

(II) SPRECHERERKENNUNG:

ARBEIT FÜR MENSCH ODER COMPUTER ?

Kurz gesagt...

Die Sprechererkennung stellt in den forensischen Wissenschaften ein ungelöstes Problem dar. Es kommen verschiedene Analysemethoden zur Anwendung: Der phonetische, der spektrographische sowie der automatische Ansatz. Der erste Teil dieses Artikels betrachtet die Validität sowie die Reliabilität dieser Methoden für eine praktische Anwendung.

Im zweiten Teil des Artikels (erscheint zu einem späteren Zeitpunkt) wird die Funktionsweise eines Modells vorgestellt werden, welches den forensischen Wissenschaften dank einer Interpretation der auf dem Bayes'schen Theorem beruhenden Beweis speziell zugeschnitten ist. Der Artikel wird einige Resultate des Auswertungsverfahrens dieses Systems präsentieren. Schliesslich... betont er die Notwendigkeit, Datenbanken zugehöriger Populationen zu schaffen. Hierzu wird die Verwendung moderner technischer Mittel zur Datenerhebung vorgeschlagen.

Vorstellung eines auf GMM basierenden automatischen Sprechererkennungs-systems

Im Hinblick auf die Anwendung einer automatischen Sprechererkennungs-Methode muss zuerst ein Stimmmodell der Stimme der verdächtigten Person konstruiert werden. Dies erlaubt den Vergleich der von einer Telefonabhörung herrührenden Aufnahme mit einer Vergleichsaufnahme.

Aufnahme

Die verdächtige Person wird von einem Telefonnetz mittels einer ISDN-Karte direkt auf Computer aufgenommen. Sofern ein Telefon vorhanden ist, kann diese Etappe damit von einem beliebigen Ort aus realisiert werden, so von einem Polizeiposten oder einer Institution des vorzeitigen Strafvollzugs aus. Zwei Arten von Aufnahmen werden gemacht. Eine erste Aufnahmeserie wird derart realisiert, dass ein Modell der verdächtigten Person resultiert. Die zweite Serie wird mit dem Zweck gemacht, die Intravariabilität der verdächtigten Person auszuwerten.

Beschreibung der Stimme der verdächtigten Person anhand eines Modells

Vorbehandlung

Die Aufnahmeserie, welche für die Stimmmodell-Errichtung der verdächtigten Person realisiert wird, erfordert eine automatische

Vorbehandlung zur Unterdrückung von Pausen, um die Sprechorte von den Sprechpausen unterscheiden zu können.

Extraktion der Merkmale

Die sprecher-abhängigen Merkmale werden anschliessend vom vorbehandelten Zeichen durch eine Methode, welche von der linearen Vorhersage abgeleitet ist (Linear Prediction Coding, LPC), extrahiert. Diese Methode erlaubt eine Charakterisierung der spektralen Hüllkurve des Zeichens in zeitlicher Hinsicht und wird lineare perzeptuelle Vorhersage (Perceptual Linear Prediction, PLP) genannt. Dies aus dem Grund, weil die frequentielle Komponente des Sprachzeichens nicht auf einer linearen Skala, sondern auf einer nicht-linearen Skala beschrieben wird, deren Auflösung ähnlich derjenigen des menschlichen Ohres ist [HERMANSKY, 1990]. In der Folge erfolgt die Extraktion auf den Signalsegmenten von 20 ms. Von jedem Parameter werden 12 PLP-Parameter extrahiert. Dies bedeutet, dass das ganze Zeichen durch 12 PLP-Parameter beschrieben wird. In jeder Serie ist die Anzahl Parameter direkt von der Länge des analysierten Zeichens abhängig.

Modellbeschreibung anhand GMM

Diese Etappe besteht in einer Modellbeschreibung der Verteilung jeder einzelnen der Serien der PLP-Parameter durch das Zusammenwirken von 64 gauss'schen Dichtefunktionen. Jede gauss'sche Dichtefunktion ist definiert durch ihren Mittelwert und ihre Standardabweichung. Jede PLP-Parameter-Verteilung wird durch 64 Mittelwerte sowie 64 Standardabweichungen modellhaft beschrieben. Das ganze Sprachzeichen – und damit die Stimme, welche das Signal erzeugt hat – wird anhand eines Modells mit 12 Serien von 64 Mittelwerten und 64 Standardabweichungen definiert.

Schliesslich wird diese Information in einer Kovarianz-Matrix organisiert, welche das Stimmmodell der verdächtigten Person konstituiert [REYNOLDS, 1995].

