

Vakbijlage

vergelijkend spraakonderzoek

Inhoudsopgave

- 1. Inleiding**
- 2. Methode van onderzoek**
 - 2.1. Toelichting op de auditief-akoestische analyse**
 - 2.1.1. *Auditieve analyse*
 - 2.1.2. *Akoestische ondersteuning*
 - 2.1.3. *Auditief-akoestische methode: Interpretatie van bevindingen*
 - 2.2. Toelichting op de blinde groepering**
 - 2.2.1. *Vorbereiding blinde groepering*
 - 2.2.2. *Groepering door de blinde onderzoeker*
 - 2.2.3. *Blinde groepering: Interpretatie van bevindingen*
 - 2.3. Integrale eindbeoordeling**
- 3. Formulering van de eindconclusie**
- 4. Kwaliteitsborging**
- 5. Verklarende woordenlijst**
- 6. Literatuur**

1. Inleiding

In een vergelijkend spraakonderzoek worden de spraakuitingen van de te onderzoeken spreker vergeleken met de spraakuitingen van een meestal bekende spreker in een andere opname. Het onderzoek wordt ook wel aangeduid met de (minder accurate) termen sprekeridentificatie, sprekerverificatie, spreker- of stemherkenning.

De spraakuitingen van de te onderzoeken spreker (het betwiste materiaal) zijn bijvoorbeeld afkomstig van getapte telefoongesprekken of van opnamen vertrouwelijke communicatie (OVC's). De spraakuitingen van de bekende spreker (het vergelijkingsmateriaal) zijn bijvoorbeeld afkomstig van niet-betwiste telefoon-

gesprekken, politieverhoren of van speciaal voor dit doel opgenomen spraakmateriaal. Ook komt het voor dat er betwist materiaal met betwist materiaal vergeleken moet worden.

De mogelijkheden voor vergelijkend spraakonderzoek worden onder andere bepaald door:

- de kwaliteit en kwantiteit van het beschikbare materiaal;
- de mate waarin het spraakmateriaal representatief geacht kan worden voor de betreffende spreker(s);
- de mate waarin de spraakonderzoekers van het NFI bekend zijn met de taal(variant);
- de mate waarin het vergelijkingsmateriaal een goede vergelijkingsbasis biedt (zelfde gespreksomstandigheden als het betwiste materiaal).

Deze factoren worden tijdens het vooronderzoek beoordeeld om de onderzoeksmogelijkheden te bepalen.

2. Methode van onderzoek

Spraakonderzoekers van het NFI beschikken over een uitgebreide kennis van de fonetiek, de dialectologie, de sociolinguïstiek en de algemene taalwetenschap. Zij maken gebruik van kennis, methoden en technieken die binnen deze wetenschapsgebieden zijn ontwikkeld.

Bij het NFI wordt een auditief-akoestische analyse en (waar mogelijk) een blinde groeperings-taak toegepast.

De auditief-akoestische methode wordt wereldwijd het meest gebruikt. [1,2] Daarbij wordt het materiaal beluisterd door getrainde spraakonderzoekers, en worden sprekeronderscheidende taal- en spraakkenmerken in kaart gebracht. Vervolgens worden verschillen en overeenkomsten tussen het betwiste materiaal en het vergelijkingsmateriaal genoteerd en geïnterpreteerd (zie verder 3.1.3). In sommige gevallen kunnen taalkundige waarnemingen worden geïnterpreteerd in het licht van gegevens uit de literatuur,

of wordt ad-hoc een kleinschalig onderzoek gedaan aan de hand van beschikbare databases. Akoestische metingen hebben een ondersteunende rol in vergelijkend spraakonderzoek. Akoestische metingen worden voornamelijk uitgevoerd indien de auditieve analyse daartoe aanleiding geeft en als de kwaliteit van de opnames het toelaat. Zie verder 3.1.

