Der gegenwärtige Stand der forensischen Sprachverarbeitung

Erschienen in: *Kriminalistik* 11/2003, S. 676 - 684

Einführung

In den letzten dreißig Jahren hat sich eine als forensische Sprachverarbeitung bezeichnete neue kriminalwissenschaftliche Disziplin etabliert. Im internationalen Bereich wird hierfür der 1997 auf der ersten ENFSI-Tagung zu diesem Thema vereinbarte Terminus *forensic speech and audio analysis* verwendet. Der vorliegende Beitrag vermittelt einen kurzen Überblick über die hauptsächlichen Fragestellungen und Arbeitsabläufe der forensischen Sprachverarbeitung, so wie sie sich im Laufe der Zeit herauskristallisiert haben und versucht, ihre Möglichkeiten und Grenzen insbesondere mit Blick auf die Bedürfnisse der Anwender bei Polizei und Justiz darzustellen. Um die wichtigsten Einzelaufgaben der forensischen Sprachverarbeitung möglichst realitätsnah darzustellen, wird im Folgenden ein Szenario angenommen, das typische Merkmale tatsächlicher Fälle enthält, einschließlich der Tatsache, daß häufig mehrere Merkmale in ein und demselben Fall vorkommen.

Beispielfall - Teil 1

Ein siebenjähriger Junge aus einer guten Kölner Wohngegend wird morgens auf dem Weg zur Schule entführt. Als seine Mutter gegen 14:00 Uhr von der Arbeit nach Hause kommt, erhält sie den Anruf eines unbekannten Sprechers (im Folg. "Anonymus"). Der Anruf besteht aus folgendem Text: "Wir haben Ihren Jungen auf dem Schulweg in unsere Obhut genommen. Es geht es ihm soweit gut. Wenn Sie schön tun, was wir verlangen, erhalten sie Ihren Sohn in ein paar Tagen wohlbehalten zurück. Besorgen Sie sich 500.000 € und warten Sie auf weitere Anweisungen. Lassen Sie die Polizei aus dem Spiel, sonst sehen Sie ihn nie wieder. Wir melden uns schon bald wieder." Die Mutter ist während des Anrufs so schockiert, daß nicht daran denkt, den Anrufbeantworter zur Aufzeichnung der unbekannten Stimme einzuschalten. Sie verständigt sofort ihren Ehemann, der Inhaber eines mittelgroßen Betriebes ist. Bereits um 18:00 Uhr geht ein diesmal vom Vater des Kindes entgegengenommener neuer Anruf ein, der über den Anrufbeantworter aufgezeichnet wird. Leider handelt es sich um ein billiges Gerät, und die Micro-Kassette ist durch jahrelangen Gebrauch abgenutzt.[1] Darüber hinaus ist das Verbindungskabel zum Telefon defekt, so daß eine Brummstörung in Form der Netzfrequenz und einige Obertöne derselben die Sprache überlagern. Der Entführer gibt Anweisungen bezüglich der Stückelung des Lösegeldes und des zum Transport zu benutzenden Behältnisses. Die Übergabe solle mittels Kfz erfolgen. Der Vater des Kindes ist nicht sicher, alle Anweisungen verstanden zu haben, da der Anonymus leise gesprochen hatte und ein Hintergrundgeräusch vorhanden war. Offenbar kam der Anruf von einer belebten Straße. Trotz der Warnung des Entführers verständigen die Eltern sofort die Polizei, die umgehend eine professionelle Telefonüberwachung vornimmt.[2]

Zu diesem Zeitpunkt ergeben sich für den Sachverständigen drei mögliche Tätigkeitsbereiche: Stimmenanalyse, phonetische Textanalyse und elektronische Sprachverbesserung.

Stimmenanalyse

Solange (nur) ein anonymer oder pseudonymer Anruf vorliegt, kann eine sog. Stimmen-analyse vorgenommen werden[3]. Man versteht hierunter die angesichts der Menge und technischen Qualität des Sprachmaterials möglichst weitgehende Beschreibung des lautsprachlichen Verhaltens eines Sprechers sowie aller weiteren Merkmale, die seiner

Identifizierung oder Lokalisierung dienlich sein können. Wir lassen zunächst die Tatsache außer Acht, daß sobald mehrere (aufgezeichnete!) Anrufe vorliegen, Stimmenvergleiche auch zwischen diesen anzustellen sind, denn für polizeiliche Maßnahmen ist die Erkenntnis wichtig, ob schon aufgrund der Sprachevidenz mit mehreren Tätern zu rechnen ist. Die bloße Tatsache, daß der Entführer den Plural "wir" benutzt, hat nach aller praktischen Erfahrung keine Bedeutung, da selbst die große Mehrzahl der Einzeltäter zwecks Verstärkung des psychologischen Drucks auf die Erpreßten im Plural sprechen. Je nach Ergiebigkeit des vorliegenden Sprachmaterial lassen sich zu diesem Zeitpunkt Aussagen zu folgenden Merkmalen treffen:

(a) Geschlecht

In Fällen ohne Stimmverstellung ist die Angabe des Geschlechts des Sprechers in der Regel unproblematisch. Dies kann sch jedoch ändern, wenn die sog. mittlere Sprechstimmlage als besonders prominenter Indikator des Geschlechts sehr hoch oder sehr tief ist. Akustisch gesehen: wenn die mittlere Stimmband-Grundfrequenz, d.h. die durchschnittliche Frequenz, mit der die im Kehlkopf gegenüberliegenden Stimmbänder pro Sekunde schwingen, bei Männern so hoch oder bei Frauen so tief ist, daß sie mit der Frequenzverteilung für das jeweils andere Geschlecht deutlich überlappt. Da die meisten forensischen Sprecher männlich sind, ergeben sich entsprechend häufiger Probleme mit hohen als mit tiefen Stimmen. In einer kleineren Anzahl von Fällen, insbesondere wenn Stimmverstellungen wie Flüstern, Preß-stimme oder Kopfstimme vorliegen, ist eine zuverlässige Angabe des Geschlechts nicht möglich.

(b) Alter

Die Angabe des Alters eines Sprecher kann vor allem dann nützlich sein, wenn es außerhalb des statistisch häufigsten Bereichs von 20 - 40 Jahren liegt. Dabei ist jedoch zu bedenken, daß eine genauere Schätzung als +/- ca. 5 Jahre selbst durch den mit forensischer Sprache vertrauten Experten nicht möglich ist, und zwar hauptsächlich aus folgendem Grunde. So wie Menschen im Verhältnis zu ihrem chronologischen Alter jung oder alt aussehen können, kann auch die Stimme aus unterschiedlichsten Gründen älter oder jünger erscheinen. Mit anderen Worten: biologisches und chronologisches Alter können voneinander abweichen. Obwohl eine beachtliche Menge von Fachliteratur mit empirischen Untersuchungen zu diesem Thema vorliegt, kann die Gesamtzahl der für die Altersschätzung relevanten Merkmale noch immer nicht benannt werden, geschweige denn die Bedeutung ihrer gegenseitigen Interaktion.(Diesen Satzteil würde ich vor „benannt werden“ stellen. Fest steht jedoch, daß weibliche Stimmen generell schwieriger einzuschätzen sind als männliche, was primär auf die generell höhere Sprechstimmlage (s.o.) und die damit verbundenen größeren Abstände zwischen ihren für den Stimmklang maßgebenden Obertönen zurückzuführen ist. Darüber hinaus konnte gezeigt werden, daß Zigarettenrauchen (1 Päckchen pro Tag oder mehr) eine Stimme um durchschnittlich 6 Jahre älter erscheinen läßt. Aus diesen Gründen wird seit jeher bei der Altersschätzung die Kautele eingefügt, daß im Falle chronischen Rauchens der betr. Sprecher bis zu 10 Jahren jünger sein könne als geschätzt.[4] Empirisch belegt ist auch, daß die Altersschätzung auch vom Gesundheitszustand des Sprechers abhängt, analog zu dem "guten" oder "schlechten" Aussehen eines Menschen. Ganz allgemein muß gesagt werden, daß beim gegenwärtigen Stand der Forschung die Altersschätzung ein subjektiver und durch akustische Untersuchungen nur wenig zu unterstützender Prozeß ist. Aufgrund neuerer Ergebnisse besteht sogar die Möglichkeit, daß (sprach-)wissenschaftliche und sogar spezielle forensische Fertigkeiten hierbei nicht maßgeblich sind, sondern allein die alltägliche Erfahrung in der sprachlichen Kommunikation[5].