Vergleich von Modell und Indiz: Bestimmung des Beweiselements

Das auf diese Weise kreierte Modell kann derart mit dem Indiz verglichen werden, dass man ihre Proximität misst. Dieses die Proximität messende Mass stellt ein Beweiselement dar. Das Indiz wird ebenfalls einer Vorbehandlung unterzogen, und zwar derart, dass die Pausen unterdrückt werden. Danach wird zu einem Verfahren der PLP-Parameter-Extraktion geschritten. Die Angleichung der Parameter, welche vom Indiz herausgefiltert wurden, sowie der Verteilung ebendieser Parameter gemäss Modellbeschreibung wird ausgerechnet und unter der Form eines Proximitäts-Indizes synthetisiert. Vom mathematischen Gesichtspunkt her berechnet das System die Wahrscheinlichkeit des Indizes, und zwar in Form der vom Modell herstammenden Hypothese. Allgemein ausgedrückt kann man sagen, dass alle künstlichen Wahrnehmungssysteme, wie die automatischen Identifizierungssysteme von Fingerabdrücken (*Automatic Fingerprint Identification System*, AFIS) oder die Sprechererkennungs-Systeme entweder ein Distanz- oder aber ein Proximitäts-Indiz zwischen zwei Vergleichsproben liefern.

Interpretation des Beweiselements

Einführung

Wenn Ingenieure in einem Gebiet wie der künstlichen Wahrnehmungszintelligenz eine wichtige Erkenntnis gezogen haben, so geht die Anwendung dieser Kenntnis in den forensischen Wissenschaften in zwei Schritten vor sich. Die erste Etappe ist theoretischer Natur und zielt darauf, vom methodischen Gesichtspunkt her eine valide Erkennungsmethode zu verwirklichen.

Ist es möglich, ausgehend vom Wert des Proximitäts-Indizes (errechnet aus einem automatischen Sprecher-Erkennungs-System wie GMM), einen Sprecher formell als mögliche Testaufnahme-Quelle zu identifizieren oder auszuschliessen?

Die Antwort hierauf lautet nein, da ein automatisches Sprechererkennungs-System ein Werkzeug in der Form eines Masses und nicht ein binäres Entscheidungsinstrument zur Identifizierung oder Verifikation darstellt. Es liefert ein Beweiselement, und zwar in Form eines numerischen Wertes. Es fehlt indes an einem Rahmen, welcher es erlauben würde, dieses Mass auch zu interpretieren.

Die Analogie ermöglicht ein besseres Verständnis dieser beiden Analysestufen. Gleich wie ein künstliches Wahrnehmungssystem ist auch das Messband ein Messwerkzeug. Es erlaubt die Beschreibung der Distanz zwischen zwei Objekten mit einer gewissen Präzision. Eine absolute Antwort auf die Frage, ob die zwei Objekte nahe beieinander liegen oder nicht, kann es indes nicht liefern. Will man diese Frage klären, so ist eine Definition eines die Interpretation der Distanz ermöglichenden Rahmens unabdingbar.

Befindet sich Lausanne in der Nähe von Luzern? Zu dieser Frage gibt es keine absolute Antwort. Falls aber als Referenz-Distanz die Entfernung zwischen Zürich und Winterthur genommen wird, so lautet die Antwort nein; nimmt man jedoch die Entfernung zwischen Zürich und Berlin als Referenz-Distanz, so heisst die Antwort ja.

In gleicher Art erfordert die Bestimmung, ob die Proximität zwischen dem Stimmmodell einer verdächtigten Person und einem Indiz gross ist oder nicht, die Definition eines Rahmens, welcher eine kohärente Interpretation dieses Beweiselements erlaubt.

Definition eines Referenz-Rahmens

Auswertung der Intravariabilität

In dieser ersten Etappe muss die Verteilung der Proximität definiert werden. Das Proximitäts-Indiz wird durch das Erkennungssystem geliefert, wenn das Stimmmodell der verdächtigten Person mit den Vergleichs-Aufnahmen ebendieser Stimme verglichen wird; dies geschieht vor dem Hintergrund, die Intravariabilität der verdächtigten Stimme auszuwerten (*within-source variability*) (Fig. 1).

