

Forensische Sprechererkennung und Tonträgerauswertung in Praxis und Forschung

- Teil 1

Einleitung

Die forensische Sprechererkennung und Tonträgerauswertung gehört zu den weniger bekannten Disziplinen in der Kriminaltechnik, sie kann aber auf eine langjährige Geschichte zurückblicken und ist gerade gegenwärtig Gegenstand vieler neuer Erkenntnisse und Entwicklungen, die diesen Bereich voranbringen. Sie wird im Bundeskriminalamt seit den 1970er Jahren verwendet und hatte in der DDR eine Tradition, die sogar noch etwas weiter zurückreicht (Koristka 1968). Die forensische Sprechererkennung und Tonträgerauswertung ist behördlicherseits auch in einigen Landeskriminalämtern vertreten und zwar in Bayern, Nordrhein-Westfalen, Brandenburg (mit regel-mäßiger Amtshilfe für andere neue Bundesländer) und Berlin. Außerdem gibt es einzelne Privatsachverständige in diesem Bereich.² Die forensische Sprechererkennung und Tonträgerauswertung wird nicht isoliert in Deutschland angewendet, sondern ist Teil eines internationalen Netzwerkes von Aktivitäten. Eine wichtige Berufs- und Forschungsorganisation ist die „International Association for Forensic Phonetics and Acoustics,“ (siehe www.iafpa.net). Sie richtet jährlich eine Tagung aus, bringt zusammen mit der „International Association of Forensic Linguists,“ (in der es u.a. um Autorenerkennung geht) die Zeitschrift *International Journal of Speech, Language and the Law* heraus und veröffentlicht auf ihrer Website einen Code of Practice, in dem Richtlinien über ethisch korrektes und wissenschaftlich hochwertiges Verhalten als Sachverständiger für dieses Fachgebiet spezifiziert werden. Weiterhin wird auf europäischer Ebene im Rahmen der Arbeitsgruppe „Forensic Speech and Audio Analysis,“ der ENFSI (European Network of Forensic Science Institutes) an der Standardisierung und Weiterentwicklung des Faches gearbeitet. Seit dem Wintersemester 2007 wird an der Universität von York ein eigener fachrelevanter Studiengang „MSc Programme in Forensic Speech Science,“ angeboten (<http://www.york.ac.uk/depts/lang/postgrad/forensic.htm>). In Deutschland gibt es noch keinen solchen Studiengang, allerdings werden an einigen wenigen Universitäten Veranstaltungen in diesem Bereich angeboten.

Fachliche Unterteilung der Sprechererkennung und Tonträgerauswertung

Wie die Bezeichnung des Faches schon sagt, wird zunächst eine Unterscheidung getroffen zwischen Sprechererkennung und Tonträgerauswertung. Die **Sprechererkennung** ist traditionell das zentrale Anliegen in der Forschung und Praxis des Faches. Die Aufgaben der Sprechererkennung werden in Tabelle 1 zusammengefasst und in den folgenden Abschnitten genauer erläutert.

Die **Tonträgerauswertung** ist eine Zusammenstellung weiterer wichtiger Aufgaben, die regelmäßig angefragt werden. Zu ihnen zählen:

I Die **Qualitätsverbesserung**: Hier wird die Anhörbarkeit und Analysierbarkeit von Sprachaufzeichnungen mit Verfahren der digitalen Sprachsignalverarbeitung, z.B. durch Filterverfahren, verbessert. Die Erfolgsmöglichkeiten solcher Verfahren variieren. Qualitätsverbesserung ist insbesondere dann schwierig, wenn die Störungen ständigen Schwankungen unterworfen sind und wenn der Frequenzbereich der Störung mit dem von Sprache überlappt. Beides zusammen ist beispielsweise der Fall bei lauten Hintergrundstimmen oder Musik.

