CAMPS DE DISPERSIÓ VOCÀLICA EN IMITACIONS DE VEU: PRIMERS INDICIS D'UN EXPERIMENT SOBRE IDENTIFICACIÓ DE LOCUTOR

RAMON CERDÀ MASSÓ
Universitat de Barcelona
rcerda@ub.edu

Casi todas las voces humanas son diferentes; […] *siempre hay un no sé qué que las matiza, y lo que es aún más extraordinario: lo difícil es demostrar que emisiones de voz de un mismo individuo realizadas en situaciones distintas proceden de la misma fuente.*

Antonio Quilis (2000)

RESUM

El treball presenta un experiment adreçat a resoldre un dels principals reptes de la fonètica forense: la identificació de locutor. El punt de referència adoptat és la noció de camp de dispersió, és a dir, el conjunt de freqüències obtingudes a partir d'una sèrie estadísticament representativa de realitzacions, en el nostre cas de les vocals tòniques [i], [a] i [u]. Aquest principi s'ha aplicat a tres nivells registrals de veu, tots ells pertanyents al català central, distribuïts així: una selecció de fragments, extreta de sengles entrevistes a tres polítics popularment molt coneguts, i dues rèpliques fonològicament idèntiques per a cada fragment, una de repetida i una altra d'imitada, executades per dos especialistes professionals amb un total de nou resultats, tres per a cada polític. Els primers resultats obtinguts són, com a mínim, tendencialment positius.

Paraules clau: *fonètica forense, identificació de locutor, imitació de veu, camp de dispersió, formants vocàlics.*

ABSTRACT

This paper presents an experiment addressed to solve one of the main challenges in forensic phonetics: the speaker recognition. The point of reference taken is the notion of range of dispersion, that is, the set of frequencies obtained from a statistically representative series of realizations, in this case from the stressed vowels [i], [a] and [u]. This principle has been applied to three voice register levels, all of them belonging to Central Catalan, distributed as follows: a selection of passages, extracted from interviews to three well-known politicians, and two phonologically identical replicas for each passage, a repeated and an imitated one, performed by two professional impersonators leading to a total of nine results, three for each politician. The preliminary results show, at least, some positive tendencies.

Keywords: *forensic phonetics, speaker recognition, voice imitation, range of dispersion, vowel formants.*

1. ANTECEDENTS

A l'actualitat es treballa arreu del món per aconseguir d'una manera cada cop més concloent i definitiva que la identificació de locutor aplicada a la lingüística forense funcioni com una prova independent, autoprobatòria i amb tantes garanties de fiabilitat com en altres aplicacions. En aquest treball ens referirem només a la identificació de locutor amb fins judicials on, com diu el lema de més amunt, les proves habituals realitzades en situacions diferents –p. ex., comparant enregistraments telefònics amb d'altres extrets d'una lectura directa– fan, ara per ara, gairebé impossible atènyer una certesa absoluta sobre la identificació (cf. alguns clàssics com Aderman 2005, Nolan 1983, Rose 2002).

El curiós és que als millors resultats només s'hi pot arribar a través d'experiments que vénen a rebatre un seguit d'intuïcions que tothom assumeix més o menys espontàniament. En primer lloc, la de pensar que, només escoltant-lo, qualsevol és identificable per la veu, almenys en condicions normals –és a dir, sense interferències p. ex. de fatiga, emocions fortes o malalties, sobretot laríngies–. També se sol atorgar una considerable fiabilitat identificativa a la multitud de vestigis descoberts d'ençà que s'ha aprofundit en l'etiquetació de fets de parla, entre ells dialectològics o sociològics (com l'ús d'expressions pròpies d'un territori o d'una professió), d'altres pragmàtics (com la repetició de certs expletius) i d'altres fonètics (com una cantarella o un tartamudeig). Per descomptat, la capacitat d'arreplegar alhora molts d'aquests vestigis augmenta la possibilitat de la identificació, però, per molt que s'atenyi un cert grau de raresa individual, el resultat no garanteix mai, almenys fins ara, una certesa prou absoluta o suficient des del punt de vista judicial. De fet, les conclusions habituals dels especialistes més qualificats es manifesten en forma de probabilitat estadística, com es pot comprovar a Aderman 2005, Harrison et al. 2007, etc.

Els enginyers especialitzats en teoria del senyal són els que més aprofundeixen en els trets acústics de les emissions, és a dir, en les característiques físiques o genèmiques dels impulsos produïts per la veu. De fet, tot i que no desdenyen en principi els vestigis que descrivim als paràgrafs anteriors, es concentren en els trets acústics, sense cap referència fonològica o lingüística, com ara en les magnituds espectrals, les freqüències dels formants, els perfils d'energia, etc.

No cal subratllar fins a quin punt es fa necessària una cooperació interdisciplinària entre tots els especialistes per augmentar i, qui sap si optimar decisivament, les possibilitats. El que segueix n'és una petita mostra. Més exactament, com veurem, només és una experiència preliminar en diversos sentits.

2. ELS SUPÒSITS IMMEDIATS

2.1. Indicis inicials

Com hem dit, l'objectiu final consisteix a trobar, a partir d'enregistraments, un o més trets característics, senzills o complexos –no importa com de complexos–, que permetin identificar d'una manera prou aproximada la font d'una emissió de veu al marge de qualsevol circumstància o fenomen que l'hagi pogut condicionar poc o molt. La fita és diferent a la d'obtenir indicis biomètrics d'identitat com els de l'ADN, les empremtes digitals o la conformació de l'iris ocular, perquè aquí les premisses són clares, tot i que l'obtenció mai no està exempta tampoc de condicionaments experimentals, és a dir, de com i amb quina pulcritud se'n fan els mesuraments.

En el nostre cas el problema prové del fet que els principals factors distorsionants s'originen en la pròpia emissió, fins i tot en l'emissió espontània. I són molts aquests factors. Al marge dels canvis evolutius de veu per l'edat, els més importants són els ocasionals, els que es poden produir durant uns dies –posem per un constipat–, unes hores –per fatiga– o uns moments –per un estat d'ànim o una simulació–. Un factor important depèn de si la comprovació es fa o no amb l'aquiescència de qui s'hi sotmet.