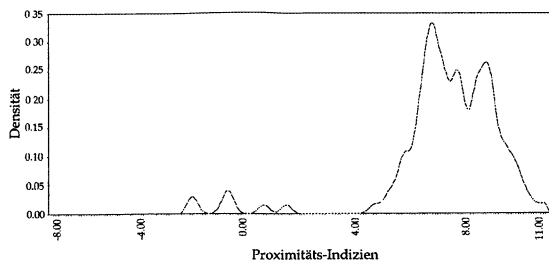


Fig. 1: Verteilungsbeispiel der auf einem Intravariabilitäts-Auswertungs-Verfahren beruhenden Proximitäts-Indizien

Um dieses Ziel zu erreichen, wird das Stimmmodell der verdächtigten Person mit dem zweiten Aufnahmetyp, welcher mit dieser Person gemacht wurde, verglichen. Dies umfasst gelesene sowie spontan gesprochene Sprachaufnahmen, die Verwendung von Mobil- sowie Festnetztelefonen und möglicherweise die Aufnahme von Sequenzen, welche einige Tage oder Wochen zurückliegen. Dieses Verfahren erfolgt mit dem Zweck, der Veränderung der Stimme in zeitlicher Hinsicht gerecht zu werden.

Auswertung der Intervariabilität

Dieser zweite Schritt beginnt mit dem Anhören des Indizes mit dem Ziel, die gesprochene Sprache, den Akzent und das Geschlecht zu bestimmen, sofern dies möglich ist. Diese Parameter dienen dazu, die in Frage kommende mögliche Population von Sprechern zu bestimmen. Als solche kommt beispielsweise die Gesamtheit aller männlichen Sprecher, welche sich auf französisch mit einem Westschweizer Akzent ausdrücken, in Frage. Weil es unmöglich ist, bei jedem dieser Westschweizer Sprecher eine Sprechprobe zu machen, wird die potenzielle Population mit Hilfe einer möglichst repräsentativen Datenbank modellhaft beschrieben. In der Schweiz besitzt die Unternehmung Swisscom[®] beispielsweise eine als Modellbeschreibung dienende Datenbank der schweizerischen Sprecher. Die Datenbank enthält 5000 französisch sprechende, 5000 deutsch sprechende sowie 5000 italienisch sprechende Personen. Für den Moment besteht dieser Typ von Datenbank nur in den ökonomisch am weitest entwickelten Ländern, währenddem die telefonischen Abhörmassnahmen generell die in den weniger entwickelten Ländern gesprochenen Sprachen betreffen.

Eine Lösung auf dieses Problem könnte darin bestehen, die von den Abhörmassnahmen herstammenden Aufnahmen nach Sprachen und Akzenten zu sortieren. Damit könnten Datenbanken kreiert werden, welche als sehr pertinente Modelle die potenzi-

ellen Populationen repräsentieren könnten. Bedauerlich ist jedoch, dass sich Rechtsvertreter über den möglichen Zugang zur diesbezüglich besten Datenbank im Hinblick auf die Modellbeschreibung verschiedener potenzieller Populationen gemäss der herrschenden Rechtslage meist nicht bewusst sind.

Wenn bereits eine Datenbank existiert oder eine kreiert wird, so wird ein darauf beruhendes Stimmmodell jedes Sprechers gemäss oben erwähnter Methode konstituiert. Diese Datenbank dient dazu, die Verteilung der durch das Sprechererkennungs-System gelieferten Proximitäts-Indizien zu definieren. Dies ist der Fall, wenn das Indiz mit einem zufällig aus der potenziellen Population ausgewählten Sprecher gleicher Sprache und gleichen Akzents verglichen wird (*between-source variability*) (Fig. 2).

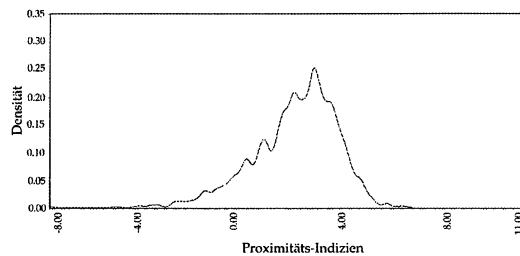


Fig. 2: Verteilungsbeispiel der aus einem Intervariabilitäts-Auswahlverfahren stammenden Proximitäts-Indizien

Interpretation des Beweiselementes

Die Auswertung der Intravariabilität einer verdächtigten Person und der Intervariabilität des Indizes in der potenziellen Population macht die Interpretation des Beweiselementes in einem kohärenten Rahmen möglich. Die Wahrscheinlichkeit des Beweiselementes wird in folgenden Alternativhypothesen kalkuliert:

- H1 :Die verdächtige Person ist die Indizquelle
- H2 :Eine andere Person aus der potenziellen Population ist Indizquelle

Die Wahrscheinlichkeit des Beweiselementes in der Hypothese H1 wird gleich wie die Wahrscheinlichkeit des Beweiselementes in der Intravariabilitäts-Verteilung berechnet. Falls beispielsweise dem Beweiselement ein Gewicht von 8.00

zukommt, so beträgt seine Wahrscheinlichkeit rund 0.17. Falls hingegen das Beweiselement 4.00 beträgt, so ist die Wahrscheinlichkeit fast 0, nämlich etwa 10^{-3} (Fig. 3).

