Forensisch-phonetischer Stimmenvergleich

Angelika Braun

Universität Trier, Fachbereich II, Phonetik

brauna@uni-trier.de

Abstract: Dieser Beitrag bietet eine zusammenfassende Darstellung der in der forensischen Phonetik entwickelten Verfahren zum Stimmenvergleich. Verschiedene Methoden werden vorgestellt und hinsichtlich ihrer praktischen Relevanz diskutiert. Dabei finden sowohl der Stimmenvergleich durch Experten als auch die sog. *voice line-ups* (Sprechererkennung durch Opfer und Zeugen) Berücksichtigung.

1 Stimme und Individuum

Intuitiv gehört die Stimme zu den persönlichkeitskonstituierenden Eigenschaften eines jeden Individuums und ist insoweit typisch für einen und nur einen Menschen. Auf solchen Überlegungen beruht das Konzept der Idiosynkrasie (in der Forensik vertreten beispielsweise von John Baldwin [1]), das letztlich davon ausgeht, dass Stimmen ähnlich wie Fingerabdrücke oder DNA-Spuren individualtypisch und damit einzigartig sind.

Ein flüchtiger Blick auf die Anatomie und Physiologie der Sprachproduktion scheint diese Sichtweise zu bestätigen: Gunnar Fants [9] *source-filter model of speech production* ist bis heute unumstritten. Es betrachtet den sprachlichen *output* als Resultat zweier Komponenten; zum einen des Quellsignals, das im Kehlkopf generiert und wesentlich durch dessen Anatomie beeinflusst wird, und zum zweiten der Filterfunktion des – individuell konfigurierten – Ansatzrohres. Insbesondere die Höhe der Resonanzfrequenzen im Ansatzrohr (Formanten) kann damit direkt aus dessen Dimensionen einschließlich der Zu- bzw. Abschaltung des Nasenraums errechnet werden. Ein solches, an Fingerabdrücke und DNA angelehntes Bild der Sprachproduktion wird der Einfachheit halber in medialer Umsetzung stimmenvergleichender Untersuchungen (z.B. im Film *Clear and Present Danger* und in verschiedenen deutschen Krimi-Serien), aber auch in der *voiceprint*-Methode (s.u.) vermittelt.

Diese mechanistische Sichtweise greift jedoch in mehrerer Hinsicht zu kurz. Zum einen ist der menschliche Vokaltrakt eben nicht unveränderlich wie die Papillarlinien auf den Fingerbeeren, zum anderen ist Sprechen nicht nur das Endprodukt physikalischer Prozesse, sondern es schließt auch die Komponente "Verhalten" ein. Der erste Aspekt betrifft die sog. Plastizität des Vokaltrakts [19, 20], der zweite bezeichnet die situationsabhängige Variabilität menschlicher Sprachproduktion.

1.1 Plastizität des Vokaltrakts

Dieser von Francis Nolan [19] geprägte Begriff beschreibt den Umstand, dass der menschliche Vokaltrakt – anders als DNA oder Fingerabdrücke – nicht lebenslang konstant bleibt, sondern sich allein schon durch die Alterung der beteiligten Strukturen verändert [18]. Hinzu kommen Änderungen in der Hohlraumgeometrie durch Entfernung von Rachen- oder Gaumenmandeln bzw. durch Zahnersatz, aber neuerdings auch durch Zahnspangen oder Piercings. Weiterhin sind die Auswirkungen von Genussgiften wie Alkohol und Nikotin auf die

Physiologie der Sprachproduktion zu bedenken [3, 15]. Selbst wenn es also möglich ist, die Resonanzeigenschaften eines Vokaltrakts exakt zu bestimmen, so gilt dies immer nur bedingt, d.h. in Abhängigkeit von den o.g. Bedingungen.

