

Forensische Sprechererkennung und Tonträgerauswertung in Praxis und Forschung - Teil 2

Mit „habituell“, werden Angewohnheiten erfasst, die eine Person im Laufe der Zeit für sich selbst entwickelt bzw. von anderen übernommen hat. Gemeint sind insbesondere solche Angewohnheiten, die mit dem grammatischen System einer Sprache nichts zu tun haben. Ein Beispiel für eine solche Angewohnheit ist das **Sprechtempo**. Einige Personen sprechen schneller als andere, und die Geschwindigkeit, mit der gesprochen wird, ist nicht etwas, das von den grammatischen Regeln einer Sprache vorgeschrieben oder eingeschränkt wird. Insofern haben verschiedene Sprecher die Freiheit, unterschiedlich schnell zu sprechen. Sprechtempo kann in Form der sog. Artikulationsrate gemessen werden, wobei innerhalb flüssiger Sprachanteile die Anzahl der Silben pro Sekunde gemessen werden (Jessen 2007). Ein anderes Beispiel für ein habituelles Merkmal ist die **Melodik** der Stimme. Diese kann als Standardabweichung, oder besser noch als Variationskoeffizient der Grundfrequenzwerte in einer Aufzeichnung (Standardabweichung geteilt durch Mittelwert) gemessen werden (Künzel 1987; Jessen et al. 2005). Je höher der Wert, umso größer die wahrnehmbare Bewegtheit der Sprechmelodie (Intonation), und je geringer der Wert, umso eher ist dies ein Zeichen einer monotonen Sprechweise. Zwar ist die Melodik der Stimme auch von Situationen, Emotionen und linguistischen Faktoren abhängig, aber wie melodisch oder monoton jemand spricht, ist auch eine sprecherspezifische Angewohnheit.³ Das **Pausenverhalten**, d.h. die Häufigkeit, Dauer und Platzierung von Sprechpausen ist ein weiteres Merkmal, anhand dessen Sprecher sich unterscheiden. Pausen können entweder reine Stillephasen sein oder sie können „gefüllt“, sein. Von „gefüllten Pausen“, spricht man bei Äußerungen, die sich als „äh“, und „ähm“, verschriften lassen. Sprecher unterscheiden sich unter anderem in der Häufigkeit, mit der sie solche gefüllten Pausen verwenden (und einige verwenden sie gar nicht) sowie in dem Häufigkeitsverhältnis und dem Kontext, in dem sie „äh“, und „ähm“, verwenden (Bauer 2007). Pausen können auch in dem Sinn gefüllt sein, dass man ein Atemgeräusch erkennen kann. Das Atemverhalten ist dann auch eine weitere Eigenschaft, die am ehesten in den Bereich der habituellen Merkmale fällt.

Im Rahmen eines Stimmenvergleichs werden in der Täteraufzeichnung und der Vergleichsaufzeichnung die vorkommenden sprecherklassifizierenden und sprecherunterscheidenden Merkmale gesammelt und verglichen, wobei aufgrund der Qualität und Quantität des Untersuchungsmaterials nicht zwangsläufig alle der hier genannten Merkmale untersucht werden können bzw. aussagekräftig sind (bzw. weitere hinzukommen können, die hier nicht genannt sind). Am Ende dieser Analyse wird eine Wahrscheinlichkeitsaussage über die Identität oder Nichtidentität der beiden Sprecher getroffen.