Naast de auditief-akoestische analyse wordt ook een auditieve 'blinde groepering' uitgevoerd door een tweede onderzoeker. De 'blinde groepering' is vrij van contextinformatie zoals het gebruik van namen, gebruik van een bepaald telefoonnummer, eerdere toekenningen door de politie etc. Voor de blinde groepering wordt een verzameling van fragmenten voorbereid door de eerste onderzoeker, waarbij elk fragment spraak bevat afkomstig van het betwiste materiaal, van het vergelijkingsmateriaal of van een afleider. Een tweede onderzoeker, die niet weet welke fragmenten al eerder aan één spreker zijn toegerekend, beluistert en analyseert deze fragmenten, om ze vervolgens per spreker te groeperen. Zie verder 3.2.

Na de auditief-akoestische en blinde analyses vormen twee onderzoekers onafhankelijk van elkaar zich een eendoordeel over het materiaal ("vier-orenprincipe"). Deze eindconclusie wordt gebaseerd op een integrale interpretatie van alle bevindingen. Zie verder 3.3.

2.1. Toelichting op de auditief-akoestische analyse

2.1.1. Auditieve analyse

Tijdens de auditieve analyse wordt eerst het betwiste materiaal en daarna het vergelijkingsmateriaal geanalyseerd.

Tijdens de auditieve analyse wordt het materiaal herhaaldelijk beluisterd om een systematische en nauwkeurige analyse te maken van de volgende isoleerbare kenmerken [3,4]:

- Taalvariant: regionale accenten en dialecten (bv. Haags, Limburgs), etnolect (bv. Turks-Nederlands), sociolect (bv. straattaal);
- Stem: type stemgeving (borstregister, kopregister, onvast, 'creaky voice'), stemkwaliteit (bv. hees, schor), klankkleur (bv. dof, helder), spreektoon-hoogte (laag of hoog);
- Suprasegmenteel: nasaliteit, spraakmelodie (levendig of monotoon), tempo (langzaam of snel), ritme (staccato), mate van vloeiendheid (bv. herhaling, verlenging);

- Taalgebruik: zinsbouw (bv. afwijkende zinsconstructies), woordkeus (bv. stopwoorden);
- Segmenteel: realisatie van afzonderlijke spraakklanken.

Deze auditieve analyse is veelomvattend maar moeilijk te kwantificeren. Daarnaast geldt dat niet alle auditieve waarnemingen in isoleerbare kenmerken te vangen zijn. Stemkwaliteit is bijvoorbeeld een belangrijk sprekeronderscheidend kenmerk, maar niet goed te omschrijven in woorden [5,6]. Het geheel (Gestalt) van het stemgeluid wordt gevormd door de trillende stembanden, en de keel-, mond- en neusholte, die per persoon verschillen. Uit onderzoek is gebleken dat zowel het waarnemen van het geheel (rechterhersenhelft) als het isoleren en apart benoemen van afzonderlijke kenmerken (linkerhersenhelft) een rol spelen bij het herkennen van sprekers [5]. Gestaltwaarneming speelt een rol bij de auditieve analyse, maar wordt vooral ingezet – en tot op zekere hoogte getoetst – bij de blinde groepering.

2.1.2. Akoestische ondersteuning

Indien de kwaliteit van de opnamen het toelaat, kunnen onderstaande aspecten van het akoestische signaal bekeken worden ter onderbouwing van auditieve waarnemingen:

- Toonhoogtemetingen worden gedaan om de waarneming van een opvallend hoge of lage stem te ondersteunen. De gemiddelde toonhoogte, de verdeling van de toonhoogte of het bereik kunnen gemeten worden.
- Articulatiesnelheid wordt gemeten indien het waargenomen spreektempo uitzonderlijk snel of langzaam klinkt. Het aantal lettergrepen per seconde kan in een dergelijk geval bepaald worden.
- Pauzes worden gemeten indien de spreker opvallend pauzegegedrag in zijn spraak vertoont. De duur in milliseconden en de frequentie van pauzes kunnen gemeten worden.
- Het verloop van de resonantiefrequenties worden in het spectrum bekeken indien waargenomen klinkers, tweeklanken of nasalen afwijken.
- De duur van klanken of lettergrepen worden gemeten indien een opvallende verlenging of verkorting wordt waargenomen.

2.1.3. Auditief-akoestische methode: interpretatie van bevindingen

Waargenomen en gemeten overeenkomsten en verschillen worden gezien in het licht van omstandigheden die van invloed kunnen zijn op hoe een spreker zijn of haar uitingen realiseert.