1.2 Sprechen ist Verhalten

Sprechen ist auch insofern nicht nur physikalische Folge anatomischer Gegebenheiten, als der Sprecher sein Verhalten in vieler Hinsicht steuern kann. Diese Steuerung erfolgt nicht immer bewusst – Stress, emotionale Befindlichkeit, aber auch Rollenkonstitution[1] wirken deutlich auf das stimmlich-sprachliche Verhalten ein. Das bedeutet, dass zwar die Grenzen des physiologisch Möglichen durch die Anatomie der Sprechwerkzeuge gesetzt werden, die individuell präferierten Bewegungsmuster jedoch habituell sind und darüber hinaus emotionalen und situativen *constraints* unterliegen. Insofern muss ein Stimmenvergleich auch immer diese Verhaltenskomponente eingeziehen. Dies gilt gleichermaßen für traditionell-phonetische wie auch für (semi)automatische Verfahren.

2 Stimmenvergleich durch Experten

Im "Normalfall" der forensischen Sprechererkennung liegt mindestens eine fragliche Aufzeichnung vor, mit der die Stimmen eines oder mehrerer Vergleichssprecher abgeglichen werden sollen. Hierzu sind verschiedene Verfahren diskutiert worden.

2.1 Voiceprints

Das *voiceprint*-Verfahren ist während des zweiten Weltkriegs parallel auf russischer[2] und amerikanischer Seite [23] entwickelt worden. Es bot erstmals die Möglichkeit, Sprache quasi-dreidimensional sichtbar zu machen. Ursprünglich war es dazu entwickelt worden, deutsche Funker zu identifizieren und damit die Bewegungen deutscher U-Boote nachzuvollziehen, Lawrence Kersta veröffentlichte einen 1962 einen Artikel in der Zeitschrift *Nature*, der ganz bewusst eine Analogie zwischen Fingerabdrücken und Stimmabdrücken postulierte [13]:

"Closely analogous to features found in people's fingerprints, voiceprint identification uses the unique features found in their utterances. Fingerprinting uses the inked impressions of the ten fingers; voiceprinting uses the spectrographic impressions of the utterances of ten words freqently used in telephone conversations." [p.1253]

Die Ausbildung zum *voiceprint*-Experten dauerte wenige Wochen mit anschließendem zweijährigen Praktikum und setzte lediglich einen High-School-Abschluss voraus. Dennoch wurden seitens der Protagonisten dieses Verfahrens Untersuchungen mit erstaunlich geringen Fehlerraten publiziert.[26] Gleichzeitig sah sich das *voiceprint*-Verfahren von Anfang an massiver Kritik ausgesetzt, in deren Zentrum der Vorwurf stand, dass es sich dabei lediglich um einen *gestalt*-basierten Mustervergleich ohne jegliche theoretische Fundierung handle. Auch die scheinbar so geringen Fehlerraten der Methode erfuhren in weiteren Publikationen eine Relativierung.[11] Die *voiceprint*-Technik kann mithin spätestens seit den 1970er Jahren als wissenschaftlich überholt gelten, obwohl sie sich hartnäckig in der Rechtsprechung einiger US-Bundesstaaten hält. In Deutschland hat sie mit Ausnahme eines kurzen Intermezzos am Bayerischen Landeskriminalamt in den 1970er Jahren nie eine Rolle gespielt.

[1] So kann man sich durch Anhebung der Grundfrequenz "klein machen"; umgekehrt ruft eine Absenkung der Grundfrequenz den Eindruck von Autorität und Macht hervor [21].
[2] Eine literarische Darstellung findet sich bei Aleksander Solshenitzyn, Der erste Kreis der Hölle.

2.2 Auditiv

Parallel zur *voiceprint*-Methode entwickelte sich in Großbritannien eine auditiv-phonetische Herangehensweise zum forensisch-phonetischen Stimmenvergleich. Sie basierte auf der starken auditiv-phonetischen Tradition der britischen phonetischen Institute und wurde u.a. durch John Baldwin und Stanley Ellis repräsentiert. Diese Forscher rekurrierten in erster Linie auf regionalsprachliche Merkmale, aber auch auf Stimmqualität im Sinne von John Laver [17]. In Deutschland ereignete sich eine analoge Entwicklung [6]; Edeltraud Knetschke und Margret Sperlbaum sowie Joachim Göschel vertraten im wesentlichen die gleiche auditiv-phonetische Methodik. Diese war durchaus valide, krankte jedoch an mangelnder akustischer Untermauerung.

