

Een gevalstudie in forensische fonetiek

Paul Corthals^{a,b}, John Van Borsel^b, Kristiane Van Lierde^b

^a Hogeschool Gent, Departement Gezondheidszorg Vesalius

^b Universiteit Gent, Faculteit Geneeskunde en Gezondheidswetenschappen

Abstract

In de forensische fonetiek gaat het niet zozeer over de herkenning van een spraakklank, maar veeleer over de herkenning van een spreker. Nochtans zijn de concepten, grootheden en meetinstrumenten waarvan men zich bedient in wezen dezelfde. Akoestische details die voldoende variëren van spreker tot spreker en tegelijk voor een gegeven spreker relatief constant zijn echter moeilijk te vinden. In de forensische fonetiek bestaat er (nog) geen definitieve set van parameters en is er evenmin een vast protocol voor spreker-identificatie of authenticatie van opnames: men moet parameters en protocol aanpassen aan de voorhanden zijnde data. Deze bijdrage gaat over een geval van dreig telefoons. De auteurs werd gevraagd om een kopie van de originele opname van een antwoordapparaat, gemaakt door politie, te verifiëren (authenticatie) en om vervolgens na te gaan of een verdachte effectief als de spreker in de opname aanzien kon worden (spreker-identificatie).

Inleiding

Daar waar de klassieke fonetiek zich richt op foneem-specifieke kenmerken, d.w.z.

klankeigenschappen die typisch zijn voor een gegeven klank en constant blijven van spreker tot spreker, hebben forensische toepassingen van de akoestische fonetiek altijd het omgekeerde als doel: zoeken naar spreker-specifieke kenmerken, die los staan van individuele klanken en verschillen van spreker tot spreker en tegelijk voor een gegeven spreker relatief constant blijven (Lu & Dang, 2008). Een voorbeeld van foneem-specifieke eigenschappen is de formantenstructuur van klinkers. Zo zullen de formanten van de doffe [ə] over alle sprekers heen een constant patroon vertonen: de tweede formant zal namelijk altijd qua frequentiegetal ongeveer het drievoud zijn van de eerste. Dit patroon is geldig voor sprekers van beide geslachten, van alle leeftijden, uit alle taalgroepen en wordt uitgelokt door de typische resonantieverschijnselen in luchtkolommen in de vorm van het aanzetstuk bij de articulatie van [ə]. Foneem-specifieke aspecten zijn het aanknopingspunt bij uitstek voor de

herkenning van spraakklanken (in de psycholinguïstiek, als het om menselijke informatieverwerking gaat; in de informatica, als het om automatische spraakherkenning gaat; in de logopedie en audiologie, als het om spraakverstaanbaarheid gaat).

In de forensische fonetiek gaat het niet zozeer over de herkenning van een spraakklank, maar veeleer over de herkenning van een spreker. Nochtans zijn de concepten, grootheden en meetinstrumenten waarvan men zich bedient in wezen dezelfde. Zo zullen stempathologen, die de habituele spreektoonhoogte of de microperturbaties in de stem bestuderen, ook op spectrale analyses via Fouriertransformaties steunen, net zoals forensische onderzoekers, die stemmen willen identificeren. Sommige experts combineren trouwens beide expertises (Hollien, 2002). Fonetici, logopedisten en audiologen kunnen door de gerechtelijke autoriteiten gevraagd worden om als expert op te treden. Het gaat dan meestal om twee concrete vragen: authenticatie van opnames en spreker-identificatie.

Authenticatie van opnames

Authenticatie van een opname houdt in dat in het geregistreerde signaal gezocht wordt naar tekenen die wijzen op een eerdere manipulatie, bijvoorbeeld om eventuele juridische beslissingen op basis van de opname te beïnvloeden. Hiervoor worden vaak andere dan puur akoestisch-fonetische technieken aangewend, bijvoorbeeld detectie van non-vocale transiënte geluiden die op aan- en uitschakelen van apparatuur kunnen wijzen (Maher, 2009). Niettemin is akoestisch-fonetische analyse van de prosodie (vb. intonatiecontouren) ook een methode om onderbrekingen op te sporen of om de gelijkheid tussen twee opnames na te gaan.