Zeugenwahrnehmung

Für den Fall, dass gar keine Aufzeichnung des Täters vorliegt, dass die Täterstimme aber durch einen Zeugen wahrgenommen wurde (welcher auch gleichzeitig das Opfer der Straftat sein kann, wie im Falle einer Vergewaltigung), muss man zwei Fälle unterscheiden. Zum einen kann es sein, dass der Täter dem Zeugen bereits bekannt ist. Wenn durch dessen Zeugenaussage oder aus unabhängigen Gründen eine Person verdächtigt wird und der Fall vor Gericht kommt, ist die Stimmenidentifizierung als reguläre Zeugenaussage zu behandeln. Aus Sicht der Sprechererkennung kann die Aussagekraft einer solchen Zeugenaussage reduziert sein, wenn bestimmte ungünstige Faktoren vorliegen. Beispielsweise kann es sein, dass es vom Täter nur sehr kurze Äußerungen gibt, dass seine Sprechweise ungewöhnlich ist (z.B. dadurch dass geschrien wird oder die Stimme verstellt wird) oder dass der Täter maskiert war und so einige Frequenzbereiche abgeschwächt werden. Solche Konstellationen können beispielsweise bei einem Banküberfall auftreten. Probleme können bereits entstehen, wenn die Begegnung mit dem Täter nicht direkt war, sondern über das Telefon stattgefunden hat.

Man weiß aus der eigenen Erfahrung, dass es Beispiele gibt, wo man einen Bekannten oder Verwandten nicht (gleich) am Telefon erkannt hat, beispielsweise weil zu wenig gesagt wurde oder in ungewöhnlicher Weise gesprochen wurde. Das Problem kann auch darin bestehen, dass der Anruf zu einer ungewöhnlichen Zeit erfolgte, zu der man einen Anruf von dieser Person nicht erwartet hätte oder zu der man auf den Anruf einer anderen Person gewartet hat, die ähnlich klingt. Dieses zeigt, dass unsere Fähigkeit, Bekannte oder Verwandte am Telefon zu erkennen, teilweise erwartungsgesteuert ist, von einem begünstigenden Kontext abhängt und in ungewöhnlichen Situationen und ohne den vertrauten Kontext zusammenbrechen kann (natürlich bis der Name genannt wird). Der bekannte und kürzlich verstorbene Phonetiker Peter Ladefoged berichtete, dass er in einem Experiment, in dem ihm kontextfrei Wörter und Sätze vorgespielt wurden, seine eigene Mutter nicht wiedererkannt hat.

Neben der am Beispiel geschilderten „falschen Zurückweisung“, d.h. der Situation, in der eine bestimmte Person nicht identifiziert wurde, obwohl sie es war, können ungünstige Umstände auch zu einer „falschen Identifizierung“, führen, wobei die falsche Person identifiziert wird. Solche Aspekte sind bei der Beurteilung von Zeugenaussagen zu berücksichtigen. Eine ausführliche Zusammenstellung von Faktoren, welche die Wiedererkennbarkeit von Stimmen negativ beeinflussen können, liefern Hammersley und Read (1990).

Die zweite Situation, die bei der Stimmwahrnehmung durch Zeugen vorliegen kann, besteht darin, dass der Täter dem Zeugen nicht bekannt ist. In einem solchen Fall ist es möglich, eine sog. Stimmenwahlgegenüberstellung durchzuführen. In verschiedenen Punkten ist eine solche auditive Gegenüberstellung analog zur bekannteren visuellen Gegenüberstellung. In beiden Fällen ist es erforderlich, eine Menge weiterer Kontrollpersonen zu präsentieren, die sich nicht in gravierenden Punkten von dem Verdächtigen unterscheiden, so dass es nicht sein darf, dass der Verdächtige schon von unbeteiligten Testpersonen als besonders auffällig oder abweichend gegenüber den anderen dargebotenen Personen wahrgenommen wird. Im Unterschied