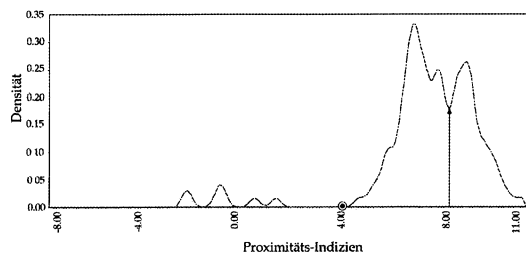


Fig. 3: Auswertung der Wahrscheinlichkeit zweier Beweiselemente (4.00 und 8.00) in der Intra-variabilitäts-Verteilung.

Die Wahrscheinlichkeit des Beweiselements in der Hypothese H2 wird wie die Wahrscheinlichkeit des Beweiselements in der Intervariabilitäts-Verteilung berechnet. Falls beispielsweise das Beweiselement 8.00 beträgt, so ist seine Wahrscheinlichkeit nahe 0, nämlich etwa 10^{-3} . Falls hingegen dem Beweiselement ein Gewicht von 4.00 zukommt, so beträgt seine Wahrscheinlichkeit rund 0.15 (Fig. 4).

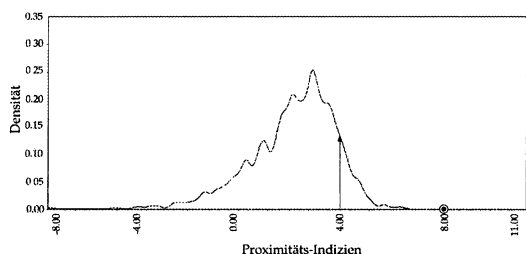


Fig. 4: Auswertung der Wahrscheinlichkeit zweier Beweiselemente (4.00 und 8.00) in der Inter-variabilitäts-Verteilung.

Diese zwei Elemente erlauben es schliesslich, einen Wahrscheinlichkeits-Zusammenhang des Beweiselementes für den Fall auszuwerten, dass entweder die Hypothese H1 oder aber Hypothese H2 wahr ist. Dieser Wahrscheinlichkeits-Zusammenhang macht eine Aussage über die Pertinenz der Wahl der verdächtigten Person als mögliche Indiz-Quelle.

Im vorerwähnten Beispiel mit einem Beweiselement von 8 ist die Wahrscheinlichkeit für dieses Beweis-element 170 Mal ($0.17 / 10^{-3}$) höher für den Fall, dass Hypothese H1 wahr ist, als für den Fall, dass Hypothese H2 wahr ist (Fig. 5). Die LR beträgt 170.

Wenn hingegen das Beweiselement 4 beträgt, so ist seine Wahrscheinlichkeit 150 Mal ($0.15 / 10^{-3}$) kleiner für den Fall, dass Hypothese H1 wahr ist, als für den Fall, dass H2 wahr ist (Fig. 5). Die LR ergibt einen Wert von $1/150$.

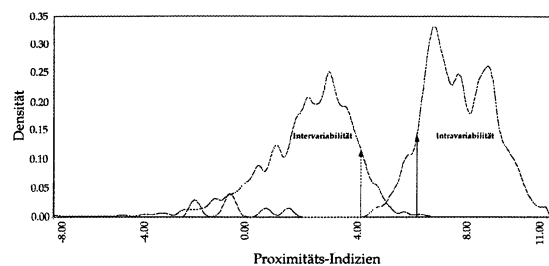


Fig. 5: Auswertung des Wahrscheinlichkeitszusammenhangs zweier Beweiselemente (4.00 und 8.00) für den Fall, dass entweder Hypothese H1 oder H2 wahr ist.

Auswertung des Erkennungssystems

Einführung

Wenn die Validität einer Methode akzeptiert ist, so besteht eine zweite, praktischere Etape darin, die Reliabilität der Methode auszuwerten. Dies hat in Bedingungen zu erfolgen, welche potenziellen reell anzutreffenden Situationen möglichst nahekommen.