La idea de l'experiment va partir inicialment de la tesi doctoral de Cerdà 1972, tot i que a Cerdà et al. 2007 s'exposa d'una manera més precisa i elaborada, però encara intuïtiva. En resum gira a l'entorn de la noció de «camp de dispersió fonològica» per a les vocals. Com se sap, es tracta d'un concepte provinent de l'estructuralisme que consisteix en el conjunt de freqüències que ocupen dins d'un sistema de coordenades per al primer i el segon formant les realitzacions obtingudes per via espectrogràfica de cadascuna de les vocals analitzades. Així es combina el pla fonològic, pel qual es reconeix la identitat de les vocals, amb el pla fonètic, amb la determinació dels punts freqüencials que n'ofereixen les realitzacions[1].

En aquest cas es comptava amb algunes evidències entre presumptes i provisionals. I és que es comparaven centenars de realitzacions vocàliques d'un informant AB, masculí d'uns 45 anys (figura 1), amb la realització aïllada –dient tot just [i], [e], [ɛ], [a], [ɔ], [o], [u], [ə]– del mateix sistema vocàlic per part d'un altre informant

[1] Entre altres treballs basats en el mateix concepte, podem citar els de Carrera-Sabaté et al. 2005 i Puig et al. 1990 per al català, i el de Fernández Planas 1993 per al castellà.

RC, també masculí i de 25 anys, pertanyent al mateix dialecte, el català oriental barceloní (no xava). Més encara, la freqüència dels formants s'obtenia a ull, com era freqüent llavors, per no dir inevitable, multiplicant per 81 Hz els mil·límetres de distància entre la base de l'espectrograma i el punt central de la zona més estable possible del formant. El cas d'[u] era el més problemàtic, ja que solia haver una massa d'energia més o menys contínua entre el F_0, el F_1 i el F_2.

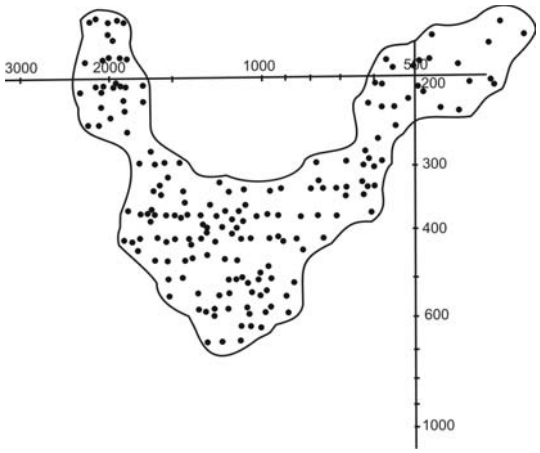


Figura 1. *Conjunt de realitzacions vocàliques de l'informant AB analitzades a Cerdà 1972 a partir dels dos primers formants, el primer (F_1) a l'ordenada i el segon (F_2) a l'abscissa.*

La comparació esdevenia especialment clara quan el primer conjunt de realitzacions també quedava reduït a un altre triangle en abstracte, obtingut en dues etapes a partir justament de la noció de camp de dispersió fonològica. En primer lloc, el continu de realitzacions quedava distribuït en una sèrie de zones delimitades per criteris fonològics –els mateixos camps– que, lògicament, corresponia a un altre triangle vocàlic ideal amb els mateixos referents de l'anterior (figura 2). En segon lloc, la transformació dels camps a punts es va fer a partir del punt geomètric central de cada camp.

En síntesi, la comparació dels triangles (figura 3) donava a entendre que les diferències eren producte de les característiques individuals dels informants i que aquestes diferències serien prou estables, i potser permanents, com per utilitzar-les

com a indicis versemblants o, encara millor, fidedignes per identificar la veu dels respectius locutors.

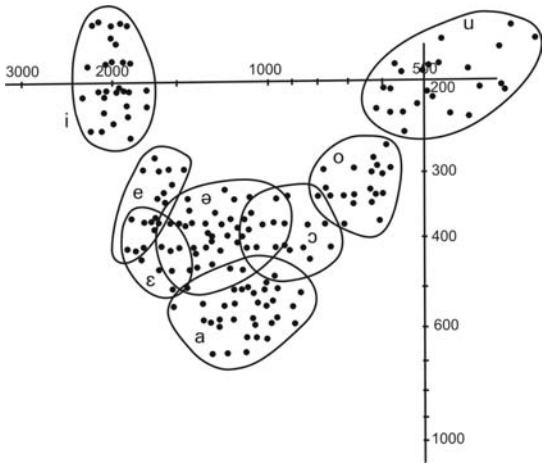


Figura 2. *Camps de dispersió de les realitzacions vocàliques obtinguts a partir de les realitzacions de l'informant AB (Cerdà 1972).*

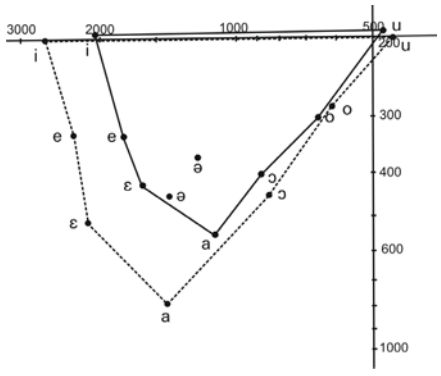


Figura 3. *Triangles vocàlics dels informants AB (amb línia contínua) i RC (amb línia discontínua). El primer prové del punt geomètric central dels camps de dispersió de la figura 2. El segon prové d'una pronúncia aïllada dels fonemes vocàlics (Cerdà 1972).*

Això confereix a l'experiment actual una sèrie de característiques. Entre les aparentment favorables:

1. En primer lloc, el procediment ocupa una zona de màxima abstracció acústica per al fonetista lingüista, un punt que, durant tot el temps en què lingüistes i enginyers del processament del senyal han treballat totalment separats, suposa una de les màximes aproximacions amb unes expectatives aprofitables per a tots.

2. Tenint en compte la naturalitat de les realitzacions vocàliques en emissions espontànies de parla, almenys en condicions relativament normals, cada conjunt de realitzacions pot ser ben representatiu de les característiques individuals de la veu del locutor.