I Die **phonetische Textanalyse**: Hier geht es darum, den Wortlaut schwer verständlicher Sprachanteile zu verschriften. Im Gegensatz zur *Sprechererkennung* geht es hier also um *Sprach-*erkennung, d.h. die Erkennung des sprachlichen Inhaltes. Vielfach schließt sich die phonetische Textanalyse an eine Qualitätsverbesserung an, die beiden Aufgabenstellungen können aber auch unabhängig sein.

I Die **Authentisierung**: Hier wird geprüft, ob eine Aufzeichnung durch Hinzufügen, Löschen oder Verändern gegenüber dem Original manipuliert wurde oder ob andere Behauptungen zum Hergang der Aufnahme (z.B. verwendetes Aufzeichnungsgerät) korrekt sind. Traditionellerweise gehören zur Authentisierung der Einsatz von magneto-optischen Verfahren zur Visualisierung und Messung von Aufnahmespuren bei Tonbändern und Audiokassetten (siehe Boss et al. 2003). Allerdings ist diese Tätigkeit durch die wachsende Verbreitung von digitalen Aufzeichnungsverfahren zurückgegangen. Stattdessen gewinnen Methoden an Bedeutung, bei denen Diskontinuitäten im Signal, die auf Manipulation hinweisen, auf phonetischer oder linguistischer Grundlage erkannt werden. Dieses ist der Fall, wenn Brüche in der Aufnahme beobachtet werden, die durch die Prinzipien der Sprachproduktion oder der Grammatik nicht erklärbar sind.

I **Nicht-sprachliche akustische Untersuchungen**: Da Experten für Sprechererkennung und Tonträgeranalyse über die entsprechende Ausrüstung verfügen, Audioaufzeichnungen wiederzugeben und zu analysieren, werden sie manchmal gefragt, auch solche akustischen Ereignisse zu analysieren, die nichts mit Sprache zu tun haben. Zu solchen nichtsprachlichen Ereignissen zählen Schussgeräusche, Umgebungsgeräusche oder die Geschehnisse in einem Cockpit. Vielfach ist dabei die Zusammenarbeit mit anderen Experten erforderlich, z.B. mit einem Ornithologen oder Bioakustiker bei der Analyse von Vogelstimmen, die Hinweise über Ort (geografisch bzw. Stadt vs. Land) und Zeit (Tageszeit bzw. Jahreszeit) der Aufnahme geben können.

	Audio-Aufzeichnung des Täters	Keine Audio-Aufzeichnung, aber Zeuge
Verdächtiger existiert	Stimmenvergleich [falls Bereitschaft des Verdächtigen zur Vergleichsaufnahme oder Vergleichsaufnahme aus anderer Quelle]	Verdächtiger ist dem Zeugen bekannt: reguläre Zeugenaussage Verdächtiger ist dem Zeugen nicht bekannt: Stimmenwahlgegenüberstellung
kein Verdächtiger	Stimmenanalyse (=Stimmenprofil) und/oder Medienpräsentation	Kommt selten vor Mögliche Zukunft: „akustisches Phantombild“ durch Sprachsynthese

