal (Hz)	267	467	550	717	640	517	367	620
DT	52	58	71	52	55	41	41	125

	[i]	[e]	[ɛ]	[a]	[ɔ]	[o]	[u]	[ə]
F2 vocal (Hz)	2267	2100	1883	1442	1100	1050	1108	1580
DT	186	173	117	150	55	55	150	84

A la taula V podem observar que la durada màxima es dona abans del fonema [i] amb 120 ms i la mínima davant d'[u] amb 96 ms. El valor amitjanat per totes les vocals resulta de 107 ms. La transició del formant F2 té una durada considerable (entre 74 i 107 ms) i cal remarcar que s'inicia ja durant el temps de fricció, encara que la major part s'estén a la vocal següent.

En general les ressonàncies de [ʃ] es veuen poc afectades per la vocal següent, a excepció del cas de [ʃu] que té el formant F3 molt baix.

Tots els formants fricatius estan bastant difuminats, amb ressonàncies poc marcades, particularment pel cas del F4.

L'efecte coarticulatori sobre els formants vocàlics és molt marcat. Sistemàticament observem un augment de freqüència tant per F1 com per F2. L'increment més notable és el del segon formant d'[u] amb un increment de 371 Hz. Resulta també considerable l'increment del F1 i del F2 corresponents a la vocal neutra amb 115 i 294 Hz respectivament.

5. CONSONANTS FRICATIVES SONORES

5.1. Paràmetres del fonema [z]

Les consonants fricatives sonores es caracteritzen per una presència simultània de soroll aleatori generat per la turbulència de l'aire i d'un to periòdic produït per la vibració de les cordes vocals. En els espectrogrames es distingeixen clarament els dos components: a la part baixa de l'espectre els harmònics de baixa freqüència, i a la part alta el soroll comprès entre 1500 Hz i 8000 Hz aproximadament. (Fig. 5).

A la taula VI es dona la recopilació de tots els paràmetres quantificats amb els seus valors amitjanats i les desviacions típiques (DT).

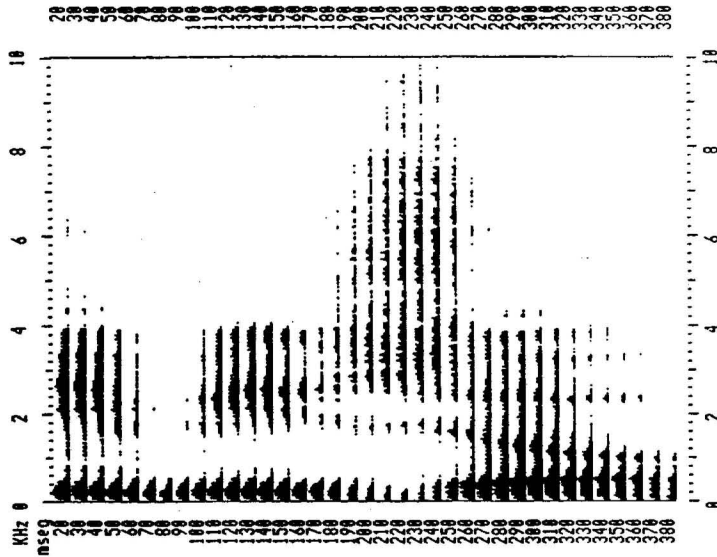


Fig. 5 Transició [zo] dins la paraula "divisor".

En aquesta taula destaca el valor baix del primr formant que s'és-tén únicament als primers harmònics de l'espectre. Aquest formant contribueix decisivament a donar el caràcter sonor, tot i que el seu nivell màxim és sempre bastant inferior al de les resonàncies fricatives. El cas de freqüència més baixa és el de [zə] amb 183 Hz, molt lluny de la mitjana general de 280 Hz.

El segon formant situat a prop dels 1600 Hz es veu poc afectat en la seva freqüència central pel corresponent F2 de la vocal següent. Únicament al final de la fricció s'observa ja l'inici de la transició, particularment pel cas de les vocals posteriors.

TAULA VI. *Paràmetres de la consonant [z] en contacte amb totes les vocals i DT.*

	[i]	[e]	[ɛ]	[a]	[ɔ]	[o]	[u]	[ə]
Durada (ms)	100	80	100	84	91	90	114	111
DT	21	13	12	8	13	14	21	16
F1 de [z] (Hz)	330	340	307	244	328	254	251	183
DT	69	45	73	73	91	46	76	41
F2 de [z] (Hz)	1528	1650	1642	1546	1464	1735	1583	1591
DT	191	112	73	90	94	85	160	102
F3 de [z] (Hz)	2590	2642	2716	2533	2365	2451	2700	2583
DT	158	230	147	140	47	76	163	143
F4 d [z] (Hz)	3583	3714	3840	3810	3740	3628	3766	3833
DT	286	227	167	336	89	214	151	137
Tranesció F2 (ms)	124	92	98	103	84	120	82	90
DT	31	14	30	8	10	28	15	15
F1 vocal (Hz)	290	463	540	720	625	520	340	540
DT	22	48	55	45	29	45	42	55
F2 vocal (Hz)	2200	1880	1780	1410	1050	980	920	1500
DT	71	110	84	143	35	91	110	0

El tercer formant fricatiu situat al voltant dels 2500-2600 Hz es veu poc afectat per la coarticulació i tampoc presenta transicions notables.

El quart formant entre 3700 i 3800 Hz es troba a la zona més realçada de l'espectre, tal com succeeix amb [s], i contribueix decisivament a la identificació del fonema.

Quant a la durada de la consonant, destaca el cas de [zu] amb 114 ms, mentre que per [ze] són únicament 80 ms. El valor amitjanat per totes les vocals resulta de 96 ms.

La transició del F2 és molt llarga, com succeeix amb tots els sons fricatiu. Quasi bé s'estén a tota la vocal següent i arranxa ja durant la part fricativa. Els temps fluctuen entre uns 85 ms i uns 120 ms amb un valor amitjanat de 99 ms.

