

La identificació de la veu humana

ÁNGEL RODRÍGUEZ BRAVO

Departament de Comunicació Audiovisual i Publicitat
de la Universitat Autònoma de Barcelona

85

1. INTRODUCCIÓ

La gran popularitat de tot tipus de tecnologies de l'àudio ha facilitat d'una manera enorme la comunicació mitjançant la veu a través del telèfon convencional, el radiotelèfon, els contestadors automàtics, els cassets, els ordinadors personals, els minidiscs, etc. Com a conseqüència directa d'això, avui és relativament fàcil aconseguir registres electromagnètics de converses entre persones relacionades amb delictes greus. Aleshores, sorgeixen força sovint situacions policials com la següent: «s'ha aconseguit una gravació determinada d'una veu i s'ha comprovat posteriorment, sense cap mena de dubte, que la seva font va ser una persona que va cometre el delictes que s'està investigant».

La qüestió que es planteja, doncs, és l'existència d'un registre en una cinta magnètica de la veu del delinqüent (veu no dubitativa), que es podria transformar en una prova conclouent de culpabilitat o d'innocència de les persones presumptament implicades en el cas. No obstant això, perquè aquesta gravació es pugui utilitzar com a prova judicial, cal demostrar sense cap risc d'error que alguna de les veus dels sospitosos (veus dubitatives) coincideix exactament amb la de la gravació.

Sovint descobrim a la nostra vida quotidiana que som capaços de reconèixer algú solament si el sentim parlar. Així, l'experiència auditiva humana ens indica que, certament, un locutor es podria identificar a partir del so de la seva veu si es disposa d'uns instruments tècnics determinats i dels coneixements científics adequats. Aquesta intuïció, és encertada? Disposem realment dels instruments i els coneixements que permetrien la identificació d'un sospitós a través de l'estudi de la seva veu? I si aquesta resposta fos afirmativa, quins són aquests instruments i en quin tipus de coneixements científics s'han de basar per ser utilitzats?

En aquestes pàgines intentaré donar algunes respostes a aquestes preguntes que, en realitat, constitueixen un territori comú a la majoria dels problemes científics que tenen com a objecte d'estudi la veu humana.

2. ELS INSTRUMENTS DISPONIBLES

El desenvolupament vertiginós de la informàtica durant els darrers vint anys ha donat un impuls molt fort a l'acústica aplicada. Com a conseqüència, les disciplines com ara la fonologia, la psicoacústica, la fonètica experimental o la lingüística computacional han avançat molt. Ara, a qualsevol laboratori d'anàlisi i tractament de la parla disposem d'instruments molt potents que possibiliten el desenvolupament d'un tipus de treballs que abans eren impensables. Certament, això ha fet avançar molt els coneixements sobre la veu.

No obstant això, la potència dels instruments d'anàlisi acústica ha avançat, en realitat, molt més ràpidament que el coneixement bàsic sobre els objectes d'estudi que és possible analitzar-hi. El que vull dir és que el desenvolupament de la tecnologia d'anàlisi comporta per si mateixa el coneixement d'allò analitzable. Avui qualsevol sistema d'anàlisi acústica és capaç de llançar milers de dades fiables a partir d'uns quants segons de gravació d'una veu humana; però aquest cúmul d'informació no és en absolut directament interpretable. Per l'estudiós de la veu, la major part d'aquesta informació encara és críptica o, en el millor dels casos, sumament confusa. I el seu esforç se centra encara a buscar models que li permetin processar-la i utilitzar-la adequadament.

És cert que actualment es comercialitzen al mercat nombrosos sistemes d'anàlisi acústica i que són de fàcil accés; podem esmentar-ne alguns que ja són clàssics i d'alt nivell, com el recent PULSE de Brüel & Kjaer, el CSL de Kay Elemetrics, o el Mac Speech Lab d'Apple. També hi ha sistemes més especialitzats en paràmetres concrets, com el SIGNAIX desenvolupat en el Laboratoire de Langage et Parole de la Universitat d'Aix-en-Provence. Al mercat n'hi ha molts altres, sempre amb la mateixa base científica de treball, per bé que amb unes prestacions i uns preus molt diversos. I també és cert que qualsevol d'aquests instruments es pot utilitzar per resoldre problemes d'identificació de la veu humana. Però hem de tenir en compte que tots són equips d'aplicació general a l'anàlisi acústica en les seves diverses facetes.