I entre les evidentment desfavorables destaca que no és fàcil descobrir amb tota la precisió desitjable quan s'ha trobat exactament el que se cerca. Entre altres raons, perquè:

1. Una mateixa vocal pot ocupar un conjunt molt elevat de contextos tant segmentals com suprasegmentals, la combinació dels quals dóna lloc a una casuística coarticulatòria molt i molt complexa sobretot tenint en compte que els contextos suprasegmentals (prosòdics, en general) són sovint, per no dir sempre, difícils de delimitar, qualificar i, per tant, de classificar (cf. Recasens et al. 2001).

2. El càlcul de les freqüències mitjanes dels formants vocàlics pot complicar-se per diverses raons, entre elles per una veu molt timbrada de l'informant (que pot donar lloc a una superposició de mini-formants amb freqüències variables), i sobretot per la raó anterior, la inestabilitat contextual. Així, p. ex., una parla ràpida pot dificultar la tria de valors freqüencials característics.

3. Les interseccions entre camps de dispersió, que s'observa clarament a la figura 2, introdueix dubtes sobre la seva interpretació i, per tant, sobre la fiabilitat dels resultats, que podria augmentar si, posem per cas, s'hi introduís el càlcul d'un tercer formant, com s'indica a Cerdà et al. 2007.

4. En conjunt, tot considerant viable o, com a mínim, digne d'interès l'experiment –com és el nostre cas–, és difícil establir el nombre de realitzacions

necessàries estadísticament representatiu, és a dir, saber quan es farà innecessari augmentar els casos experimentats perquè els resultats es mantindran estables.

2.2. El nou plantejament

Tornant a les dades de Cerdà 1972, a partir de les quals s'extrauen les que van donar lloc a la nostra certesa, cal remarcar que ofereixen algunes deficiències importants, com ara la de comparar dos triangles obtinguts de manera massa diferent, un a partir de camps de dispersió i l'altra per emissions úniques. No hi ha dubte que el triangle de les emissions úniques, el de traç discontinu a la figura 3, tendeix a presentar una major separació entre les respectives realitzacions perquè van ser diferenciades d'una manera conscient. Una altra deficiència important prové de determinar el punt més representatiu dels camps a partir del seu centre geomètric en les cartes de freqüència, en comptes de fer-ho a partir de la mitjana dels valors freqüencials recollits en cada camp.

L'important, però, d'aquelles dades és que van suggerir l'experiment i prou. I que ara, després de tant de temps, s'hagin creat unes condicions aparentment òptimes per dur-lo a terme. Això ha estat possible gràcies a la Tesi de Doctorat de Mireia Farrús (Farrús 2008) –de la qual l'experiment només és un esquitx– i al conjunt de circumstàncies favorables que l'envolten.

Entre altres estratègies orientades a la identificació de locutor, és a dir, cap a la possibilitat de determinar d'una manera conclusiva si dos enregistraments de veu pertanyen o no a un mateix informant, una d'elles és òptima per executar el nostre experiment. Es basa en la performança de dos reconeguts imitadors de veu que actuen sobre diversos personatges polítics igualment reconeguts prenent com a referència dotze paràmetres. Nou d'aquests paràmetres estan inspirats en els utilitzats en el treball de Peskin et al. 2003 per tal de complementar, amb informació prosòdica, un sistema espectral convencional de reconeixement de locutor. Sis d'ells estan basats en mesures de la freqüència fonamental (mitjana, valor màxim, valor mínim, rang, i dues mesures sobre el pendent) i els altres tres estan relacionats amb la durada de segments i paraules (longitud mitjana de les paraules, i longitud dels fragments sords i sonors en cada paraula). Els tres paràmetres que completen el total de dotze es basen en mesures del *jitter* i del *shimmer*, dos paràmetres de qualitat de veu força utilitzats en la detecció de patologies de la veu que han demostrat, en un estudi recent a Farrús et al. 2007, ser d'utilitat en el reconeixement de locutor.

Una breu anàlisi d'un sistema d'identificació del locutor basat en aquests dotze paràmetres a Farrús 2008 demostra que l'error d'identificació per a cadascun dels paràmetres –exceptuant el rang de la freqüència fonamental– augmenta considerablement quan s'utilitzen les imitacions de veu en comptes de les veus originals dels imitadors. És a dir, a grans trets, els paràmetres prosòdics són susceptibles de ser imitats –almenys per imitadors professionals–, de manera que les imitacions de veu utilitzades a l'experiment comporten una alteració dels paràmetres prosòdics prou gran perquè el funcionament d'un sistema d'identificació de locutor basat únicament en paràmetres prosòdics disminueixi de forma considerable.

En aquest punt és rellevant referenciar la metodologia i el material utilitzat a la Tesi de Doctorat d'Elisabeth Zetterholm (2003), en la qual s'analitzen diverses característiques corresponents a imitadors professionals i les veus originals dels imitadors. Un aspecte de la tesi rellevant per al nostre experiment és l'indici segons el qual, en les veus imitades, es tendeix a alterar les freqüències dels formants, especialment en les vocals tòniques. És possible, però, que la diferència fos més apreciable perquè allí els imitadors i els imitats parlaven dialectes diferents.

Per al nostre experiment, hem partit d'una simplificació que s'ha concentrat en els quatre següents materials d'estudi:

1. Una selecció d'emissions espontànies dotades d'una pronúncia, unes característiques prosòdiques i una qualitat d'enregistrament modèliques extreta d'entrevistes periodístiques fetes a tres polítics força reconeguts en especial per la seva dicció (PM, XT i JS).

2. Les mateixes emissions pronunciades pels imitadors (CC sobre JS i QN sobre PM i XT) amb la veu pròpia, encara que, per raons de naturalitat, simulant l'entonació original.

3. Les mateixes emissions pronunciades pels imitadors mirant ara de reproduir al màxim totes les característiques fòniques del personatge (igual que abans, CC sobre JS i QN sobre PM i XT).

4. Una selecció de les mateixes paraules en les diferents versions que tenien una pronúncia clara –i, per tant, en un context màximament coincident– de vint realitzacions de cadascuna de les vocals [i], [a] i [u], les extremes del sistema fonològic comú a tots els implicats (que, casualment, és el mateix que hem pres a 2.2 com a referència: el barceloní). Es tracta de les vocals encerclades a la figura 4.