zur visuellen Gegenüberstellung ist es allerdings bei der Stimmenwahrgegenüberstellung schwieriger zu definieren und zu erkennen, was im stimmlich-sprachlichen Bereich als „ähnlich,,, „auffällig,, usw. zu betrachten ist. Hierzu ist ein umfangreiches Fachwissen in Gebieten wie Linguistik und Phonetik erforderlich, das bei der Auswahl von Kontrollpersonen und verschiedenen anderen Aspekten der Planung, Durchführung und Auswertung von Stimmenwahlgegenüberstellungen eingesetzt wird. Es ist deswegen sehr zu empfehlen, in Fällen, in denen eine Stimmenwahlgegenüberstellung beabsichtigt ist, einen Experten für Sprechererkennung zu Rate zu ziehen. Werden methodische Fehler gemacht, die später angefochten werden, kann eine Stimmenwahlgegenüberstellung nicht erneut durchgeführt werden, da dann das Problem der „sekundären Identifikation,, besteht, d.h. es ist möglich, dass sich der Zeuge bei einer zweiten Gegenüberstellung nicht an die Stimme in der Tatsituation erinnert, sondern an die Verdächtigenstimme aus der ersten Gegenüberstellung. Es gibt einen umfangreichen Katalog von Richtlinien zur Durchführung von Stimmenwahlgegenüberstellungen, die von internationalen Behörden im Rahmen von ENFSI anerkannt sind (Broeders und van Amelsvoort 1999). Eine Zusammenfassung dieser Richtlinien liefert Gfroerer (2006). *Die Autorin wurde in einem Fall beauftragt, die wissenschaftliche Qualität einer Stimmenwahlgegenüberstellung zu bewerten, die bereits durchgeführt worden war. Dabei hatte ein maskierter Täter nach Dienstschluss einen Supermarkt überfallen, den noch anwesenden Schlüsselhaber zur Herausgabe des vorhandenen Geldes gezwungen, ihn anschließend in die Kühlkammer eingesperrt und den Supermarkt in Brand gesteckt. Der Zeuge konnte glücklicherweise rechtzeitig befreit werden. Im Rahmen einer Stimmenwahlgegenüberstellung hatte er den Täter identifiziert. Allerdings wurden mehrere methodische Fehler gemacht. Der grundsätzlichsste Fehler bestand darin, dass die Gegenüberstellung „live,, durchgeführt wurde, d.h. es wurden der Verdächtige und die Kontrollpersonen in einem Nebenraum einbestellt und jeweils zum Reden aufgefordert, so dass der Zeuge die Stimmen durch die angelehnte Tür hören, die Personen aber nicht sehen konnte. Richtig wäre es gewesen, die Stimmen des Verdächtigen und der Kontrollpersonen vorher aufzuzeichnen und die aufgezeichneten Stimmen dem o.g. „Fairness-Test,, mit unbeteiligten Hörern auszusetzen, bevor die eigentliche Gegenüberstellung beginnt. Dieser und weitere wichtige Mängel mussten von der Autorin vor Gericht erwähnt werden.*