Die Variablen, welche sich eignen, die Reliabilität einer in den forensischen Wissenschaften angewandten automatischen Erkennungsmethode zu beeinflussen, sind weiter oben ausführlich dargelegt worden. Eine Darstellung der Resultate für einige dieser analysierten Variablen erlaubt es, die Robustheit der GMM-Methode besser abschätzen zu können.

Prinzip

Das Auswertungsprinzip der Methode besteht darin, die Verteilung der Wahrscheinlichkeits-Beziehungen abzuschätzen und zu vergleichen. Letztere können ausgehend vom Beweiselement erhältlich gemacht werden. Zwei Möglichkeiten stehen hierfür zur Verfügung. So einerseits, wenn sich die Hypothese H1 bestätigt, d.h. wenn die Modellquelle und das Indiz identisch sind. Eine zweite Möglichkeit steht andererseits zur

Disposition, wenn die Hypothese H2 bestätigt wird, also wenn die Modellquelle und das Indiz verschieden sind.

Vorstellung der Resultate

Für die Präsentation der Resultate beruht die Wahl auf einer Vorstellungsmodalität, wie sie durch EVETT UND BUCKLETON auf dem Gebiet der Interpretation der genetischen Analyse vorgeschlagen worden war [EVETT UND BUCKLETON, 1996]. Die Abszisse ist in wachsenden Werte-Termen von Wahrscheinlichkeitsbeziehungen skaliert. Die Ordinate gibt die Wahrscheinlichkeit an, dass das Testresultat einen gegebenen LR-Wert überschreitet. Jede graphische Darstellung besteht aus zwei Kurven; die erste berücksichtigt die Evolution der geschätzten Wahrscheinlichkeitsbeziehungen, wenn die Hypothese H1 sich bestätigt. Die zweite Kurve gibt die Verteilung der geschätzten Wahrscheinlichkeitsbeziehungen an, wenn die Hypothese H2 bestätigt wird. N bezeichnet dabei die Anzahl realisierter Experimente.

Resultate

Ein erstes Experiment zeigt den Einfluss des Faktors „Zeit“ zwischen dem Zeitpunkt der Indiz-Aufnahme und dem Stimmmodell der verdächtigten Person. Das Indiz wurde vor dem Modell aufgenommen, und die zwei Aufnahmen wurden vom selben Telefon aus gemacht. Logischerweise sind die mittels der Methode ausgewerteten Wahrscheinlichkeitsbeziehungen höher, wenn die Hypothese H1 sich bestätigt, als wenn die Hypothese H2 verifiziert wird. Zugleich nimmt aber die Leistungsfähigkeit langsam ab, wenn das Intervall zwischen dem Aufnahme-Indiz und dem Modell sich erhöht (von 0 bis 60 Tagen) (Fig. 6).

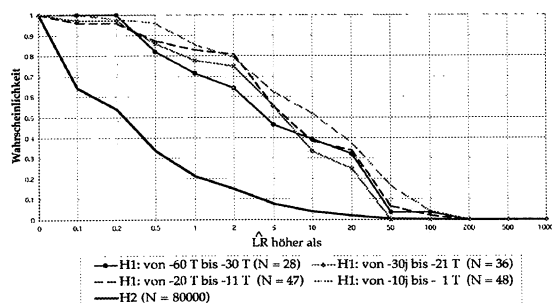


Fig. 6: Evolution der Wahrscheinlichkeitsbeziehungen in Funktion des Zeitintervalls, welches die Aufnahme des Modells von der Aufnahme des Indizes trennt.

Ein zweites Experiment zeigt den Einfluss einer Stimmverstellung im Indiz. Wenn dasselbe Telefon sowohl für die Aufnahme des Modells sowie für die Aufnahme des Indizes verwendet wird, so bleiben die ausgewerteten Wahrscheinlichkeitsbeziehungen höher, wenn sich die Hypothese H1 bestätigt, als wenn die Hypothese H2 verifiziert wird. Diese Feststellung trifft sogar trotz einer beträchtlichen Verminderung der Leistungsfähigkeit durch eine derartige Stimmverstellung zu (Fig. 7).