Tabelle 1: Unterschiedliche Aufgaben der forensischen Sprechererkennung

Stimmenanalyse

Es wird jetzt auf die einzelnen Gebiete der Sprechererkennung, die in Tabelle 1 zusammengefasst sind, eingegangen. Eine Stimmenanalyse wird typischerweise dann durchgeführt, wenn ein tatrelevanter anonymer Anruf aufgezeichnet wurde, wenn es aber keinen Tatverdächtigen gibt, der diesen Anruf getätigt haben könnte. Bei solchen Anrufen handelt es sich unter anderem um Erpressungen, Bedrohungen oder Stalking. Ziel der Stimmenanalyse ist es, den Ermittlern anhand der stimmlichen und sprachlichen Eigenschaften des anonymen Anrufers Hinweise zu geben über dessen biologische, soziale oder regionale Zuordnung oder über sprachliche Besonderheiten, die auch ohne sprachwissenschaftliche Ausbildung von den Ermittlern oder Personen aus der Bevölkerung (wieder)erkannt werden können. Wenn es auf diese Weise und in Zusammenhang mit anderen Ermittlungstätigkeiten gelingt, einen Verdächtigen zu finden, schließt sich in der Regel für die Beweisführung ein Stimmenvergleich an. Sind die Voraussetzungen für eine Stimmenanalyse gegeben, besteht grundsätzlich auch die Möglichkeit zu einer Medienpräsentation. Dabei wird die Tatraufzeichnung ganz oder in Ausschnitten (und ggf. mit Qualitätsverbesserung) einem breiten Publikum zugänglich gemacht, also über Rundfunk, Fernsehen oder Internet. Eine solche Medienpräsentation durch das BKA wurde z.B. im Fall der Reemtsma-Entführung durchgeführt. Je nach Dauer, Ergiebigkeit und Qualität der Tatraufzeichnung können im Rahmen einer Stimmenanalyse Aussagen getroffen werden über die folgenden Eigenschaften des Sprechers:

a. Geschlecht: Das Geschlecht eines Sprechers ist in der Regel eindeutig. Es gibt aber auch Grenzfälle, insbesondere dort, wo die Sprechstimmlage (wahrnehmbare Stimmtonhöhe) entweder nicht erkennbar ist (z.B. weil geflüstert wurde) oder genau in dem Bereich liegt, in dem sie uneindeutig ist, also für eine typische Männerstimme zu hoch oder eine typische Frauenstimme zu niedrig ist. Es kann auch sein, dass der Sprecher seine Stimme absichtlich so verstellt, dass das Geschlecht verschleiert werden soll.

Eine solche Situation lag vor bei einer Serie von Bombendrohungen im September 2003, in denen gefordert wurde, sämtliche Flugbewegungen am Düsseldorfer Flughafen zu stoppen. Diese Drohungen mussten angesichts des spezifischen Inhaltes erst genommen werden (u.a. wurde eine Mitgliedschaft zur Al Kaida behauptet), so dass der Flughafen tatsächlich für ungefähr sieben Stunden gesperrt wurde. Die Stimme des Anrufers war nicht eindeutig einem Mann oder einer Frau zuzuordnen. Es sprach jedoch das meiste dafür, dass es sich um eine Frau handelte, die ihre Stimme verstellte, indem sie die Lippen beim Sprechen rundete. Durch eine solche Lippenrundung kommt es zu einer Absenkung von Resonanzfrequenzen, die eine weibliche Stimme in die Nähe einer männlichen rückt. Es handelte sich tatsächlich um eine Frau, die später zugab, die Stilllegung des Flughafens nur gefordert zu haben, um nicht mit ihrem Freund in Urlaub fliegen zu müssen.

b. Alter: Grobeinteilungen in Kindheit, Jugend, (normales) Erwachsenenalter und Greisenalter sind anhand der Stimme relativ leicht möglich aber forensisch nur wenig relevant, da es sich in den meisten Fällen um Täter im Erwachsenenalter handelt. Unterteilungen innerhalb dieser Altersgruppe sind nur mit eingeschränkter Genauigkeit möglich, u.a. deswegen, weil verschiedene Personen mit ihrer Stimme unterschiedlich umgehen. Einige Personen achten schon berufsbedingt darauf, ihre Stimme „pfleglich“, zu behandeln (z.B. Lehrer, Sänger), während andere ihre Stimme durch Rauchen, Alkohol oder häufiges und ungeübtes Lautsprechen stark strapazieren, so dass die Stimme älter als erwartet klingt. Auch möglich sind unfreiwillige krankheitsbedingte Alterungsprozesse der Stimme. Man spricht in solchen Fällen auch davon, dass das „biologische Alter“, höher ist als das „chronologische Alter“, d.h. das tatsächliche Alter. Immerhin aber haben Untersuchungen gezeigt, dass eine Alterseinschätzung, in der ein Altersintervall von ungefähr 10 bis 15 Jahren angegeben wird, eine hohe Zuverlässigkeit hat. Solche Intervallangaben entsprechen dann auch der Praxis in der Stimmenanalyse.