2.3 Auditiv-akustisches Kombinationsverfahren

Als Antwort auf die Nachteile des *voiceprint*-Verfahrens und der auditiven Analyse wurde am Bundeskriminalamt in den 1980er Jahren ein integratives Verfahren entwickelt, das traditionell-auditive und akustische Verfahren kombiniert [5, 10, 12, 14]. Es trägt dem Umstand Rechnung, dass auditiv-phonetische Befunde mit Hilfe akustisch-phonetischer Verfahren dokumentiert, quantifiziert und damit präzisiert werden können. So lässt sich der auditive Eindruck einer "ungewöhnlich hohen" Stimme exakt als Mittelwert der Stimmbandgrundfrequenz ausdrücken. Dieser Wert kann anschließend mit dem Messwert einer Vergleichsaufzeichnung abgeglichen werden. Dank vorliegender Hintergrundstatistiken zur Verteilung der mittleren Sprechstimmlage [12] ist es ggf. möglich, Aussagen zur Auftretenswahrscheinlichkeit einer gefundenen Übereinstimmung zu treffen.

Folgende Merkmalskomplexe finden bei der auditiv-akustischen Kombinationsmethode Verwendung [5, 10, 14]: Stimme, Sprache und Sprechweise. Im Bereich der **Stimme** interessiert vor allem die bereits erwähnte mittlere Sprechstimmlage sowie deren Standardabweichung, Varianz und Spannbreite (*range*), wobei die gemessenen Werte der fallspezifischen Diskussion bedürfen [4]. Darüber hinaus finden verschiedene Rauhigkeitsindices wie Jitter, Shimmer und HNR Berücksichtigung.

Im Bereich der **Sprache** geht es im wesentlichen um regionale bzw. fremdsprachliche, soziale und individuelle Kennzeichen. Insbesondere für die regionalsprachlichen Merkmale ist die Sprachprägephase von Bedeutung. Diese entspricht in etwa der Schulzeit und bezeichnet denjenigen Lebensabschnitt, während dessen sich die regionaltypischen Artikulationsmuster verfestigen. Erfahrungsgemäß ist es schwierig, nach der Sprachprägephase neue Artikulationsmuster (z.B. in einer Fremdsprache) zu erlernen, aber auch antrainierte Bewegungsmuster fallen zu lassen. Die Erweiterung des Lexikons ist hingegen auch jenseits der Sprachprägephase problemlos möglich. Aus diesem Grund kann bei Vorliegen eines Mischakzents häufig die Abfolge des Erwerbs festgelegt werden.

Genau so wenig wie ein einmal erlernter Regionalakzent verliert sich ein fremdsprachiger Akzent vollständig. Mit Hilfe der vergleichenden Phonetik ist es nicht nur möglich, die Sprachfamilie festzustellen, der die Erstsprache eines Sprechers angehört, sondern meist sogar die individuelle Sprache. Hierbei findet nicht nur auf die Qualität, sondern auch der Ausprägungsgrad des fremdsprachigen Akzents Berücksichtigung.

Soziolektale Merkmale beziehen sich im Deutschen primär auf die linguistische, d.h. suprafonetische Ebene. Es geht dabei um die Komplexität des Satzbaus, den verwendeten Wortschatz und den Grad der phonetischen Reduktion, d.h. den Unterschied zwischen kanonischer (d.h. von der Normlautung geforderter) Aussprache und der tatsächlichen Realisierung.

Individuelle Merkmale finden sich vor allem im Bereich der Sprachpathologie, d.h. bei den Fehlbildungen einzelner Laute. Hier geht es z.B. um verschiedene Lispelfehler, aber auch um

Störungen des Redeflusses. Ferner werden Angewohnheiten (Stereotypien) im Bereich der Lexik berücksichtigt, wie z.B. "eben(t)" oder "also".