L'efecte coarticulatori de [z] sobre els formants de les vocals següents resulta poc important, tret de la llarga transició del F2. Els valors finals concorden bastant bé amb els formants característics de les vocals. Solament cal remarcar un increment notable del F2 en el cas de les vocals [u, ə]: 183 i 214 Hz respectivament. Precisament aquestes dues vocals són les més afectades per modificacions coarticulatories en qualsevol tipus de transició CV.

Els nivells de la consonant no han estat quantificats a partir dels enregistraments, sinó que, com en la resta dels sons fricatiu, s'han deduït a partir de proves de percepció dins la banda de 0 — 4 KHz. Aquí s'han observat nivells molt uniformes de la part fricativa, independentment de la vocal següent. Únicament pel cas de [zu] ha calgut rebaixar 3 dB el nivell fricatiu. En general els nivells són molt semblants als d'[s]: entre 18 i 21 dB per sota de la part estacionària de la vocal. Altra cosa seria si la síntesi es fes incloent també l'energia situada més enllà dels 4 KHz.

A l'hora de sintetitzar aquest fonema amb el MEA 8000 ens hem trobat amb la dificultat que el sintetitzador no pot generar simultàniament sons periòdics i aperiòdics. El problema s'ha solventat introduint un fragment sonor de 40 ms al principi del fonema, amb les mateixes ressonàncies que la part fricativa i amb un nivell més baix (uns 6 dB menys). El sistema dona una percepció bastant correcta dels sons fricatiu sonors.

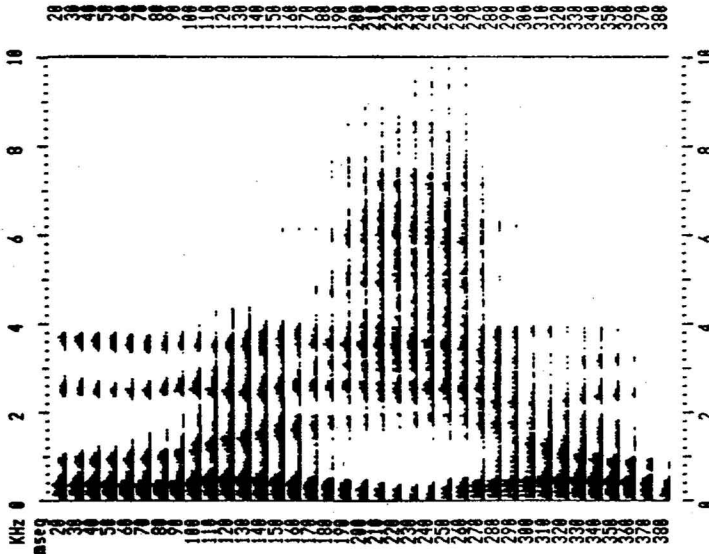


Fig. 6 Transició [ʒo] corresponent a l'expressió "la jove"

5.2. Paràmetres del fonema [ʒ]

Podem dir que aquest fonema és la versió sonora de [ʃ] així com [z] ho és de [s]. El paral·lisme entre les dues parelles de consonants és bastant exacte, amb algunes petites diferències. Així, per exemple, la part sonora de l'espectre destaca més en el cas de [ʒ] que en el de [z], i en general tot el nivell de [ʒ] és també més alt, amb uns formants fricatius molt marcats que s'estenen fins a 4 KHz. Més enllà d'aquesta freqüència l'espectre continua fins a uns 8 KHz però més difús i més atenuat. Per la part baixa l'energia del soroll comença aproximadament a partir dels 1700 Hz. A la figura 6 es pot veure un espectrograma de la transició [ʒo].

TAULA VII. *Paràmetres de [ʒ] en contacte amb totes les vocals i DT*

	[i]	[e]	[ɛ]	[a]	[ɔ]	[o]	[u]	[ə]
Durada (ms)	98	90	92	80	78	88	83	88
DT	15	6	8	9	9	13	12	10
F1 de [ʒ] (Hz)	257	222	262	264	314	268	238	178
DT	61	77	87	103	69	100	49	40
F2 de [ʒ] (Hz)	2114	2207	2047	2065	1971	1908	1900	1966
DT	90	154	142	189	125	123	224	82
F3 de [ʒ] (Hz)	2792	2885	2814	2600	2585	2525	2600	2642
DT	164	227	135	0	38	99	179	140
F4 de [ʒ] (Hz)	3528	3633	3666	3366	3471	3214	3714	3533
DT	189	151	82	151	150	234	107	103
Transició F2 (ms)	80	95	108	101	95	97	84	105
DT	17	14	13	12	10	8	10	17
F1 vocal (Hz)	320	438	625	710	610	490	340	490
DT	45	48	50	55	22	22	42	22
F2 vocal (Hz)	2240	1925	1810	1480	1160	1010	1040	1510
DT	89	50	55	84	55	22	55	89

A la taula VII es dóna la mitjana dels paràmetres mesurats i la seva desviació típica (DT).

El primer formant se situa al voltant dels 250 Hz. Destaca el valor particularment baix de [ʒə] amb 178 Hz com succeïa també en el cas de [zə].

El segon formant se situa molt clarament al voltant dels 2000 Hz amb desviacions molt petites si considerem els difonemes corresponents a la mateixa vocal; però presenta una certa tendència al creixement per les vocals anteriors. En tots els casos constitueix el punt d'arrencada de la transició vers el F2 de la vocal. Aquesta transició resulta particularment marcada pel cas de les vocals posteriors, de manera que afecta quasi bé a tot el temps d'aquests fonemes.

El tercer formant situat al voltant dels 2700 Hz resulta encara més definit que el segon, per les baixes desviacions observades. També s'observa aquí una tendència a la pujada per les vocals anteriors. El quart formant es manté molt proper al valor típic de 3500 Hz.

Els efectes coarticulatoris d'aquesta consonant sobre totes les vocals es manifesten amb un augment sistemàtic del dos primers formants d'aquestes últimes. Destaquen particularment l'augment de 62 Hz del primer formant de [i] i els augments del segon formant de [a, o, u, ə] amb 115, 146, 303 i 224 Hz respectivament.