Verificatie of identificatie

Er wordt traditioneel een onderscheid gemaakt tussen spreker-verificatie en sprekeridentificatie (Nolan, 1999). In geval van spreker-verificatie, vraagt de spreker zelf om herkend te worden, bijvoorbeeld om toegang te krijgen tot geprivilegieerde gegevens of

privéruimten. In tegenstelling tot spreker-verificatie, komt de vraag naar herkenning bij spreker-identificatie niet van de spreker zelf: de spreker moet als “auteur” van een spraakstaal geselecteerd worden uit een min of meer beperkte lijst van mogelijke sprekers.

Automatische spreker-verificatie of –identificatie gebeurt d.m.v. computeralgoritmen die, meestal na een trainingsfase, sprekers herkennen door bepaalde kenmerken uit het spraaksignaal te extraheren (Jayanna & Prasanna Mahadeva, 2009). Deze kenmerken worden opgezocht in het spectrum (vb. cepstrale coëfficiënten of formantenbandbreedte om er de dimensies van het aanzetstuk uit af te leiden en inverse LPC-filtering om er eigenschappen van de larynx en de glottale cyclus mee op te sporen) of in de suprasegmentele laag van het signaal (vb. habituele spreektoonhoogte en articulatiesnelheid). Vervolgens worden de resultaten getoetst aan referentiewaarden in een referentiedatabase van geautoriseerde gebruikers.

In een forensische setting, heeft sprekeridentificatie meestal als doel te bevestigen of te ontkennen dat stem en spraak van een beschuldigde overeenkomen met de stem en de spraak in een opname. In dergelijke omstandigheden is er geen sprake van een referentiedatabase, noch van een bereidwillig meewerkende spreker. Opnames bedoeld voor forensisch onderzoek zijn bovendien meestal van slechte kwaliteit, veroorzaakt door distorsie afkomstig van de apparatuur, opzettelijk toegevoegde stoorsignalen of pogingen om de identiteit te maskeren door wijzigingen in de manier van stemgeven of spreken. Zelfs als de spreker vlot meewerkt om bepaalde uitingen voor te lezen of in te spreken, kunnen er nog altijd subtiele verschillen bestaan tussen spontane spraak en de leesopdracht of spreekopgave, alleen al omdat de emotionele spanning in beide situaties kan verschillen en stem of spraak tijdens de opname kan beïnvloeden.

Perceptuele versus technische sprekeridentificatie

Men kan een spreker ofwel langs technische ofwel langs perceptuele weg proberen identificeren. In de perceptuele benadering wordt onbevooroordeelde luisteraars gevraagd om geluidsopnamen te vergelijken. Er zijn nadelen aan een perceptuele vergelijking verbonden, o.m. dat de geheugencapaciteit van de luisteraars meespeelt (hoe lang kan men, met het oog op een vergelijking, de akoestische details van een uiting blijven onthouden en hoeveel kan men er onthouden?). Andere moeilijkheden bij perceptuele identificatie zijn dat het soort opgave en de instructies een rol spelen (een “*gelijkaardige stem*” zoeken in een reeks is anders dan een spreker “*aanwijzen*” door hem te selecteren uit een reeks tijdens een zogenaamde “*voice line-up*”, etc.) en dat zelfs de samenstelling van de reeks waaruit een stem gekozen moet worden de uitslag kan beïnvloeden. De perceptuele identificatie van een spreker kan inderdaad anders uitdraaien naargelang zijn stem en spraak herkend moeten worden te midden van twee verschillend samengestelde reeksen van gematchte sprekers. Daarbij komt nog dat er subtiele psycholinguïstische fenomenen kunnen interfereren met de auditieve beoordeling, zoals bijvoorbeeld *cue trading* (één akoestisch kenmerk kan een ander overstemmen of vervangen als het perceptueel meest relevante gegeven) of *verbal overshadowing* (het fenomeen waarbij de labels gebruikt in een eerder opgestelde omschrijving de herkenning en auditieve analyse van die stem inperken, iets wat overigens ook voor beeldmateriaal optreedt) (Nolan, 1999; Vanags et al., 2005).