A diverses universitats s'està treballant directament amb el desenvolupament de sistemes automàtics per al reconeixement del locutor; en el cas d'Espanya podríem esmentar les investigacions que s'estan fent en el Departament de Matemàtica Aplicada de la Universitat de Múrcia, en el Departament de Teoria del Senyal i de Comunicacions de la Universitat Politècnica de Catalunya, o en el Departament d'Enginyeria Audiovisual i Comunicacions de la Universitat Politècnica de Madrid. Tanmateix, els resultats que s'han obtingut fins ara encara són molt parcials. En realitat no existeixen encara sistemes comercials orientats a resoldre un problema tan concret com la identificació del locutor, que proporcionin resultats d'identificació satisfactoris amb qualsevol tipus de veu. Podem trobar sistemes més o menys ben adaptats a l'amidament d'uns o altres paràmetres, o relativament especialitzats en l'anàlisi del soroll industrial o de la veu; però els que són realment fiables són sempre instruments d'anàlisi genèrica que requereixen amplis coneixements tant d'acústica com del fenomen concret que s'està analitzant perquè puguin ser explotats adequadament.

3. ELS CONEIXEMENTS NECESSARIS

Com he dit més amunt, encara que les investigacions sobre la veu humana han tingut un impuls molt gran en les darreres dècades, la parla s'ha anat revelant progressivament com un objecte d'estudi amb una complexitat extraordinària i que només està resolt molt parcialment.

A hores d'ara, la investigació aplicada a la parla s'orienta majoritàriament cap a la síntesi i el reconeixement automàtic, buscant l'immens camp d'aplicació que es preveu per a aquest tipus de sistemes en la telefonia, la traducció automàtica i la comunicació home-màquina (màquines que escriguin al dictat, màquines que parlin, màquines que entenguin la parla, etc.).

Aquesta orientació té, a parer meu, un problema important des del punt de vista de la investigació policial. Per a aquesta mena d'aplicacions és molt més interessant l'estudi dels factors que són comuns a totes les veus, que no pas investigar i buscar les característiques específiques i úniques de cada locutor concret. És a dir, la perspectiva d'investigació sobre lingüística aplicada que és actualment majoritària avança justament en el sentit contrari al que necessitaria la identificació de locutors. A més, veig que aquest no és un problema conjuntural que depengui d'una moda passatgera. La lingüística mateixa és, en el seu sentit més profund, una disciplina orientada a la localització i la formalització de les estructures sonores comunes que comparteixen els parlants per intercomunicar-se.

Així, ni la lingüística ni les seves disciplines derivades —fonètica experimental, fonologia, sociolingüística, etc.— tendeixen a observar i sistematitzar els trets acústics que caracteritzen la individualitat del locutor (que són els que necessita la investigació policial), sinó que els ocupa precisament el contrari. És a dir, allò que és intercanviable, allò que comparteixen tots els parlants.

És indiscutible que molts coneixements desenvolupats per la fonètica experimental són directament aplicables al problema de la identificació del locutor. Un exemple molt clar sobre aquesta qüestió és tota la investigació duta a terme al voltant de l'anàlisi de l'espectre mitjà a llarg termini o SMLT (cfr. HARMEGNIES, 1992). Evidentment, també és aplicable a la identificació el coneixement sobre certes característiques concretes de la composició espectral dels fonemes, o tota la informació ja disponible sobre trets acústics sociolingüístics. No obstant això, com veurem detalladament una mica més endavant, per a l'acústica forense aquest tipus de coneixements genèrics encara és insuficient.