Im Bereich der **Sprechweise** werden vor allem Merkmale des Sprechtempos, der Rhythmisierung und der Atmung untersucht. Bei Messungen des Sprechtempos ist es wichtig, die phonetischen (tatsächlich artikulierten) im Gegensatz zu den linguistischen (standardsprachlich vorgesehenen) Silben zu berücksichtigen, weil nur diese die tatsächlich ausgeführten artikulatorischen Gesten repräsentieren [25]. Auch bevorzugte Intonationsintervalle sind sowohl akustisch als auch auditiv beurteilbar. Schließlich interessiert das Respirationsverhalten. Normale Sprechatmung ist meist nicht hörbar. Wenn dies dennoch der Fall ist, stellt sich die Frage nach der Ursache bzw. die Anschlussfrage, ob Atemformanten vorliegen, die sich dann ggf. vergleichen lassen.

2.4 Automatisch

Seit den 1980er Jahren wurden (halb)automatische Sprechervergleichssysteme mit forensischem und kommerziellem Hintergrund erarbeitet. Als Parameter dienten Schallspektren. Diese Systeme scheiterten zunächst am Fehlen einer Kanalkompensation, aber auch den geschilderten forensischen Randbedingungen. Seit den späten 1990er Jahren nahm die Leistungsfähigkeit der Prozessoren enorm zu, so dass die für einen Stimmenvergleich erforderlichen komplexen Rechenoperationen beschleunigt oder sogar erst ermöglicht wurden. Zum zweiten wurden die Methoden zur Parameterextraktion und vor allem der Kanalnormalisierung optimiert.

Zum gegenwärtigen Zeitpunkt existieren verschiedene Systeme[3], die alle nach demselben Prinzip operieren. Dieses wird im folgenden kurz skizziert. – Die Bearbeitung läuft dabei in drei Schritten ab: Parameterextraktion, Modellierung und Berechnung der Ähnlichkeiten bzw. Unähnlichkeiten. In einem ersten Schritt werden aus dem Sprachsignal, im konkreten Fall aus der fraglichen Aufzeichnung und der Vergleichsaufzeichnung, Merkmale extrahiert. Dabei handelt es sich anders als bei den frühen Systemen um Cepstralkoeffizienten[4]. Aus diesen Merkmalen wird ein statistisches Modell der Vergleichsstimme berechnet. Für die weiteren Berechnungen wird ein weiteres, vom konkreten Fall unabhängiges Modell benötigt, das eine möglichst große Bandbreite an Sprechereigenschaften repräsentiert. Dieses ist das Weltmodell oder UBM. Für die Modellierung werden Gaußsche Mischverteilungsmodelle (GMMs) verwendet. Durch einen Vergleich der Werte der Tatprobe mit dem Sprechermodell und dem Weltmodell werden Ähnlichkeitswerte (*scores*) berechnet. Mittels Statistikmodulen werden aus diesen *scores* die *Likelihood Ratios* (LRs) im Sinne des Bayes-Theorems berechnet. Darunter versteht man Verhältniszahlen, welche die Hypothesen "Tat- und Vergleichsaufname stammen vom selben Sprecher" (Hypothese der Staatsanwaltschaft) und "Tat- und Vergleichsaufnahme stammen von verschiedenen Sprechern" (Hypothese der Verteidigung) gegeneinander abwägen. Ein Wert von 1 bedeutet, dass beide Hypothesen gleich wahrscheinlich sind, ein Wert zwischen Null und 1 deutet auf Nichtidentität hin, während ein Wert über 1 die Hypothese unterstützt, dass Sprecheridentität vorliegt. (Werte nahe 1 sind außerdem von geringerer Beweiskraft als sehr hohe oder sehr niedrige.)