Pel que fa als nivells del soroll, podem afirmar que són més alts que els de la resta de consonants fricatives. Aquesta diferència s'accentua si, com en el nostre cas, limitem la banda a 4 KHz; ja que en aquesta zona és on es donen les principals ressonàncies espectrals. No hem notat diferències de nivell segons la vocalsegüent i en tot els casos hem utilitzat un nivell 3 dB més alt que el que s'ha utilitzat en el cas de [ʃ] en contacte amb les vocals més obertes.

També aquí hem tingut el mateix problema que en el cas de [z] per a la síntesi dels sons fricatius sonors i l'hem resolt pel mateix sistema: amb un fragment sonor de 40 ms a l'inici del fonema, que en aquest cas hem hagut de deixar a un nivell 9 dB més baix que el soroll posterior.

6. ESTUDI COMPARATIU DE LES CONSONANTS FRICATIVES

6.1. Durada

És interessant concloure aquí amb una visió conjunta de les cinc fricatives estudiades anteriorment. Començarem per la durada del temps de fricció mesurat des de l'instant que apareix el contingut d'alta freqüència característic de la turbulència de l'aire, fins el moment que desapareix per donar lloc al so periòdic. La presència o absència de periodicitat és un tret decisiu per, determinar la frontera en el cas de les fricatives sordes. Per a les sonores, el criteri no és vàlid i hem acudit més aviat a la presència o absència de continguts d'alta freqüència.

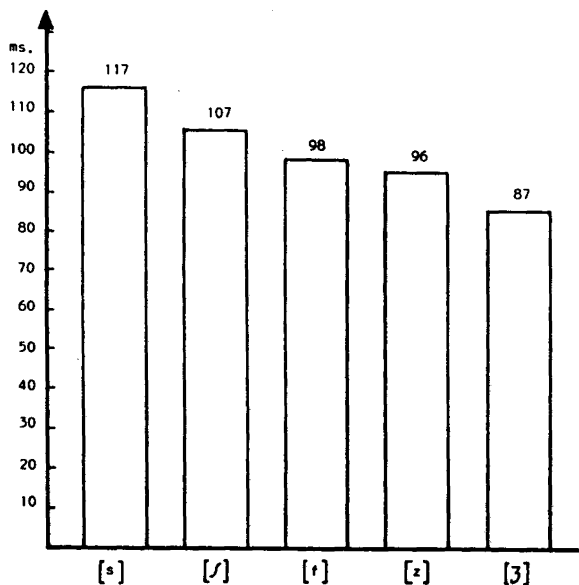


Fig. 7 Durada de les consonants fricatives catalanes.

Els resultats comparatius es veuen representats en el gràfic de la figura 7 en l'ordre decreixent de durada: [s] > [ʃ] > [f] > [z] > [ʒ] que suposa sempre un temps més llarg per les fricatives sordes i més curt per les sonores. Sembla com si el fet de la sonoritat fos ja un tret distintiu que permetés estalviar temps d'emissió, sense manca de percepció correcta del fonema.

6.2. Banda de soroll

Els límits inferior i superior de la banda de soroll constitueixen un tret distintiu de les consonants fricatives. A la figura 8 es poden veure els histogrames corresponents als límits inferior i superior de l'espectre de soroll per a cada consonant. Com que els límits mai resulten massa definits hem agrupat els casos a intervals de 1000 Hz, i donem simplement les freqüències estadístiques observades.

Cal advertir que, per a la determinació del límit inferior, no considerem la «barra de sonoritat» de les fricatives sonores, sinó únicament l'espectre de soroll. La mesura s'ha fet sobre els registres espectrogràfics dels informants que han pronunciat els mots del test de referència, desestimant alguns casos en els quals la lectura es feia molt dubtosa.

Tot i tractar-se d'una mesura orientativa, resulta molt il·lustrativa de la tendència corresponent a cada consonant. Podem observar en primer lloc que quasi bé tots els espectres comencen entre 1000 i 2000 Hz, excepte [f] que té components molt més baixes. Pel tipus de distribució estadística es pot deduir que [s] i [z] comencen més amunt que [ʃ] i [ʒ], i entre elles la versió sorda no es distingeix pràcticament de la versió sonora.

Quant al límit superior, hi ha més dispersió de resultats. El fonema [s] ocupa quasi bé tota la banda fins a 10 kHz. [z] i [f] també arriben a valors molt alts, encara que tenen algunes realitzacions amb un límit bastant inferior. Pel cas de [ʃ] tenim un límit superior més probable entre 8 i 9 KHz i per [] entre 7 i 8 KHz. De totes formes les diferències entre aquests dos últims fonemes són escasses.