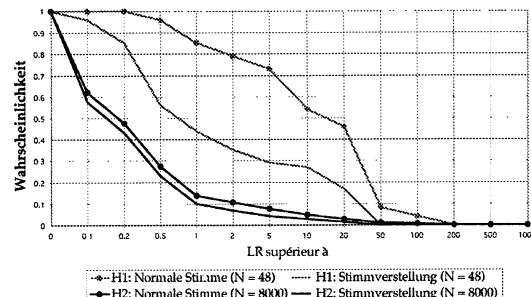


Fig. 7: Evolution der Wahrscheinlichkeitsbeziehungen in Funktion des Fehlens oder des Bestehens einer Verstellung im Indiz (Verwendung desselben Telefons für die Aufnahme des Modells sowie des Indizes)

Wenn indes für die Aufnahme des Modells sowie des Indizes nicht identische Telefone verwendet werden, und sofern zusätzlich eine Verstellung der Stimme im Indiz besteht, so sind die mittels der Methode ausgewerteten Wahrscheinlichkeitsbeziehungen äquivalent. Unabhängig davon, ob sich die Hypothese H1 oder H2 bestätigt, kann unter derartigen Bedingungen kein beweiskräftiges Ergebnis erbracht werden (Fig. 8).

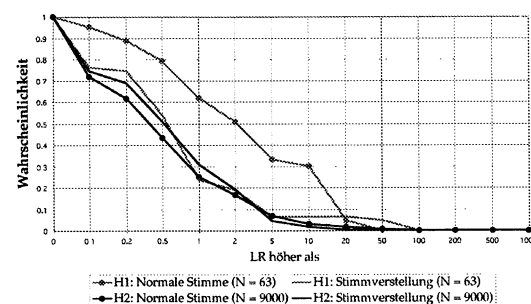


Fig. 8: Evolution der Wahrscheinlichkeitsbeziehungen in Funktion des Fehlens oder des Bestehens einer Verstellung im Indiz (Verwendung verschiedener Telefon für die Aufnahme des Modells sowie des Indizes)

Ein drittes Experiment zeigt den Einfluss der Präsenz eines von einer Stimme herrührenden Grundgeräusches im Indiz.

Das Signal/Rausch-Verhältnis (auf englisch „Signal to Noise Ratio“, SNR) einer normalen Telefonlinie liegt in der Größenordnung von 40 dB. Bedingungen starken Grundgeräusches (SNR = 12 dB oder 6 dB) verschlechtern die Leistungsfähigkeit der Methode zwar beträchtlich, aber die mittels der Methode ausgewerteten Wahrscheinlichkeitsbeziehungen bleiben höher, wenn sich Hypothese H1 bestätigt, als wenn sich H2 verifiziert. Wenn hingegen das Grundgeräusch ebenso stark wie das Signal ist (SNR von 0 dB), so sind die mittels der Methode ausgewerteten Wahrscheinlichkeitsbeziehungen äquivalent, unabhängig von der Bestätigung der Hypothese H1 oder H2. Damit kann unter derartigen Bedingungen kein beweiskräftiges Resultat erwartet werden (Fig. 9).

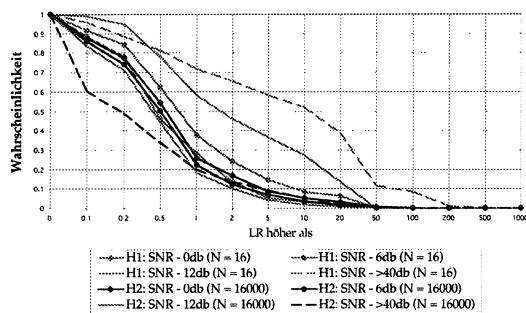


Fig. 9: Evolution der Wahrscheinlichkeitsbeziehungen in Funktion des Fehlens oder des Bestehens eines Grundgeräusches anlässlich der Aufnahme des Indizes.

Ein letztes Experiment zeigt den Einfluss des für die Aufnahme des Indizes gebrauchten Systems. In zwei Fällen wird das Indiz durch die Vermittlung des Telefonnetzes aufgenommen; im ersten Fall wird es direkt auf Computer aufgenommen, und zwar durch den Einsatz einer ISDN-Karte. Im zweiten Fall wird es auf dem analogen Registriergerät einer Kantonspolizei aufgenommen. Die Resultate zeigen, dass das analoge Aufnahmesystem die Informationsqualität dermassen verschlechtert, dass die mittels der Methode ausgewerteten Wahrscheinlichkeitsbeziehungen äquivalent sind, egal, ob Hypothese H1 oder H2 sich bestätigt. Dies heisst, dass pertinente Information bei der Datenerhebung infolge des verwendeten Aufnahmesystems verloren geht, währenddem diese Information im Fall der Verwendung einer numerischen Aufnahme erhalten bleibt (Fig. 10).