Te denken valt aan de relatie tot de gesprekspartner, inhoud van het gesprek, tijd van de dag en emotionele toestand van de spreker.

Van de waargenomen overeenkomsten worden zeldzaamheid en sprekeronderscheidendheid ingeschat, in het licht van de taalachtergrond van de spreker (bijvoorbeeld consequent 'de' zeggen in plaats van 'het' is minder sprekeronderscheidend bij sprekers van Marokkaans Nederlands dan bij autochtone Nederlanders). De onderscheidendheid van een bepaald kenmerk staat dus niet vast, maar verschilt per spreker en gesprekssituatie. In sommige gevallen kunnen achtergrondgegevens uit de literatuur of uit een kleinschalig ad-hoc onderzoek meer licht werpen op de onderscheidendheid van een bepaald kenmerk.

Van de waargenomen verschillen wordt ingeschat of ze waarschijnlijker zijn binnen het repertoire van één spreker (intra-sprekerverschillen), of dat ze waarschijnlijker zijn onder de hypothese dat het materiaal afkomstig is van twee sprekers (intersprekerverschillen). Voorbeelden zijn respectievelijk een verschil in toonhoogte dat verklaarbaar is door een verschil in mate van stemverheffing, tegenover een dialectverschil dat wijst op verschillende sprekers.

2.2. Toelichting op de blinde groepering

2.2.1. Voorbereiding blinde groepering

Eerst wordt een geluidsbestand samengesteld door een eerste onderzoeker; dat is een andere onderzoeker dan degene die de blinde groepering uitvoert (de 'blinde' onderzoeker). [6] Dit samengestelde geluidsbestand bevat gewoonlijk niet meer dan 20 geluidsfragmenten, ieder met een duur van ongeveer 20 seconden (indien beschikbaar). De eerste onderzoeker plaatst fragmenten uit zowel het betwiste als het vergelijkingsmateriaal plus fragmenten van geschikte gesprekspartners (indien beschikbaar) in willekeurige volgorde in een geluidsbestand. De fragmenten van de gesprekspartners fungeren als afleiders.¹ Er worden geen spraakuitingen gebruikt met inhoudelijke informatie zoals eigennamen, die een aanwijzing kunnen vormen voor de identiteit van de spreker.

¹ Met name getapte telefoongesprekken en opnames van vertrouwelijke communicatie (OVC's) bevatten vaak spraakmateriaal van geschikte afleiders, aangezien de gesprekspartners vaak sprekers zijn uit dezelfde *peer group*.

2.2.2. Groepering door de blinde onderzoeker

De blinde onderzoeker beluistert de fragmenten en deelt deze in groepen in. Er worden geen beperkingen opgelegd wat betreft aantal groepen. Fragmenten die de blinde onderzoeker waarneemt als van dezelfde spreker worden in één groep geplaatst. Fragmenten die de blinde onderzoeker waarneemt als van verschillende sprekers worden in verschillende groepen ingedeeld. De blinde onderzoeker kan zowel Gestaltwaarneming als kenmerkanalyse gebruiken om tot een groepering te komen. Het kan voorkomen dat fragmenten helemaal niet in groepen (kunnen) worden ingedeeld, bijvoorbeeld omdat de duur als te kort wordt ervaren of de kwaliteit als te slecht.

De blinde onderzoeker duidt de onderlinge overeenkomsten van de fragmenten *binnen* een groep – en daarmee de overtuiging dat het om dezelfde spreker gaat – aan als 'sterk', 'normaal' of 'zwak'. Ook worden de onderlinge verschillen tussen de fragmenten in *verschillende* groepen – en daarmee de overtuiging dat het om verschillende sprekers gaat – aangeduid als 'sterk', 'normaal' of 'zwak'. Het is ook mogelijk voor de blinde onderzoeker om 'geen oordeel' te geven.

De groepering die de blinde onderzoeker maakt heeft en de sterkte van de onderlinge verbanden worden in het rapport opgenomen.

2.2.3. Blinde groepering: interpretatie van bevindingen

In een blind onderzoek is de 'blinde' onderzoeker niet bekend met de context van de zaak, noch bekend met wat het betwiste of het vergelijkingsmateriaal omvat, noch bekend met het aantal verschillende sprekers waar het om kan gaan.