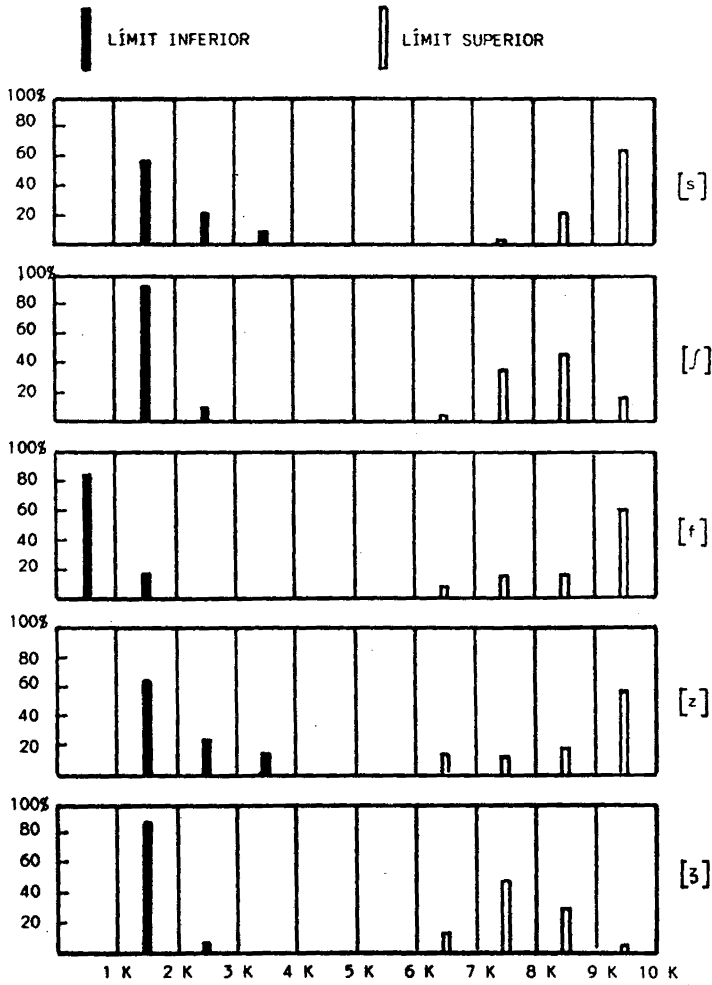


Fig. 8 Límit inferior i superior de l'espectre de soroll de les consonants fricatives catalanes.

Aquestes dades no corresponen necessàriament a la percepció subjectiva, ja que no puntualitzen la distribució d'energia dins tota la banda de soroll, concretament pel cas d'[f] no hem observat concentracions d'energia particularment importants, mentre que per la resta de fricatives es donen zones de ressonància molt localitzades que són decisives per a la identificació del timbre característic de cada fonema. Com que aquestes ressonàncies estan situades preferentment a la part baixa de l'espectre, podem afirmar que el límit inferior coincidirà normalment amb una d'elles. En aquest sentit deduïm que aquest límit inferior és més important que el límit superior, per a una correcta percepció de les consonants fricatives.

6.3. Transicions dels formants F1, F2, F3

A partir dels valors obtinguts dels formants fricatiu per cada una de les consonants [s, ʃ, z, ʒ], amb les corresponents vocals, i amb els valors estacionaris dels mateixos formants en les 8 vocals considerades hem construït els gràfics de les figures 9 i 10 que donen una visió molt completa de totes les transicions. Hem prescindit del cas d'[f] ja que pràcticament no dona transicions apreciables. Hem separat les transicions corresponents a les vocals posteriors de les anteriors, ja que cada grup presenta un comportament molt peculiar: amb un F2 creixent o quasi bé horitzontal per les anteriors i fortament decreixent per les posteriors. El cas de les vocals centrals [a, ə] s'aproxima més aviat al cas de les vocals posteriors. Les transicions del F1 i F3 són molt semblants en tots els casos estudiats, mentre que la del F2 sofreix variacions molt diferents segons els casos. El pendent és molt més fort per les consonants palatals [ʃ, ʒ] que per les alveolars [s, z], per tant aquesta evolució és decisiva per a la determinació del punt d'articulació, mentre que no aporta cap informació per a la distinció sorda-sonora. Entre les transicions corresponents a [s] i a [z] es dona una identitat quasi bé perfecta, com succeeix també entre [ʃ] i [ʒ].

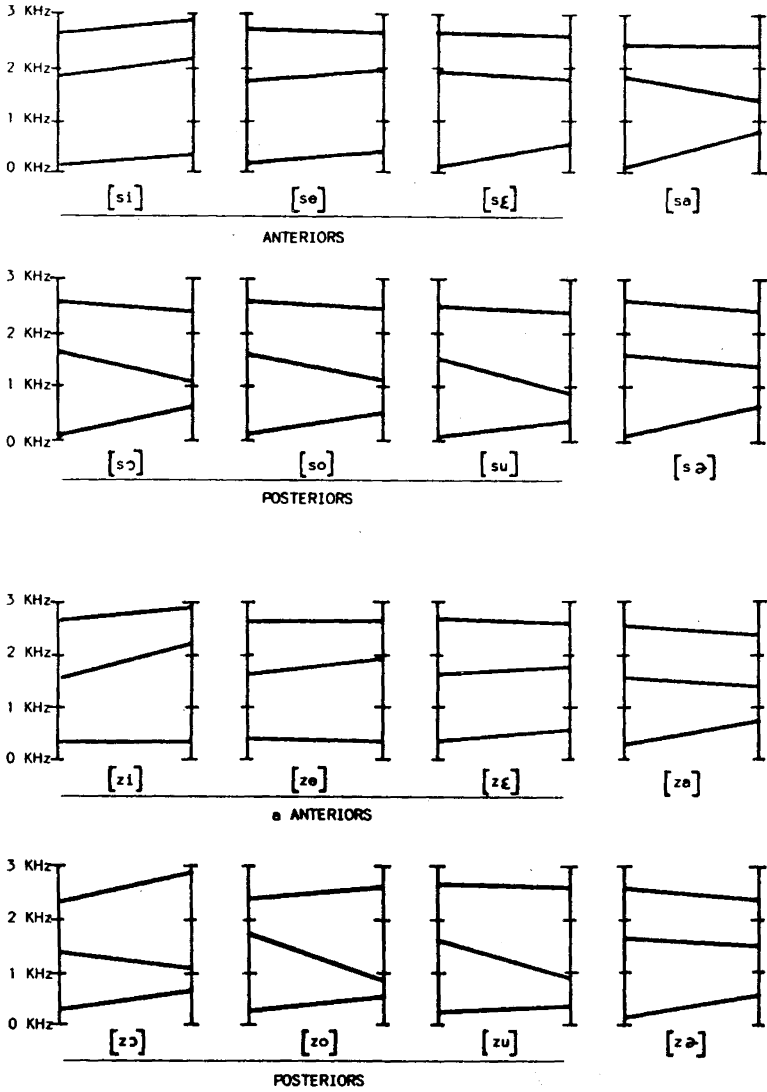


Fig. 9 Transicions dels tres primers formants de les consonants [s,z] en contacte amb totes les vocals.