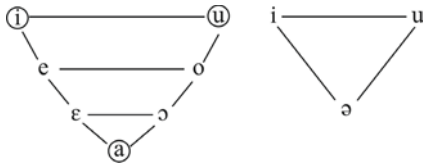
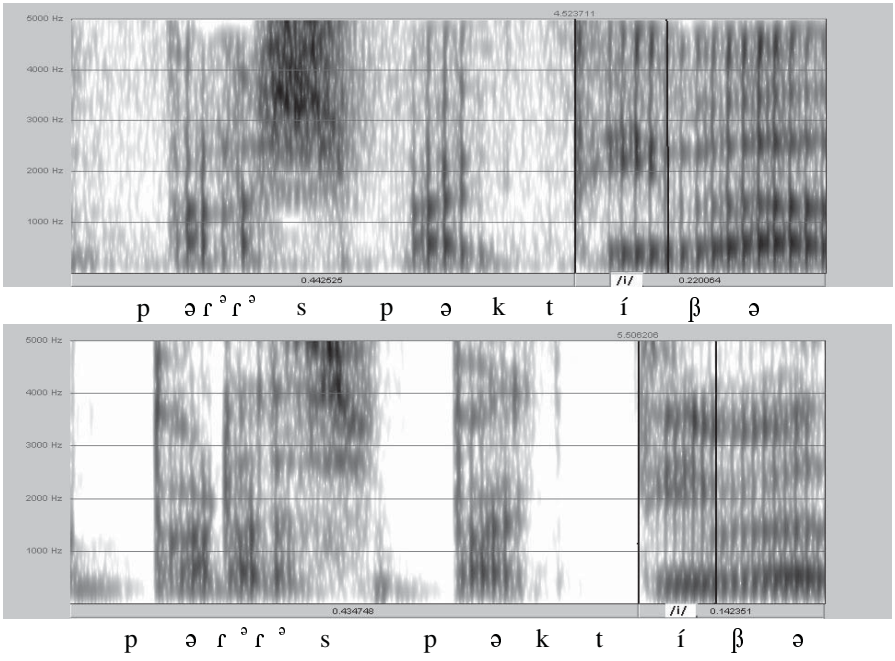


Figura 4. *Representació fonètica dels subsistemes vocàlics, tònic i àton, del català oriental barceloní, inclosos en un sistema de coordenades on l'ordenada indica l'altura de la llengua (més alta per a* [i]-[u] *i més baixa per a* [a]*), i l'abscissa la posició anteroposterior de la llengua (més avançada per a* [i] *i més endarrerida per a* [u]*).*

Cadascuna de les realitzacions vocàliques es troben en la posició tònica de les paraules seleccionades, entre elles: *sigui*, *xifres* i *aquí* per a la vocal [i], *càrrec*, *descomptat* i *candidata* per a la vocal [a], i *Catalunya*, *preguntes* i *tingut* per a la vocal [u]. A la figura 5 tenim un exemple d'[i] amb la paraula *perspectiva*.



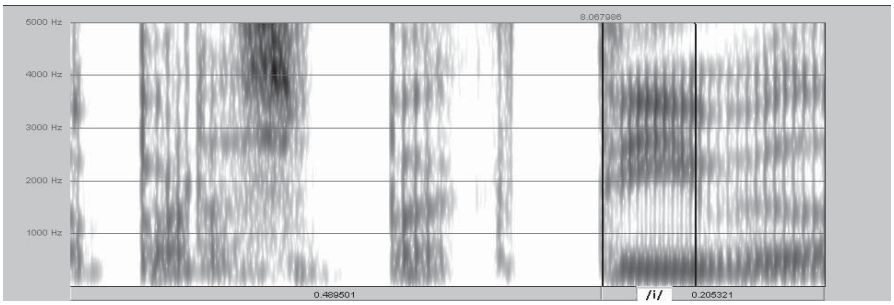


Figura 5. *Espectrogrames de la paraula* perspectiva *pronunciada, respectivament de dalt a baix, per JS en expressió espontània, per CC llegint JS sense imitar i per CC imitant JS. De pas, es pot observar, a partir de les vibracions glòtiques (les franges verticals), un tret que no tenim en compte a l'experiment, i és que la veu de JS és més greu –hi ha menys franges en la mateixa quantitat de temps– que la de l'imitador CC en totes dues modalitats.*

2.3. Anàlisi de les hipòtesis

Abans de continuar, convé establir com es poden manifestar els resultats. Si pretenem prendre el camp de dispersió com un punt de referència almenys relativament constant perquè es pugui identificar el seu locutor, estem admetent que ha d'oferir uns resultats freqüencials també relativament constants. Aquesta certesa és essencial, almenys en la pronunciació espontània d'un parlant en condicions estables de salut i dintre d'una mateixa edat. Si en aquestes condicions aquells resultats freqüencials varien –sense que puguem determinar tampoc quant, abans de fer un seguit raonable de comprovacions–, el supòsit bàsic i les possibles derivacions se'n van en orris. A partir d'això, les dues següents qüestions obren unes perspectives encara més indeterminades:

1. Si un parlant distorsiona la seva veu, com quedaran afectats els valors dels camps de dispersió?

2. I si simula la veu d'un altre parlant fins al punt d'induir a la confusió, de provocar la pèrdua d'indicis diferencials en la percepció de molts testimonis, quins seran els resultats relatius de les tres veus?

Val la pena afegir alguns aspectes de subjectivitat en tot això. I és que, per un costat, els imitadors declaren no ser gens conscients de les modificacions que introdueixen en la seva articulació. Fan provatures fins que l'efecte queda prou reeixit, per descomptat comptant amb el reconeixement de persones alienes. Això resulta molt més exigent en el nostre cas per diverses raons:

1. Moltes d'aquestes imitacions s'han popularitzat sobretot per TV, on intervé, sens dubte d'una manera primordial, la caracterització física dels personatges.

2. En les imitacions habituals, tant a la TV com a la ràdio, els imitadors també caricaturen el discurs dels personatges exagerant sovint la dicció i fins i tot introduint expressions pròpies.

En el nostre experiment no s'ha pogut apel·lar a cap d'aquest recursos complementaris. La imitació no sols s'ha limitat a la veu –una imitació que, en tots els casos, ha quedat plenament reconeguda per tothom que l'ha sentit–, sinó que s'ha cenyit estrictament al mateix material lingüístic, expressat, a més a més, de la mateixa manera.

Cal presentar l'experiment que proposem afirmant en primer lloc que, almenys a partir d'ell mateix, no disposem de cap hipòtesi sobre si existeix o no un conjunt de trets que permeten establir d'una manera prou aproximada –recordem-ho, dins dels requeriments judicials– la identificació de locutor. Com a molt, això només es podrà albirar si es perfila cap manifestació en aquest sentit.