In tegenstelling tot perceptuele spreker-identificatie wordt technische sprekeridentificatie gedaan aan de hand van apparatuur, i.h.b. computeralgoritmen, waardoor er een zekere “immuniteit” voor psycholinguïstische effecten ontstaat. Tenzij het gaat om volledig geautomatiseerde, tekstonafhankelijke sprekerherkenningssystemen (Campbell et al., 2009), is de tussenkomst van een expert wel nog vereist voor het selecteren van relevante fragmenten

en het interpreteren van de resultaten. Het sterke punt van deze akoestisch-fonetische aanpak zit in de objectivering van de parameters en de standaardisering van de extractie- en vergelijkingsprotocollen.

In 1962 werd de term “*voice print*” gelanceerd als synoniem voor spectrogram. Deze woordkeuze insinueerde dat via spectrografische analyse dezelfde zekerheidsgraad als voor vingerafdrukken (“*finger print*”) bereikt kan worden. Het is duidelijk dat dit nooit bevestigd is geworden (Boë, 2000; Campbell et al., 2009; Zhang & Tan T., 2008; Maher, R.C., 2009).

Immers, vergeleken bij vingerafdrukken, zijn de spraaksignalen van een zelfde spreker oneindig meer variabel. In de forensische fonetiek mondt identificatie daarom meestal uit in een probabiteit, d.w.z. een statistische uitspraak over de kans dat het bekomen resultaat op het toeval terug te voeren is, anders gezegd: de kans op een type II-fout (Champod & Meuwly, 2000). De research naar stabiele spreker-specifieke akoestische details gaat nog steeds verder.

Spreker-specifieke kenmerken

Strikt persoonlijke aspecten van iemands stem en spraak hebben meestal te maken met sociolinguïstische factoren (zoals dialect, beroepsklasse en socio-economische status) en met fysiologisch bepaalde factoren (zoals leeftijd, geslacht en morfologie van het aanzetstuk).

Nochtans zijn akoestische details die voldoende variëren van spreker tot spreker en tegelijk voor een gegeven spreker relatief constant blijven moeilijk aan te wijzen. Nolan (1999) wijst erop dat de stem- en spraakkenmerken van dezelfde spreker kunnen variëren als gevolg het taalregister (aanpassing van spraak en taal aan de setting en de status van de gesprekspartner), als gevolg van stem-spraakpathologie of vermoeidheid en zelfs als functie van het moment van de dag.

De meest effectieve forensische spreker-afhankelijke kenmerken komen voort uit de anatomische eigenschappen van larynx en mondkeelkanaal, meer bepaald van onderdelen die weinig of geen variatie (kunnen) vertonen tijdens stemgeving en articulatie en daarom ook relatief stabiel blijven bij een gegeven spreker, los van spreekstijl of gespreksinhoud.

Verscheidene auteurs stellen dergelijke kenmerken voor, zoals bijvoorbeeld Animo & Arai (2009), die in de stijging van de akoestische energie binnen nasaal-klinker-syllaben, gemeten in stappen van 10 ms. van het syllabebegin tot de klinkerkern, een spreker-specifieke hellingsgraad terugvonden. Ze concludeerden dat velo-faryngale activiteit en nasale resonantie weinig intra-sprekervariabiliteit vertonen en toch aanzienlijke verschillen van spreker tot spreker, hetgeen het ideaal is met het oog op spreker-identificatie. Lu & Dang (2008) suggereren vier mogelijke kenmerken: (a) anti-resonanties uitgelokt door de paranasale holten, of (b) door de fossa piriformis, (c) de vierde klinkerformant omdat die tot op zekere hoogte van de dimensies van iemands larynx afhangt en (d) the frequentiegamma waarbinnen de grondtoon valt, omdat die tot op zekere hoogte van de massa van iemands stembanden afhangt. Net zoals de grondtoon afhangt van de anatomie van de stembanden, valt ook de eerste formant, hoewel verschillend van klinker tot klinker bij een gegeven spreker, toch binnen een frequentie-interval, waarvan de grenzen gedicteerd worden door de persoonlijke dimensies van het mondkeelkanaal. Hollien (2002) haalt inderdaad de formantenstructuur van klinkers aan als akoestisch-fonetische parameter met spreker-specifieke waarden, samen met het langetermijnspectrum van de spraakproductie, de grondtoonwaarde van de stem en typische manieren van accentueren. Ook Van den Heuvel & Kraayeveld (1997) opperen de grootteorde grondtoon van de stem als kandidaat, vooral op het einde van de toonhoogtecontour van een uiting. Op segmenteel vlak wijzen ze naar de verlenging van de [a:] in [a:r]-sequensen als spreker-specifiek kenmerk in het Nederlands.