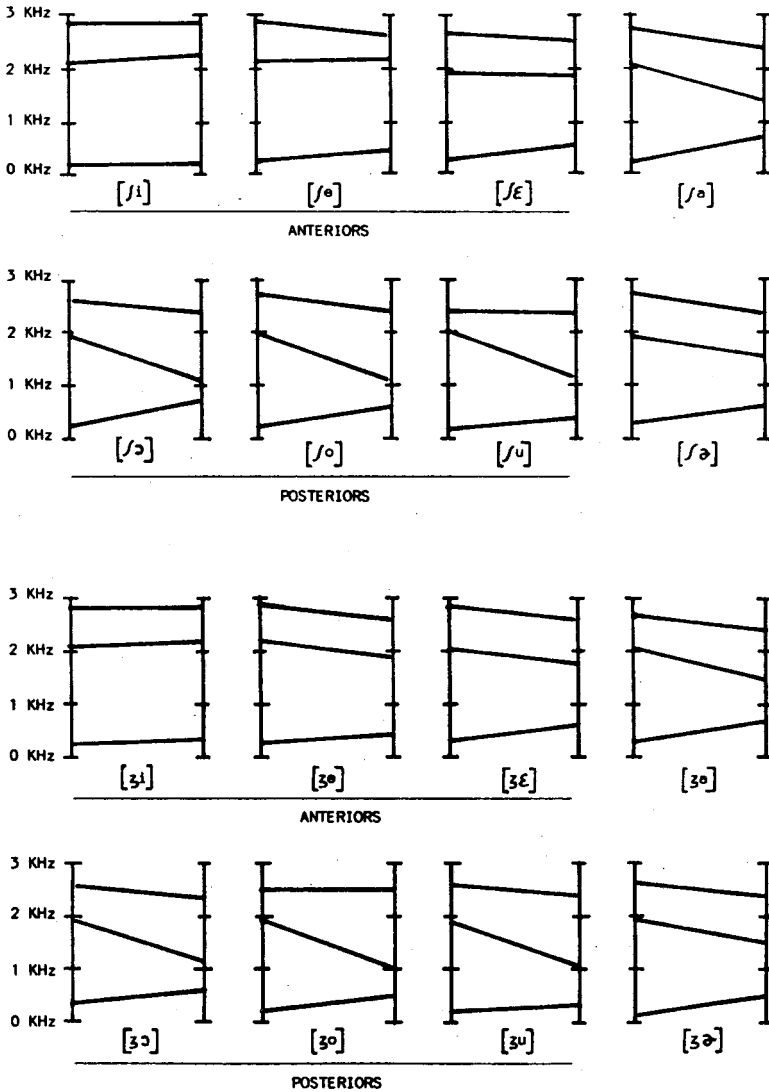


Fig. 10 Transicions del tres primers formants de les consonants $[f, \text{ɸ}]$ en contacte amb totes les vocals.

6.4. Nivells de soroll

L'estudi comparatiu del nivell s'ha fet mitjançant la percepció de sons sintetitzats amb el sintetitzador MEA 8000. A partir de proves successives de percepció hem intentat uniformar al màxim la sensació de «naturalitat» en el nivell de les consonants fricatives seguides de les respectives vocals. Una de les constatacions més immediates ha estat la falta de proporcionalitat entre el nivell objectiu de l'estímul i la percepció del grau de naturalitat de la consonant. Això s'ha concretat fonamentalment en una necessitat de baixar el nivell quan la vocal següent és més tancada i augmentar-lo quan és oberta. De totes formes, no sempre ha estat necessari fer aquesta distinció: les consonants [f, ʒ] s'han sintetitzat sempre al mateix nivell, independentment de la vocal; encara que a nivells molt diferents per una i altra consonant (Fig. 11). La distinció entre vocals obertes i tancades s'ha fet únicament per [s, ʃ, z] amb una diferència de +3 dB en les transicions amb vocal oberta, respecte de les transicions amb vocal tancada.

Volem deixar constància aquí també de la necessitat d'assolir aquests nivells d'una forma progressiva a l'inici de la consonant. Altrament l'inici sobtat de la fricació dóna una sensació d'oclusió prèvia que correspon a la percepció de les consonants africades. Aquest fet porta a confusions de fricatives per africades, quan aquelles es troben a principi de paraula, es a dir, després d'un silenci.

Per a l'estudi comparatiu s'han escollit les transicions amb vocals obertes. Les diferències són molt notables entre els casos extrems de [ʒ] i [z]. De totes formes, insistim en el fet que la síntesi s'ha realitzat únicament fins a 4 KHz, amb la qual cosa hem privilegiat les consonants amb bandes de soroll més baixes com són [ʒ, ʃ]. Amb una síntesi feta fins a 10 KHz segurament s'obtidrien resultats molt diferents.

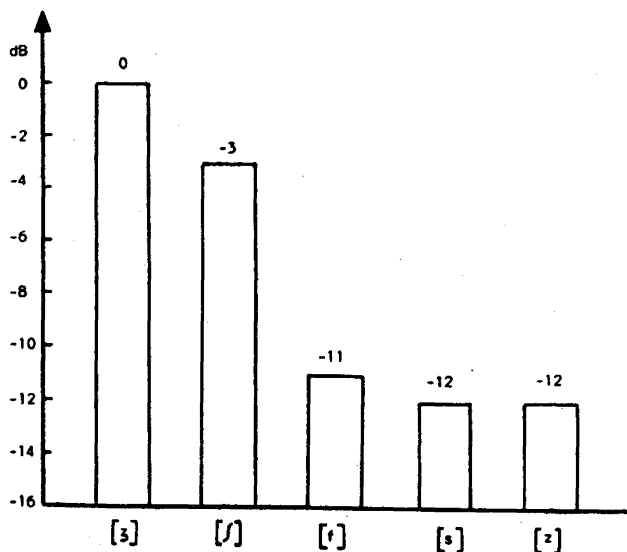


Fig. 11 Nivells comparatius de les cinc consonants fricatives estudiades, seguides de vocals obertes.