4. TÈCNiques APPLICABLES A LA IDENTIFICACIÓ DEL LOCUTOR

Certament, existeixen una sèrie de procediments d'anàlisi acústica que són aplicables com a punt de suport científic per decidir si una veu dubitativa determinada va ser emesa o no per un locutor concret. A continuació, revisarem a grans trets els corrents d'estudi que proporcionen instruments útils per a la identificació a través de la veu.

Podríem establir tres grans línies de treball al voltant de la veu que ja han estat utilitzades amb força regularitat i que aporten solucions útils per a la identificació del locutor:

- l'anàlisi sociolingüística
- l'anàlisi manual de zones puntuals de l'espectre
- l'anàlisi automatitzada global de l'espectre

4.1 L'anàlisi sociolingüística

88

Aquest procediment es basa en la localització i la sistematització dels trets acústics i articuladoris que estan associats a la manera de parlar dels diferents grups a què pertany l'individu, com ara el medi geogràfic, l'origen social, el grau de cultura, la professió, etc. A banda dels trets estandarditzats que es configuren a cada llengua, a cada zona geogràfica, a unes professions determinades (per exemple, en el *singing formant* o formant del cant), als diferents dialectes, etc., apareixen unes característiques sonores específiques que només són localitzables en un grup concret. Aquests caràcters acústics sociolingüístics són observables en les formes entonatives, en l'espectre d'unes vocals determinades, en maneres concretes d'articular algunes consonants, etc.

Òbviament, l'ús adequat d'aquest tipus d'informació per part d'un expert permet observar tècnicament els trets sociolingüístics de la veu no dubitativa i determinar si aquests apareixen en les veus dubitatives. És evident que el procediment és molt útil per descartar les veus en què no estan presents els trets sonors que s'estan buscant. Però no hi ha prou garanties únicament a partir d'aquest tipus d'anàlisi per afirmar inequívocament que el propietari de la veu en què hem localitzat les coincidències és qui va emetre la veu no dubitativa. El mateix caràcter de la metodologia amb què es treballa parteix de la premissa que aquests trets acústics no són específics d'un sol individu, sinó que són característics d'un grup determinat. Per tant, el resultat d'aquest tipus d'anàlisi no ens identifica d'una manera específica el subjecte, perquè sabem positivament que n'hi ha molts altres amb aquestes mateixes característiques a la seva veu.

Així, el procediment només és útil, en realitat, com a suport complementari de la globalitat de la informació sobre el cas.

4.2 L'anàlisi manual de zones puntuals de l'espectre

Aquest procediment consisteix a estudiar detalladament unes parts molt concretes d'un fragment de la veu no dubitativa, de les quals sabem que aporten informació específica del subjecte. És el cas, per exemple, de la «distància» entre els dos primers formants dins dels marges de variabilitat de cadascuna de les vocals. La dada concreta que aporta informació significativa és la que s'obté quan es resta la freqüència del primer formant a la del segon, en sons vocàlics concrets, és a dir, fent servir l'índex ($F_2 - F_1$). D'acord amb la informació de què jo dispo en aquests moments, aquest tipus de dades s'ha d'estudiar comparant sempre entre

la mateixa vocal de les mateixes paraules. Aquesta informació reflecteix les posicions internes de la boca específiques del locutor.

El procediment de la distància entre formants ha demostrat la seva utilitat per identificar actors que estaven construint personatges diferents, és a dir, que emmascaraven i manipulaven la seva veu voluntàriament. Dit d'una altra manera, els resultats de l'anàlisi estadística de les dades que s'obtenen quan es calcula l'índex ($F_2 - F_1$) vocal per vocal, mostren que en contrastar el resultat de l'anàlisi acústica de locutors diferents, apareixien diferències significatives. Contràriament, quan la contrastació es feia entre les distintes interpretacions d'un mateix locutor, l'índex ($F_2 - F_1$) no reflectia diferències significatives (cfr. RODRÍGUEZ, 1989). Aquest índex també l'han utilitzat a Itàlia altres investigadors i n'han obtingut resultats molt semblants.