(c) Regionale Herkunft

In Sprachräumen mit ausgeprägten Dialekten, im Falle des Deutschen in erster Linie Deutschland, Österreich, und die Schweiz, hat die Erfahrung gezeigt, daß die je nach Materiallage möglichst weitgehende Beschreibung der dialektalen und damit regionalen Herkunft eines Sprechers in einer sehr großen Anzahl Fällen entscheidend zur Identifizierung bzw. zum Ausschluß beigetragen hat und in der Praxis daher mit Abstand das für sich betrachtet bedeutendste Ergebnis einer forensischen Stimmenanalyse darstellt. "Regionale Herkunft" bezeichnet dabei nicht notwendigerweise den - von Seiten der Polizei i.d.R. leicht feststellbaren - Geburtsort, sondern das dialektologisch definierbare Gebiet, in dem ein Individuum seine sog. Sprachprägephase verbracht hat. Diese Zeitspanne stimmt im allgemeinen gut mit der Schulzeit überein und berücksichtigt auch die Tatsache, daß die sprachliche, und zwar auch dialektale, Flexibilität um so länger gegeben ist, je länger der Schulbesuch oder auch eine weiterführende Ausbildung, z.B. an der Universität, andauert. Bei günstiger Materiallage besteht sogar die Möglichkeit, bei ein und demselben Sprecher mehrere dialektale Merkmalskomplexe und damit (längere) Aufenthalte in verschienen regionalen Räumen nachzuweisen und somit dem Ermittler einen weiteren Ansatzpunkt zu geben.

In den meisten Fällen spricht ein forensisch relevanter Sprecher nicht Dialekt oder Mundart im engen wissenschaftlichen Sinne, sondern eine mehr oder weniger stark dialektal gefärbte Umgangssprache. Die Aufgabe des Sachverständigen besteht in der Identifizierung und Klassifizierung dieser regionalen Elemente auf allen sprachlichen Ebenen. Selbstverständlich ist die Aussprache hier der wichtigste Bereich, aber auch Grammatik, Wortwahl und Satzbau können regionale Eigenheiten enthalten. Die Technik der regionalen Beschreibung eines Sprechers ist eine klassische Anwendung der traditionellen Phonetik, die heute von modernen, rechnergestützten Werkzeugen in der Weise ergänzt werden kann, daß die subjektiven Erkenntnisse des Sachverständigen auf eine objektive Basis gestellt werden und somit auch vor Gericht höheren Beweiswert gewinnen. Ein Werkzeug besonderer Art, ein sog. Expertensystem, ist die unter Leitung des Verfassers am BKA entwickelte rechnerbasierte Datenbank regionaler Umgangssprachen des Deutschen (DRUGS), mit deren Hilfe phonetisch festgestellte lautliche Merkmale bestimmten Dialekten und damit Regionen des deutschen Sprachraums zugeordnet werden können.[6] Die Grundidee für DRUGS bestand in der Erkenntnis, daß auch der kundigste dialektologische Fachmann nicht sämtliche regionalen Varietäten des Deutschen beherrschen kann, zumindest nicht gleich gut. Angesichts des Anwenderkreises bei den Fachdezernaten von BKA und LKÄ wurde die Datenbank in einer Weise konzipiert, daß der Nutzer nicht selbst über erhebliches dialektologisches Fachwissen verfügen, sondern lediglich mit der auditiv-phonetischen Analysemethode vertraut sein muß. Er gibt das Ergebnis dieser Analyse in Form von Lautschriftsymbolen (nicht zu verwechseln mit normaler Buchstabenschrift!) in eine Suchmaske ein und erhält auf dem Bildschirm eine Karte des deutschen Sprachraums mit farblicher Kennzeichnung der Region bzw. Regionen, in denen genau die eingegebene Aussprache zu finden ist. Wiederholtes Suchen unterschiedlicher Lautkombinationen im Sinne einer Rasterung erlaubt eine schnelle Eingrenzung der in Frage stehenden Region.

(d) Fremde Muttersprache

Die Bestimmung der fremden Muttersprache eines unbekannten Sprechers kann ebenfalls von großer Bedeutung für die Identifizierung sein. Sie wird genau wie die Bestimmung der regionalen Herkunft mit Hilfe der auditiv-phonetischen Methode durchgeführt. Dabei ist jedoch eine besondere Tatsache zu beachten, nämlich daß erfahrungsgemäß in etwa der Hälfte der Fälle, in denen Sprecher mit - meist laut Angabe der Polizei, aber auch von Opfern und Zeugen - "ausländischem Akzent" aufgetreten ist, dieser durch die phonetische Analyse als Fälschung entlarvt wird: eine für die Ermittlungsarbeit wertvolle Information. So mag beispielsweise ein anonymer Erpresser vorgeben, Mitglied der Mafia zu sein, um seiner Drohung Nachdruck zu verleihen und zu diesem Zwecke einen Akzent vortäuschen, den er für italienisch hält. In den meisten Fällen kann der Experte ohne große Probleme herausfinden, ob ein ausländischer Akzent echt oder aufgesetzt ist, indem er die lautlichen und grammatischen sog. Interferenzen zwischen den involvierten Sprachen untersucht, also hier Deutsch und Italienisch. Von besonderer Bedeutung ist dabei der Einfluß des muttersprachlichen Lautsystems auf das fremdsprachliche. In diesem Sinne kann gerade auch das Fehlen von Interferenzen an Stellen, bei denen sie bei Annahme der fremden Muttersprache im Deutschen zu erwarten wären, von Bedeutung sein.[7] Voraussetzung für die Analyse ist natürlich, daß der Experte mit dem lautlichen System der betreffenden Sprachen vertraut ist. In diesem Zusammenhang hat sich eine Datenbank mit Proben aus 40 Sprachen als nützlich erwiesen. Ein Muttersprachler spricht jeweils einen Standardtext in der betreffenden Sprache und danach auf Deutsch. Der erste Teil enthält somit die Aussprache der Laute und Lautkombinationen sowie der "Sprechmelodie" der Fremdsprache, der zweite die daraus entstehenden Interferenzen im Deutschen.

(e) Weitere individualtypische Merkmale

Auf der Grundlage der drei hier dargestellten Hauptbereiche der Stimmenanalyse sind fallweise weitere Untersuchungen zur Individualisierung möglich. Es versteht sich dabei von selbst, daß ein Merkmal um so individualtypischer ist, je seltener es vorkommt. Hier sind in erster Linie pathologische Merkmale der Stimme, Sprache und Sprechweise zu nennen, von denen die unterschiedlichen Arten des Lispelns (Fehlaussprache des Lautes /s/) die häufigsten sind. Auch die verschiedenen Formen der Heiserkeit sind individualtypisch geprägt. Mit geeigneten instrumentellen Verfahren können solche Merkmale auch objektiv dargestellt werden, z.B. Lispeln mittels Spektralanalyse und Heiserkeit mittels der sog. Jitter-Shimmer-Analyse. Auch die Atmung eines Sprechers kann individuelle Züge aufweisen. Hierbei werden Frequenz, Tiefe (Dauer) und vor allem Klang (spektrale Zusammensetzung) der Atemzüge analysiert, wobei allerdings die Amplitudenauflösung einer Telefonaufzeichnung hier Grenzen setzt. Im Bereich der Sprechweise sind vor allen sog. Stereotypien von Interesse, d.h. der - vor allem der übertriebene - Gebrauch von Wörtern, Floskeln, ja ganzen Sätzen an Stellen, wo sie nicht hingehören. Diese sprachlichen Angewohnheiten ("habits") können auditiv und akustisch nachgewiesen werden, vorausgesetzt, das Sprachmaterial ist so lang, daß sie auch auftreten. Darüber hinaus gibt es auch sog. nicht-sprachliche Angewohnheiten wie z.B. Schmatzen beim Trennen der Lippen. Häufigste Ursache hierfür sind schlecht sitzende Prothesen und bestimmte Medikamente, die die Konsistenz des Speichels verändern (z.B. Beta-Blocker).[8] Der Vorteil dieses Merkmals für den Experten besteht darin, daß sich der Sprecher selbst dessen nicht bewußt ist und es somit nicht zu unterdrücken versucht. Das Merkmal fällt zwar ungeschulten Personen nicht von allein auf, kann jedoch von diesen nach Hinweis des Experten ohne weiteres wahrgenommen werden. Von Bedeutung kann auch das von einem Sprecher benutzte Vokabular sein, das z.B. auf Zugehörigkeit zu einer sozialen oder beruflichen Gruppe hinweist.[9]

In Abhängigkeit von der Komplexität der Aufgabe, insbesondere von Quantität und Qualität des sprachlichen Materials, benötigt ein erfahrener Sachverständiger in aller Regel maximal einen Tag, um ein Gutachten zur Stimmenanalyse in schriftlicher Form zu erstellen. Die Tatsache, daß die Aufzeichnung der Stimme eines Täters im frühen Stadium einer Entführung oder Erpressung häufig die einzige direkte Spur darstellt, macht ihre große kriminalistische Bedeutung offensichtlich.