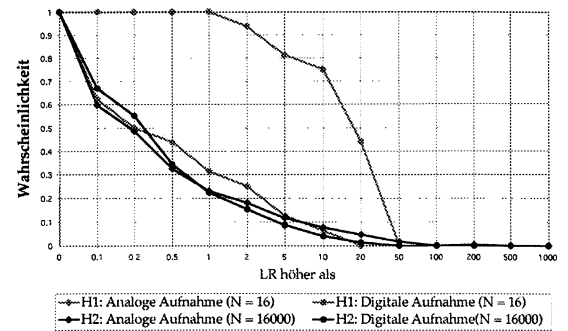


Fig. 10: Evolution der Wahrscheinlichkeitsbeziehungen in Funktion des für die Aufnahme des Indizes verwendeten Systems

Schlussfolgerung

Natürliche oder künstliche Wahrnehmung?

Die Stimme ist in erster Linie eine menschliche Verhaltensweise und erst in zweiter Linie ein Signal i.S. einer nicht perfekten Übertragung dieses Verhaltens. Das natürliche Wahrnehmungssystem, das menschliche auditive System, ist komplizierter als jedes künstliche Wahrnehmungssystem. Dank einer parallelen und simultanen Analyse mehrerer Kommunikationsstufen unterscheidet es zugleich die Verständlichkeit des Mitteilungsinhalts, den emotionalen Zustand des Sprechers sowie seine Absicht. Dazu kommt, dass das auditive System das einzige effiziente Wahrnehmungssystem unter Bedingungen mit extremen Grundgeräuschen bleibt.

Subjektive oder objektive Analyse?

Die Behandlung auditiver Information durch das menschliche Gehirn erweist sich hingegen als ungeeignet für die Sprechererkennung unter forensischen Bedingungen.

Wird diese Aufgabe durch Nicht-Experten – Opfer oder Zeuge – ausgeführt, so haben die Ergebnisse einen indikativen Wert, dessen Ungewissheit vergleichbar ist mit einer Zeugenaussage.

Wird die Aufgabe auf der Grundlage phonetischer Analysen (auditiv-perzeptive und phonetisch-akustische) durch Experten ausgeführt, so erfordert diese Aufgabe eine perfekte Kenntnis der Indizien-Sprache des Experten, will man die von der Sprache oder dem Akzent abhängigen Charakteristiken nicht mit den wirklich sprecherabhängigen Merkmalen verwechseln. Weil der analog funktionierende menschliche Denkmodus sich statistischen Wahrscheinlichkeiten schlecht anpasst, ist die subjektive Auswertung seltener sprecherabhängigen Charakteristiken von der jeweiligen Person des Experten abhängig. Daraus resultieren potenziell sehr unterschiedliche Evaluationen. Eine kohärente Interpretation der Resultate in einem kohärenten Rahmen kann nur unter erschwerten Umständen ins Auge gefasst werden. Die Kosten des phonetischen Ansatzes an Zeit/Personal sind nämlich beträchtlich, sofern die Interpretation nach den Regeln der Kunst erfolgt. Schliesslich ist festzustellen, dass ein Spezialist ausreichende Kenntnisse nur in seiner Muttersprache sowie einer beschränkten Anzahl anderer Sprachen erwerben kann.

Die Verwendung automatischer Methoden zur Erfüllung dieser Aufgabe der Sprechererkennung in den forensischen Wissenschaften sieht sich in erster Linie noch immer mit der Unvollkommenheit künstlicher Wahrnehmungssysteme konfrontiert, welche in der Praxis empfindlich auf diverse Variationstypen des Sprachsignals reagieren. Der Vorteil der Verwendung dieses Methodentyps liegt in der Behandlung der Daten. Erstens ist jede beliebige Sprache zugänglich, da die Grundkenntnis vom Vorhandensein einer Sprecher-Datenbank abhängt; die Methode ist somit geeignet, sich sehr schnell neuen Kriminalitätsbedingungen anzupassen. Zweitens wird die Seltenheit herausgefilterter Charakteristiken mittels einer automatischen Methode gemäss statistischen Wahrscheinlichkeiten ausgewertet, was der Wirklichkeit näher kommt als subjektive Wahrscheinlichkeiten. Drittens ist eine Interpretation in einem kohärenten Rahmen zwar aufwendig, was den zeitlichen Kontext und die Infrastruktur angeht, nicht aber was den zeitlichen Kontext und die menschlichen Ressourcen betrifft. Auch dies dürfte indes angesichts der aktuellen sowie künftigen Leistungsfähigkeit von Computern kein Hindernis darstellen. Endlich erlaubt der automatische Ansatz, schnell und sehr kostengünstig Simulationen zu machen. Dies geschieht vor dem Hintergrund, die Wahrscheinlichkeit, dass die Analyse zu einem beweiskräftigen Ergebnis führt, *a priori* bestimmen zu können, und zwar indem nur das Indiz analysiert wird (*preassessment*).