Die Berechnung der *Likelihood Ratios* erfordert eine Untersuchung der Frage, welche Aussagekraft die im konkreten Fall gemessenen Ähnlichkeitswerte haben. Dies geschieht mit Hilfe einer Referenzpopulation, von der ebenfalls Modelle gerechnet und mit dem Weltmodell verglichen werden. In der forensischen Praxis ist es besonders wichtig, diese Referenzpopulation

[3] Diese Systeme wurden zumeist im Rahmen von Forschungsvorhaben oder im Auftrag von Behörden entwickelt; eines (BATVOX) wird in der forensischen Welt kommerziell vermarktet (vgl. z.B. [16]).
[4] Becker hat im Rahmen seiner Dissertation [2] ein formantbasiertes System programmiert (FGMM), dessen Performanz jedoch deutlich hinter den cepstral basierten Systemen zurück blieb.

genau an den jeweiligen Fall anzupassen, um eine Verfälschung der Ergebnisse auszuschließen.[5] Gegebenenfalls muss eine fallspezifische Referenzpopulation aufgezeichnet werden.

Um die Performanz eines (semi)automatischen Systems zu evaluieren, wird meist die Irrtumswahrscheinlichkeit getestet. Dies geschieht normalerweise entweder durch die Angabe des Prozentsatzes der korrekt beurteilten Proben, gemittelt über große Zahlen von Tests und ggf. spezifiziert nach unterschiedlichen akustischen Bedingungen[6] oder durch Angabe der sog. *Equal Error Rate*. Dieses Maß trägt dem Umstand Rechnung, dass jedes automatische System zwei Typen von Fehlern machen kann: Falsche Identifikationen und falsche Zurückweisungen. Diese Fehlertypen sind insofern nicht unabhängig voneinander, als, je größer der eine Fehler ist, umso geringer der jeweils andere. Da mithin die Angabe der Häufigkeit lediglich eines der beiden Fehlertypen nicht aussagekräftig ist, ist es üblich, die *Equal Error Rate* (EER) anzugeben. Diese EER ist in sehr starkem Maße von der Qualität und Quantität des Materials abhängig. Es ist daher nicht ausreichend, **eine** EER zu beziffern, sondern es ist immer nach der Irrtumsrate **unter den konkreten Materialvoraussetzungen des jeweiligen Falles** zu fragen. Hier geht es um Dauer, Aufnahmekanal, Sprechweise, gesprochene Sprache, Sprechergeschlecht usw. [2] Ein System zur Sprechererkennung besitzt also nicht eine einzige Irrtumsrate, sondern viele verschiedene.

Es verbleiben allerdings etliche ungelöste und teilweise auch unlösbare Probleme, die eine allzu optimistische Sicht auf die Anwendbarkeit (semi)automatischer Systeme derzeit unangemessen erscheinen lassen. Hier geht es zunächst um die quantitativen und qualitativen Materialvoraussetzungen für den Einsatz (semi)automatischer Systeme, die in der forensischen Wirklichkeit häufig nicht gegeben sind. Unter den international tätigen Experten herrscht Einigkeit darüber, dass eine Netto-Sprechdauer von 20 bis 30 Sekunden die Untergrenze für eine Verwendung in maschinellen Systemen bildet [27]. Ein solcher Materialumfang ist in der forensischen Praxis häufig nicht gegeben. Zudem sind die Systeme nicht auf wechselnde Übertragungskanäle ausgelegt. Auch verstelltes Material kann nicht untersucht werden [22]. Schließlich erweisen sich menschliche Hörer als überlegen, wenn Störgeräusche, Hintergrundrauschen oder instabile Telefonkanäle auftreten [24].

Zusammenfassend kann festgestellt werden, dass automatische Verfahren niemals allein, also ohne parallele Überprüfung durch den geschulten Phonetiker, angewendet werden sollten. Sie finden daher zur Zeit allenfalls als ergänzende Methode Anwendung.