In theorie valt elk van deze spreker-specifieke kenmerken in een zekere frequentieband. Lu & Dang (2008) onderzochten 60 frequentiebanden op hun discriminerende waarde bij het onderscheiden van spraakstalen van 4 seconden afkomstig uit 5 opeenvolgende opnamesessies met 35 sprekers. Er kwamen drie spectrale zones naar voor, corresponderend met (a) de frequentiegamma van de stemgrondtoon (100 Hz – 300 Hz), (b) de resonantie-effecten van de fossa piriformis (4 kHz – 5.5 kHz) en (c) een zone rond 7,5 kHz die toegeschreven werd aan verschillen in articulatiegewoonten voor obstruenten. Ook het gebied rond 500 Hz zou in de dataset van Lu & Dang aangeduid kunnen worden als spreker-specifiek. De logica achter dit laatste is dat het, althans bij volwassenen, de basisresonantiefrequentie is van het mondkeelkanaal bij volwassenen. Voor de neutrale klinker doffe [ə] lijkt het mondkeelkanaal op een holle buis met constante diameter en één open uiteinde, die als eerste resonantiefrequentie (lees: formant) de waarde

$$f = c/4L \quad (\text{vergelijking 1})$$

zal hebben (c is hier de snelheid van geluid en L is de lengte van het mondkeelkanaal). Voor andere klinkers treden er uiteraard andere resonantie-effecten (formanten) op, maar de waarde van L blijft altijd een rol spelen de uitkomst. Dit betekent dat voor een gegeven klinker, de interindividuele verschillen in de eerste klinkerformant samengaan met interindividuele verschillen in de lengte van het aanzetstuk.

Samengevat: in de forensische fonetiek bestaat er (nog) geen definitieve set van parameters en is er evenmin een vast protocol voor spreker-identificatie of authenticatie van opnames. Men moet parameters en protocol aanpassen aan de voorhanden zijnde data. In deze bijdrage wordt een studie voorgesteld van een geval van dreigtelefoons. In dit soort rechtszaken gaat het erom de spreker te identificeren, maar ook om uit te sluiten dat de opname op een of andere wijze gemanipuleerd zou zijn geweest om de rechtsvoortgang te beïnvloeden. Daarom werd

de auteurs eerst gevraagd om een digitale kopie van de originele opname van een antwoordapparaat, gemaakt door politie, te verifiëren (studie I) en om vervolgens (studie II) na te gaan of een verdachte effectief als de spreker in de opname aanzien kon worden.

Studie I: Authenticatie van de opname.

Werkwijze studie I

Twee opnames van 14 seconden van een telefonische boodschap, t.t.z. een origineel en een politie-kopie waren ter beschikking gesteld. Om de authenticiteit van de politie-kopie na te gaan werd door de auteurs een nieuwe kopie gemaakt in wave-formaat met een bemonsteringsfrequentie van 44 kHz met behulp van Praat-software (Boersma & Weeninck, 2008, versie 4.3) en een personal computer. De politie-kopie en de eigen kopie werden op twee manieren vergeleken. Om te beginnen werden toonhoogte- (fig. 1) en sterktecontouren (fig. 2) van beide opnames over elkaar gelegd en visueel onderzocht op verschillen die op ingrepen in de eerste opname zouden kunnen wijzen. Er werd gelet op het aantal en de duur van de stemhebbende fragmenten en het patroon van de intonatie per uiting. Daarnaast werd het geluidsstrekteverloop van ogenblik tot ogenblik vergeleken. De weergave van analoge antwoordapparaten op magneetband meestal beïnvloed is door ruis en is beperkt tot de typische frequentieband voor telefonie (grotweg van 300 Hz tot 3500 Hz. Acht fragmenten, die 11 woorden bevatten en in totaal 7 seconden opname vertegenwoordigden (i.e. 50 % van de totale boodschap op het antwoordapparaat) werden geselecteerd voor verdere analyse. Deze fragmenten werden gekozen omwille van hun relatief grote intensiteit en hun redelijke signaalruisverhouding. Met SPSS-software (versie 10.1) werd de product-moment-correlatie berekend tussen de ogenblikkelijke geluidsstrektewaarden van elk van de 11 fragmenten in beide opnames.