c. Regionale Herkunft und Muttersprache: Ein wichtiges Sprechermerkmal ist seine regionalsprachliche Prägung, die er/sie ungefähr bis zum Ende des Schulbesuchs erfährt. Hierunter versteht man in erster Linie den *Dialekt*, mit dem eine Person aufwächst. Die Dialektbestimmung ist also ein wichtiges Element in der forensischen Stimmenanalyse. Dialekt im klassischen Sinn umfasst alle sprachlichen Ebenen, so neben der Aussprache auch die Wortstruktur, den Satzbau und den Wortschatz. Es gibt aber auch eine leichtere Form von dialektalem Einfluss, die sich vor allem auf die Aussprache beschränkt. Hierbei spricht man von „regionaler Umgangssprache“, oder „regional gefärbter Standardsprache“. Während es Kriminalfälle gibt, die Dialekt

im klassischen Sinn betreffen (z.B. bei gegenseitigen Bedrohungen von verfeindeten Parteien im lokalen Milieu), sind in vielen Fällen überregionaler und organisierter Verbrechen die leichteren Dialektformen beteiligt. In diesem Zusammenhang (z.B. im organisierten Drogenhandel) tritt häufig auch *fremdsprachiger Akzent* auf. Dieses sind sprachliche Einflüsse, die beim Zweitspracherwerb entstehen. Es ist nur mit Einschränkungen möglich, von den spezifischen Eigenschaften eines fremdsprachigen Akzents auf die Muttersprache zu schließen. Manchmal kann angesichts der eingeschränkten Quantität und Qualität des Materials nur eine grobe Schätzung angegeben werden, z.B. indem auf ganze Gruppen verwandter Sprachen verwiesen wird (z. B. slawische Sprachen) oder auf geographische Regionen (z. B. Sprachen im südosteuropäischen Raum). Merkmale von Dialekt und Zweitspracherwerb können auch zusammen auftreten, was die Menge der in Frage kommende Sprecher deutlich eingrenzt und deswegen ein wichtiger Ermittlungshinweis sein kann. Beispiele aus der Fallpraxis sind eine Kombination von griechischer Muttersprache mit schwäbischem Dialekteinfluss oder eine Kombination von russischer Muttersprache mit ostfälischem Dialekteinfluss.

d. Weitere Sprechereigenschaften: Es gibt weitere Kategorien von Sprechereigenschaften, die nur in einigen Fällen auftreten. In einigen Fällen liegen Auffälligkeiten vor, die auf *Sprach-, Sprech- oder Stimmstörungen* hinweisen. Wichtig für die Stimmenanalyse sind solche Störungen, die nicht temporär, sondern dauerhaft oder langfristig sind, also nicht etwa Veränderungen aufgrund einer Erkältung. Beispiele für relevante Störungen sind Stottern oder Sprachentwicklungsstörungen, die zur Fehlbildungen bestimmter Laute führen, wie das Lispeln (Fachbegriff ist Sigmatismus). Ein anderer Typus von Auffälligkeiten sind Beispiele von Fachsprache, die darauf hinweisen, dass sich ein Sprecher beruflich oder in anderer Weise intensiv mit einem bestimmten Fachthema auseinandergesetzt hat. Im Zusammenhang mit dem Thema Ausbildung und Beruf kann anhand der Eloquenz, grammatischen Komplexität und Wortwahl u.U. darauf geschlossen werden, wie hoch das Ausbildungsniveau des Sprechers ist und ob er es gewohnt ist, regelmäßig auf professionellem Niveau zu sprechen (Sprechberuf) oder nicht.