Phonetische Textanalyse

In dem obigen Beispiel war sich der Vater des entführten Kindes wegen der leisen Sprechweise und des vernehmbaren Hintergrundgeräuschs nicht sicher, den Wortlaut des Anonymus vollständig verstanden zu haben. Das Brummgeräusch als Folge des defekten Verbindungskabels kommt noch hinzu. In diesem Stadium ist die Ermittlung des gesprochenen Wortlautes gefordert. Hierzu ist eine geschulte, d.h. mit den besonderen Gegebenheiten forensischer Sprachaufzeichnungen vertraute Person erforderlich, am besten also ein forensicher Sprachsachverständiger. Für diese im Prinzip als Sprach-Erkennung anzusehende Aufgabe wurde im Deutschen der Begriff *phonetische Textanalyse* eingeführt[10]. Im englischen Rechtssystem mit seinem Prinzip der beiden gegnerischen Parteien hat der Begriff *disputed utterances* ("strittige Äußerungen") Verbreitung gefunden[11]. Der deutsche Terminus wie im übrigen auch der in Amerika gebräuchliche Terminus *speech decoding* sind eher "generisch" bestimmt und tragen der Tatsache Rechnung, daß vielfach weite Passagen einer fraglichen Sprachaufzeichnung völlig unkontrovers und leicht zu transkribieren sind. Aus offensichtlichen Gründen ist die phonetische Textanalyse häufig mit elektronischer Sprachverbesserung verbunden (s.u.).

Durch ein erstes Anhören der gesamten Sprachaufzeichnung verschafft sich der Sachverständige zunächst einen globalen Eindruck ihrer akustischen und sprachlichen Charakteristik. Danach wird das Sprachmaterial mittels digitaler akustischer Editiertechnik bearbeitet. Insbesondere können beliebig lange Endlosschleifen hergestellt werden, die eine Konzentration auf nicht eindeutige und daher möglicherweise kontroverse Stellen gestatten. Ist auch dann noch keine Eindeutigkeit zu erreichen, listet der Experte sämtliche aus phonetisch-phonologischen Gründen möglichen Alternativen auf.[12] Diese Tätigkeit verlangt nicht nur ein geschultes Ohr, sondern gründliches sprachliches Wissen und vor allem Wissen über Sprache. In der Regel wird eine wörtliche Transkription des fraglichen Texts in normaler Orthographie erstellt, die ohne weiteres von Polizei oder Gericht verwendet werden kann. Für seine eigenen Aufzeichnungen fertigt der Sachverständige meistens zusätzlich zumindest partiell eine sog. phonetische Transkription in Lautschrift an, in welcher lautliche Feinheiten dokumentiert sind. Dadurch besteht die Möglichkeit, erforderlichenfalls die Grundlagen der eigenen Hörentscheidungen offenzulegen, z.B. in der Auseinandersetzung mit einem "gegnerischen" Sachverständigen, wie es im angelsächsischen Rechtssystem der Regelfall ist.

Ein besonderes Problem bei der Verschriftung sind Orts- Personen- und andere Eigennamen, also Elemente der Sprache, die unter schlechten akustischen Bedingungen nicht einmal zum Teil durch den Kontext erschlossen werden können. Hier hat sich in der Praxis die Hinzuziehung eines mit dem Fall und insbesondere den Örtlichkeiten vertrauten Ermittlers außerordentlich bewährt.[13] Gleichzeitiges Sprechen mehrerer Personen ist ein weiteres Problem bei forensischen Sprachaufzeichnungen. Es tritt häufig auf, wenn die akustischen Begleitumstände ohnehin schon ungünstig sind, wie z.B. bei einer verdeckten Aufzeichnung in einem schallharten Raum mit Nachhall (z.B. Gefängniszelle) und sich verändernden Entfernungen der Sprecher zu dem Mikrofon. Hier muß sich der Sachverständige auf jeweils eine Stimme konzentrieren und alle anderen bewußt unterdrücken, den von der Stimme gesprochenen Wortlaut notieren und diese Prozedur mit jeder anderen Sprecherstimme wiederholen. Die beste Art der Dokumentation gleichzeitigen Sprechens mehrerer Sprecher ist die Notation senkrecht untereinander wie bei einer Notenpartitur. Dadurch erhält der Leser nicht nur einen kontinuierlichen Eindruck vom Verlauf des Gesprächs (horizontal), sondern kann auch die zeitlichen Verhältnisse der Redeteile der einzelnen Sprecher (Vertikal) zueinander nachvollziehen.

Elektronische Sprachverbesserung

Seit Beginn der 1980er Jahre sind immer feinere Techniken der zunächst analogen, später dann digitalen Behandlung gestörter Sprachsignale entwickelt worden, insbesondere mit Blick auf die forensischen Verhältnisse. Viele Verfahren, vor allem solche, die für die Musikindustrie von Interesse sind, sind heute als relativ preiswerte Software-Pakete frei erhältlich. Einige der potentesten Verfahren sind jedoch - soweit überhaupt für den kommerziellen Bereich freigegeben - noch immer sehr teuer und daher den staatlichen Labors vorbehalten. Obwohl die meisten Algorithmen rein technisch betrachtet leicht zu handhaben sind, sind einige Vorsichtsmaßregeln angebracht. Generell gilt: Der ideale Algorithmus zur Sprachverbesserung existiert genauso wenig wie das ideale Medikament in der Medizin: keine Hauptwirkung ohne ungewollte Nebenwirkungen! Es ist nämlich nicht möglich, alle und nur die Störungen eines Signals zu beseitigen, ohne dabei die nicht gestörten Passagen völlig intakt zu lassen. In der Praxis ist daher immer ein Kompromiß zwischen beiden Extremen zu suchen, wobei der Zweck der Maßnahme entscheidend ist: Für ungeschulte Hörer wie Polizisten, Zeugen oder Mitglieder eines Gerichts bedeutet das Anhören einer längeren durch laute Brummtöne verzerrten Sprachaufzeichnung eine große Belastung, die alsbald Ermüdung bewirkt. Hier wird die robuste Beseitigung der Störungen die Anhöraufgabe erleichtern, allerdings sollten die Hörer darauf aufmerksam gemacht werden, daß der Klang der involvierten Stimmen verändert worden ist. Solche Warnungen werden besonders dann wichtig, wenn stark invasive Algorithmen wie adaptive Filterung oder Spektralsubtraktion angewandt werden. Falls nicht phonetische Textanalyse, also Sprach-Erkennung, sondern Sprecher-Erkennung das oberste Ziel ist, wird der geübte Sachverständige zur Wahrung des unverfälschten Klangs der zu analysierenden Stimme immer mittels Kopfhörer das originale, gestörte Sprachsignal auf dem einen Ohr und das gefilterte Signal auf dem anderen Ohr abspielen und dabei die Störung auditiv zu unterdrücken versuchen.

Hier wird deutlich, daß auch die so technisch und objektiv anmutende elektronische Sprachverbesserung eine erhebliche subjektive Komponente enthält, angefangen mit der Entscheidung, ob in einem gegebenen Fall die Anwendung von Sprachverbesserung überhaupt angezeigt ist. Zwar arbeitet ein einmal ausgewählter Filteralgorithmus immer exakt und in identischer Weise, aber sämtliche Schwellenwerte, Filtereigenschaften, Zeitkonstanten etc. sind vom Sachverständigen auf Grund seiner Sachkunde und Erfahrung einzustellen. Insgesamt lautet das Fazit, daß bei Störungen wie Verkehrslärm, Netzbrummen, Musik, GSM-Impulsen sowie - und das wäre der schlimmste Fall - weitere interferierende Stimmen aufweisenden forensischen (Mono-)Aufzeichnungen von Mikrofon- oder Telefonsprache auch von modernsten elektronischen Sprachverbesserungsverfahren keine Wunder im Sinne eines Alles-oder-Nichts-Effekt zu erwarten sind: Der kategoriale Sprung von der völligen Unverständlichkeit eines Originals zur störungsfreien Verständlichkeit der elektronisch verbesserten Kopie bleibt fast immer Illusion.[14]

Fallbeispiel - Teil 2

Einen Tag nach Erhalt des Anrufbeantworterbandes mit der Botschaft des Anonymus übermittelt der Sachverständige für Sprachverarbeitung das Ergebnis seiner Stimmenanalyse der Polizei. Nach Anwendung elektronischer Sprachverbesserung konnte der Wortlaut der Nachricht zweifelsfrei ermittelt werden. Das stimmlich-sprachliche Profil des Anonymus lautet: männlich, Muttersprache Deutsch mit mäßig starken Merkmalen einer ripuarischen (rheinischen) Dialektfärbung, Alter ca. 20 - 40 Jahre, für männliche Erwachsene durchschnittlich hohe Sprechstimmlage. Auf Grund der Melodieführung und des Rhythmus der Sprache hat Anonymus den Text wahrscheinlich nicht frei gesprochen, sondern abgelesen.