7. ESTUDI DE LA CONSONANT AFRICADA [c]

Seguint el criteri de Daniel Recasens (RECASENS, 1984) hem estudiat separadament la consonant africada [c] que correspon aproximadament al conjunt [tʃ], però de tal forma que quasi bé es pot considerar com un sol fonema. Per exemple, utilitzarem la transcripció [c] pels sons finals de les paraules «mig», «lleig», «veig», etc. En considerar aquesta successió de dos sons, un oclusiu i l'altre fricatiu, com una sola unitat hem aconseguit una major naturalitat en l'expressió oral sintetitzada.

En els espectrogrames d'aquest tipus de so seguit d'una vocal, com és el cas de la figura 12, hem pogut observar com clarament es distingeixen els trets característics de [t] i de [ʃ] successivament. Com en el cas de [t] observem el temps d'oclusió inicial que hem considerat d'uns 80 ms, seguit de l'explosió durant uns 8 ms i atenuat durant 8 ms més. A continuació s'inicia la zona de fricció amb els mateixos formants fricatius que [ʃ] i una durada d'uns 64 ms. Al final d'aquesta zona es nota un decreixement del nivell abans d'entrar a la part sonora corresponent a la vocal, mentre s'inicien ja les transicions que es perllongaran durant l'emissió de la vocal. Aquest fenomen d'atenuació és semblant al que s'ha detectat en el cas de les fricatives sordes [s] i [ʃ]. Amb tot, l'efecte sembla aquí més accentuat, tant en nivell com en durada, provocant una major separació entre consonant i vocal. La forma de les transicions, particularment del segon formant, són molt semblants a la de les corresponents transicions [ʃV].

Coincidim perfectament en les observacions de Peter HOWELL i Stuart ROSEN (1983) quant a la durada més breu de la zona fricativa de [c] respecte a la de la consonant [ʃ], amb un temps de pujada («rise time») també més curt. És a dir que en el cas de la consonant fricativa l'inici del soroll aleatori es fa d'una forma més suau i progressiva, mentre que per la consonant africada l'inici és més abrupte i assoleix el seu nivell estacionari amb un temps més breu.

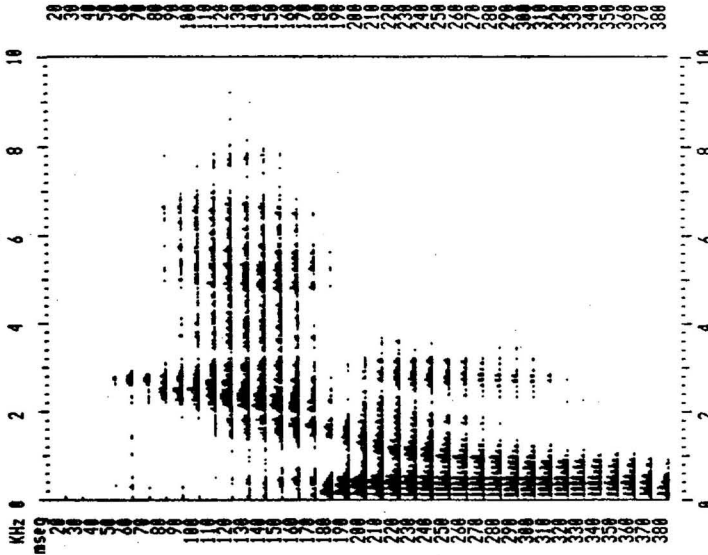


Fig. 12 Evolució espectral del difonema [co].

8. LA SÍNTESE DE SÍLLABES [CV]

A partir de les dades estadístiques de cada consonant fricativa [f, s, ʃ, z, ʒ] en connexió amb les vuit vocals [i, e, ε, a, o, u, ə] hem preparat un inventari de 40 síl·labes [CV], codificades segons els paràmetres del sintetitzador Philips MEA 8000 i guardades en memòria digital. Per a cada una d'aquestes unitats s'ha preparat una edició acurada de l'evolució espectral que comprèn els sis intervals següents:

- a) Temps de pujada del nivell fricatiu durant els primers 60 o 70 ms per les fricatives sordes i durant uns 40 ms per les sonores. El canvi de nivell durant aquest temps ha estat d'uns 12 o 15 dB pels dos tipus de consonants fricatives.
- b) Fragment fricatiu estacionari durant uns 50 o 60 ms més per les fricatives sordes i d'uns 30 ms per les sonores.
- c) Atenuació del soroll durant uns 24 ms per les sordes i durant 16 ms per les sonores.
- d) Interval de transició entre la consonant i la vocal amb l'evolució de formants que s'ha estudiat per a cada síl·laba. La durada d'aquest temps és molt variable en cada cas i el nivell és sempre creixent. Cal advertir, de totes formes, que les transicions dels formants s'inicien ja lleugerament durant el final de la zona fricativa.
- e) Zona estacionària de la vocal amb un nivell constant. En principi s'ha deixat un temps d'uns 50 ms. Encara que, pel cas de les vocals tòniques, aquest interval es modifica notablement, afegint un període de 64 ms a un nivell 3 dB més alt i una inflexió del to que augmenta 16 Hz durant els primers 32 ms i que torna al to normal durant els 32 ms restants.
- f) Finalment es deixa un petit interval de 24 ms per reduir el nivell progressivament de 18 dB.

A la figura 13 es pot veure l'oscillograma de la síl·laba sintetitzada [sa] en versió tònica. Sobre ella es poden distingir les diferents fases anteriorment esmentades.

Aquesta tasca de generació de les síl·labes s'ha dut a terme amb l'ajut d'un programa editor de veu dissenyat al Laboratori d'Acústica de l'Escola Universitària de Telecomunicació La Salle Bonanova de Barcelona, que permet modificacions àgils dels paràmetres acústics sobre un terminal d'ordinador i l'audició immediata del so generat. Així hem pogut completar els resultats de l'anàlisi anterior amb els avantatges de la percepció directa dels sons sintetitzats.