Forschung

In der forensischen Sprechererkennung und Tonträgerauswertung besteht ein sehr großer Bedarf an Forschung und Entwicklung. Teilweise kann dieser Bedarf dadurch gedeckt werden, dass es Firmen gibt, die Lösungen für einige der relevanten Aufgaben anbieten. Dieses ist insbesondere der Fall im Bereich der Qualitätsverbesserung, wo es eine Vielzahl entsprechender Hard- und Softwareangebote im Bereich der Audiofiltertechnologie gibt. Hier ist ein genaues Beobachten und Testen der Produkte erforderlich, denn es gibt große Unterschiede im Leistungsumfang, in den methodischen Grundlagen und im Preis solcher Produkte. Eine andere Herangehensweise ist das Sichten der sehr umfangreichen Forschungsliteratur über die Themen Sprache und Stimme. Vor allem bei der Sprecherklassifikation, die für die Stimmenanalyse zentral ist, kann auf bestehende Forschungsergebnisse zurückgegriffen werden, denn die meisten Bereiche der Sprecherklassifikation sind durch eigenständige akademische Disziplinen vertreten. So beschäftigt sich die Dialektologie mit regionalen Variationen, die Soziolinguistik mit sozialen Variationen (die das Thema Alter und Geschlecht teilweise einschließen), das Fach Deutsch als Fremdsprache bzw. die Zweitspracherwerbsforschung mit Deutschsprechern einer fremden Muttersprache und die Phoniatrie bzw. klinische Linguistik mit Sprach-, Sprech-, und Stimmstörungen. Zwar gibt es auch in diesen Bereichen Bedarf an Forschung, in der die forensischen Gegebenheiten besonders berücksichtigt werden, aber der größte Forschungsbedarf liegt im Bereich des Stimmenvergleichs und der in Tabelle 2 aufgeführten sprecherspezifischen Eigenschaften. Dieses ist auch der Bereich, für den sich andere akademische Disziplinen am wenigsten zuständig fühlen. Um die Eignung eines Aussprachemerkmals für die Sprechererkennung zu prüfen, ist es erforderlich, die Frage der *intraindividuellen* Variation und die der *interindividuellen* Variation zu klären. Eine große **interindividuelle Variation** ist die Grundvoraussetzung für ein sprecherspezifisches Merkmal. Dieses gilt natürlich nicht nur für die Sprechererkennung, sondern auch für andere Bereiche der Personenerkennung wie die Daktyloskopie und die DNA-Analyse. Das heißt, Personen sollten sich möglichst stark bzgl. eines Merkmales oder einer Merkmalskombination unterscheiden, so dass bei einem Vergleich nur noch wenige Personen oder überhaupt nur eine Person in Frage kommt. Das individualisierende Potential in der Sprechererkennung hängt zunächst von dem Sprecher ab. Es gibt Sprecher, die sich am Rande einer Häufigkeitsverteilung befinden (z.B. eine ungewöhnlich hohe Stimme haben) oder die ungewöhnliche Merkmale im Bereich der Sprecherklassifikation haben, wie z.B. eine bestimmte Sprechstörung, die nur selten vorkommt. Andere Sprecher hingegen haben eine eher „durchschnittliche,, Stimme und werden auch von Laien häufiger mit anderen verwechselt. Was durchschnittlich und was ungewöhnlich ist, ist allerdings nicht immer von Laien einsichtig und wird auch für die Experten erst im Laufe der Zeit durch Erfahrung und durch Forschung immer klarer. Nicht alle solcher ungewöhnlichen Merkmale und Merkmalskombinationen sind durch das Gehör erfassbar und klassifizierbar. Deshalb ist es erforderlich, akustische Messungen an einer größeren Anzahl von Sprechern vorzunehmen und eine statistische Häufigkeitsverteilung zu erstellen. Wichtige Grundlagen zu diesem Ansatz wurden am BKA in den 1980er Jahren anhand einer Untersuchung der mittleren Grundfrequenz bei 100 Männern und 50 Frauen erarbeitet (Künzel 1987). Eine neuere Untersuchung wurde 2001 am BKA durchgeführt. In dem als „Pool 2010,, bezeichneten Sprachkorpus wurden Sprachproben von 100 Männern erhoben, die in verschiedenen Sprechstilen (Lesen, Spontansprache) und Versuchsbedingungen (normal, laut, Telefon) sprachen. An diesem Korpus wurden zunächst die mittlere Grundfrequenz und andere Grundfrequenzparameter (Standardabweichung, Variationskoeffizient, Kurtosis usw.) gemessen (Jessen et al. 2005; Becker und Kreuzer 2008). Abbildung 2 zeigt, wie sich die mittlere Grundfrequenz bei 100 männlichen erwachsenen Sprechern des Deutschen verteilt.

Zunächst ist das vordere Histogramm relevant (blaue Säulen), was die Messergebnisse aus dem Teil des Pool 2010 wiedergibt, in dem spontan und normal-laut gesprochen wurde. Der Schwerpunkt der Verteilung liegt dort zwischen 110 und 120 Hz, was

somit als eine normale Sprechstimmlage zu klassifizieren ist. Spricht hingegen jemand mit einer mittleren Grundfrequenz von ca. 85 Hz oder einer von ca. 165 Hz, zeigt die Abbildung, dass es sich respektive um eine besonders tiefe bzw. besonders hohe Stimme handelt.