Indizienhebung und Kreierung von Datenbanken

Von einem allgemeinen Standpunkt aus gesehen ist die Nachfrage seitens der Strafverfolgungs- und Strafjustizbehörden in bezug auf das Gebiet der Sprechererkennung steigend. Aktuell betragen die Nachfragen rund 40 pro Jahr. Ohne Zweifel gibt es darüber hinaus eine Anzahl nicht explizierter Anfragen, und zwar aus dem Grund, weil den Fachleuten die fehlenden Reliabilität der vorgeschlagenen Methoden bekannt ist.

Es bleibt an dieser Stelle jedoch zu erwähnen, dass die Resultate – egal von welchem Analysetyp sie erwartet werden können – in erster Linie von der Qualität der analysierten Probe abhängen. Die von den schweizerischen Polizeikorps für die Datenerhebung verwendete Technologie entspricht den aktuellen Qualitätsstandards in keiner Weise mehr. Es ist deshalb nötig, die Polizei mit modernem Material auszurüsten, sofern man zu Sprechererkennungs-Analysen schreiten will.

Ein ideales System müsste erstens geeignet sein, die zwei Sprecher eines Telefongesprächs auf zwei unabhängigen Radiokanälen aufzunehmen. Diese Trennung von Kanälen könnte einerseits eine durch Lärm bedingte, unnötige Qualitätsverminderung vermeiden. Andererseits könnten damit alle Zweifel in bezug auf den Ursprung eines bestimmten Ausspruchs aus dem Weg geschafft werden. Zweitens könnte das System numerisch Bit für Bit das auf dem Telefonsystem übermittelte Signal aufnehmen. Dadurch könnte jeder infolge der Komprimierung auftretende Datenverlust vermieden werden, und die Vollständigkeit der Information wäre gewährleistet.

Die Interpretation von Analyseresultaten erfordert die Existenz von Datenbanken. Ideal wäre es, die von Telefonabhörungen herrührenden Aufnahmen zur Kreierung von Datenbanken verwenden zu können, da auf diese Weise die potenziellen Sprechergruppen modellhaft beschrieben werden könnten. Die meisten Indizien

Bibliographie

HERMANSKY
H (1990) *Perceptual linear predictive (PLP) analysis of speech*. *J. Acoust. Soc. Am.* 1738 - 1752.

REYNOLDS
DA (1995) *Automatic Speaker Recognition Using Gaussian Mixture Speaker Model*. *The Lincoln Laboratory Journal*. 8 : 173 - 191.

EVETT IW and
BUCKLETON
JS (1990) *Statistical analysis of STR data in: Advances in Hoemogenetics*
(eds: Carraredo A, Brickmann B,
Bär W) Springer-Verlag,
Heidelberg
Germany.

stammen nämlich von Sprachen her, für welche bis anhin keine Datenbanken bestehen.

Aktuell läuft diese wertvolle Information über die Polizeibehörden, welche mit der Datenerhebung betraut sind. Die Information wird jedoch nicht als Primärmaterie zur Kreierung von Datenbanken ausgenutzt.

Eine Verwendung der vorhandenen Information zu diesem Zweck erfordert das Vorhandensein juristischer und organisatorischer Einrichtungen. Dies ist denn auch der Preis, welcher bezahlt werden muss, damit wirklich reliable Sprechererkennungs-Methoden den Institutionen der Strafverfolgung sowie der Strafjustiz vorgeschlagen werden könnten.

Verfasser dieser Nummer: Didier Meuwly

*Ein spezieller Dank gebührt Herrn Martin Rindlisbacher,
Forschungsassistent am Institut de police scientifique et de criminologie,
für die Qualität seiner Übersetzung des französischen Originaltextes auf deutsch.*

Redaktion: Prof. P. Margot und Prof. M. Killias, IPSC, UNIL, 1015 Lausanne

Bitte senden Sie Ihre Bemerkungen und Mitteilungen an:

Sekretariat *Crimiscope*
UNIL - Institut de police scientifique et de criminologie
CH-1015 LAUSANNE

☎ (021) 692 46 42
Fax (021) 692 46 05
Int. (+ 41 21) 692 46 42