Sigui com sigui, cal tenir en compte que l'índex ($F_2 - F_1$) no va ser eficaç en el 100 % dels casos. Es van estudiar vuit interpretacions de quatre actors diferents (dues interpretacions per actor) i el procediment es va mostrar eficaç per identificar el locutor a totes les combinacions comparatives possibles, tret de la que contrastava les dues interpretacions d'una de les locutores. En aquest cas, la diferència entre totes dues versions va ser prou bona per enganyar el mètode. En resum, el marge d'incert de l'índex era del 96,4 %. Certament molt robust si tenim en compte el tipus de prova a què s'havia sotmès.

És perfectament possible trobar altres índexs d'aquesta mena i fer-los servir combinats, cosa que reforçaria les garanties dels resultats. De totes maneres, considero que aquest tipus de treball requereix línies d'investigació bàsica que siguin específiques, és a dir, les investigacions han de tenir la identificació del locutor com a objectiu concret, en lloc de recollir únicament els resultats complementaris d'altres projectes que treballen amb objectius essencialment lingüístics.

4.3 L'anàlisi automatitzada global de l'espectre

Aquest és, sens dubte, el procediment que té més adeptes entre els enginyers. Es basa a obtenir de forma automàtica una sèrie d'anàlisis espectrals, de les quals es faran mitjanes. Es duen a terme, per exemple, unes cent anàlisis per segon de la composició de freqüències i se n'obtenen aproximadament quatre-cents valors freqüencials a cada espectre (aquestes xifres depenen del mètode d'anàlisi que s'apliqui). L'anàlisi es fa durant uns quants segons de gravació i, posteriorment, se sumen i es fan les mitjanes de totes les anàlisis a fi d'obtenir un espectre mitjà. La similitud o dissimilitud entre els espectres mitjans respectius de dues veus diferents és l'índex que es fa servir per estudiar si aquestes veus pertanyen o no al mateix locutor. Per estudiar quantitativament el grau de similitud a partir dels espectres mitjans, se sol utilitzar la distància euclidiana. D'altra banda, hi ha investigadors que han desenvolupat els seus propis índexs per objectivar la comparació.

L'espectre mitjà a llarg termini s'ha revelat, efectivament, com un índex molt útil per estudiar aspectes de la veu no dependents de la llengua. Aquest procediment s'ha aplicat per estudiar la influència en la parla dels estats emocionals, dels

trets de personalitat, del sexe i l'edat, de problemes fisiològics com la sordesa, de les patologies de la veu, etc., i se n'obtenen bons resultats (cfr. HARMEGNIES, 1992). Naturalment, també s'ha fet servir per estudiar la individualitat vocal i s'han aconseguit unes taxes d'encert en la identificació del locutor entre el 90 % i el 95 % (cfr. FURUI, 1978; TOSI, 1977 i 1979).

La base teòrica d'aquest tipus d'aproximació al problema és la convicció que cada locutor té unes dimensions específiques i invariables del tracte vocal que es reflecteixen constantment en l'espectre acústic de la seva veu. Si això fos cert, n'hi hauria prou de trobar el mètode adequat d'anàlisi espectral i aplicar-lo en un fragment suficientment llarg de la veu, per subratllar en l'espectre mitjà aquests caràcters sonors específics de l'individu. De fet, l'enginyeria del tractament de senyal persegueix des de fa molt temps desenvolupar un sistema automàtic basat en aquest tipus de procediment, que permeti comparar dues veus i que ens torni com a resposta inequívoca si pertanyen o no al mateix locutor. Tanmateix, segons la informació que en tinc, cap dels sistemes desenvolupats fins ara no ha aconseguit rebaixar les taxes d'error, que ja fa una dècada que segueixen mantenint-se sistemàticament entre el 5 % i el 10 %.

