

15. Deutscher EDV-Gerichtstag Saarbrücken 2006

Sitzung des Arbeitskreises

„Sprechererkennung - Fingerabdruck der Stimme?“

Universität Saarbrücken, Gebäude 16, Hörsaal 117,

Donnerstag, den 14. September 2006, 13 Uhr 10 bis 15 Uhr 00

Die Sitzung wurde durch den Vorsitzenden des Arbeitskreises, Herrn Prof. Dr. Manfred Pausch, vor ca. 30 teilnehmenden Personen eröffnet. Der Vorsitzende stellte den Referenten vor und gab eine kurze Einleitung zum Thema des Vortrags.

Der Vortrag befasst sich mit der Sprechererkennung unter kriminalistischen bzw. forensisch-strafrechtlichen Aspekten. Zeitgeschichtlich sei es in diesem Zusammenhang interessant, dass das Wiedererkennen einer Stimme bereits im Prozess um die Entführung des Lindbergh-Sohns im Jahre 1936 eine Rolle spielte. Dort bekundete der im Zeugenstand angehörte Vater, dass die Stimme des Angeklagten auch diejenige gewesen sei, die ihm am Telefon Anweisungen für die Lösegeldübergabe erteilt habe. Diese Aussage trug – allerdings nur neben anderen Beweismitteln - dazu bei, dass der Angeklagte schuldig gesprochen und hingerichtet wurde. Aus der Sicht des heutigen deutschen Strafprozessrechts würde dies allerdings nicht zur Überführung eines Straftäters ausreichen. Von einem Wiedererkennen einer Sprechstimme durch einen Zeugen kann nur dann ausgegangen werden, wenn der Zeuge die Sprechstimme des Angeklagten aus einer Mehrzahl von Proben ähnlicher Stimmen zu identifizieren vermag, insofern gelten ähnliche Grundsätze wie bei der Gegenüberstellung oder bei Lichtbildern. In der forensischen Praxis spricht man infolgedessen auch von einer "Stimmgegenüberstellung".

Beim Sprechen handelt es sich um einen Bestandteil des Kommunikationsprozesses, also um die Übermittlung von Information. Der Sprecher ist in diesem Modell der Kommunikationstheorie der Sender, das Gesprochene ist das Signal, die Sprache der Code und der Hörer nimmt die Position des Empfängers ein. Unter physikalischer Betrachtungsweise handelt es sich bei Sprachsignalen um stehende akustische Wellen mit einem definierten Festpunkt, so dass sich neben der Grundfrequenz Oberwellen in ungradzahligen Vielfachen beobachten lassen.

Die Phonetik befasst sich mit einem Teil dieses Prozesses, nämlich mit der Substanz der gesprochenen Sprache.

Im Vortrag wurde zunächst grundlegend dargelegt, wie die Sprechstimme im Körper des Menschen physiologisch entsteht. Hieran sind die Atmung, die Erzeugung akustischer Wellen im Kehlkopf und die abschließende Lautbildung im Mundraum beteiligt. Das Einwirken aller drei Prozesse auf die Sprechstimme durch Muskel- und Nerventätigkeit führt zu Mustern und Charakteristiken, die bei jedem einzelnen Menschen verschieden ausgeprägt sind und sich kaum willentlich beeinflussen lassen.

Liegen Aufzeichnungen der Sprechstimme auf Tonträgern vor, so lassen sich diese Muster und Charakteristiken, sei es z. B. der Atemrhythmus beim Sprechen oder die Grundfrequenz des Sprechtons, durch technische Verarbeitung des Sprachsignals herausfiltern. Hierbei spielen elektronische Verarbeitungsmethoden, insbesondere die Errechnung von zeitabhängigen Frequenzspektren im Wege der Fourier-Analyse eine besondere Rolle.