Über die Grundfrequenz hinaus wurde am Pool 2010 das Sprechtempo untersucht (Jessen 1997), das Vorkommen von gefüllten Pausen (Bauer 2007), einige linguistisch-phonetische Detailmerkmale (Jessen 2008) und vor kurzem die Formantenfrequenzen (Moos 2008; Becker et al. 2008). Das Pool 2010 hat auch bei der Entwicklung der maschinellen Sprechererkennung am BKA in Zusammenarbeit mit der FH-Koblenz (Prof. Dr. Franz Broß) eine Rolle gespielt. Die maschinelle Sprechererkennung ist ein neuerer Ansatz, der sich derzeit in Entwicklung befindet. Er wird, insofern das Material in einem Fall dazu geeignet ist, ergänzend (aber nicht ersetzend) zu den bisher geschilderten phonetisch-sprachwissenschaftlichen Methoden eingesetzt und wird immer einer Prüfung durch den forensisch-sprachwissenschaftlich ausgebildeten Sachverständigen unterzogen.⁴

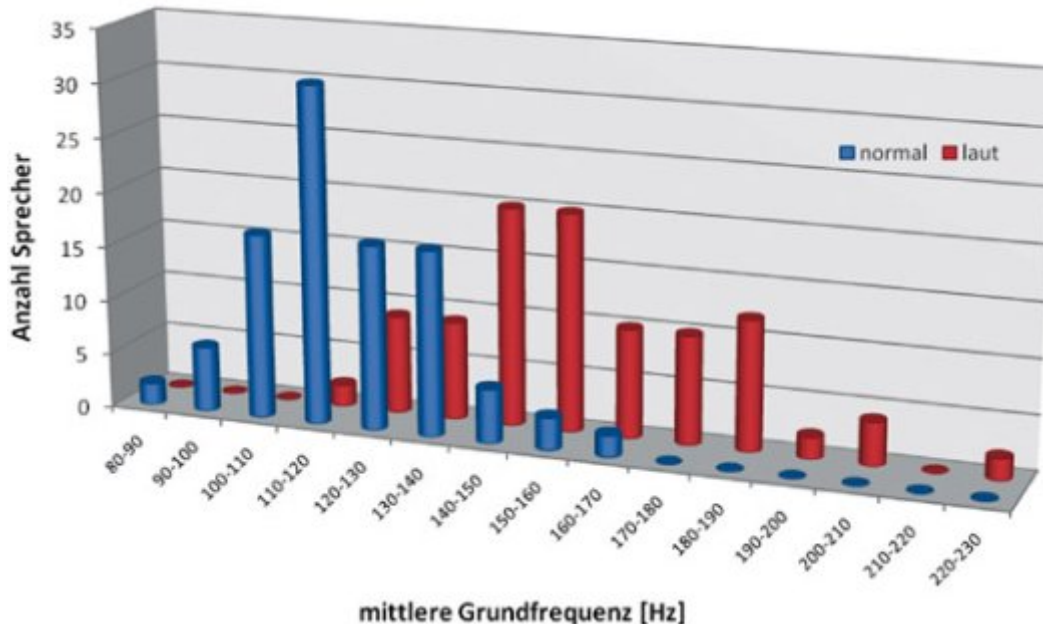


Abbildung 1: Histogramm der mittleren Grundfrequenz bei 100 Männern. Vordere Reihe (in blau): Spontansprache in normaler Sprechlautstärke. Hintere Reihe (in rot): Spontansprache, wenn laut gesprochen wird.

Während die interindividuelle Variation eine Grundvoraussetzung in der Sprechererkennung ist, stellt die **intraindividuelle** Variation eine Herausforderung dar. Mit dem Begriff der intraindividuellen Variation wird die Problematik erfasst, dass die Stimme und die sprachlichen Muster eines Sprechers sich langfristig und kurzfristig verändern können. Dabei stellen vor allem die kurzfristigen Veränderungen ein Problem dar. Eine häufige Konstellation in der Fallarbeit ist die, dass der fragliche Sprecher laut und gestresst spricht, während der mutmaßlich gleiche Sprecher, wenn mit ihm eine Vergleichs-aufnahme erhoben wird, ruhig und unemotional ist. Diese Unterschiede in der Sprechlautstärke und dem Grad an Stress bzw. Emotion haben einen Einfluss auf die Merkmale, die in einem Stimmenvergleich analysiert werden. Passen die Sprechweisen in den zu vergleichenden Proben nicht zusammen, ist es erforderlich, genügend Wissen darüber zu haben, welchen Einfluss solche Faktoren auf die Merkmale haben.