Es liegen aber keine Hinweise auf das Abspielen einer sog. Konserve von einem Taschendiktiergerät oder Walkman vor (keine Ein- und Ausschaltimpulse, reduzierter Frequenzgang oder Laufgeräusche). Da die Nachricht nicht frei gesprochen wurde, ist die Beurteilung der verbalen Ausdrucksfähigkeit (Eloquenz) des Sprechers schwer zu beurteilen, allerdings sind Satzbau und Grammatik perfekt. Das einzige auffällige Merkmal ist eine Besonderheit bei der Aussprache des Lautes /sch/, der dreimal vorkommt und dabei zweimal einen auch im Spektrogramm als Resonanz bei 2,8 kHz nachweisbaren leichten Pfeifton enthält Die häufigste physiologische Ursache für dieses Merkmal ist eine Lücke zwischen den ersten oberen Schneidezähnen. Im vorliegenden Fall ist irrelevant, ob das Merkmal eine Normvariante darstellt oder bereits als Sprechfehler anzusehen ist, es ist in jedem Fall relativ selten und damit individualtypisch.

Die Synopse der Stimmenanalyse mit Ergebnissen aus anderen kriminalistischen Bereichen ergibt als Sprecher einen Mann aus der Kölner Gegend, südlich bis Bonn, aber nördlich nicht mehr Düsseldorf, mit detaillierten Kenntnissen über das Opfer und die Familie. Daher wird der Kindesvater von der Polizei um eine Liste von Geschäftspartnern und Kunden gebeten, mit denen es möglicherweise einmal Schwierigkeiten gegeben hatte. Vier Personen werden benannt und unter einem Vorwand angerufen, um auf diese Weise eine Sprachprobe zu gewinnen. Auf dieser Basis sind nun Stimmenvergleiche mit dem anonymen Anrufer möglich.

Stimmenvergleich (Sprecher-Erkennung)

In Hinsicht auf die wichtigste und wissenschaftlich anspruchsvollste Aufgabe, die Identifizierung einer Person an Hand ihres lautsprachlichen Verhaltens (Sprecher-Erkennung, Sprecher-Identifizierung), ist eine Vorbemerkung angebracht. Aufgrund des *nemo tenetur* -Grundsatzes kann ein Beschuldigter die eine aktive Mitwirkung erfordernde Abgabe einer Stimmprobe verweigern und damit einen Stimmenvergleich von vornherein verhindern. Dies gilt auch für die unten beschriebene Identifizierung durch Laien, die sog. akustische Gegenüberstellung[15]. In manchen Fällen besteht in der Praxis immerhin die Möglichkeit, mehr oder weniger gut geeignetes Sprachmaterial des Sprechers aus anderer Quelle zu benutzen, z.B. aus Telefonüberwachungsmaßnahmen. Praktische Schwierigkeiten entstehen auch, wenn ein Sprecher lediglich einen Teil der abverlangten Sprachprobe produziert, also beispielsweise bereitwillig über sein Hobby erzählt oder einen Text aus einer Zeitschrift vorliest, das Vorlesen oder Nachsprechen des betreffenden Tattexts ablehnt. In solchen Fällen ist ein Wort-für-Wort-Vergleich nicht möglich, der wissenschaftlich gesehen am ergiebigsten ist und dem Gutachter unabhängig von der Art der verwendeten Wahrscheinlichkeitsskala die höchstmögliche Absicherung seines Endurteils gestattet. In geringerem Maße (d.h. beim Vorliegen von sehr wenig Sprachmaterial) gilt diese Erschwernis auch für die unabhängig vom Wortlaut der Sprachprobe, rein akustisch arbeitenden Verfahren der automatischen rechnergestützten Sprecher-Identifizierung.

Aus der Sicht der Strafverfolgung hat die menschliche Stimme unter bestimmten Umständen einen einzigartigen Vorteil gegenüber allen anderen Beweismitteln, einschließlich Fingerabdrücken oder DNA-Spuren, indem sie den Sprecher unzweifelhaft mit der Tat in Verbindung bringt. Anders als z.B. ein theoretisch unabhängig von einer Tat am Tatort hinterlassener Fingerabdruck setzt ein inhaltlich auf die Tat bezogener Anruf wie der typische Bekenneranruf oder in unserem Fall der Übermittler der Lösegeldforderung zumindest Mitwisserschaft voraus[16]. Die entscheidende Frage - und gleichzeitig das größte Handikap gegenüber Fingerabdruck- oder DNA-Spuren - ist nur, mit welcher Sicherheit die stimmlich -

sprachliche Evidenz einem Individuum zugeordnet werden kann. Dieses Problem der Sprecher-Erkennung hat weltweit Sprachwissenschaftler und Ingenieure immer wieder dazu angespornt, existierende Verfahren zu verbessern und neue zu entwickeln. Seit den 1970er Jahren gab es in diesem Zusammenhang zahlreiche Versuche einer rechnergestützten, weitgehend vollautomatischen Methode, die aus dem fraglichen und dem Vergleichsmaterial möglichst stark sprecherspezifische akustische Merkmale extrahiert und ihre Ähnlichkeit statistisch bewertet. Obwohl diese Ansätze in "kommerziellen" Anwendungen, hauptsächlich biometrischen Zugangskontrollsystemen (Zugang zu sicherheitsempfindlichen Anlagen wie Rechenzentren, Waffenlagern, neuerdings auch beim Zugriff auf das Bankkonto) durchaus erfolgreich waren und sind, scheiterten sie bis vor kurzem im forensischen Bereich. Hauptgrund dafür ist anerkanntermaßen nicht etwa mangelnde Individualisierungspotenz der akustischen Merkmale oder Unzulänglichkeiten der zu ihrer Auswertung benutzten Algorithmen, sondern die Unmöglichkeit, die typischen forensischen Umweltbedingungen wie Telefonübertragung, Stimmverstellung, sehr kurzes bzw. ungeeignetes Sprachmaterial und andere Merkmale eines sog. nicht-kooperativen Sprechers in ihrer störenden Wirkung zu kontrollieren. Nachdem in den letzten Jahren die mit der Telefonübertragung als (in den meisten Fällen) hauptsächlichem Störfaktor zusammenhängenden Probleme im Prinzip gelöst werden konnten (sog. *channel equalising*, Kanalnormalisierung) ist endlich ein Durchbruch der automatischen Sprecher-Identifizierungssysteme auch im forensischen Bereich möglich. Zur Zeit werden weltweit mindestens ein halbes Dutzend Verfahren getestet bzw. schon in der Praxis angewendet, darunter das weiter unten vorgestellte. Zuvor ist jedoch kurz auf die zur Zeit benutzten konventionellen Verfahren einzugehen.[17]

Das früher in vielen Ländern gebräuchliche sog. Stimmabdruckverfahren (Voiceprint-Verfahren) ist glücklicherweise heute stark in den Hintergrund getreten. Es besteht aus dem visuellen Vergleich einer bestimmten Darstellungsweise dreidimensionaler Spektrogramme (Kurzzeit-Breitband-Spektrogramme), die von jeweils ca. 2,5 Sekunden langen Abschnitten der zu vergleichenden Sprachproben hergestellt werden. Dabei werden vom mehr oder weniger geschulten Beobachter, z.B. Polizisten nach einem zweiwöchigen Trainingskurs (!), Ähnlichkeiten und Unterschiede der als Graustufungen auf Spezialpapier dargestellten Amplituden-Frequenzmuster von Lauten und Lautkombinationen der fraglichen und der Vergleichssprachprobe betrachtet und auf einer subjektiven Skala bewertet. Die Tatsache, daß solche Spektren heutzutage in hoher technischer Qualität und mit Hilfe digitaler Signalverarbeitung herstellbar sind, ändert nichts an der Tatsache, daß das Verfahren vom Ansatz her so erhebliche Schwächen enthält, daß es nach praktisch einhelliger Meinung der Wissenschaft zur Sprecher-Identifizierung völlig untauglich ist. Dies wurde in den USA bereits 1979 in einer wissenschaftlichen Studie im Auftrag des Justizministeriums und zahlreichen weiteren Studien bewiesen[18]. Insbesondere wurde auf die mangelnde sprecherspezifische Potenz der Merkmale verwiesen: Die angeblich individuellen Merkmale unterscheiden sich z.T. innerhalb ein und desselben Individuums stärker als zwischen verschiedenen Individuen. Um Mißverständnissen vorzubeugen: Das Breitbandspektrum an sich ist seit seiner Erfindung vor über 50 Jahren ein unverzichtbares Verfahren in der Analyse sprachlicher und anderer akustischer Signale (z. B. Herztöne, Fließgeräusche in Blutgefäßen, Tierstimmen), nur ist seine Verwendung als Kernstück der Sprecher-Erkennung abzulehnen.