Die biologischen, sozialen oder regionalen Zuordnungen, die hier angesprochen wurden, werden auch als *Sprecherklassifikation* bezeichnet. In der Ermittlung können solche Sprecherklassifikationen direkt verwendet werden, d.h. auch ohne dass sich der Ermittler mit der Stimme an sich beschäftigt, beispielsweise indem die Tätersuche auf eine bestimmte Region konzentriert wird oder indem der soziale und berufliche Status von in Frage kommenden Personen überprüft wird.

Über solche sprecherklassifikatorischen Merkmale hinaus kann eine Stimmenanalyse auch auf **stimmlich/sprachliche Besonderheiten** aufmerksam machen, die zwar keine Sprecherklassifikation und damit keine stimmunabhängige Ermittlungsansätze erlauben, die aber allein durch ihre Auffälligkeit und relative Seltenheit bei der Ermittlung helfen können, indem diese Besonderheiten vom Ermittler oder anderen Personen an einer verdächtigen Person (wieder)erkannt werden. Solche Auffälligkeiten müssen natürlich vom sprachwissenschaftlichen Laien nachvollziehbar sein, denn sonst würden sie bei der Suche nach einem Verdächtigen wenig helfen. Beispiele solcher Auffälligkeiten sind:

- eine besonders hohe oder tiefe Stimme
- eine ungewöhnliche Stimmqualität wie z.B. starkes Knarren oder starke Behauchung in der Stimme
- eine besonders schnelle oder langsame Sprechweise

Außerdem sind auch verschiedene der oben genannten Sprecherklassifikationsmerkmale zumindest teilweise von Laien nachvollziehbar, was bei der Ermittlung helfen kann. Dieses gilt z.B. für das ungefähre stimmliche Alter, das Vorkommen eines Lispelns oder das Sprechen mit einem fremdsprachigen Akzent.

organisch	grammatisch	habituell
1. Sprechstimmlage	4. individuelle Aspekte von Dialekt, Soziolekt und Akzent	7. Sprechtempo
2. Vokaltraktresonanzen	5. linguistisch-phonetische Details	8. Melodik
3. Stimmqualität	6. forensisch-linguistische Merkmale	9. Pausenverhalten

Tabelle 2: Wichtige sprecherunterscheidende Merkmale bei Stimmenvergleichen

Stimmenvergleich

Existiert neben der Täteraufzeichnung auch ein Verdächtiger und ist dieser bereit, eine Stimmprobe abzugeben bzw. gibt es eine Vergleichsstimmprobe aus unabhängiger Quelle, kann ein Stimmenvergleich durchgeführt werden. Wenn entschieden wurde, eine Vergleichsaufnahme mit dem Verdächtigen zu erstellen, ist es sinnvoll, dieses von dem Sachverständigen durchführen zu lassen, der auch die stimmvergleichende Begutachtung durchführen wird, denn es sind eine Reihe von technischen und sprachwissenschaftlichen Aspekten zu bedenken.

Vereinfachend kann ein Stimmenvergleich als Durchführung und vergleichende Gegenüberstellung zweier Stimmenanalysen betrachtet werden. Beispielsweise ist eine Sprecherklassifikation nach Merkmalen wie dem Dialekt auch bei Stimmenvergleichen relevant. Auch ist das Konzept der stimmlich/sprachlichen Besonderheiten beim Stimmenvergleich wichtig.

Im Unterschied zur Stimmenanalyse ist man bei einem Stimmenvergleich allerdings nicht mehr an die Nachvollziehbarkeit durch sprachwissenschaftlichen Laien gebunden. Stattdessen können auch phonetische und linguistische Fachkonzepte herausgezogen werden oder es können akustische Messungen vorgenommen werden, die mehr erfassen, als durch bloßes Hören erkennbar ist. Aus diesem Grund ist auch der Katalog möglicher Sprecherunterscheidender Merkmale umfangreicher als in der Stimmenanalyse. Es ist sinnvoll, diesen Katalog genauer zu unterteilen. Künzel (1987) hat eine Unterteilung in Stimme – Sprache – Sprechweise vorgeschlagen. Eine vergleichbare, jedoch in einigen Aspekten revidierte Einteilung wird zusammen mit einigen häufig in der Sprechererkennung verwendeten Merkmalen in Tabelle 2 gezeigt (der Katalog ist also nicht komplett und es kommt außerdem noch die oben besprochene Sprecherklassifikation hinzu).