Zu diesem Zweck wurde bei der Erhebung des bereits erwähnten Pool 2010 eine Versuchsbedingung aufgenommen, bei der die Sprecher zum Lautsprechen angeregt wurden. Dieses geschah, indem sie einen Kopfhörer aufsetzten, über dem ein lautes Rauschen dargeboten wurde. Dass man bei einer lauten Schallquelle (z.B. auch Musik oder Straßenlärm) unwillkürlich lauter spricht (sog. „Lombard-Effekt“), entspricht der Alltagserfahrung und wurde erstmals 1911 von dem französischen Hals-Nasen-Ohren-Arzt Etienne Lombard, der der Namensgeber dieses Phänomens ist, publiziert. Das hintere Histogramm in Abbildung 2 zeigt das Ergebnis der Lombard-Bedingung auf die mittlere Grundfrequenz. Wie dort zu sehen ist, verschiebt sich die Grundfrequenzverteilung beim Lautersprechen in deutlich höhere Wertebereiche als bei normal lautem Sprechen.

Eine weitere Untersuchung zur intraindividuellen Variation wurde von Jessen (2006) durchgeführt. Hierbei wurden zwei Gruppen von Versuchspersonen jeweils mehreren Arten von Stress ausgesetzt. Bei den Versuchspersonen handelte es sich zum einen um zehn männliche Polizeianwärter der Hessischen Fachhochschule der Polizei und zum anderen um zehn männliche Mitglieder des deutschen Sondereinsatzkommandos GSG 9.

Das Experiment fand in den Räumlichkeiten der GSG 9 statt. Bei der ersten Aufgabe bekamen die Probanden die Tonaufzeichnung einer Geiselnahme vorgespielt und sollten sich möglichst viele Details davon merken. Gleichzeitig bekamen sie Bögen mit Rechenaufgaben vorgelegt, die aus den Aufnahmeprüfungen der Polizei stammen, und sollten diese schnell und vollständig lösen. Diese Art von Stress wurde als „kognitiver Stress“, bezeichnet. Für die zweite Aufgabe wurden die Probanden in das Schießkino der GSG 9 geführt. Dort wurden auf einer Doppelleinwand aus mehreren Diaprojektoren verschiedene Szenen einer Geiselnahme in einer Bank projiziert. In den Dias waren jeweils bewaffnete, unbewaffnete und bedrohende Täter zu sehen sowie Bankangestellte, Zivilisten und Mitglieder eines Sondereinsatzkommandos. Die Instruktion lautete, nur auf jene Geiselnahmer zu schießen, die eine Zivilperson unmittelbar mit einer Waffe bedrohten; alle anderen sollten nicht berücksichtigt werden. Die Versuchspersonen benutzten ihre eigenen Dienstwaffen, die mit scharfer Munition geladen waren. Diese Art von Stress wurde als „physischer Stress“, bezeichnet. Jeweils unmittelbar nach diesen Bedingungen waren sprachliche Aufgaben aus