Door het toevoegen van afleiders wordt de blinde groepering een vorm van zaakgebonden validatie van het oordeel van de blinde onderzoeker, zij het in omvang beperkt. In het resultaat van de blinde groepering moet dan ook eerst worden bekeken of de afleiders correct apart gegroepeerd zijn. Is dit niet het geval, dan is het kennelijk voor de 'blinde' onderzoeker niet mogelijk om de sprekers in kwestie goed van elkaar te onderscheiden, althans niet op grond van de korte fragmenten aangeleverd in de blinde analyse. Of dit aan de beperkingen van het materiaal ligt, of aan de selectie van de fragmenten, of aan de beperkingen van de capaci-

teiten van de onderzoeker, kan niet worden vastgesteld.²

Als de afleiders wel correct apart gegroepeerd zijn, kan bekeken worden hoe het betwiste materiaal is gegroepeerd ten opzichte van het vergelijkingsmateriaal. Indien betwist materiaal en vergelijkingsmateriaal bij elkaar wordt gegroepeerd, levert dit steun op voor de hypothese dat het door dezelfde spreker is geproduceerd. Indien betwist materiaal en vergelijkingsmateriaal apart wordt gegroepeerd, levert dit steun op voor de hypothese dat het door verschillende sprekers is geproduceerd. Indien de 'blinde' onderzoeker geen oordeel heeft kunnen geven, levert dit niet meer steun op voor de ene of de andere hypothese.

Het voordeel van de blinde groepering is dat de blinde onderzoeker op geen enkele wijze kan worden gestuurd door *confirmation bias* [7]. De blinde onderzoeker weet immers niet welke fragmenten al eerder aan één spreker zijn toegerekend, noch hoeveel afleiders in de blinde groepering zijn gestopt. Ook kan worden gezien of elk betwist fragment op zichzelf voldoende kenmerken bevat om beoordeeld en gegroepeerd te worden.

Een beperking van de blinde groepering is dat de onderzoeker een gefragmenteerd beeld krijgt van het materiaal; zo kan het vergelijkingsmateriaal niet als geheel beoordeeld worden. Immers, het vergelijkingsmateriaal wordt net als het betwiste materiaal in losse fragmenten voorgelegd aan de blinde onderzoeker. Een tweede beperking van de blinde groepering is dat de onderzoeker bepaalde contextgegevens niet kan gebruiken om bepaalde bevindingen te kunnen wegen en interpreteren, zoals de gesprekssituatie en de relatie met de gesprekspartner. De blinde groepering is dan ook geen eindoordeel, maar dient ten eerste als een vorm van zaakgebonden calibratie en ten tweede als een 'blinde' beoordeling (vrij van bepaalde contextinformatie) die wordt meegenomen in de eindbeoordeling.

² Een tweede blinde groepering (door een derde onderzoeker) zou in zo'n geval mogelijk duidelijkheid kunnen scheppen, maar dit is meestal niet haalbaar omdat het aantal spraakonderzoekers werkzaam bij het NFI zeer klein is.

2.3. Integrale eindbeoordeling

Na de auditief-akoestische analyse en de blinde analyse vindt nog een integrale eindbeoordeling plaats, om de bewijskracht van het totaal aan bevindingen te bepalen. Twee onderzoekers kunnen in dit stadium kennis nemen van alle beschikbare informatie, inclusief contextgegevens die nodig zijn om de bevindingen te interpreteren.

De Gestaltwaarneming, elk van de auditieve kenmerken, de akoestische waarnemingen en het resultaat van de blinde groepering zijn niet onafhankelijk, en laten zich dus niet 'optellen' tot een conclusie. De conclusie van het onderzoek is gebaseerd op een integrale interpretatie van alle bevindingen. De conclusie heeft het karakter van een deskundig oordeel t.a.v. de bewijskracht, dat de onderzoeker op grond van zijn kennis en ervaring in een concreet geval heeft bereikt.