Mit „organisch“, sind solche Sprechermerkmale gemeint, die zumindest zu einem großen Teil von anatomischen Unterschieden zwischen Personen abhängen. **Sprechstimmlage** bezeichnet individuelle Unterschiede auf der Achse zwischen einer tiefen und einer hohen Stimme. Diese Unterschiede hängen zum Großteil von der Länge der Stimmlippen ab, d.h. dem Teil des Kehlkopfs, an dem die Stimme entsteht. Die Sprechstimmlage kann akustisch als durchschnittliche „Grundfrequenz“, gemessen werden. Diese entspricht der Schwingungsfrequenz der Stimmlippen pro Sekunde (gemessen in Hertz) und beträgt bei Männern im Mittel ca. 120 Hz. Die **Vokaltraktresonanzen** entstehen vor allem durch Unterschiede in der Vokaltraktlänge, d.h. dem Abstand zwischen den Stimmlippen und der Mundöffnung (durchschnittlich ca. 17,5 cm bei Männern). Diese Vokaltraktresonanzen sind in Form von „Formantenfrequenzen“, messbar, d.h. besonders energiereichen Schallanteilen, die mit Methoden der Spektralanalyse sichtbar gemacht werden können. In Abbildung 1 wird die Messung der Formantenfrequenzen illustriert.

Mit **Stimmqualität** werden bestimmte Schwingungsmuster der Stimmlippen bezeichnet, wie z.B. Knarren, Rauigkeit, Behauchung oder Gepresstheit. Auch werden mit Stimmqualität im weiteren Sinne besondere Einflüsse des Vokaltrakts bezeichnet, wie die Nasalität (Luftstrom durch den Nasenraum aufgrund von Absenkung des sog. Gaumensegels). Einen Überblick über die Stimmqualitätsanalyse in der Forensik liefern Köster und Köster (2004).