Des del meu punt de vista, aquest procediment arrossega un error fonamental en la seva base conceptual: ignora que tot el tracte vocal és globalment tou, elàstic i enormement dinàmic, i que pot ser alterat d'una manera important en totes les seves zones per factors psicològics, emocionals i d'estrès, i, a més, per la pròpia voluntat del locutor si hi té prou traça. És a dir, l'anàlisi automatitzada global de l'espectre ignora que la configuració del tracte vocal pot resultar molt modificada per a un mateix individu amb una facilitat relativa.

5. EL PROBLEMA POLICIAL DAVANT DEL PROBLEMA CIENTÍFIC

Després d'haver revisat els tres grans corrents científics d'anàlisi de la parla aplicables a la identificació, crec que és important posar en relleu les diferències que hi ha entre la concepció científica i la concepció policial del problema de la identificació del locutor a partir de la veu.

Des del punt de vista científic, un procediment d'anàlisi que doni un grau d'encert en la identificació del locutor al voltant del 95 % se sol considerar molt satisfactori, perquè demostra que les hipòtesis des de les quals partia la investigació han estat contrastades i que, per tant, no hi ha hagut errors fonamentals. Per exemple, si allò que s'estudia és l'índex ($F_2 - F_1$), un resultat d'aquest tipus significa que, efectivament, aquest paràmetre proporciona la informació que busquem encara que hi hagi alguns factors no controlats que ens distorsionen el resultat amb un error del 5 %. L'estudi es considera, doncs, finalitzat i les conclusions accepten com a científicament vàlid l'índex localitzat.

En canvi, des del punt de vista de l'aplicació policial, les condicions en què es troba aquest índex encara són força insatisfactòries. Els resultats obtinguts suposen que de cada cent processos judicials en què s'apliqui aquesta prova, cinc persones podrien ser acusades i condemnades injustament. Un marge d'error altíssim

per basar-s'hi exclusivament. Així mateix, cal afegir que mentre la qualitat acústica de les mostres amb què es treballa en un laboratori científic sol ser excel·lent (veus sense sorolls i amb una alta gamma dinàmica i espectral), les mostres que es poden aconseguir en les investigacions policials acostumen a ser de molt baixa qualitat, gravades en males condicions i mesclades amb molts sorolls. Aquest canvi radical en les condicions de les mostres analitzades fa augmentar amb facilitat el marge d'error en la identificació a un 15 % o un 20 %, la qual cosa converteix l'índex que, en un principi, era científicament vàlid en totalment inacceptable des del punt de vista pericial.

No obstant això, les males condicions de la mostra pel que fa als sorolls són un problema clàssic de l'enginyeria acústica; hi ha diverses línies d'investigació que treballen en sistemes especialitzats a millorar la identificació de la veu a partir del reconeixement i la sostracció dels sorolls que hi puguin estar barrejats (cfr. ORTEGA-GARCÍA I GONZÁLEZ-RODRÍGUEZ, 1997).

De totes maneres, les proves pericials que es fan servir en els casos en què es recorre a la identificació a través de la veu continuen basant-se en el judici d'experts, perquè encara no hi ha cap protocol sistemàtic d'anàlisi acústica que es pugui acceptar de manera universal. El problema sorgeix, aleshores, quan el dictamen d'un perit es contraposa al d'un altre que —a partir de la mateixa prova— afirma justament el contrari.

Tot comptat, l'aplicació policial dels procediments d'anàlisi acústica a la identificació, pel fet que té uns objectius molt concrets i específics, exigeix unes garanties de fiabilitat que estan molt per sobre de les que s'apliquen habitualment en els projectes d'investigació científica bàsica. Evidentment, això suposa que si es vol aconseguir el tipus de resultats que requereix la investigació policial sigui imprescindible desenvolupar projectes d'investigació amb els seus objectius explícitament orientats a les necessitats de l'acústica forense.

6. LA VARIABILITAT DE LA PARLA

Abans d'exposar les conclusions a què ens condueixen els plantejaments d'aquest treball, em sembla que cal fer una darrera reflexió global sobre els problemes objectius que ha d'afrontar la recerca de mètodes d'anàlisi per aconseguir la identificació del locutor.