Resultaten studie I

Visuele vergelijking van de gesuperponeerde toonhoogte- en sterktecontouren leverde geen duidelijke verschillen op. De correlaties tussen de geluidsstrektewaarden van de 11 corresponderende fragmenten waren allemaal significant ($p < .001$) en lagen tussen 0.804 en 0.987. De conclusie was dan ook dat de politie-kopie authentiek was.

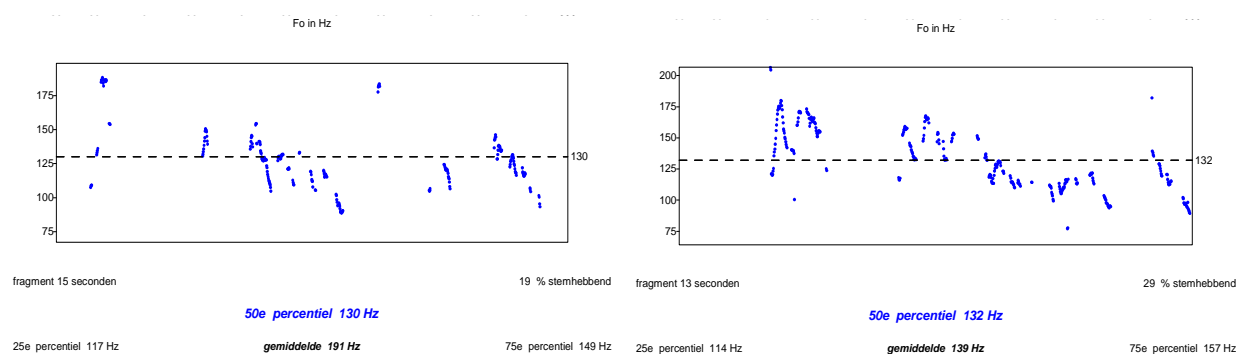


Fig.1

Extract uit de toonhoogtecontouren van twee corresponderende stukken uit de eigen digitale kopie (links) en de politie-kopie (rechts) van de dreigtelefoonopname. De stippellijn geeft het 50e percentiel weer (130 Hz en 132 Hz respectievelijk).

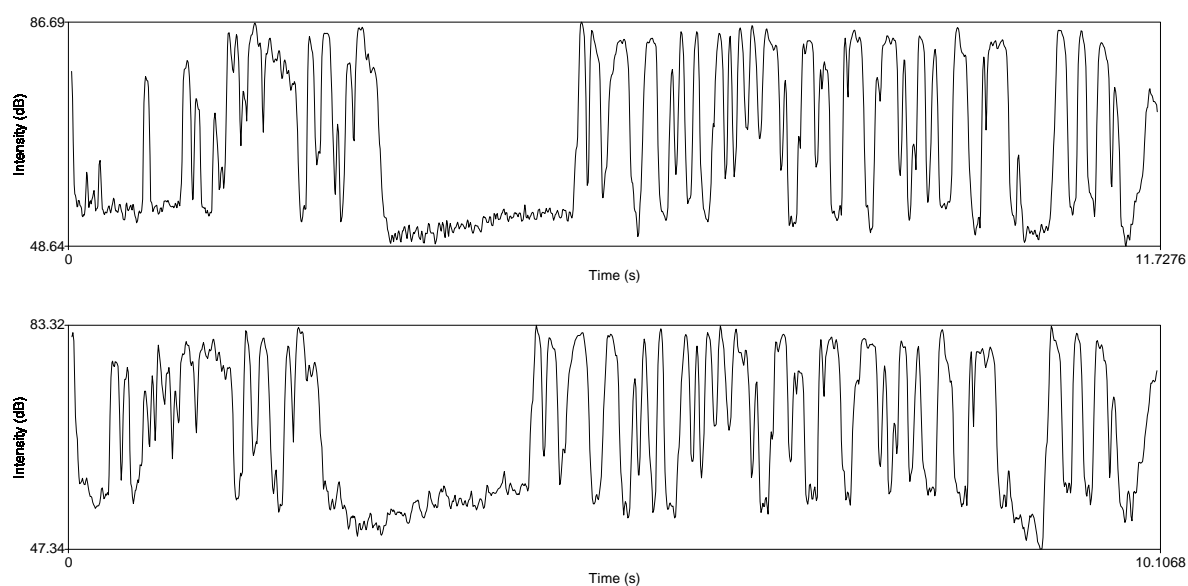


Fig. 2.