En aquest punt, hi ha, però, un important problema metodològic afegit. I és que la naturalesa de l'experiment que proposem no en deixa interpretar, almenys fàcilment, els resultats. Val a dir, per tant, que es tracta d'un repte epistemològic insòlit i, ens atreviríem a dir, irresistible, atès que no partim, com és habitual (per no dir gairebé inevitable), de cap hipòtesi inicial...

De fet, hi ha diverses hipòtesis extremes i contraposades que van des de creure fins a no creure en l'existència en la veu de trets biomètrics que apunten d'una forma prou versemblant la identitat del locutor, almenys a partir de confrontacions entre veus no alterades per factors aliens com ara el temps, un trastorn o una simulació.

Si ho creiem, es tracta de trobar aquells trets a partir de principis i procediments més o menys complexos, que poden anar des d'una única comprovació de dades

úniques i definides, com les que proposem aquí a partir del camp de dispersió fonològica de les vocals, fins a una estimació conjunta de pocs, molts o moltíssims trets més o menys diversos. En aquest aspecte, cal remarcar dues qualitats bàsiques sobre el nostre experiment:

1. En primer lloc, que el conjunt i la distribució de les dades de què hem disposat per a una comparació així, a partir de tres versions sobre textos idèntics, són òptimes, pràcticament immillorables. Com a contrapartida, això mateix les fa pràcticament irrepetibles en comprovacions reals, almenys en àmbits forenses.
2. En segon lloc, per això mateix l'experiment constitueix per si sol una possible confirmació, encara que no definitiva, de la hipòtesi i dels plantejaments potencials que admet, però no del seu refús.

En efecte, aplicant la hipòtesi a l'experiment, hi ha almenys quatre alternatives extremes pel que fa a les possibles estimacions sobre el camp de dispersió fonològica:

1. Que els imitadors presentin unes dades constants tant si s'expressen amb la seva veu espontània com si n'imiten qualsevol altra d'un parlant distint, les dades del qual es diferenciïn sempre d'una manera igualment inequívoca.
2. Al contrari, que les dades vinguin a confirmar allò que percep l'oïda i que els valors freqüencials oscil·lin d'una manera corresponent.
3. Que apareguin estimacions poc o molt ininterpretables.
4. Que es produeixi una mescla de dues de les alternatives anteriors o de totes tres.

Al marge de la seva respectiva versemblança, vistes de fora estant, és evident que la primera alternativa confirmaria tota sola la hipòtesi sobre l'existència de trets identificadors de locutor, mentre que la segona, tot i que no la contradiria d'una manera conclusiva, descartaria l'experiment actual almenys com a prova única.

De fet, les tres últimes alternatives, i fins i tot la primera, no descarten la possibilitat d'introduir alguna correcció o millora metodològica afinant al màxim el

moment freqüencial en què es fa la comprovació, posant en consideració un tercer formant (i un quart, etc.), combinant les dades amb altres estimacions, etc.

Creiem, en canvi, que la tria de només tres vocals és prou representativa de tot el que podria oferir el conjunt vocàlic del català oriental barceloní, el que utilitzen tots els participants en l'experiment.

3. L'EXPERIMENT

3.1. Alguns punts de referència

Presentarem les dades en forma tant numèrica com gràfica. I prendrem com a punt inicial de referència els triangles de la figura 3 tenint en compte que les seves dades ofereixen un valor només aproximat –però, sens dubte, prou vàlid– que deriva de l'observació directa sobre la taula de coordenades. Així, si observem la taula 1, podem fer les següents estimacions (que vénen a precisar numèricament el que ja s'observa visualment a la figura 3):

AB				
[i]-[u]	F_1	270	270	0
	F_2	2000	480	1520
[i]-[a]	F_1	270	580	310
	F_2	2000	1100	900
[a]-[u]	F_1	580	270	310
	F_2	1100	480	620

RC				
[i]-[u]	F_1	200	200	0
	F_2	2700	470	2230
[i]-[a]	F_1	200	800	600
	F_2	2700	1500	1200
[a]-[u]	F_1	800	200	600
	F_2	1500	470	1030

Taula 1. *Valors freqüencials aproximats (en Hz) dels dos primers formants i de les respectives diferències, a la columna de la dreta, d'*[i], [a] *i* [u] *a partir dels triangles d'AB i RC.*

1. Els valors d'F_1 són gairebé iguals o similars, excepte per a [a], molt més oberta a RC.

2. Els valors d'F_2 són similars a la sèrie velar, però molt més alts a [a] i a la sèrie palatal a RC. Les diferències són poc o gens significatives tenint en compte l'extracció, tan diferent, de les respectives dades (cf. 2.1).

3. Més important, potser, és l'equidistància relativa de tots els valors vocàlics.

4. I encara ho sembla més la simetria també relativa d'ambdós triangles, que en el seu moment es podia –i es va– interpretar com un fet derivat de la disposició logarítmica dels valors freqüencials en les coordenades per tal de donar una imatge considerada 'natural' des dels tractats clàssics de fonologia.

3.2. Les dades

A títol d'exemple, i a partir del que hem exposat a 2.1, a la figura 6 tenim una mostra de la constel·lació de vint realitzacions vocàliques d'[a] per triplicat i la situació de cadascuna en un sistema logarítmic de coordenades similar als de les figures anteriors, on l'ordenada correspon al primer formant (F_1) i l'abscissa al segon (F_2).

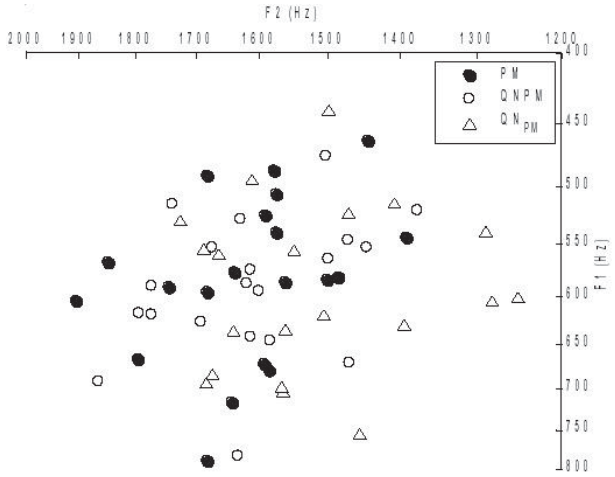


Figura 6. *Realitzacions d'*[a] *distribuïdes per parlants, d'acord amb els símbols del requadre. Les negres, del personatge PM, provenen*

d'una pronúncia espontània extreta d'una entrevista; les blanques són de l'imitador QN i, entre elles, les triangulars extretes d'una repetició literal, però espontània, de l'anterior, i les circulars d'una imitació plena.