Si volem obtenir uns resultats realment aplicables a la identificació, em sembla que és imprescindible que tant els serveis policials com els investigadors científics i tècnics prenguin consciència que acarem dos problemes de molta envergadura:

- a) la pròpia tendència natural de la investigació amb base lingüística a l'estudi dels trets acústics compartits i comuns, en lloc dels trets peculiars de l'individu;
- b) la dificultat objectiva que implica l'extraordinària variabilitat acústica de la parla.

Ja he parlat més amunt del primer problema, per això en aquest apartat ens dedicarem al segon, és a dir, a la variabilitat.

La veu humana configura formes acústiques a partir d'un doble sistema de formes sonores que evolucionen paral·lelament en el temps i que articulen dos fluxos d'informació complementaris: d'una banda, l'arbitrari i normalitzat del *sistema lingüístic* i, de l'altra, el del sistema de l'*expressió fonoestèsica*, compost per índexs acústics relacionats, per exemple, amb el sexe, l'edat, la configuració física, la personalitat i l'estat emocional del locutor, o sobre els seus mateixos moviments i tot allò que el parlant descriu oralment (cfr. RODRÍGUEZ, 1989).

Així, per exemple, imaginem-nos que hem aconseguit gravar la veu d'un delinqüent intervenint una trucada telefònica. En parlar, certament, es construiran formes sonores basades en l'estructura fonemàtica i lèxica de la llengua. I aquestes formes fonemàtiques estaran organitzades globalment amb unes variacions tonals i d'intensitat sistematitzades, també, pels codis lingüístics. Tanmateix, tota estructura lingüística resulta sempre alterada, matisada o, fins i tot, molt modificada per una sèrie de condicions prèvies que no depenen gens de la llengua mateixa.

Ara suposem que aquest delinqüent, quan feia aquesta trucada, per alguna raó que desconeixem, estava furios. Això significa que durant aquella locució es deu haver modificat la seva manera habitual de parlar.

L'agressivitat fa que el to i la intensitat de la veu siguin molt més alts del normal, la tensió articulatòria es torna també força més gran que de costum i la duració de les pauses s'escurça. A més, l'agressivitat modifica el tracte vocal i augmenta la tensió muscular, la qual cosa influeix sobre l'espectre acústic, potencia les freqüències altes i redueix la seva harmoniositat. Com a resultat, les variacions acústiques que depenen de l'estat emocional del locutor (expressió fonoestèsica) alteren tots els trets sonors lingüístics i provoquen una modificació en la veu gens previsible des d'una perspectiva estrictament fonètica.

Ara suposem que una gravació d'aquest tipus és la nostra veu de referència, és a dir, la veu no dubitativa que ens cal comparar amb les veus dels sospitosos. Encara que un d'aquests sospitosos fos realment el criminal, és molt difícil que durant les noves gravacions de la seva veu aquest estigui en unes condicions emocionals semblants a les que tenia quan va fer la trucada que estem fent servir com a prova. Per tant, pel fet que una veu emocionada pot resultar molt alterada en totes les seves dimensions sonores, ens enfrontaríem a un problema de variabilitat acústica que dificultaria moltíssim la identificació.

En suma, l'acústica de la parla es configura amb dos sistemes paral·lels, complementaris i perfectament articulats. En primer lloc, el parlant selecciona el seu to mitjà o modulador, la intensitat mitjana de la seva veu i la col·locació global del seu tracte vocal en funció de la situació comunicativa en què es troba. I a partir d'aquí comença una feina paral·lela de construcció de formes sonores segons el sistema de la llengua, que, a mesura que es van construint, es van desviant acústicament pel sistema de l'expressió fonoestèsica en un sentit o en un altre, d'acord amb l'estat emocional del parlant, la seva voluntat expressiva, l'actitud del receptor, etc.