dem Bereich Spontansprache und Lesen zu erfüllen, wobei davon auszugehen war, dass der erlebte Stress der Bedingungen sich auf das stimmliche und sprachliche Verhalten auswirken würde. Die Aufnahmen wurden in Hinblick auf mehrere Merkmale ausgewertet, die in der Sprechererkennung verwendet werden, so u.a. die Sprechstimmlage, die Melodik, das Sprechtempo und das Pausenverhalten (siehe Tabelle 2). Die Ergebnisse waren teilweise sehr komplex, aber zu den robusteren Ergebnissen gehörte, dass für viele der Sprecher die mittlere Grundfrequenz unter Stress gegenüber der Normalbedingung signifikant anstieg. Dieses war bei der Bedingung mit physischem Stress etwas deutlicher der Fall als in der mit kognitivem Stress. Außerdem zeigte eine Wahrnehmungsstudie anhand des Materials, dass die GSG 9-Beamten als deutlich weniger stressbelastet und deutlich konzentrierter wahrgenommen wurden als die Polizeischüler. Abbildung 3 illustriert die zweite Stressbedingung. Dort ist auch zu erkennen, dass die Probanden an weitere portable Messgeräte angeschlossen waren, die zur Messung von physiologischen Stresskorrelaten verwendet wurden.



Abbildung 2: Stressbelastung im Schießkino der GSG 9 (Bedingung „physischer Stress“)

Über Experimente dieser Art bekommt der Experte ein genaueres Bild über Art und Grad intraindividuelle Variationen und ist für Stimmenvergleiche in solchen Fällen besser vorbereitet, in denen sich die fraglichen Aufzeichnungen und die Vergleichsaufzeichnungen bezüglich dieser und ähnlicher Sprechweisen und Sprechbedingungen unterscheiden. In diesem Zusammenhang ist u.a. auch eine Untersuchung von Künzel et al. (1992) über den Einfluss von Alkohol auf das Sprechen zu nennen sowie die Untersuchung von Künzel (2000) über den Einfluss verschiedener Stimmverstellungen. Zum Abschluss sei noch einmal auf Tabelle 1 eingegangen. Dort wurde die Möglichkeit erwähnt, dass es einen Zeugen gibt aber keine Sprachaufzeichnung und keinen Verdächtigen. Im visuellen Bereich wäre diese Situation ein möglicher Anlass für die Erstellung eines Phantombildes. Gibt es so etwas wie ein Phantombild auch für Stimmen? Was die Praxis der Sprechererkennung in Deutschland betrifft, lautet die Antwort nein. (Basierend auf den existierenden Veröffentlichungen in der forensischen Sprechererkennung gibt es dieses auch nicht in anderen Ländern.) Technisch ist die Erstellung eines „akustischen Phantombildes“, das bereits von Nolan (1983) als wünschenswert herausgestellt wurde, ein Fall für die Sprachsynthese. Erst seit ungefähr 10 Jahren ist es möglich, sehr natürlich klingende Stimmen zu synthetisieren und kontinuierlich zu verändern. Einen Überblick über moderne Verfahren der „voice transformation“, und „voice conversion“, liefert Stylianou (2008). Es könnte sich also lohnen, diese technischen Möglichkeiten auszuloten, um eines Tages ein akustisches Phantombildverfahren zu realisieren.

Fußnoten:

1 URL <http://www.stimmenvergleich.de>

2 Einen anderen und in verschiedenen Punkten umfangreicheren Überblick über das Gebiet der forensischen Sprechererkennung und Tonträgerauswertung, als in diesem Artikel möglich ist, liefert Gfroerer (2006).

3 Auch die anderen der in Tabelle 2 aufgeführten und hier besprochenen Merkmale sind nicht ausschließlich Sprechermerkmale, sondern können in einem gewissen Umfang durch Faktoren wie Situation oder Emotion variieren. Auch ist die Zuordnung der

einzelnen Merkmale zu den drei Faktoren nur eine Vereinfachung. Beispielsweise kann zwar eine raue Stimmqualität eine Folge der anatomischen Struktur (und Funktion, d.h. Physiologie) der Stimmlippen sein, aber eine raue Stimme kann auch durch Angewohnheit verstärkt oder abgeschwächt werden oder kann sogar absichtlich der Imagepflege dienen (z.B. bei Männern, die das Image eines „tough guy“, vermitteln wollen). Solche Möglichkeiten sind bei der Analyse zu berücksichtigen.