3 Stimmenvergleich durch Opfer und Zeugen

3.1 Methoden

Bei der akustischen Wahlgegenüberstellung (sog. *voice line-up*) handelt es sich um eine in der forensischen Phonetik relativ neue Fragestellung, die seit ca. 1995 systematisch und nach fest etablierten, innerhalb der forensisch-phonetischen *community* unumstrittenen Regeln durchgeführt wird [7]. Sie findet in solchen Fällen Anwendung, in denen das Opfer/der Zeuge den Täter zwar nicht gesehen hat, aber dennoch Angaben zu dessen sprachlichem Verhalten machen kann. Eine akustische Wahlgegenüberstellung kann selbstverständlich nur dann erfolgen, wenn ein Tatverdächtiger vorhanden ist, der zudem kooperativ an der Aufzeichnung einer Vergleichsprobe mitwirkt.

[5] Dies bezieht sich auf das Sprechergeschlecht, die jeweils gesprochene Sprache und die Materialqualität des jeweiligen Falles. Zu unähnliche Referenzpopulationen erhöhen u.U. die LR, so dass der Verdächtige unangemessen belastet wird.

[6] So ist der Anteil korrekter Urteile bei HiFi-Qualität erheblich höher als bei Telefonübertragung und dort wiederum höher als bei GSM-Qualität (Mobilfunk-Standard).

Die wichtigsten methodischen Schritte bei der Erstellung einer akustischen Wahlgegenüberstellung sind folgende:

3.1.1 Befragung des/der Opfer/Zeugen durch die Sachverständige

Hierbei werden die sprachliche und metasprachliche Kompetenz des Opfers/Zeugen exploriert, und das Opfer/der Zeuge wird aufgefordert, das stimmlich-sprachliche Verhalten des Täters so detailliert wie möglich zu beschreiben. Wenn dieses Gespräch ergibt, dass Opfer/Zeuge keine verwertbare Beschreibung der Täterstimme abgeben (können), wird das Vorhaben abgebrochen. Gleiches gilt, wenn das Opfer/der Zeuge aussagt, in dem Täter eine vertraute Stimme wiedererkannt zu haben. In diesem Fall könnte ein *line-up* allenfalls belegen, dass das Opfer/der Zeuge tatsächlich in der Lage ist, die in Frage stehende Person an ihrem stimmlich-sprachlichen Verhalten zu erkennen, nicht jedoch, dass diese Person in der Tatsituation gehört wurde.

3.1.2 Gespräch mit dem Tatverdächtigen seitens der Sachverständigen

Ein solches Gespräch dient dem Abgleich zwischen der Beschreibung der Täterstimme durch Opfer/Zeugen und der sachverständigen Einschätzung derselben. Wenn deutliche und nicht – z.B. durch die Situation – erklärbare Unterschiede festgestellt werden zwischen dem sprachlichen Täterverhalten, wie es das Opfer schildert, und dem Sprachverhalten des Tatverdächtigen, so bedeutet dies das Ende des Vorhabens, weil es nicht sinnvoll wäre, einen Zeugen mit einer Identifikationsaufgabe zu konfrontieren, die der eigenen Schilderung der Täterstimme entgegensteht.

3.1.3 Aufzeichnung des Tatverdächtigen sowie einer größeren Zahl von Distraktoren (*foils*)

Die Vergleichssprecher werden im Hinblick auf eine Vergleichbarkeit mit der Stimme des Tatverdächtigen ausgewählt. Dies betrifft Geschlecht, wahrgenommenes Alter, regionalen / fremdsprachigen Akzent, wahrgenommene Stimmhöhe, Sprechfehler, Sprechtempo etc. Bei Bedarf wird die Beurteilung der Ähnlichkeit durch akustisch-phonetische Messungen unterstützt. Dieser Schritt ist entscheidend für die Konstruktion der Wahlgegenüberstellung: Weicht die Stimme eines Distraktors zu stark von der Stimme des Tatverdächtigen ab, so reduziert sich damit die effektive Zahl der Vergleichsstimmen, ist die Stimme eines Distraktors der Tatverdächtigen extrem ähnlich (z.B. durch professionelle Imitation), so wird die Aufgabe für Opfer/Zeugen über Gebühr erschwert. Aus diesem Grund werden deutlich mehr Vergleichssprecher aufgezeichnet, als für die Gegenüberstellung letztendlich erforderlich sind. Aufgrund psychophonetischer Forschungsergebnisse hat sich für die Gegenüberstellung eine Gesamtzahl von fünf bis sechs Stimmen als optimal erwiesen (vier bis fünf Distraktoren plus Tatverdächtiger).