Concretament ens hem hagut d'emmotllar al sistema de codificació que utilitza l'esmentat sintetitzador MEA 8000. Fonamentalment consisteix en definir cada 8 ms la freqüència central i l'amplada de la banda de tres ressonàncies o formants distribuïts entre 150 i 3400 Hz; més un quart formant situat sempre a 3500 Hz, del qual podem definir únicament la seva amplària. Igualment cada 8 ms es pot modificar la intensitat o nivell del so en una escala de 42 dB a increments de 3 dB, i el to fonamental a increments que poden oscil·lar entre 1 i 15 Hz també cada 8 ms, tant en sentit creixent com decreixent. El mateix sistema de formants s'utilitza tant per als fragments sonors com per als sords, indicant en cada cas si el tipus d'excitació és periòdica o aleatòria.

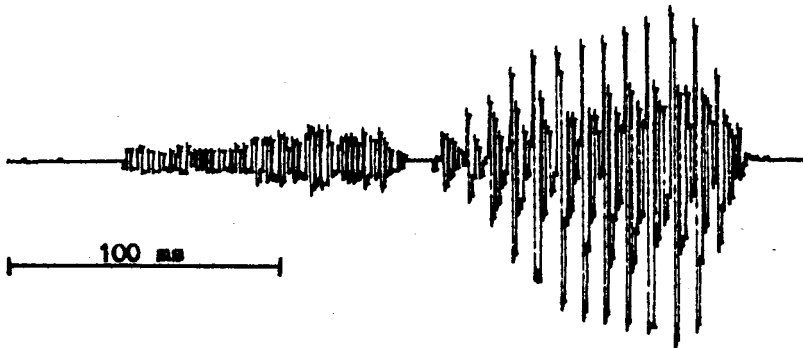


Fig. 13 Oscil·lograma de la síl·laba [sa] sintetitzada amb el MEA 8000.

9. PROVES DE PERCEPCIÓ

Un cop acabada la fase de síntesi, es van sotmetre totes les síl·labes [CV] a l'audició de persones experimentades en transcripció fonètica. La presentació es va fer en un ordre totalment aleatori per tal que donessin la seva percepció en una opció oberta. És a dir que, el que s'ha demanat als experts és que donessin la seva transcripció fonètica sense condicionar la seva opció a consonants fricatives ni a qualsevol altre tipus de consonant.

Concretament s'han fet dues proves de percepció per retocar en cada cas alguns paràmetres de les síl·labes segons els resultat observats, i una tercera prova per donar els resultats finals.

El primer test va ser aplicat a 9 experts. Els resultats es poden veure resumits en la matriu de confusions de la taula VIII de doble entrada. En vertical es donen les 5 consonants sotmeses a prova en tots els contextes vocàlics possibles (40 síl·labes en total), i en horitzontal les diferents respostes obtingudes, en tant per cent, com també el casos no reconeguts.

Cal remarcar que el reconeixement a partir de síl·labes desprovistes de significat presenta sempre bastants errors, fins i tot en el reconeixement de veu natural. La manca de context significatiu fa que les confusions es donin amb més facilitat. De totes formes, si aquestes confusions es repeteixen d'una forma sistemàtica cal sospitar que la síntesi no resulta prou assolida. Aquest és el cas en el nostre test del fonema [ʃ] interpretat en un 25 % dels casos com a [s] o, pitjor encara, el fonema [z] interpretat també com a [s] en un 42 % de casos. El fonema [ʒ] presenta una confusió menys definida amb un 13 % de casos no reconeguts.

TAULA VIII. *Matriu de confusions del primer test*

Consonants del test (estímul)	Respostes en %									
	[f]	[s]	[ʃ]	[z]	[ʒ]	[t]	[λ]	[j]	[θ]	No rec.
[f]	67	16	1	3	1	3			7	1
[s]		70	14	9		1			1	4
[ʃ]		3	25	49	7	7	3			7
[z]			47	4	36	10	4			4
[ʒ]			4	10	8	59		1	4	13

Després d'una observació detinguda dels resultats d'aquest primer test es va procedir a una revisió de les síl·labes sintetitzades segons els següents criteris:

- Provocar un creixement de nivell més progressiu a l'inici de les consonants fricatives sordes per evitar la sensació d'oclusió inicial que podria donar una versió més aviat africada.
- Augmentar el temps de sonoritat de les dues fricatives sonores [z, ʒ] i reduir el nivell de la part estrictament fricativa, per tal d'evitar la confusió d'aquestes consonants amb les corresponents versions sordes [s, ʃ].
- Accentuar el nivell de la ressonància més baixa (al voltant dels 2000 Hz) de les fricatives palatals [ʃ, ʒ] i atenuar les ressonàncies més altes.

Després d'aquesta revisió es va tornar a passar el test a un segon grup d'experts diferents dels primer. Aquest cop van ser 11. Els resultats foren els de la taula IX.

TAULA IX. *Matriu de confusions del segon test*

Estímul	Respostes en %																		
	[f]	[s]	[ʃ]	[z]	[ʒ]	[t]	[k]	[d]	[g]	[n]	[l]	[λ]	[δ]	[γ]	[θ]	[c]	[v]	[j]	No rec.
[f]	81	3		8	1										2	2	2		2
[s]	2	87	1	8	1										1				
[ʃ]	2	2	62	2	9		9		3									1	10
[z]	5	54	4	24	8					1									3
[ʒ]	1	7	16	8	48	2		4	1		2	2	1	1	1				4

Com es pot observar, comparant amb la taula anterior, els resultats han millorat molt pel cas de les consonants sordes, però no les sonores que resulten menys reconegudes que en el cas anterior. Segons això, es va procedir encara a una nova reedició de les síl·labes amb les fricatives [z, z̄, ʃ] segons els següents criteris:

- Allargar encara més la part sonora del temps de fricció a costa del temps de soroll estrictament aleatori de [z, z̄].
- Provocar un creixement més progressiu del nivell a l'inici d'aquestes mateixes consonants, tal com es va fer per les fricatives sordes.
- Pujar de nivell la ressonància fricativa que es troba al voltant dels 2700 Hz per les dues consonants palatals [ʃ, z̄] tal com es va fer anteriorment per la ressonància de 2000 Hz.
- Suavitzar el canvi de nivell en passar de la consonant a la vocal, fent-lo més progressiu.