In Deutschland wurden aus der erwähnten amerikanischen Studie frühzeitig Konsequenzen gezogen und neue methodische Wege beschritten. Als Ergebnis entstand Mitte der 1980er Jahre das heute auch international am häufigsten benutzte phonetisch-instrumentelle Kombinationsverfahren.[19] Es basiert auf der Extraktion sprecherspezifischer Merkmale aus den drei Bereichen des lautsprachlichen Verhaltens: Stimme (im engeren Sinne), Sprache und Sprechweise. Die Summe der Übereinstimmungen und Unterschiede wird auf einer Rangskala bewertet. Zu einigen Merkmalen stehen statistische Verteilungen in der Bevölkerung zur Verfügung, die eine teilweise Quantifizierung der Verhältnisse im Einzelfall zulassen. Ein wichtiges Charakteristikum des Verfahrens im Hinblick auf die Benutzung als Beweismittel ist die doppelte Analyse jedes Merkmals, einerseits mit Hilfe klassischer phonetischer bzw. medizinischer Diagnoseverfahren und andererseits als akustische Darstellung mittels moderner digitaler Signalanalyseverfahren. Damit ist eine Kontrolle sowohl durch den mit der Analyse beschäftigten Experten als auch durch andere Experten (Zweit- oder Gegengutachter) sowie bis zu einem gewissen Grade auch durch Laien (Gericht) möglich. Tabelle 1 enthält eine grobe Übersicht der häufigsten sprecherspezifischen Merkmale.[20] Die Schwäche dieses Verfahrens liegt zum einen in der unvermeidbaren und erheblichen subjektiven Komponente, die vor allem bei der Bewertung der Übereinstimmungen bzw. Unterschiede der Merkmale involviert ist. Zum anderen ist festzustellen, daß die extrahierten Merkmale in ihrer individualtypischen Potenz schwanken und im konkreten Falles auch nicht immer auswertbar sind, wie beispielsweise das Merkmal der sog. mittleren Stimmfrequenz, wenn ein Sprecher flüstert. Insgesamt ist festzuhalten, daß das durch den kompetenten Experten angewandte phonetische Kombinationsverfahren eine wertvolles Mittel zur Sprecheridentifizierung unter den schwierigen forensischen Bedingungen darstellt. Es erlaubt jedoch, von seltenen Fällen abgesehen, keine Identifizierung mit an Sicherheit grenzender Wahrscheinlichkeit. Ein sicherer Ausschluß ist dagegen häufiger möglich. Die mittlerweile von fast allen in Deutschland mit einschlägigen Gutachten befaßten Sachverständigen traditionell benutzte Rangskala für Wahrscheinlichkeit wurde aus dem Bereich der forensischen Handschriftenidentifizierung entliehen, die im Wesentlichen von denselben methodischen und sachlichen Problemen und Imponderabilien betroffen ist.[21]

Wie bereits erwähnt, werden seit neuestem auch automatische Verfahren zur forensischen Sprecher-Erkennung eingesetzt. Das amerikanischen Normungsinstitut NIST veranstaltet regelmäßig Qualitätstests, die eine Evaluierung der Systeme an genormtem Sprachmaterial (Mikrofon- und Telefonsprache) erlaubt. In Spanien, Frankreich und der Schweiz haben diese Verfahren bereits Eingang in den Gerichtsalltag gefunden, in den USA wird ein System getestet. Allerdings ist angesichts der weiterhin bestehenden grundsätzlichen und methodischen Unzulänglichkeiten und ungeklärten Fragen davon auszugehen, daß sie auf absehbare Zeit nicht an Stelle von, sondern als Ergänzung zu konventionellen Verfahren eingesetzt werden. In der Zukunft werden daher Gutachten zur Sprecher-Erkennung aller Voraussicht nach aus zwei Komponenten bestehen: einem "konventionellen" und einem automatischen Stimmenvergleich.

Die automatischen Verfahren unterscheiden sich durchaus in Hinsicht auf die akustischen Parameter, Hintergrund-Datenbasis und statistische Algorithmen. Allen gemeinsam sind jedoch die weitgehende Reduzierung der durch den Systembenutzer eingeführten subjektiven Komponente, die automatische Extraktion akustischer Merkmale aus den Reflexionseigenschaften der zur Spracherzeugung benutzten Hohlräume des Hals-Nasen-Rachenraums, die Bildung statistischer Modelle der involvierten Sprecher und einer großen Menge anderer Sprecher, der sog. Welt-Population, und schließlich die Berechnung der Wahrscheinlichkeit, daß der zwischen dem Modell des fraglichen Sprechers und des Vergleichssprechers berechnete Abstand (Ähnlichkeit) in der Hintergrundpopulation vorkommt. Je geringer diese Wahrscheinlichkeit, desto größer die Wahrscheinlichkeit, daß Identität vorliegt.

Aufgrund der Tatsache, daß die von automatischen Verfahren benutzten Merkmale nicht wie bei dem phonetisch-instrumentellen Verfahren auf linguistische Einheiten wie Wörter, Silben und Laute sondern, wie erwähnt, auf die globalen Resonanzeigenschaften der für das Sprechen relevanten Strukturen in Mund, Nase und Rachen ausgerichtet sind, werden sie auch als Total-Voice-Systeme bezeichnet. Die betreffende Sprache ist somit irrelevant, und der das System bedienende Sachverständige muß diese nicht einmal beherrschen. Aus dieser Tatsache ergibt sich ein für die Arbeit der Polizei aber auch der Gerichte enormer Vorteil der automatischen Sprecher-Erkennungssysteme. Man stelle sich dazu folgenden Fall vor: In einem Verfahren liegt eine große Menge, möglicherweise Hunderte, abgehörter Telefongespräche in einer "exotischen" Sprache wie Albanisch, Kurmandschi, Ibo, Chinesisch etc. vor, für die nicht immer leicht ein Dolmetscher zu finden ist. Es sind ein Dutzend Personen bekannt, die als Sprecher in Frage kommen und von denen Vergleichsaufzeichnungen vorliegen, z.B. TÜ-Gespräche, bei denen ihre Identität nicht bestritten wird. Das automatische System kann nun innerhalb von Tagen herausfinden, an welchen Gesprächen welche dieser Personen mit mehr oder weniger großer Wahrscheinlichkeit beteiligt sind. Man wird dann diese - und vermutlich schon aus Kosten- und Zeitgründen nur diese - Gespräche übersetzen lassen: eine außerordentliche Ersparnis von Arbeit, Zeit und Kosten! Diese Funktion des automatischen Systems wird auch interne Klassifikation genannt.