3.1.4 Vortests mit Unbeteiligten

Um herauszufinden, ob der Tatverdächtige *per se* eine erhöhte Erkennungswahrscheinlichkeit aufweist, werden Vortests durchgeführt mit Versuchspersonen, die den Fallzusammenhang nicht kennen. Sie werden mit den für die Wahlgegenüberstellung zusammengestellten Stimuli konfrontiert und sollen beurteilen, welche(r) der Sprecher ein entsprechendes Delikt begangen haben könnte(n).

Die Befragung zielt ausschließlich auf die Ermittlung von stimmlich-sprachlichen Stereotypen ab – selbstverständlich ist es anhand der Stimme genauso wenig möglich wie anhand der Physiognomie, zwischen Straftätern und Unbescholtenen zu unterscheiden. Wenn allerdings ausgerechnet der Tatverdächtige eine Stimme hat, die von Laien als "typische Verbrecherstimme" wahrgenommen wird und dies bei den Distraktoren nicht der Fall ist, so wäre die

Gegenüberstellung dem Tatverdächtigen gegenüber unfair und sollte so nicht durchgeführt werden. Die Urteile sollten im Idealfall zufällig verteilt sein. Wenn sie sich deutlich auf eine Vpn. konzentrieren, so scheidet diese aus dem Experiment aus; sollte es sich dabei um den Tatverdächtigen handeln, so ist der *line-up* neu zu gestalten.

In einem zweiten Vortest wird in einem paarweisen Test geprüft, wie ähnlich naive Hörer die beteiligten Stimmen empfinden. Wenn sich dadurch erweist, dass eine der Stimmen als extrem divergierend von allen anderen wahrgenommen wird, so ist dieser Sprecher durch einen anderen zu ersetzen.

3.1.5 Testdurchführung

Der Test sollte durch eine Person administriert werden, die nicht darüber informiert ist, bei welchem Sprecher es sich um den Tatverdächtigen handelt (Doppelblind-Verfahren). Das Opfer/der Zeuge erhält Gelegenheit, sich die verschiedenen Stimmen in beliebiger Reihenfolge auch wiederholt anzuhören. Es ist unbedingt auf die Möglichkeit hinzuweisen, dass sich der Täter **nicht** unter den Sprechern befindet. Der Ablauf der Wahlgegenüberstellung kann per Videoaufzeichnung dokumentiert werden, um ggf. in der Hauptverhandlung die Reaktionen des Opfers/Zeugen nachvollziehbar zu machen. Unter Umständen ist es erforderlich, einen psychologischen Betreuer vorzuhalten.

Insbesondere angesichts des außergewöhnlich hohen Arbeitsaufwands und der selten gegebenen Voraussetzungen wird die Konstruktion eines *voice line-up* immer einen seltenen Sonderfall forensischer Stimmenvergleichung bilden.

Literatur

[1] Baldwin, J. / French, P.: Forensic Phonetics. London / New York: Pinter 1990.

[2] Becker, T.: Automatischer forensischer Stimmenvergleich. Diss. Trier 2011.

[3] Braun, A.: The Effect of Cigarette Smoking on Vocal Parameters. In: Proceedings of the ESCA Conference on Speaker Identification Verification Recognition, Martigny 1994, pp. 161 - 164.

[4] Braun, A.: Fundamental Frequency - How Speaker-Specific Is It? In: Braun, A. / Köster, J.-P. (Eds.): Studies in Forensic Phonetics. Trier: Wissenschaftlicher Verlag 1995, pp. 9-23.

[5] Braun, A.: Die forensische Analyse von Stimme und Sprache. In: Gundermann, H. (Ed.): Die Ausdruckswelt der Stimme. Erste Stuttgarter Stimmtage. Heidelberg: Hüthig 1998, pp. 88-102.