Cal afegir, a més, que tots aquests trets acústics que, en un principi, estan determinats per la situació contextual, són també susceptibles de ser alterats per voluntat del mateix locutor si aquest hi està entrenat adequadament. Així, cal tenir

en compte que tot parlant està capacitada fisiològicament per alterar la seva veu, variant a voluntat el seu to mitjà, el grau d'harmoniositat, les inflexions tonals, la tensió articulatòria, etc., i així es mostraria més o menys agressiu, jove, vell, alegre, trist, afectuós, acriaturat, femení...

Aquesta gran variabilitat fa que el so de la veu —recollit per un micròfon fora de la boca—, sigui una massa acústica compacta i interdependent, que configura un tipus de senyal en què és extraordinàriament difícil trobar alguna característica que no hagi estat modificada diverses vegades durant el seu recorregut del tracte vocal. En conseqüència, no podem ignorar que qualsevol característica del senyal, que suposem dependent únicament d'un locutor, sempre pot arribar a ser alterada per alguna de les moltes possibilitats de modificació de l'aparell fonador.

7. CONCLUSIONS

En primer lloc, es conclou que, si bé és perfectament possible basar-se en instruments d'anàlisi i en coneixements d'acústica i de lingüística per a la identificació a través de la veu, encara no existeixen protocols d'anàlisi ni sistemes automàtics que puguin fer una identificació amb prou confiança i fiabilitat en tots els casos.

En segon lloc, és important indicar que la gran variabilitat acústica de la parla no suposa que la identificació a través de la veu no sigui possible, sinó que exigeix una nova perspectiva de treball.

Malgrat que més amunt afirmava que no hi ha paràmetres estables en la veu humana, la meua experiència com a educador en el camp de la locució també m'ha ensenyat que, en realitat, no és gens senzill adquirir la capacitat de canviar la qualitat acústica de la veu i l'estil de la locució. Encara que s'aconsegueixi un bon domini de l'aparell fonador, sempre ens descuidem d'alterar certs trets i aquests romanen constants la majoria de les nostres locucions. El problema és que aquests trets acústics estables varien d'un locutor a un altre. Mentre que en un locutor allò que roman invariable és una certa inflexió tonal característica, en un altre és la posició una mica elevada del dors de la llengua (palatalització), i en un tercer, la durada màxima de les vocals.

És a dir, segons el meu parer, el problema de la identificació de la veu s'ha de tractar més aviat com una qüestió d'identificació d'estil, o sigui, com si tractéssim de saber si una pintura ha estat o no pintada per un artista determinat, i no pas d'assignar-li un propietari a una empremta digital o a una seqüència d'ADN. El so de la veu no és el fragment físic d'un individu produït per la seva programació genètica i sobre el qual ell no té cap control racional, sinó que és el simple resultat d'un treball expressiu sobre les formes sonores de la parla, en el qual el subjecte emissor influeix conscientment. I cal tractar-lo com a tal.

Personalment, penso que efectivament sí que és possible resoldre el problema de la identificació analitzant acústicament la veu. Tanmateix, estic convençut que per fer-ho s'ha de tenir en compte la multidimensionalitat de la matèria fònica i treballar amb un model considerablement complex. Aquest model hauria de tenir

en compte coneixements de fonètica, sociolingüística, psicoacústica, acústica de les emocions, etc. i dur a terme anàlisis tant de paràmetres sonors segmentals com suprasegmentals, així com contemplar dimensions com l'entonació, els atacs i les caigudes. És a dir, el model hauria de trobar un procediment d'anàlisi estadística de n variables acústiques, que hagin estat seleccionades en diverses investigacions prèvies. El model desembocaria, finalment, en un *protocol d'anàlisi acústicoestadística*, que caldria posar a prova analitzant gravacions de locutors experts que hagin modificat la seva veu i també veus en què s'hagi destruït tècnicament una part de la seva informació sonora.

En aquesta línia, actualment hi ha en marxa un projecte interuniversitari en què col·labora també el Servei de Policia Judicial de la Guàrdia Civil; l'objectiu és la posada a punt d'un mètode per verificar sistemes automàtics d'identificació per la veu (cfr. ORTEGA-GARCÍA i d'altres, 1998). La base del projecte és construir un corpus molt ampli amb veus no dubitatives, que contingui i sistematitzi moltes variacions de la parla i de la qualitat de gravació en cada locutor.