A la taula 2 tenim tots els valors freqüencials mitjans de cada constel·lació vocàlica per a cada parlant acompanyats del respectiu índex d'oscil·lació. Aquí i en general, les abreviatures (a part dels personatges PM, XT i JS) equivalen al següent: QN_{PM}, QN llegint PM; QNPM, QN imitant PM; QN_{XT}, QN llegint XT; QNXT, QN imitant XT; CC_{JS}, CC llegint JS; CCJS, CC imitant JS.

	[i]		[a]		[u]	
	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)
QN_{PM}	323 ± 34	2127 ± 94	599 ± 82	1524 ± 143	441 ± 97	1418 ± 486
PM	388 ± 84	2026 ± 120	588 ± 82	1625 ± 130	494 ± 78	1503 ± 353
QNPM	402 ± 166	2023 ± 121	595 ± 71	1618 ± 131	496 ± 111	1454 ± 432
QN_{XT}	337 ± 31	2184 ± 170	639 ± 74	1548 ± 152	423 ± 108	1355 ± 424
XT	337 ± 110	2326 ± 167	604 ± 104	1564 ± 155	409 ± 209	1281 ± 473
QNXT	349 ± 42	2331 ± 208	685 ± 80	1632 ± 201	441 ± 121	1226 ± 504
CC_{JS}	329 ± 37	2214 ± 199	641 ±46	1390 ± 130	397 ± 30	1136 ± 345
JS	368 ± 28	1971 ± 232	661 ± 86	1399 ± 224	409 ± 65	1106 ± 286
CCJS	272 ± 26	2223 ± 196	618 ± 61	1429 ± 85	375 ± 66	1096 ± 371

Taula 2. *Valors freqüencials mitjans, amb l'índex d'oscil·lació o desviació típica, dels dos primers formants de les vocals* [i], [a] *i* [u], *respectivament, a partir de les realitzacions analitzades.*

A partir d'aquests valors mitjans obtenim –d'una manera similar a com hem procedit més amunt per passar de la figura 2 al triangle de la figura 3–, el conjunt de triangles de la figura 7 acompanyats dels respectius valors numèrics a la taula 3 (anàloga a 1), amb les freqüències i la diferència, positiva o negativa, a la dreta.

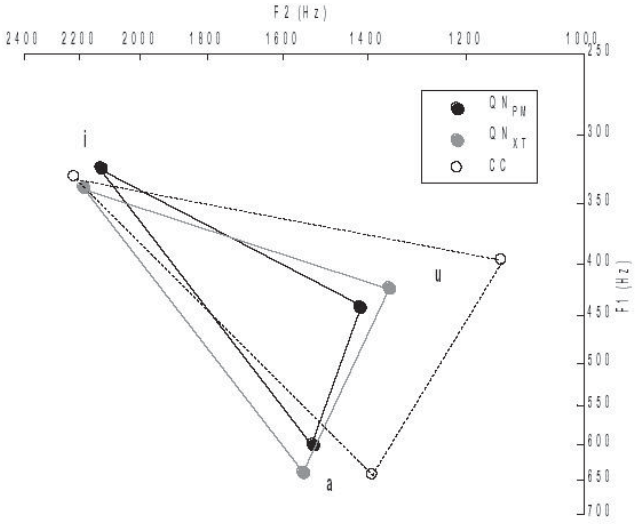


Figura 7. *Triangles de QN i CC en pronúncia espontània –és a dir, sense imitació, excepte en la prosòdia, per raons de naturalitat– repetint textos del personatges imitats. Així, QN_{PM} (cercle negre) i QN_{XT} (cercle gris) són els triangles de QN reproduint textos de PM i XT, respectivament, i CC és el triangle de CC reproduint textos de JS.*

		QN_{PM}			QN_{XT}			CC_{JS}		
[i]-[u]	F_1	323	441	118	337	423	86	329	397	68
	F_2	2127	1418	709	2184	1355	829	2214	1136	1.078
[i]-[a]	F_1	323	599	276	337	639	302	329	641	312
	F_2	2127	1524	603	2184	1548	636	2214	1390	824
[a]-[u]	F_1	599	441	158	639	423	216	641	397	244
	F_2	1524	1418	106	1548	1355	193	1390	1136	254

Taula 3. *Valors numèrics (en Hz) de QN_{PM}, QN_{XT} i CC_{JS} amb les seves diferències freqüencials.*

En una primera observació es comprova d'immediat que els triangles derivats de la parla espontània de QN són molt afins, al costat del que mostra la parla espontània de CC. Encara que això sembli lògic, s'imposen almenys dues observacions immediates.

La primera prové de preguntar-nos per què, tanmateix, no coincideixen plenament els triangles de QN. La diferència no sembla que pugui derivar dels textos llegits, excepte que la prosòdia –o una imitació subliminar– hagin influït en l'execució vocàlica. És molt probable també que els resultats tendirien a ser més coincidents a mesura que augmentés el nombre de comprovacions.

La segona observació deriva sobretot de l'estructura poc simètrica dels triangles de QN. I no pas per una raó purament estètica, sinó perquè cal tenir en compte que entre [i] i [a] hi ha dues realitzacions vocàliques –[e] i [ɛ]–, de la mateixa manera que també n'hi ha dues més entre [u] i [a] –[o] i [ɔ]–. De fet, els valors del F_1 d'[i] i [u] són extremadament diferents al costat no sols dels que es desprenen de la figura 3 i de les taules 1 i 2, sinó també dels que sempre s'han considerat teòricament «normals». Aquesta diferència, que segueix confirmant-se en les altres comprovacions, fa pensar en la possibilitat d'haver utilitzat mitjans, i potser criteris i tot, diferents per extreure els valors formàntics d'[u] (recordeu el que hem dit a 2.1 sobre això). Podríem fins i tot sospitar d'alguns errors produïts en l'extracció automàtica mitjançant l'analitzador acústic Praat (Boersma et al. 1992), la qualitat de l'enregistrament o la influència reiterada del context de la vocal [u] com a algunes de les causes.