Els resultats d'aquestes modificacions foren sotmesos novament a l'audició, aquest cop de 8 experts, que oferiren una percepció tal com la que reflecteix la taula X. Les transicions amb [f, s] restaren inalterades respecte a les de l'última edició.

TAULA X. *Matriu de confusions del tercer test*

Estímul	Respostes en %						
	[f]	[s]	[ʃ]	[z]	[ʒ]	[θ]	[X] No ident.
[f]	84	5	5			2	4
[s]		88	7	5			
[ʃ]	6		87		2	2	4
[z]		44		54			2
[ʒ]	7		30		59		4

La percepció de [f, s, ʃ] resulta àmpliament satisfactòria. És de remarcar particularment la millora de la [ʃ] respecte del test anterior (passem d'un 62 % a un 87%) únicament per la pujada de la ressonància de 2700 Hz. La percepció de les fricatives sonores es pot considerar també satisfactòria si atenem al sistema de síntesi que ha calgut emprar, alternant l'emissió sorda i l'emissió sonora. Precisament la confusió d'aquestes dues consonants [z, ʒ] es dona quasi exclusivament a favor de les corresponents version sordes [s, ʃ]. Per tant deixarem ja com a definitiva aquesta última edició de les transicions entre fricatives i vocals.

10. CONCLUSIONS

Amb aquest estudi hem intentat donar una primera descripció de les consonants fricatives catalanes; una descripció que no pretén ser exhaustiva, sino únicament la base per a la síntesi automàtica de les transicions entre fricatives i vocals. Hem quantificat els paràmetres referits a l'evolució dels formants i del nivell global o energia acústica. La utilització d'aquestes dades permet simplificar la informació del senyal

acústic d'una forma molt manejable informàticament, i amb el mateix grau de condensació que suposa el text escrit.

Una constatació força interessant fa referència a la necessitat de completar les informacions objectives de les evolucions espectrals amb l'aportació subjectiva de la percepció auditiva humana. Mercès a la utilització d'un sistema de síntesi artificial de la veu, hem pogut constatar la gran importància dels petits retocs acústics, fets en un sentit predeterminat, per a canviar radicalment la sensació auditiva.

Després de les experiències de percepció es pot tornar a l'anàlisi dels espectrogrames amb un sentit de realimentació molt enriquidor que fa descobrir uns trets acústics fonamentals i que, tal volta, no han estat captats en una primera observació dels enregistraments gràfics.

Aquest sistema d'anàlisi i síntesi s'està aplicant actualment als nostres Laboratoris per a un estudi semblant de tots els sons consonàntics del català, en vistes a un sistema automatitzat de conversió text-veu que esperem tenir a disposició en un termini molt breu.

11. AGRAÏMENTS

Al final d'aquest treball voldríem fer constar l'agraïment als professors, col·laboradores i alumnes. del Departament d'Acústica de l'Escola Universitària de Telecomunicació La Salle Bonanova que estan recolzant i col·laborant activament en aquesta tasca de recerca en l'àmbit del tractament de la veu. Particularment cal fer menció de les valuoses aportacions de l'Enginyer Pere Buhigas en l'enregistrament de la veu de locutors professionals i en l'anàlisi espectral dels diferents sons aquí estudiats. La tasca d'edició de veu a partir del sintetitzador MEA 8000 ha resultat molt fàcil mercès al sistema dissenyat per Francesc Gudayol, enginyer i col·laborador d'aquest Departament.

Igualment ens cal d'agrair les aportacions com a experts en Fonètica Acústica dels professors Dr. Eugenio Martínez Celdrán i Lluís Izaguirre que personalment i amb els seus alumnes de Fonètica han aportat la seva percepció acústica dels sons sintetitzats.

També cal esmentar l'encoratjament de tots els membres del Grup de Comunicació Oral, creat recentment dins l'àmbit de l'Institut d'Estudis Catalans, amb les seves valuoses aportacions en les sessions de treball sobre anàlisi de fricatives catalanes.

REFERÈNCIES BIBLIOGRÀFIQUES

- BORZONE, A. M. & MASSONE, M. I.: «Acoustic analysis and perception of Spanish fricative consonants». *JASA* 69, (1981), pp. 1145-1153.
- DELATTRE, P.; LIBERMAN, A. M.; COOPER, F. S.: «Formant transitions and loci as acoustic correlates of place of articulation of American fricatives» *STUDIA LINGUISTICA*, 26, (1962), pp. 104-121.
- GURLEKIAN, J. A.: «Recognition of the Spanish fricatives /s/ and /f/» *JASA*, 70, (1981), pp. 1624-1627.
- HOWELL, P. & ROSEN, S.: «Production and perception of rise time in the voiceless affricate/fricative distinction». *JASA*, 73, (1983), pp. 976-984.
- MANN, V. A. & REPP, B. H.: «Influence of preceding fricative on stop consonant perception». *JASA*, 69, (1981), pp. 548-558.
- MARTÍ, J.: «Paràmetres vocàlics del català». *FOLIA PHONÈTICA*. Revista del Laboratori de Fonètica «Pere Barnils», 1, (1984), pp. 23-43.
- RECASENS, D.: «Producció i coarticulació de vocals i consonants del català en el decurs». Tesi doctoral presentada a la Universitat de Barcelona (1984).
- REPP, B. H. & MANN, V. A.: «Fricative-stop coarticulation: acoustic and perceptual evidence». *JASA*, 71 (1982), pp. 1562-1567.
- SOLI, S. D.: «Second formant in fricatives: Acoustic consequences

of fricative - vowel coarticulation». *JASA*, 70, (1981), pp. 976-984.

STEVENS, K. N.: «Airflow and turbulence noise for fricative and stop consonants: static considerations». *JASA*, 50, (1971), pp. 1180-1192.

WHALEN, D. H. «Effects of vocalic formant transitions and vowel quality on the English [s] - [ʃ] boundary». *JASA*, 69, (1981), pp. 275-282.

YENI-KOMSHIAN, G. H. & SOLI, S. D.: «Recognition of vowels for information in fricatives: Perceptual evidence of fricative-vowel coarticulation». *JASA*, 70, (1981), pp. 966-975.