En suma, considero que la solució estarà en el disseny d'un protocol d'anàlisi múltiple amb un nombre sistemàtic de controls acústics prou ampli i diversificat a fi que no pugués ser «enganyat» per manipulacions tècniques ni per la capacitat expressiva de cap locutor. Lògicament, aquest protocol exigiria prèviament unes condicions molt concretes de qualitat acústica perquè les gravacions poguessin ser sotmeses a la prova.

Finalment, penso que les institucions interessades a aplicar l'anàlisi de la veu humana a la investigació de delictes han de ser conscients que, en majoria, les línies actuals d'investigació no avancen en absolut en el camí que requereix la identificació.

Precisament pel fet que a la lingüística i a l'enginyeria els sol resultar molt més interessant i rendible l'estudi dels factors que són comuns a totes les veus que investigar i formalitzar les característiques acústiques de la individualitat, el coneixement i la tecnologia tendeixen a avançar justament en el sentit contrari al que necessita la identificació de locutors. De fet, quan s'aconsegueixen coneixements o eines aplicables a la identificació per la veu, aquests solen resultar insatisfactoris per a la investigació policial, justament perquè només són conclusions complementàries. És a dir, perquè són resultats que, en realitat, provenen de projectes amb objectius clarament orientats a allò que avui s'anomena les *indústries de la llengua*.

Com a investigador i com a coneixedor del rendiment real del mètode científic, sé que només és possible trobar la solució d'un problema de coneixement quan els objectius de la investigació defineixen amb total claredat i exactitud què és el que es busca. I sé també que l'exploració científica només obté alguna cosa que no està buscant quan de manera fortuïta i inesperada intervé l'atzar.

Considero, en conseqüència, que si es vol aconseguir un avenç real en la investigació de delictes aplicant l'anàlisi de la veu, és imprescindible que les institucions interessades donin suport, estimulin i proporcionin els recursos econòmics necessaris per finançar projectes que tinguin la identificació del locutor com un objectiu únic, específic i concret.

BIBLIOGRAFIA

- FURUI, S. «Effects of Long-Term spectral Variability on Speaker recognition». *J.A.S.A.*, 64, suppl. 1, paper NN28, 183, 1978.
- HARMEGNIES, B. «Les sources de variation du spectre a long terme de parole: revue de la literature». Dins: *Acoustique Canadienne*, 20 (2), pàg. 9-35, 1992.
- ORTEGA-GARCÍA, J.; MARRERO-AGUIAR, V.; GONZÁLEZ-RODRÍGUEZ, J.; DÍAZ-GÓMEZ, J.; GARCÍA-GIMÉNEZ, R.; LUCENA-MOLINA, J.; SÁNCHEZ-MOLERO, J.A.G^a: «Speaker Recognition Oriented "AUMADA" Large Speech Corpus». Dins: First International Conference on Language Resources and Evaluation, Granada, maig de 1998.
- ORTEGA-GARCÍA, J. i GONZÁLEZ-RODRÍGUEZ, J. «Providing Single and Multi-chanel Acoustical Robustness to Speaquer Identification Systems». *IEE International Conference on Acoustics Speech and Signal Processing. ISASSP-97*, vol. II, pàg. 107-110, Munic, abril de 1977.
- RODRÍGUEZ, A. *La construcción de una voz radiofónica*. Tesi doctoral. Departament de Comunicació Audiovisual i Publicitat. Universitat Autònoma de Barcelona, 1989.
- TOSI, O.; PISANI, R.; DUBES, R.; JAIM, A. «An Objective Metod of Voice Identification». Dins: International Phonetic Sciences Congress, Miami, 1977.
- TOSI, O. *Voice identification: Theory and legal Aplications*. Baltimore: University Park Press, 1979.