La qüestió primordial, tanmateix, no és comparar les dades triangulars de l'experiment actual amb les de Cerdà 1972, ni els criteris respectius, sinó les dades que ofereix l'actual experiment entre elles mateixes. Per aquesta raó, prescindim de raonar sobre trets com ara la simetria del triangle.

A partir del que observema la figura 8 s'imposa també un parell d'observacions. Primer de tot, veiem que s'accentua encara més l'apropament del F_1 entre [u] i [a] en el triangle de PM.

I pel que fa als canvis que experimenta el triangle de QN quan imita PM (figura 9), cal remarcar que s'aproxima molt i molt en tots els valors, amb una coincidència gairebé plena per a [a]. Segons això, es confirmaria que la imitació afecta de ple els resultats dels camps de dispersió vocàlica i, de pas, que les variacions més o menys esporàdiques i/o voluntàries de veu els modifiquen.

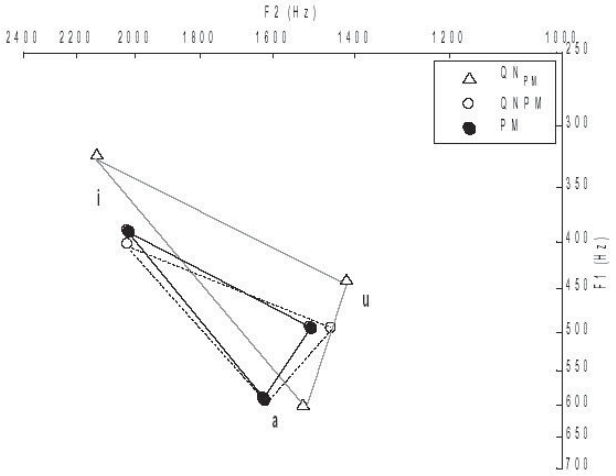


Figura 8. *Triangles de PM (cercle negre) en parla espontània, de l'imitador QN (triangle blanc) repetint també espontàniament el mateix text, i de QN (cercle blanc) imitant plenament PM també amb el mateix text.*

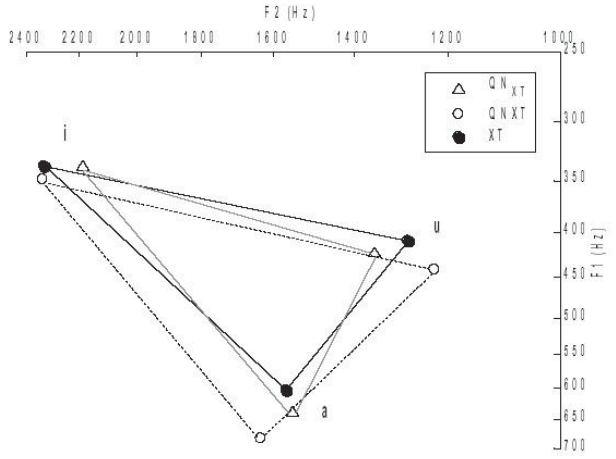


Figura 9. *Triangles de XT (cercle negre) en parla espontània, de l'imitador QN (triangle blanc) repetint també espontàniament el mateix text, i de QN (cercle blanc) imitant plenament XT també amb el mateix text.*

Aquí els resultats són només en part diferents i inesperats. I és que els triangles espontanis de XT i de QN s'assemblen més que no pas els de XT i de QN imitant-lo, especialment pel que fa a [a]. De fet, tots tres triangles són relativament similars des del punt de vista freqüencial, cosa que no descarta que hi hagi una important diferenciació entre les respectives veus espontànies i entre elles i la imitació, que sol basar-se en moltíssims més trets que no pas els que considerem aquí. Pensem en la pronúncia de la vocal neutra, de les consonants –especialment d'algunes, com ara [ɾ] i [r], en el cas d'XT– i sobretot de les modulacions prosòdiques. Les diferències entre tots tres triangles són mínimes, en definitiva.

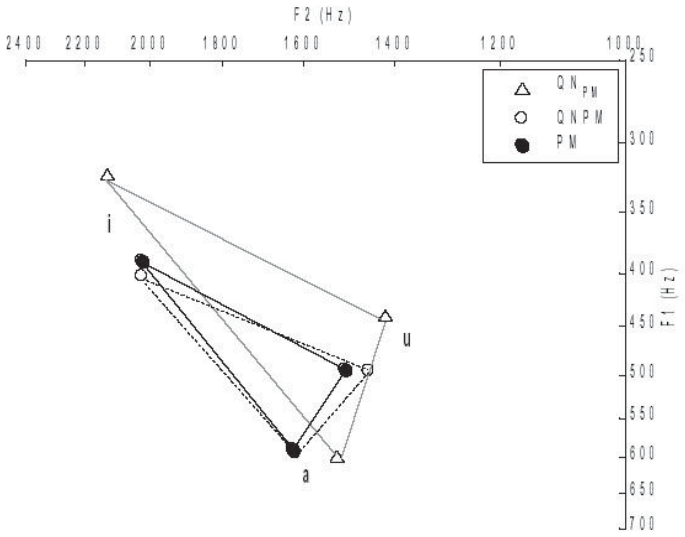


Figura 10. *Triangles de JS (cercle negre) en parla espontània, de l'imitador CC (triangle blanc) repetint també espontàniament el mateix text, i de CC (cercle blanc) imitant plenament PM també amb el mateix text.*

Per sort, aquí (figura 10) els resultats vénen a contradir el que semblava albirar-se fins ara, i és que, pel que fa només a la [i], quan CC imita JS s'allunya dels valors freqüencials no sols de JS sinó fins i tot dels propis. I és una sort perquè almenys evita incórrer en la interpretació immediata que es desprenia de les comprovacions anteriors i obre la necessitat de recórrer a altres consideracions.