Trotz der stetig verbesserten technischen Möglichkeiten müssen Verfahren der vollautomatischen Stimmerkennung nach derzeitigem Stand als nicht hinreichend zuverlässig angesehen werden. Eine für den einzelnen Menschen charakteristische, aus der Sprechstimme hergeleitete und maschinell aus einer Probe ermittelbare Kenngröße, die -ähnlich der DNA-Analyse- nahezu ausschließlich ihn als Spurenleger identifiziert und andere ausschließt, wurde bis jetzt nicht gefunden. Die höchsten erzielten Wahrscheinlichkeiten für eine Identifizierung erreichten in entsprechenden Versuchsreihen maximal 88 %. Selbst dieses Ergebnis setzt optimale Proben (ausreichend lange Sprechdauer und volle Erfassung des akustischen Spektrums ohne störende Umgebungsgeräusche) voraus, die unter Laborbedingungen erreicht werden können, aber oft nicht bei realen Spuren oder entsprechenden Asservaten. Zudem unterliegt die Sprechstimme eines Menschen darüber hinaus auch Schwankungen, z. B. unter Stressbedingungen. Versuche, sich diese Veränderungen kriminaltechnisch zur Beurteilung des Wahrheitsgehalts einer Aussage oder Einlassung zu Nutze zu machen, werden mittlerweile als gescheitert angesehen. Die Konstruktion eines Lügendetektors auf der Basis der Sprechanalyse ist demnach nicht möglich.

Der für Sprechererkennung maßgebliche internationale Fachverband IAFPA hat sich in seinen Verlautbarungen von den vollautomatischen Verfahren distanziert und erkennt mittlerweile nur noch objektivierende Verfahren mit einem hybriden Ansatz an.

Bei forensischen Fragestellungen zum Sprechstimmenvergleich, die im Rahmen eines gerichtlichen Sachverständigengutachtens beantwortet werden, kommen somit nur diese Verfahren zum Einsatz. Sie zeichnen sich dadurch aus, dass zwar die Sprechstimmenproben

elektronisch analysiert und die Ergebnisse der zeitaufgelösten Frequenzanalyse einander gegenübergestellt werden, die so gefundenen Muster und individuell variierenden Charakteristiken aber noch durch zusätzliche Kriterien, etwa Syntax, Morphologie und Semantik ergänzt und modifiziert werden. Diese ergänzenden Schritte können nur von Spezialisten vorgenommen werden, die über die nötige Kenntnis der vom Sprecher verwendeten Sprache verfügen. Am Ende der gutachterlichen Stellungnahme steht dann eine Äußerung zur Wahrscheinlichkeit, mit der sich die Identität des Spurenlegers aufgrund der zu vergleichenden Sprechstimmenproben annehmen oder ausschließen lässt.

Bei der sich anschließenden Diskussion ging es um die relativ häufig auftretende Problemstellung, bei einem nur als Tonträgeraufzeichnung vorliegenden Gespräch die einzelnen Äußerungen den verschiedenen Beteiligten zuzuordnen. Dies sei grundsätzlich möglich, wenn auch zeitaufwendig.

Ein Diskussionsbeitrag wies auf die Verwandtschaft zwischen Sprechererkennung und Biometrie hin.

Auf eine abschließende Frage hin wurde das Phänomen des „Bauchredners“ aus phonetischer Sicht erläutert. Dieses beruht auf der Ausnutzung normaler Variationsmöglichkeiten der menschlichen Stimme sowie der Kunstfertigkeit, die Lippen möglichst wenig zu bewegen und die Aufmerksamkeit des Publikums von der Beobachtung des Gesichts des Sprechers auf andere Objekte, z. B. eine bewegte Handpuppe, zu lenken.

Kaiserslautern, den 15.09.2006

Protokollführer: Dr. rer. nat. Jan Fritz Geiger

Rechtsanwalt und Diplomphysiker

Universität Saarbrücken und

RAe Scheidel & Scheidel, Kaiserslautern