[6] Braun, A. / Köster, J.-P. / Künzel, H. J. / Odenthal, H.-J.: Speaker Recognition in Germany. In: Wolf, D. (Ed.): Beiträge zur Geschichte und neueren Entwicklung der Sprachakustik und Informationsverarbeitung. Werner Endres zum 90. Geburtstag. Dresden: w.e.b. 2005, pp. 78-86.

[7] Broeders, A.P.A. & van Amelsvoort, A.G.: Line-up Construction for Forensic Earwitness Identification: a Practical Approach. In: Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco 1999, pp. 1373-1376.

[8] Champod, Chr. / Meuwly, D.: The inference of identity in forensic speaker recognition. In: Proceedings of RLA2C, Avignon 1998, pp. 125-134.

[9] Fant, G. (1960): Acoustic Theory of Speech Production. Den Haag: Mouton.

[10] Gfrörer, St.: Sprechererkennung und Tonträgerauswertung. In: Widmaier, Gunter (Ed.): Münchener Anwalts-Handbuch Strafverteidigung. München: Beck 2006, pp. 2505-2526.

[11] Hazen, B.: Effects of differing phonetic contexts on spectrographic speaker identification. In: The Journal of the Acoustical Society of America 54, 1973, pp. 650-660.

[12] Jessen, M.: Forensic Phonetics. In: Language and Linguistics Compass, 2, 2008,pp. 1-41.
[13] Kersta, L.: Voiceprint identification. In: Nature, 196, 1962, pp. 1253-1257.
[14] Künzel, H.J.: Sprechererkennung. Grundzüge forensischer Sprachverarbeitung. Heidelberg: Kriminalistik-Verlag 1987.
[15] Künzel, H.J. / Braun, A. / Eysholdt, U.: Einfluß von Alkohol auf Sprache und Stimme. Heidelberg: Kriminalistik-Verlag 1992.
[16] Künzel, H.J., Gonzalez-Rodriguez, J., Ortega-García, J.: Effect of Voice Disguise on the Performance of a Forensic Automatic Speaker Identification System. Paper presented at A Speaker Odyssey: The Speaker and Language Recognition Workshop, Toledo 2004, 4p.
[17] Laver, J.: The Phonetic Description of Voice Quality. Cambridge: CUP 1980.
[18] Linville, S.E.: Vocal aging. San Diego: Singular 2001.
[19] Nolan, F.: The Phonetic Bases of Speaker Recognition. Cambridge: CUP 1983.
[20] Nolan, F.: Speaker Recognition and Forensic Phonetics. In: Hardcastle, W.J. / Laver, J. (Eds.): The Handbook of Phonetic Sciences. Oxford: Blackwell 1997, pp. 744-767.
[21] Ohala, John J.: An Ethological Perspective on Common Cross-Language Utilizatian of F_0 of Voice. In: Phonetica 41, 1984, pp. 1-16.
[22] Perrot, P. / Chollet, G.: The question of disguised voice. In: Acoustics08, Paris 2008, pp. 5682-5685.
[23] Potter, R. / Kopp, G. / Green, H.: Visible Speech. New York: Van Nostrand Reinhold 1947.
[24] Schmidt-Nielsen, A. / Crystal, T. H.: Human vs. Machine Speaker Identification with Telephone Speech. In: Proceedings of the International Congress on Speech and Language Processing 1998, 9 p.
[25] Schilz, J.: Wie individualtypisch ist die Artikulationsrate unter Berücksichtigung phonetischer bzw. linguistischer Silben? Eine empirisch vergleichende Studie. Magisterarbeit Trier 2008.
[26] Tosi, O. / Oyer, H. / Lashbrook, W. / Pedrey, C. / Nichol, J. / Nash, E.: Experiment on voice identification. In: The Journal of the Acoustical Society of America 51, 1972, pp. 2030-2043.
[27] Van Leeuwen, D.A. / Martin, A.F. / Przybocki, M.A. / Bouten, J.: NIST and NFI-TNO evaluations of automatic speaker recognition. In: Computer Speech and Language, 20, 2006, pp. 128-158.