Intensiteitscontouren van de eigen kopie (bovenaan) en de politie-kopie (onderaan) van de originele opname.

Studie II: Sprekeridentificatie.

Werkwijze studie II

E werd een spraakstaal van de verdachte persoon opgenomen in een geluidsarme ruimte in wave-formaat met behulp van een *Computerized Speech Lab* CSL4300 met een bemonsteringsfrequentie van 44 kHz. De verdachte persoon werd gevraagd woorden in een neutrale context luidop te lezen. De taalinhoud van het in de kliniek gebruikelijke protocol werd aangepast opdat het staal 10 welbepaalde woorden zou bevatten, die ook in de opname van de dreigtelefoons voorkwamen. Precies dezelfde leesopdracht werd gegeven aan 3 andere sprekers van hetzelfde geslacht en dezelfde leeftijd. Zodoende kwamen er 4 opnames met dezelfde doelwoorden ter beschikking. Hieruit werden telkens 9 verschillende klinkerfragmenten met verschillende kaakhoek en tongstand geëxtraheerd. Deze 4 reeksen van 9 klinkerfragmenten werden op drie aspecten onderling vergeleken (fig. 3) met een t-test in SPSS 10.1. Voor elk van de drie aspecten werden er 63 parameterwaarden genoteerd, corresponderend met 63 momenten op de tijds-as. De 12 vergelijkingen (drie aspecten, vier sprekers) met de originele opname stonden allemaal in verband met de eerste formant. Het ging per klinker om (1) de absolute waarde van de eerste formant, (2) de relatieve positie van de eerste formant ten opzichte van de stemgrondtoon, (3) de relatieve positie van de eerste en de tweede formant. Er werd geopteerd voor kenmerken gerelateerd aan de eerste formant omdat dit blijktens de literatuur een akoestisch-fonetisch kenmerk met een zeker potentieel is (want een functie van de dimensies van het aanzetstuk) en omdat de eerste formant het meest opvalt in de opname omwille van zijn relatief grote energie.

Uitgaande van de hypothese dat de verdachte persoon *niet* de spreker op het antwoordapparaat is, zouden alle 12 vergelijkingen moeten resulteren in significante verschillen. Anderzijds, als de verdachte persoon *wel* de spreker op het antwoordapparaat is, zouden er logischerwijze

geen significante verschillen mogen optreden in de vergelijkingen van het type A (figuur 3), maar wel in de vergelijkingen van de stalen afkomstig van de drie geselecteerde vrijwilligers (vergelijkingen van type B, C, en D in figuur 3).