Das folgende Beispiel illustriert einen Stimmenvergleich mit Hilfe eines automatischen Sprecher-Erkennungssystems. Es handelt sich um das von der Polytechnischen Universität Madrid im Auftrag der spanischen Polizei entwickelte IdentiVox[22], das dem Autor im Rahmen eines gemeinsamen Forschungsprojekts zur Verfügung steht und seit einem Jahr "außer Konkurrenz" parallel zum phonetischen Kombinationsverfahren in der Gutachtenerstellung benutzt wird. In den bislang 12 bearbeiteten Fällen wurden ausnahmslos kompatible Ergebnisse erzielt. Wegen der unterschiedlichen Wahrscheinlichkeitsskalen ist eine quantitative Vergleichbarkeit beider Methoden nicht möglich, doch läßt sich feststellen, daß in keinem Falle eine entgegengesetzte Tendenz resultierte.[23]

Bei zwei Drohanrufen von zusammen nur 18 s Netto-Dauer z. N. einer Firma wurde der Mitarbeiter A als Verdächtiger ermittelt. Das phonetische Stimmenvergleichsgutachten ergab eine "hohe Wahrscheinlichkeit" für die Identität von A und Drohanrufer, unter anderem wegen einer charakteristischen, im Gerichtssaal auch vom Laien nach entsprechendem Hinweis des Sachverständigen deutlich wahrnehmbaren Mißbildung des Sprachlautes /f / bei beiden Sprechern, die sich in der Spektralanalyse als Verlust sämtlicher Frequenzen oberhalb von 2,5 kHz belegen ließ. Als kuriose Ursache ergab sich bei dem Angeklagten A, daß sämtliche vier oberen Schneidezähne fehlten, ein extrem seltenes und daher stark individualtypisch zu wertendes Merkmal. Die Analyse mittels IdentiVox ergab in diesem Falle eine Wahrscheinlichkeit von 381 zu 1 für eine Identität. Dieses Ergebnis ist statistisch gesehen völlig unabhängig von demjenigen des phonetischen Gutachtens, weil es, wie erwähnt, ganz andere Merkmale benutzt und darüber hinaus noch in einem Frequenzbereich, der unter dem liegt, in welchem sich das erwähnte seltene Merkmal manifestiert. Insofern wird ein echter Zugewinn an Sicherheit für das Identitätsurteil erzeugt und es wird deutlich, daß die bloße Ersetzung des "subjektiven" phonetischen Verfahrens durch ein "objektives" automatisches nicht sinnvoll wäre.

Stark vereinfacht ist IdentiVox in diesem Fall wie folgt eingesetzt worden: Zunächst wurden repräsentative Sprachproben von 100 männlichen deutschen Sprechern als "Weltpopulation" zusammengestellt. Im Anschluß an eine Normalisierung der jeweils benutzten Telefonkanäle wurden für jeden Sprecher die akustischen Merkmale (38 sog. Mel-Cepstrum-Koeffizienten) extrahiert. Auf dieser Basis wurden schließlich 100 Sprecher-Modelle berechnet. Sodann wurde für den Verdächtigen A aus einem Teil seiner Sprachprobe ebenfalls ein Sprecher-Modell berechnet. Die Variabilität, also die persönliche Schwankungsbreite seiner akustischen Merkmale, wurde aus einem abgetrennten Teil seiner Sprachprobe berechnet (sog. Kontrollprobe). Schließlich wurde der anonyme Drohanruf akustisch analysiert.

Abbildung 1 zeigt das Ergebnis der Analyse. Die beiden Verteilungen links und rechts geben die Variabilität der akustischen Merkmale zwischen unterschiedlichen Sprechern bzw. innerhalb ein und desselben Sprechers wieder (sog. Inter- bzw. Intra-Sprecher-Verteilung). Die senkrechten Balken auf der linken Seite repräsentieren die Ergebnisse für den Drohanruf des Anonymus. Die drei Balken auf der rechten Seite repräsentieren drei Teilstücke der Sprachprobe des Verdächtigen Das Ergebnis für den Drohanruf entspricht der kleinen senkrechten Linie auf der X-Achse, die die Intra-Verteilung bei x = 0,72 schneidet. Aus dem Quotienten der beiden Verteilungen an dieser Stelle ergibt sich ein Likelihood Ratio. Es bedeutet, daß auf der Grundlage der hier benutzten Hintergrundpopulation eine Identität der Stimmen des Drohanrufers und des Verdächtigen 381,7 mal wahrscheinlicher ist als die Nicht-Identität. Die reine Rechenzeit für die Herstellung der Hintergrundpopulation und den eigentlichen Stimmenvergleich betrug mit einem modernen Rechner (Pentium IV Prozessor) mehrere Tage.

Fallbeispiel - Teil 3

Typischerweise würde der Beispielfall folgendermaßen verlaufen: Die professionell aufgezeichneten Sprachproben der vier Vergleichssprecher weisen eine gute technisch-akustische Qualität auf. Die Menge der sprachlichen Äußerungen und damit die Ergiebigkeit des Materials ist aber jeweils unterschiedlich. Es kommen nicht viele Wörter vor, die auch Anonymus benutzt hatte, so daß entsprechend wenige Wort-für-Wort-Vergleiche möglich sind. Bei zwei Vergleichssprechern ergeben sich starke Unterschiede und keine signifikanten Übereinstimmungen. Besonders die Verschiedenheit der dialektalen Färbung ist ein sog. Killer-Kriterium, das eine Identität ausschließt. Ein weiterer Sprecher zeigt bei einigen Merkmalen Übereinstimmungen, bei anderen deutliche Unterschiede. Der vierte Sprecher weist nicht nur signifikante Übereinstimmungen bei den angesichts der Kürze des Materials begrenzt auswertbaren Merkmalen auf, sondern auch die als selten eingestufte Auffälligkeit bei dem Sprachlaut /sch/. Der spektralanalytische Vergleich ergibt sogar eine quantitative (frequenzmäßige) Übereinstimmung. Der Sachverständige bewertet alle vier Ergebnisse auf einer Wahrscheinlichkeitsskala. Wenn ein automatisches Verfahren benutzt wird, würden an dieser Stelle entsprechende Wahrscheinlichkeitskoeffizienten (Likelihood Ratios) angegeben. Die Ergebnisse des Gutachtens veranlassen die Polizei zum Ausschluß der ersten beiden Sprecher. Angesichts der angespannten Personalsituation wird auch die Spur von Sprecher 3 einstweilen nicht weiterverfolgt. Die Ermittlungen werden vielmehr auf Sprecher 4 konzentriert. Die Überwachung seines Festnetz- und Mobiltelefonverkehrs ergibt dann z.B. neues Sprachmaterial für einen weitergehenden Stimmenvergleich, im günstigen Fall liefert sie sogar einen direkten Beweis für die Täterschaft (neuer Täteranruf).

Akustische Gegenüberstellung (Sprecher-Identifizierung durch Laien)

Abschließend ist kurz auf eine in dem Beispielfall nicht vorkommende Variante der Sprecher-Erkennung einzugehen, die im forensischen Alltag in erster Linie bei Bankraub, Vergewaltigung oder anderen Straftaten angezeigt ist, bei denen der Täter maskiert ist oder aus anderen Gründen (Dunkelheit, Angriff von hinten) keine visuelle Identifizierung möglich ist. Da keine tatrelevante Sprachaufzeichnung zur Verfügung steht, ist eine Sprecher-Erkennung durch den Experten daher nicht möglich. Hier bietet sich das formalisierte Verfahren der sog. akustischen Gegenüberstellung, wissenschaftlich ausgedrückt: Sprecher-Erkennung durch Laien (Opfer, Zeugen) an. Die Analogien zur wohlbekannten visuellen Gegenüberstellung liegen dabei auf der Hand, doch besteht bei der Identifizierung über die Stimme ein gravierender Nachteil, nämlich der transitorische Charakter des Sprachsignals, das im Gegensatz zu einem Lichtbild oder einer Person nicht zeitlich unbegrenzt betrachtet werden kann. Die Forschung hat denn auch gezeigt, daß die Identifizierung über den visuellen Bereich besser funktioniert als über den auditiven - zumindest im Laborexperiment. Zwar wurde in den letzten Jahren viel auf dem Gebiet der akustischen Gegenüberstellung geforscht und auch Richtlinien mit Richtlinien zur Durchführung von Tests entwickelt[24], doch bleiben viele wissenschaftliche Fragen ungeklärt, z.B. nach Art und Dauer des Erstkontakts mit der fraglichen Stimme (Familiarisierung), Zeitdauer zwischen dem fraglichen Ereignis und dem Wiedererkennungstest, Rolle von Angst und Stress, Darbietungsform der zu beurteilenden Stimmen (einfach oder auch mehrfach). Daher ist sich die Fachwelt einig, daß akustische Gegenüberstellung als alleiniges Identifizierungsmittel nicht in Frage kommen sollte.[25] Sie ist aber in jedem Falle der im kriminalistischen Alltag noch immer anzutreffenden Identifizierung einer einzelnen fraglichen Stimme überlegen ("Ist das die Stimme des Mannes, der Sie überfallen hat?"), u.a. weil sie den Vorteil bietet, (die meisten) Falsch- Identifizierungen aufzudecken und somit die Einschätzung der generellen Sprecheridentifizierungsfähigkeit der Beurteilerperson erlaubt. Bei der Organisation einer akustischen Gegenüberstellung sind, ähnlich wie bei der visuellen, zahlreiche Bedingungen einzuhalten. Wegen der Komplexität der Konzeption, der Interpretation der Erkennungsurteile und insbesondere der sprachlich-stimmlichen Struktur der auszuwählenden Referenzstimmen sollte der Test zwecks Vermeidung im Nachhinein nicht mehr zu heilender Fehler dem Sachverständigen vorbehalten bleiben.