4. DUES CATEGORIES DE CONCLUSIONS

4.1. Conclusions específiques arran de l'experiment

Resumim: de molt s'ha lamentat que les proves que solen fer-se als tribunals sobre identificació de locutor pequen d'incoherència perquè solen treballar, d'una manera gairebé inevitable, sobre enregistraments extrets en circumstàncies diferents, que sovint es basen en converses telefòniques i en lectures. I encara que la lectura, generalment forçada, d'un text reprodueixi el mateix contingut lingüístic d'una conversa telefònica, hi ha tot un seguit de condicionaments massa desfavorables perquè la comprovació resulti vàlida, en especial pels canvis prosòdics, tan involuntaris com intencionalment provocats, al costat de molts altres factors incidentals com els que hem vingut assenyalant.

Per tot plegat, des de fa molt de temps s'ha proclamat la necessitat de comparar diferents enregistraments d'un mateix locutor. És el que hem aportat en el nostre experiment, amb condicionaments específics i resultats tendencials cap a la cerca d'una constant biomètrica entre la munió de factors que intervenen en la veu humana. Per això, des del principi hem parlat d'indicis i no de resultats.

Comparant-los amb els resultats de Farrús (2008), obtinguts a partir d'experiments totalment distints, sembla ben clar que els trets prosòdics s'alteren molt més en les imitacions de veu que no pas els triangles vocàlics, que es mantenen més estables.

Sembla confirmat, en tot cas, que l'execució de vocals és prou subliminar perquè els camps de dispersió constitueixin un punt de referència relativament fiable per a la identificació de locutor. De moment almenys, no és prudent anar més lluny en la conclusió.

4.2. Conclusions d'interès lingüístic

La comparació de dades entre locutors i elocucions tan pròximes amb resultats absoluts tan diversos entre les freqüències analitzades ens recorda no poques interpretacions maximalistes publicades des de fa anys a partir de resultats fins i tot més ajustats. Ens referim a interpretacions que arriben a «descobrir» fronteres no sols dialectals sinó fins i tot lingüístiques a partir de diferències freqüencials procedents de proves independents. S'ha arribat a concloure, p. ex., que una fron-

tera dialectal passa per una diferència d'uns 200 Hz entre dos formants vocàlics obtinguts de veus diferents o que en tal contrada no es parla la mateixa llengua que en una altra perquè els triangles vocàlics de dos informants considerats modèlics o ideals dels respectius indrets no coincideixen, essent un dels triangles més o menys allargat, ample, asimètric, etc. que l'altre.

Esperem que el nostre experiment serveixi també per aclarir alguns conceptes fins i tot fora del seu àmbit específic.

AGRAÏMENTS: Aquest treball no hauria estat possible sense la contribució de Mireia Farrús Cabeceran, becària i doctoranda del Departament de Teoria del Senyal de la Universitat Politècnica de Catalunya en el moment de redactar això, juny de 2008. Ella ha ajudat a validar la hipòtesi de treball, ha subministrat les dades de l'experiment, aprofitant els materials de la tesi, i ha contribuït activament a la redacció, fins i tot del títol.

5. REFERÈNCIES BIBLIOGRÀFIQUES

ADERMAN, T. B. (2005): *Forensic Speaker Identification. A Likelihood Ratio-based Approach Using Vowel Formants*, Munich, Lincom GmbH.

BOERSMA, P. i D. WEENIK (1992): *Praat: Doing phonetics by computer,* Institute of Phonetic Sciences, Universitat d'Amsterdam, Holanda.

CARRERA-SABATÉ, J. i A. M. FERNÁNDEZ-PLANAS (2005): *Vocals mitjanes tòniques del català. Estudi contrastiu interdialectal*, Quaderns per a l'Anàlisi 18, Barcelona, Horsori.

CERDÀ MASSÓ, R. (1972): *El timbre vocálico en catalán*, Madrid, CSIC, Instituto Miguel de Cervantes.

CERDÀ, R.; M. FARRÚS; J. HERNANDO i M. VEYRAT (2007): «Propuesta de experimento sobre la noción de campo de dispersión fonemática», *III Congreso da Sociedade Española de Acústica Forense*, Xunta de Galicia, Santiago de Compostela, pp. 127-39.

FARRÚS, M.; J. HERNANDO i P. EJARQUE. (2007): «Jitter and shimmer measurements for speaker recognition», *Actes de l'Eurospeech*, Antwerp, Bèlgica, pp. 778-781.

FARRÚS, M. (2008): *Fusing Prosodic and Acoustic Information for Speaker Recognition*, tesi doctoral, Universitat Politècnica de Catalunya, TALP Research Center, Barcelona.

FERNÁNDEZ PLANAS, A. M. (1993): «Estudio del campo de dispersión de las vocales castellanas», *Estudios de Fonética Experimental*, V, pp. 129-162.

HARRISON, P. i FRENCH (2007): «Forensic Speech Analysis and the Logically Coherent Expression of Conclusions», *Proceedings of the 2nd European IAFL Conference on Forensic Linguistics / Language and the Law*, IULA, Universitat Pompeu Fabra, Barcelona, pp. 57-69.

NOLAN, F. (1983): *The Phonetic Bases of Speaker Identification*, Cambridge, Cambridge University Press.

PESKIN, B.; J. NAVRÁTIL; J. ABRAMSON; D. JONES; D. KLUSÁČEK; D.A. REYNOLDS i B. XIANG (2003): «Using prosodic and conversational features for high-performance speaker recognition», *Actes de l'ICASSP*, Hong-Kong, Xina, pp. 792-795.

PUIG I RIERA, J. i J. FREIXA I AYMERICH (1990): «El camp de dispersió de les vocals catalanes des del punt de vista de la percepció», *Estudios de Fonética Experimental*, IV, pp. 123-46.

QUILIS, A. (2000): «El reconocimiento de la voz en la investigación judicial. La experiencia del lingüista», a Carbonero Cano, P. *et al.* (coord.): *Lengua y discurso. Estudios dedicados al Profesor Vidal Lamíquiz*, Madrid, Arco/Libros, S. L., pp. 783-9.

RECASENS, D. i MA. D. PALLARÈS (2001) *De la fonètica a la fonologia. Les consonants i assimilacions consonàntiques en català*, Barcelona, Ariel.

ROSE, P. J. (2002): *Forensic Speaker Identification*, Londres, Taylor & Francis.

ZETTERHOLM, E. (2003): *Voice Imitation. A phonetic study of perceptual illusions and acoustic success*, tesi doctoral, Departament de Lingüística i Fonètica, Universitat de Lund, Suècia.