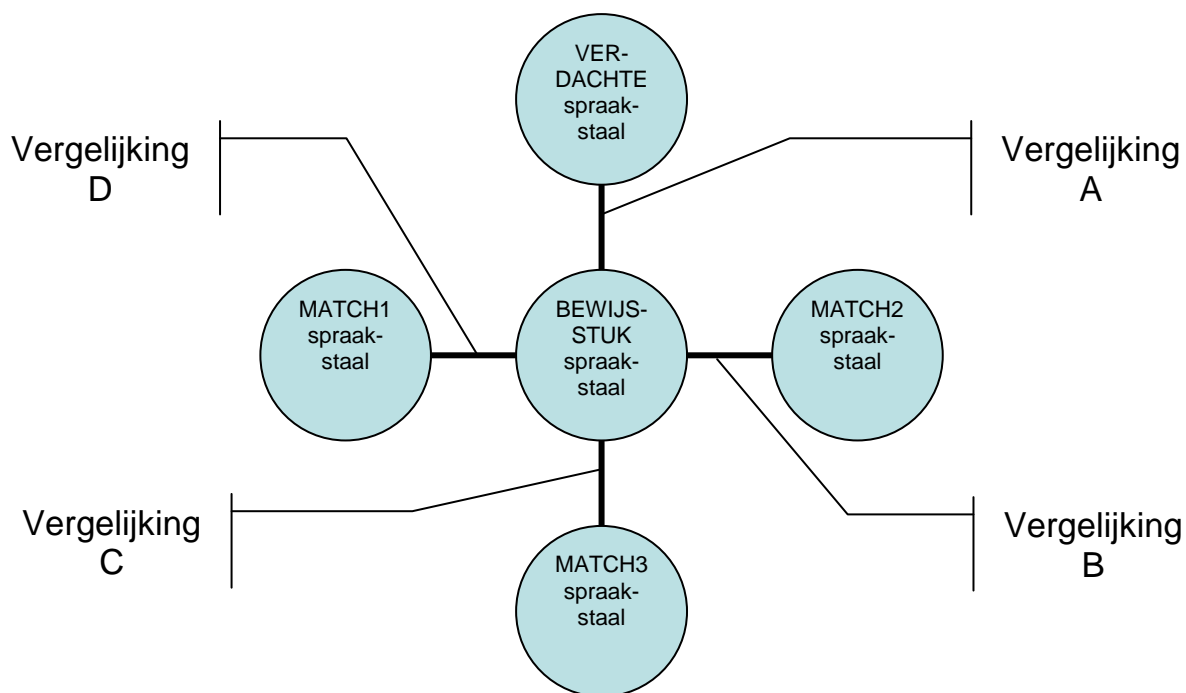


Fig. 3. Synopsis van de vergelijkingen tussen spraakstalen. Per vergelijking gaat het telkens om 3 akoestisch-fonetische kenmerken, elk vertegenwoordigd door 63 meetpunten op de tijds-as.

Resultaten studie II

Tabel 1 biedt een overzicht van het resultaat van de 12 vergelijkingen. De cellen met tekst in vetjes (i.e. 10 op een totaal van 12 tests) bevestigen de hypothese dat de verdachte persoon de spreker op het antwoordapparaat is. De kans om 10 “hits” op 12 pogingen te vinden bedraagt minder dan 5% (de exacte binomiale probabiliteit is 1.6%).

Discussie en besluit

Voor de verificatie van de politie-opname werd visuele en statistische vergelijking van intensiteits- en toonhoogteverloop in equivalente fragmenten toegepast. Er werd voor deze parameters geopteerd omdat ze eventuele verschillen ten gevolge van een ingreep achteraf in

de opname aan het licht zouden kunnen brengen. Er werden echter geen significante verschillen gevonden, zodat de opname authentiek verklaard werd, waarna studie II, de spreker-identificatie, kon beginnen.

Tabel 1. Significantie van de verschillen tussen de drie akoestisch-fonetische parameters in de 4 vergelijkingen. Significante verschillen zijn met (**) aangeduid. Zie ook fig. 3 voor de legende van de labels A, B, C en D.

Kenmerk	Vergelijking A (verdachte persoon - bewijsstuk)	Vergelijking B (spreker 1 - bewijsstuk)	Vergelijking C (spreker 2 - bewijsstuk)	Vergelijking D (spreker 3 - bewijsstuk)
Absolute waarde 1 ^e formant	<i>GEEN significant verschil p=0.300**</i>	<i>Significant verschil p<0.001**</i>	<i>Significant verschil p<0.001**</i>	<i>Significant verschil p<0.005**</i>
Relatieve positie 1 ^e formant en grondtoon	<i>GEEN significant verschil p=0.851**</i>	<i>Significant verschil p<0.001**</i>	<i>Significant verschil p<0.001**</i>	<i>Significant verschil p<0.005**</i>
Relatieve positie 1 ^e formant en 2 ^e formant	<i>GEEN significant verschil p=0.131**</i>	<i>Significant verschil p<0.001**</i>	<i>GEEN significant verschil p=0.112</i>	<i>GEEN significant verschil p=0.076</i>

De keuze van de akoestisch-fonetische kenmerken is in deze gevalstudie bepaald door de mogelijkheden die de voorhanden zijnde opnames openlieten. Dit is een typische situatie in de forensische fonetiek. Er werd voor de spreker-identificatie beroep gedaan op akoestisch-fonetische methodes omdat perceptuele vergelijkingen tussen spraakstalen verstoord kunnen worden door geheugencapaciteit en storende psycholinguïstische fenomenen. Het accent werd gelegd op kenmerken die met de 1^e formant in verband staan, omdat deze resonantie, voor een gegeven klinker, een functie is van de dimensies van het aanzetstuk van de vergeleken sprekers en bijgevolg een potentieel interessant kenmerk voor spreker-identificatie.