Abbildung 1
Ergebnis eines Stimmenvergleichs mit Hilfe des automatischen Sprecher-Erkennungssystems IdentiVox. Aus drucktechnischen Gründen ist die Darstellung in Schwarz-Weiß gehalten.

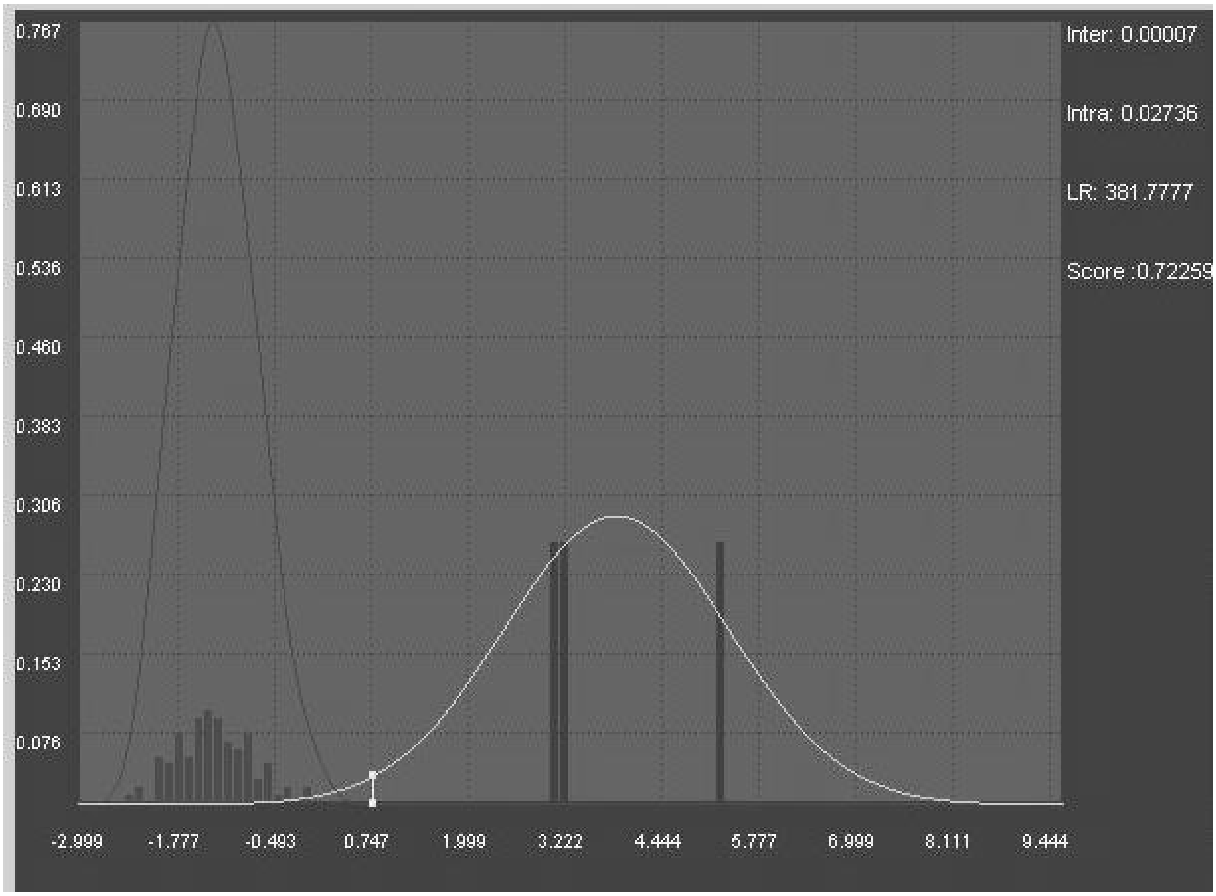

Tabelle 1

Sprecherspezifische Merkmale im Stimmenvergleichsgutachten

Termini links vom Schrägstrich bezeichnen die phonetischen, linguistischen oder medizinischen Merkmale; rechts davon sind jeweils die akustischen, physikalischen, oder statistischen Korrelate aufgeführt.

Stimme:

- mittlere Sprechstimmlage / Grundfrequenz (Fo);
- Intonationsverlauf (Melodik) / mathematisches Streuungsmaß der Grundfrequenz
- Stimmqualität / Parameter der Schwingungsform der Stimmbänder; z.B. vocal jitter (Heiserkeit), shimmer (Lautstärke-schwankungen von einer Schwingung zur nächsten) etc.

Sprache:

- Dialekt: Art und Grad der Färbung / Summe der Abweichungen von der standardsprachlichen (Aussprache-)Norm
- fremdsprachiger Akzent / Summe der Abweichungen von der Aussprachenorm einer für den Sprecher fremden Sprache
- Idiolekt (Summe der individualtypischen Merkmale) / Summe der Abweichungen vom Regelapparat der Standardsprache
- Soziolekt (Summe der schichtenspezifischen Merkmale; berufstypische Sprachen, sog. Jargons, etc.) / Abweichungen auf der lexikalischen, u.U. auch auf weiteren grammatischen Ebenen der betr. Sprache

Sprechweise:

- Sprechrhythmus (Aufeinanderfolge betonter und unbetonter Silben) / Verteilung von Lautdauer, Intensität ("Lautstärke")
- Intonation / Grundfrequenzmuster
- Artikulationsgeschwindigkeit (Sprechtempo) / Silbenrate (Anzahl der Silben pro Sek.); Artikulationsrate (Anzahl der Silben pro Sek. nach Abzug aller Sprechpausen)
- Atemverhalten / Frequenz, Dauer, Spektrum der Atemzüge

Bei allen Merkmalen außer dem Dialekt können pathologische Formen und Normvarianten auftreten, die eine Steigerung des sprecherspezifischen Wertes eines Merkmals bewirken (Sprech-, Stimm- und Sprachstörungen).

[1] Heutzutage werden digitale Geräte immer beliebter, weil sie preisgünstig sind und keine magnetischen Tonträger mehr benötigen. Dabei entstehen folgende Probleme: Um die Herstellungskosten niedrig zu halten, wird die Abtastrate und damit der benötigte Speicherplatz extrem klein gehalten. Häufig ist dann der Frequenzgang sogar deutlich schlechter als der bereits auf 300 bis 3400 Hz reduzierte Bandpaß einer (Festnetz-) Telefonverbindung. Ebenfalls aus Kostengründen werden keine Dauerspeicher benutzt, so daß beim Unterbrechen der Netzspannung die Aufzeichnung verloren geht. Solche Geräte dürfen also nie vom Netz getrennt werden, z.B. um sie einem Sachverständigen zur Auswertung zu schicken! Die Polizei umgeht dieses Problem normalerweise durch Anfertigen von Kopien vor Ort. Da aber die Geräte fast nie Ausgangsbuchsen enthalten, wird meist über den eingebauten schlechten Lautsprecher per Luftschall und Mikrofon auf einem anderen Gerät aufgezeichnet - meist einen einfachen Kassettenrekorder. Man kann sich leicht vorstellen, welche Menge additiver Störungen das Sprachsignal beeinträchtigen, bevor es dem Sachverständigen vorliegt.

[2] Die moderne Form der Aufzeichnung geschieht digital, meist auf einem magneto-optischen Laufwerk (MOD). Dabei sollte aber keine Datenkompression zwecks Ersparnis von Speicherplatz vorgenommen werden. Die moderne Vermittlungstechnik erlaubt die Trennung von eingehendem und ausgehendem Signal. Hierdurch können für die Lokalisierung des Anrufers eventuell wichtige Hintergrundgeräusche eindeutig zugeordnet werden.

[3] Diese Bezeichnungen wurde 1987 eingeführt (H.J. Künzel, Sprecher-Erkennung. Grundzüge forensischer Sprachverarbeitung, Heidelberg, S. 77 ff); im englischsprachigen Bereich hat sich hierfür der Begriff *speaker profiling* durchgesetzt.