4 Eine genauere Erläuterung der maschinellen Sprechererkennung ist für einen Artikel in der Zeitschrift Kriminalistik in Arbeit.

Literatur:

- Bauer, Dominik (2007): Sprecherspezifik der Stimmtonhöhe in gefüllten Pausen und am Ende von Intonationsphrasen. Magisterarbeit, Universität des Saarlandes, Fach Phonetik und Phonologie.
- Becker, Timo und Wolfgang Kreuzer (2008): Automatische Sprechererkennung basierend auf Stimmgrundfrequenz-Merkmalen mittels Hauptkomponentenanalyse. Jahrestagung *Deutsche Gesellschaft für Akustik*, Dresden.
- Becker, Timo, Michael Jessen und Catalin Grigoras (2008): Gaussian Mixture Models based on formant frequencies. Tagung der *International Speech Communication Association*, Brisbane.
- Boss, Dagmar, Stefan Gfroerer und Nikolai Neoustroev (2003): A new tool for the visualization of magnetic features on audiotapes. *International Journal of Speech, Language and the Law* 10: 255–276.
- Broeders, A.P.A. und A.G. van Amelsvoort (1999): Lineup construction for forensic earwitness identification: a practical approach. Arbeitsberichte *International Congress of Phonetic Sciences*, San Francisco, Band 2, 1373–1376.
- Gfroerer, Stefan (2006): Sprechererkennung und Tonträgerauswertung. In: G. Widmaier (Hrsg.) *Münchener Anwaltshandbuch Strafverteidigung*. München: Beck. S. 2505–2526.
- Hammersley, Richard und J. Don Read (1990): Das Wiedererkennen von Stimmen. In: G. Köhnken und S. L. Sporer (Hrsg.) *Identifizierung von Tatverdächtigen durch Augenzeugen*. Stuttgart: Verlag für angewandte Psychologie. S. 113–133.
- Jessen, Marianne (2006): *Einfluss von Stress auf Sprache und Stimme unter besonderer Berücksichtigung polizeidienstlicher Anforderungen*. Idstein: Schulz-Kirchner Verlag.
- Jessen, Michael (2007): Forensic reference data on articulation rate in German. *Science and Justice* 47: 50–67.
- Jessen, Michael (2008): Categorical vs. continuous variations between speakers. Jahrestagung *International Association for Forensic Phonetics and Acoustics*, Lausanne.
- Jessen, Michael, Olaf Köster und Stefan Gfroerer (2005): Influence of vocal effort on average and variability of fundamental frequency. *International Journal of Speech, Language and the Law* 12: 174–213.
- Koristka, Christian (1968): *Magnettonaufzeichnungen und kriminalistische Praxis*. Berlin: Ministerium des Inneren der DDR, Publikationsabteilung.
- Köster, Olaf und Jens-Peter Köster (2004): The auditory-perceptual evaluation of voice quality in forensic speaker recognition. *The Phonetician* 89: 9–37.
- Künzel, Hermann J. (1987): *Sprechererkennung: Grundzüge forensischer Sprachverarbeitung*. Heidelberg: Kriminalistik Verlag.
- Künzel, Hermann J., Angelika Braun und Ulrich Eysholdt (1992): *Einfluß von Alkohol auf Sprache und Stimme*. Heidelberg: Kriminalistik Verlag.
- Künzel, Hermann J. (2000): Effects of voice disguise on speaking fundamental frequency. *Forensic Linguistics* 7: 149–79.
- Moos, Anja (2008): *Forensische Sprechererkennung mit der Messmethode LTF (long-term formant distribution)*. Magisterarbeit, Universität des Saarlandes, Fach Phonetik und Phonologie.
- Nolan, Francis (1983): *The phonetic bases of speaker recognition*. Cambridge: Cambridge University Press.
- Stylianou, Yannis (2008): Voice transformation. In: J. Benesty, M.M. Sondhi und Y. Huang (Hrsg.) *Springer Handbook of Speech Processing*. Berlin: Springer-Verlag. S. 489–503.