De vergelijking van de opname van de bedreigende telefonische boodschap met het spraakstaal van de verdachte persoon leverde in 3 gevallen op 3 geen significante verschillen op, terwijl de vergelijking met de spraakstalen van geselecteerde vrijwilligers in 7 van de 9 gevallen wel significante verschillen opleverde. Er zijn dus slechts 2 vergelijkingen op 12 die niet stroken met de hypothese dat de verdachte persoon de auteur van de dreigtelefoon is. Het gaat in die twee vergelijkingen om een test waar de tweede formant bij betrokken is, een kenmerk dat als minder spreker-specifiek beschouwd mag worden dan de twee andere (grondtoon en eerste formant). Hoe dan ook, de eindconclusie moet in termen van een kans geformuleerd worden: er is volgens de binomiale kansverdeling 1.6% kans dat een dergelijk resultaat door toeval kan ontstaan. In de gerechtelijke wereld wordt deze kans echter niet geïnterpreteerd zoals dat in wetenschappelijk onderzoek de gewoonte is (waar de significantiegrens ligt op 1 of 5%). Juridisch gesproken blijft de minste twijfel nog altijd een argument. Vandaar dat er stemmen opgaan om in gevallen als dit volgens te redeneren volgens de Bayesiaanse statistiek met a priori- en a posteriori-kansen. A-priori-kansen geven aan in hoeverre een verdachte persoon op basis van andere argumenten, nog voor het forensisch onderzoek, al aan een bepaald strafbaar feit geassocieerd kan worden. De a posteriori-kans is dan een vanuit die beginsituatie bijgestelde afweging van de significantie van de forensische bevindingen.

Een laatste, niet onbelangrijke conclusie is dat de klinisch welbekende apparatuur (digitale opnames, oscillografische en spectrografische analyse) en de traditionele parameters voor stem en resonantie (grondtoon, formantenstructuur) ook in forensische context een rol spelen.

- Nolan, F. (1999): Speaker recognition and forensic phonetics. In: Hardcastle, W. J. & Laver, J.: The handbook of phonetic sciences. Oxford: Blackwell (p. 744-767).
- Lu, X., Dang, J. (2008): An investigation of dependencies between frequency components and speaker characteristics for text-independent speaker identification. *Speech Communication* (50), p. 312-322.
- Boë, L-J. (2000): Forensic voice identification in France. *Speech Communication* (31), p. 205-224.
- Champod, C., Meuwly, D. (2000): The inference of identity in forensic speaker recognition. *Speech Communication* (31), p. 193-203.
- Animo, K., Arai, T. (2009): Speaker-dependent characteristics of the nasals. *Forensic Science International* (185), p. 21-28.
- Campbell, J.P., Wade, S., Campbell, W.M., Schwartz, R., Bonastre, J.-F., Matrouf, D. (2009): Forensic speaker recognition: a need for caution. *Signal Processing Magazine*. March 2009, p. 95-103.
- Zhang, C., Tan, T. (2008): Voice disguise and automatic speaker recognition. *Forensic Science International* (175), p. 118-122.
- Maher, R.C. (2009): Audio forensic examination: authenticity, enhancement, and interpretation. *Signal Processing Magazine*. March 2009, p. 84-94.
- Vanags, T., Carroll, M., Perfect, T.J. (2005): Verbal Overshadowing: A Sound Theory in Voice Recognition? *Applied Cognitive Psychology* (19), p. 1127–1144.
- Jayanna, H.S., Prasanna Mahadeva, S.R. (2009): Analysis, feature extraction, modelling and testing techniques for speaker recognition. *IETE Technical Review* (26), p. 181- 190.
- Boersma, P., Weenink, D. (2008): Praat: doing phonetics by computer (Version 4.3) [Computer program]. Retrieved August 30, 2008, from <http://www.praat.org/>
- Hollien, H. (2002): Forensic voice identification. London: Academic Press.
- van den Heuvel, H., Kraayeveld, H. (1997): Sprekend jezelf: een verslag van twee onderzoeken naar sprekerkenmerken. *Gamma TTT tijdschrift voor taalwetenschap* (6), p. 41-52.