[4] Ein Sprecher kann nicht selten auch durch den Stimmensachverständigen als Raucher erkannt werden, wenn nämlich während eines Telefongesprächs an der Zigarette gezogen wird. Dies ist z.B. häufig bei "entspannten" Dialogen der Fall, wie sie in TÜs anfallen.

[5] A. Braun und A. C. A. Rietveld (1995): The effect of cigarette smoking on perceived age. In: Proceedings 13th Intern. Congress of Phonetic Sciences Stockholm, vol.2, S. 296.

[6] S. die ausführliche Darstellung der Struktur, Möglichkeiten und Grenzen von DRUGS bei H.J. Künzel (2001): Eine Datenbank regionaler Umgangssprachen des Deutschen (DRUGS) für forensische Anwendungen. In: Zeitschrift für Dialektologie und Linguistik ZDL 2, 129-154.

[7] In dem angenommenen Beispiel würde sich die Situation folgendermaßen darstellen: Die meisten Sprecher mit Italienisch als Muttersprache haben enorme Schwierigkeiten, Kombinationen von zwei oder drei Konsonanten am Ende eines Wortes auszusprechen, weil diese eben im Italienischen nicht existieren, und fügen dann in der Regel einen Vokal ein, der dem unbetonten /e/ im Deutschen (z.B. an Ende von "Sonne" entspricht. Ein deutsches Wort wie "erst" wird dann als "erste" ausgesprochen. Wird es dagegen korrekt ausgesprochen, sind Zweifel an der italienischen Muttersprache angebracht.

[8] In der Literatur ist ein Mordfall beschrieben, in dem die auch spektralanalytisch nachweisbaren Schmatzlaute des Täters auf beide Ursachen zugleich zurückgeführt werden konnten

(H. J. Künzel (1987): Sprecher-Erkennung. Grundzüge forensischer Sprachverarbeitung, Heidelberg, S. 67).

[9] Wenn z.B. in einem Fall von Erpressung z. N. der Deutschen Bahn der anonyme Anrufer einen so speziellen Ausdruck wie "Langsamfahrstrecke" benutzt, liegt der Verdacht nahe, daß er selbst Eisenbahner ist. Dieser Fall ist tatsächlich vor Jahren in Süddeutschland aufgetreten.

[10] H. J. Künzel (1987): Sprecher-Erkennung. Grundzüge forensischer Sprachverarbeitung, Heidelberg, S. 6f.

[11] P. French (1990): Analytic procedures for the determination of disputed utterances. In: H. Kniffka (Hg.), Texte zur Theorie und Praxis der forensischer Linguistik. Tübingen, S. 201-213.

[12] Dies ist z.B. häufig der Fall, wenn ein Sprecher undeutlich oder leise spricht, mehrere Personen zur gleichen Zeit sprechen, oder ein lautes Nebengeräusch auftritt. Diese Beeinträchtigungen können leider auch zusammen auftreten.

[13] Hierzu ein Beispiel: In einem Fall des mehrfachen Mordes war auf einer verdeckt hergestellten Tonbandaufzeichnung von dem Hauptverdächtigen angeblich der Verbringungsort einer der Leichen genannt worden, wegen des Hintergrundgeräuschs (Aufnahmeort war eine Pizzeria) aber nicht verständlich. Der phonetische Sachverständige konnte mit Hilfe der o.g. modernen Editiertechniken zwar die Anzahl der Silben ermitteln, das fragliche Wort selbst aber auch nicht verstehen. Daraufhin wurde die einsendende Dienststelle in Bayern um Entsendung eines aus der betr. ländlichen Gegend stammenden Beamten gebeten. Zunächst war auch das gemeinsame Abhören der Endlosschleife im Labor vergeblich. Dabei verfestigte sich der Höreindruck der fraglichen Stelle im Kopf des Beamten jedoch derart, daß ihm während der Rückfahrt nach Bayern spontan der Ortsname einfiel. Die Leiche wurde dort unter einem frischen Betonfundament entdeckt.

[14] Mit den in den letzten Jahren aufgekommenen, im Auftrag von Nachrichtendiensten entwickelten und nicht frei erhältlichen Geräten auf der Basis von sog. Mikrofon-Arrays (Anordnung vieler Mikrofone in exakt definierten Abständen auf einem Geräteträger) und einer komplexen Auswertesoftware können auch unter den genannten ungünstigen forensischen Bedingungen teilweise perfekte Ergebnisse erzielt werden. Allerdings handelt es sich hierbei nicht um eine Sprachverbesserung im klassischen Sinne, sondern um die Vermeidung von Störungen bereits während der Aufzeichnung. So können u.a. analog zum optischen Zoom einer Kamera akustische Ziele exakt angepeilt, akustisch "vergrößert" und dadurch gleichzeitig in der Nähe des Ziels vorhandene Störquellen wirksam ausgeblendet werden. Auf eine weitergehende Darstellung wird hier verzichtet. Interessierte mögen sich an den Verfasser wenden.

[15] S. dazu H.J. Odenthal ([3]1999): Die Gegenüberstellung im Strafverfahren, Stuttgart, S. 71ff.).

[16] Eine seltene Ausnahme bildet das "Ausleihen" einer fremden Stimme exakt zu dem Zweck, diese Tat-Täter-Verbindung zu vermeiden. Ein Beispiel bietet der Fall der Entführung und Ermordung z. N. Fiszman (Frankfurt), in dem der Haupttäter mit Hilfe seines Computers eine Stimme aus einem Sprachlernprogramm benutzte, um den nachweislich aus zahlreichen Fragmenten zusammengefügten Text einer später am Telefon als "Konserve" abgespielten

Botschaft zusammenzustellen. Dieses Verfahren hat für den Anwender ähnliche Vor- und Nachteile wie das beliebte Zusammenstellen einer schriftlichen Nachricht mittels aus einer Zeitung ausgeschnittener Wörter - einschließlich der Gefahr, genau daran entlarvt zu werden! Im Fall Fiszman wurden nämlich auf dem beschlagnahmten Computer einschlägige Dateien gefunden. Wie sich später herausstellte, hatte das Stimmengutachten im übrigen die geliehene Stimme zutreffend charakterisiert, einschließlich der Bestimmung der regionalen Herkunft. Genau dieser Umstand hatte ironischerweise die Ermittlungen zeitweise in eine falsche Richtung geführt, als intensiv aber vergeblich nach einem "im nördlichen Bereich des westfälischen Teils von NRW aufgewachsenen" Bekannten oder Komplizen des bereits frühzeitig in Verdacht geratenen, aus Wiesbaden stammenden Täters gesucht wurde.

[17] Eine ausführliche Darstellung derselben findet sich z.B. bei H.J. Künzel (1998): Zur Entwicklung der forensischen Sprecher-Erkennung in Deutschland. In: Bundeskriminalamt (Hg.), Festschrift für Horst Herold zum 75. Geburtstag, Wiesbaden, S. 369 - 395.

[18] H. Bolt et al. (1979): On the theory and practice of voice identification. Washington DC.

[19] H.J. Künzel (1987): Sprecher-Erkennung. Grundzüge forensischer Sprachverarbeitung, Heidelberg.

[20] Entnommen aus Künzel (1998), Zur Entwicklung der forensischen Sprecher-Erkennung in Deutschland. In: Bundeskriminalamt (Hg.), Festschrift für Horst Herold zum 75. Geburtstag, Wiesbaden, S. 369 - 395.

[21] S. dazu M. Hecker (1993): Forensische Handschriftenuntersuchung. Heidelberg.

[22] J. Gonzalez-Rodriguez, J. Ortega-Garcia and J.L. Sanchez-Bote (2002): Forensic Identification reporting using automatic biometric systems. In: Zhang (ed.), Biometrics Solutions for Authentication in an E-World, Kluiver.

[23] Ein erster Bericht ist kürzlich vorgelegt worden: H.J. Künzel and J. Gonzalez-Rodriguez (2003): Combining Automatic and Phonetic-Acoustic Speaker Recognition Techniques for Forensic Applications. In: Proceedings 15th Intern. Congress of Phonetic Sciences, Barcelona, Spain.

[24] H. Hollien et al. (1995): Criteria for earwitness lineups. In: Journal of Forensic Linguistics 2, 143 – 153. S. auch G.L. Wells et al. (1998): Eyewitness identification procedures: recommendations for line-ups and photo-spreads. In: Law &Human Behavior 22, S. 603-647.

[25] S. die Diskussion bei A.P.A. Broeders (2001): Forensic Speech and Audio Analysis / Forensic Linguistics. 13th Interpol Forensic Science Symposium, Lyon, France.