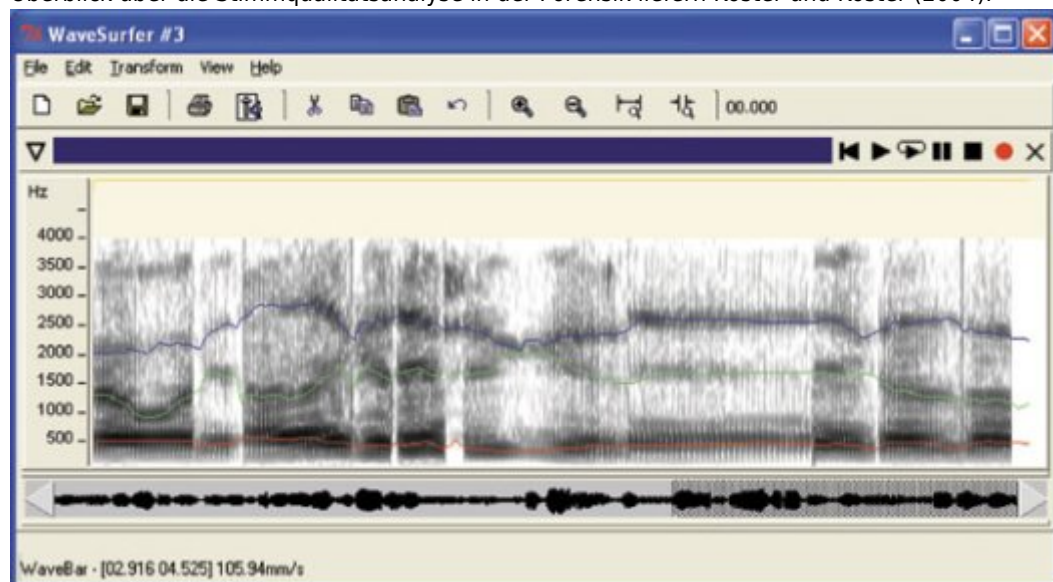


Abbildung 1: Illustration der Formantenanalyse. In der oberen Signaldarstellung, dem Spektrogramm, zeigen die LPC-Spuren (Linear Predictive Coding) die Formantenfrequenzen an, und zwar den ersten Formanten „F1“ (untere Spur, in rot), den zweiten Formanten „F2“ (mittlere Spur, in grün) und den dritten Formanten „F3“ (obere Spur, in blau). In der kleinen unteren Signaldarstellung wird das Oszillogramm der ganzen Aufnahme im Zeitverlauf gezeigt (hier: Zusammenschnitt aller Laute mit messbarer Formantenstruktur). Für die Analyse werden alle einzelnen Werte der drei Formanten, die im Zeitverlauf entstehen, gemittelt.

Die Bezeichnung „grammatisch“, in der zweiten Spalte von Tabelle 2 erfasst Unterschiede zwischen Sprechern, in denen die Variationsmöglichkeiten und Freiheitsgrade zum Tragen kommen, die bei der Anwendung des linguistischen Regelwerks einer Sprache (der Fachbegriff dafür ist Grammatik) offen bleiben. Bei den **individuellen Aspekten von Dialekt, Soziolekt und Akzent** geht es nicht allein um die regionalen u.a. Zuordnungen, die im Rahmen einer Stimmenanalyse untersucht werden, sondern es geht darüber hinaus um Unterschiede, die auch innerhalb von Dialekten und anderen Sprecherklassen möglich sind. Beispielsweise können sich auch innerhalb eines Dialektes Personen darin unterscheiden, welche spezifischen Dialektmerkmale sie verwenden und welche sie nicht verwenden. Auch können sich verschiedene Sprecher, die den gleichen Dialekt sprechen, allgemein darin unterscheiden wie stark ihre dialektale Ausprägung ist. Weitere deutlich individualisierende Muster können sich durch die spezifische Kombination von Dialekt, Soziolekt und Akzent ergeben, wie in dem o.g. genannten Beispiel eines russischen Akzents kombiniert mit ostfälischem Dialekt. **Linguistisch-phonetische** Details beziehen sich auf Aussprachemerkmale, die zu detailliert sind, als dass sie von der Grammatik der Aussprache (der sog. Phonologie) erfasst werden. Ein Beispiel ist die Einfügung eines kleinen Plosivlautes in Wörtern mit der Lautsequenz „ns“, oder „ls“, wie z.B. „Hans“, oder „Hals“. Einige Sprecher fügen hier ein „t“, ein, was als „Hants“, oder „Halts“, geschrieben werden kann (Jessen 2008). Dieses Aussprachemerkmal ist so unscheinbar, dass es kaum gehört werden kann. Es kann aber durch akustische Messungen nachgewiesen werden. **Forensisch-linguistische Merkmale** sind sprecherunterscheidende Muster, die über die reine

Aussprache hinausgehen. Diese sind Eigenschaften aus Bereichen wie Morphologie (Wortstruktur), Syntax (Satzbau) und Lexik (Wortschatz, Wortwahl). Solche und andere Bereiche sind auch bei der Autorenerkennung relevant, welche ein wichtiges Gebiet der „Forensischen Linguistik„ ist. Lexik im weiteren Sinn umfasst auch die Phraseologie. Forensisch relevant sind dabei insbesondere Redefloskeln wie „nicht wahr,,“, die von einigen Sprechern verwendet werden, oft ohne dass sie es bemerken.

(Fortsetzung folgt)