Twee onderzoekers vormen zich een eindoordeel. Dit kunnen de twee onderzoekers zijn die het auditief-akoestische onderzoek respectievelijk het blinde onderzoek hebben gedaan, maar dat hoeft niet. Indien onderzoekers verschillen in hun uiteindelijke conclusie, wordt dit gerapporteerd.

3. Formulering van de eindconclusie

In het rapport wordt met gebruikmaking van een verbale schaal aangegeven hoeveel meer steun de bevindingen bieden aan de ene hypothese dan aan de andere. Deze hypothesen worden standaard als volgt geformuleerd:

Hypothese 1: de spreker van het betwiste materiaal is dezelfde als de spreker van het vergelijkingsmateriaal.

Hypothese 2: de spreker van het betwiste materiaal is een ander (van dezelfde sekse en met een vergelijkbare taalachtergrond) dan de spreker van het vergelijkingsmateriaal.

N.B. Als er sprake is van een andere alternatieve hypothese, bijvoorbeeld bij een *closed set* aan mogelijke sprekers of als er sprake is van broers, dan is dit van wezenlijke invloed op de uitvoering van het onderzoek. Dit dient dan ook altijd bij de aanvraag te zijn vermeld.

De onderzoeker doet deze uitspraak conform de richtlijnen gegeven in de NFI-vakbijlage "De

reeks waarschijnlijkheidstermen van het NFI en het Baysiaanse model voor interpretatie van bewijs". [8] Om de transparantie voor de lezer en de uniformiteit tussen verschillende deskundigen te bevorderen, zijn de gebruikte verbale termen ook numeriek gedefinieerd. De gebruikte reeks waarschijnlijkheidstermen is als volgt:

De bevindingen van het onderzoek zijn...

- *ongeveer even waarschijnlijk...*
(*ordegrootte bewijskracht 1-2*)
 - *iets waarschijnlijker...*
(*ordegrootte bewijskracht 2-10*)
 - *waarschijnlijker...*
(*ordegrootte bewijskracht 10-100*)
 - *veel waarschijnlijker...*
(*ordegrootte bewijskracht 100-10.000*)
 - *zeer veel waarschijnlijker...*
(*ordegrootte bewijskracht 10.000-1.000.000*)
 - *extreem veel waarschijnlijker...*
(*ordegrootte bewijskracht >1.000.000*)
- ...wanneer hypothese 1 juist is, dan wanneer hypothese 2 juist is.*

Gegeven de aard van vergelijkend spraakonderzoek is een bewijskracht hoger dan 10.000 zelden of nooit te verwachten. Dit heeft vooral te maken met de beperkte mate waarin statistische gegevens beschikbaar zijn over de sprekeronderscheidendheid van kenmerken (in de betreffende populatie en onder vergelijkbare communicatieve omstandigheden). De conclusies 'zeer veel waarschijnlijker' en 'extreem veel waarschijnlijker' (bewijskracht >10.000 resp. >1.000.000) zullen bij een standaard set van hypothesen dus vrijwel nooit worden gebruikt.

Anderzijds is de bewijswaarde van een vergelijkend spraakonderzoek voor het strafrechtelijk onderzoek vaak hoog. Immers, het gegeven of een persoon wel of niet te horen is in een bepaalde opname (bronniveau), is door de inhoud en context vaak ook een gegeven op activiteitsniveau. [9] Het bepalen van de bewijswaarde is uiteindelijk aan de rechter.

4. Kwaliteitsborging

De forensisch onderzoekers van het NFI hantieren de beschreven methode van vergelijkend spraakonderzoek met het bijbehorende systeem voor kwaliteitsborging van het NFI.

De validatie van de conclusies op basis van het uitgevoerde onderzoek is onder andere gebaseerd op deelname aan internationale prestatie-

niveautoetsen (*proficiency tests*) en de NFI-opleiding tot forensisch onderzoeker.

De toepassing van blinde groepering en het feit dat elk vergelijkend onderzoek dubbel wordt uitgevoerd draagt verder bij aan de betrouwbaarheid van het uitgevoerde onderzoek.

5. Verklarende woordenlijst

Auditief - met het gehoor

Akoestisch signaal - fysische representatie van geluid

Borstregister - modaal gebruik van de stem, in tegenstelling tot kopregister

Creaky voice - bepaalde manier van stemgeven waardoor een krakerigheid hoorbaar is

Fonetiek - wetenschap van spraakproductie en spraakwaarneming

Kopregister - vorm van stemgeving met als resultaat een hogere toonhoogte. Ook bekend als falsetto

Monotoon - vlakke intonatie

Nasaliteit - mate waarin het spraakgeluid 'door de neus' klinkt

Representatief - spraak kan meer of minder representatief zijn voor wat een spreker onder gewone omstandigheden aan spraak produceert. Voorgelezen spraak is bijvoorbeeld alleen representatief voor die taak, maar niet voor spontane spraak. Een dialoog waarin een spreker zowel rustig als geanimeerd spreekt, geeft wel een goed beeld van het spraakrepertoire van een spreker.

Resonantiefrequentie - het geluid dat geproduceerd wordt door de stembanden wordt gefilterd door de keel- neus- en mondholte. Sommige frequenties worden door geluidswaerkaatsing (resonantie) versterkt en andere worden verzwakt. De resonantiefrequenties zijn de versterkte frequenties.

Segmenteel - heeft betrekking op de individuele segmenten in spraak: klanken

Spectrum - visuele weergave van het akoestisch signaal waarin onderlinge sterkte-verhoudingen zichtbaar worden

Spraakuitingen - uitingen van een spreker die bestaan uit spraak (in tegenstelling tot bijvoorbeeld huilen of lachen)

Staccato - ritmepatroon waarbij alle lettergrepen ongeveer even lang zijn. Het normale ritmepatroon in het Nederlands bestaat uit lange en korte lettergrepen die elkaar afwisselen

Suprasegmenteel - kenmerken van de spraak die op een ander niveau liggen dan klanken: bijvoorbeeld zinsmelodie

Taalgebruikssituatie - een situatie die bepalend is voor de manier waarop mensen spreken (taalgebruiken). Een politieverhoor of een toespraak wijkt qua taalgebruikssituatie af van een informeel telefoongesprek

Validatie - het vaststellen van de prestaties van een meetinstrument, in dit geval van de onderzoeker

Beyond the Segments. Experimental linguistics and phonetics. Amsterdam: John Benjamins Publishing Company, 14-27.

7. Saks, M. J., Risinger, D. M., Rosenthal, R., & Thompson, W. C., 2003. Context effects in forensic science: A review and application of the science of science to crime laboratory practice in the United States. *Science & Justice*, 43, 77-90.
8. NFI vakbijlage waarschijnlijkheidstermen versie 2.1 oktober 2014. Te vinden op: www.forensischinstituut.nl/pers_en_publicaties/publicaties/Vakbijlagen%20en%20Informatiebladen.aspx
9. Kurpershoek, J., 2012. *Niet alleen wat, maar ook wie en hoe.* Blauw – Opsporing, 7 januari 2012, nummer 1.

6. Literatuur

De geïnteresseerde lezer kan meer informatie over de behandelde onderwerpen vinden in:

1. Cambier-Langeveld, T., 2007. Current methods in forensic speaker identification: Results of a collaborative exercise. *The International Journal of Speech, Language and the Law* 14(2), 223-243.
2. Gold, E. & French, P., 2011. International practices in forensic speaker comparison. *The International Journal of Speech, Language and the Law* 18(2), 293-307.
3. Broeders, A.P.A., 2008. Vergelijkend spraakonderzoek. In: A.P.A. Broeders & E.R. Muller (eds), *Forensische wetenschap: studies over forensische kennis en organisatie*. Deventer: Kluwer, 508-513.
4. Jessen, M., 2012. *Phonetische und linguistische Prinzipien des forensischen Stimmenvergleichs*, Lincom Europa Academic publications, München.
5. Kreiman, J. & Sidtis, D., 2011. *Foundations of voice studies*, Wiley-Blackwell Malden, U.S.A.
6. Cambier-Langeveld, T., van Rossum, M. & Vermeulen, J., 2014. Whose voice is that? Challenges in forensic phonetics. In: J. Caspers, Y. Chen, W. Heeren, J. Pacilly, N.O. Schiller & E. van Zanten (eds), *Above and*