

## Untersuchung usbekischer Sprachsignale im erweiterten Frequenzbereich

Bernd Wiedecke<sup>1</sup>, Mirabbos Payasov<sup>2</sup>, Visola Hamrayeva<sup>3</sup>

<sup>1</sup> *Taschkenter Universität für Informationstechnologien, Usbekistan, E-Mail: drwdk@web.de*

<sup>2</sup> *Taschkenter Universität für Informationstechnologien, Usbekistan, E-Mail: mpayasov@mail.ru*

<sup>3</sup> *Taschkenter Institut für Pädiatrie, Usbekistan, E-Mail: jaloliddin75@mail.ru*

### Charakteristik der Ausgangssituation

Sprachsignale haben eine sehr komplexe Struktur, was dazu führt, dass im Prozess ihrer Identifizierung und Erkennung Signalabweichungen oder das völlige Fehlen einer der Komponenten, fatale Folgen hat. Für gewöhnlich ordnet man Sprachsignale dem Frequenzbereich von 50 bis 5.000 Hz zu. Was in der Regel für romanische und auch slawische Sprachen durchaus zutreffend ist, führte aber bei der usbekische Sprache, die aufgrund ihrer gutturalen Artikulation als Signale mit tiefer Frequenz wahrgenommen wird, zu völlig fehlerhaften Schlussfolgerungen. Mit den nachstehenden Untersuchungen wurde einwandfrei nachgewiesen, dass es sich bei der usbekischen Sprache eindeutig um Signale handelt, die überwiegend höheren Frequenzen (5.000 – 17.000 Hz) zuzuordnen sind.

Entsprechend der oben genannten allgemeinen Annahmen über den Frequenzbereich von Sprachsignalen, wurden in der Vergangenheit in Usbekistan audiometrische Untersuchungen auch nur in diesem Frequenzbereich durchgeführt. Da hierbei logischerweise keine zufriedenstellenden Ergebnisse erzielt wurden, sowie Hörschäden, wenn überhaupt, so nur fehlerhaft diagnostiziert werden konnten, führten audiometrische Untersuchungen über einen längeren Zeitpunkt hinweg nur ein Schattendasein. Darüber hinaus nahm aufgrund der Besonderheiten der usbekischen Sprache die Mehrzahl der Patienten Hörschäden nur begrenzt wahr, d.h. konnte nicht eigenständig die Notwendigkeit eines Facharztbesuches definieren.

Über einen längeren Zeitraum führten die Autoren anhand von Studioaufzeichnungen mit Sprechern unterschiedlichen Geschlechts verschiedener Altersgruppen entsprechende Untersuchungen mit Hilfe des Programms „Praat“ ([www.praat.org](http://www.praat.org)), welches für die Untersuchungen die besten instrumentalen Voraussetzungen mit sich brachte, durch.

Ebenso wie die deutsche hat auch die usbekische Sprache mehrere regional geprägte Dialekte, wobei diese in ihrer Vielfalt und Heterogenität gleichermaßen zu charakterisieren sind.

### Besonderheiten der usbekischen Sprache

Wie bereits oben dargelegt, sind die Komponenten der usbekischen Sprachsignale wesentlich höheren als allgemein angenommenen Frequenzen zuzuordnen. Dem allgemein üblichen Frequenzbereich von 50 bis 5.000 Hz entspricht nur mit vielen Vorbehalten in einigen Fällen der Grundfrequenz. Und das auch nur in Fällen, wenn Laute, die den romanischen und slawischen Sprachen nahestehen, artikuliert werden.

Wenn man die Grundfrequenz des erwachsenen männlichen Sprechers, die aus der Gesamtheit der Sprecher natürlich die tiefste Lage hatte, als Nullpunkt zum Vergleich ansetzt, ergeben sich in der usbekischen Sprache folgende Abstufungen der Stimmlagen zwischen den Sprechern nach Alter und Geschlecht (Tabelle 1):

**Tabelle 1:** Abstufung der Stimmlagen

Sprecher	Abstufung, Hz
Männlicher Erwachsener	Nullpunkt
Weiblicher Erwachsene	+ 25 – 30 Hz
Männlicher Teen	+ 30 – 35 Hz
Weiblicher Teen	+ 40 – 50 Hz
Männliches Kind	+ 95 – 105 Hz
Weibliches Kind	+ 120 – 135 Hz

In der Überwiegenden Mehrheit der Sprachsignale erreichte die Grundfrequenz Werte, die weit über 5.000 Hz lagen. Z.B. bei der Artikulation sprachtypischer Vokale, stimmhafter Konsonanten und Zischlauten wurden Werte der Grundfrequenz zwischen 5.000 und 7.000 Hz fixiert. Bei stimmlosen Konsonanten erreichte die Grundfrequenz sogar Werte zwischen 14.000 und 16.000 Hz.

Bei der Analyse der Formanten wurde festgestellt, dass lediglich der erste Formant teilweise (überwiegend nur in den minimalen Werten des Signals) in den Bereich unterhalb von 5.000 Hz fällt. Spitzenwerte der ersten Formanten unterhalb von 5.000 Hz sind die Ausnahme, und nur präsent, insoweit den romanischen und slawischen Sprachen ähnliche Signale produziert werden. Die Maximalwerte des ersten Formants erreichten in vielen Fällen Werte von bis zu 7.000 Hz und unterscheiden sich damit erheblich von den allgemein angenommenen physiologischen Eigenschaften der Sprecher in den einzelnen Altersgruppen.

Der zweite Formant lag ebenfalls nur in den Minimalwerten und im Mittel noch unterhalb 5.000 Hz. Die Maximalwerte im Signalverlauf erreichten Werte von 10.000 bis 11.000 Hz. Wobei die beiden ersten Formanten sich bei allen Signalen und allen Sprechern als die dynamischsten erwiesen haben. Die in der usbekischen Sprache vergleichsweise weit über den Werten romanischer und slawischer Sprachen liegenden Werte stellen hierbei keine der dort vorkommenden Pegelanhebung mit Auswirkung auf die Klangempfindung dar. Die usbekische Sprache wird generell als Signal mit niedriger Klangempfindung wahrgenommen.

Die dritten und vierten Formanten lagen durchweg in Bereichen mit hohen Frequenzen. So wurde der dritte Formant durch Werte zwischen 5.500 Hz (Minimalwert) und 13.600 Hz (Maximalwert) charakterisiert. Der vierte Formant schwankte zwischen Werten von 8.500 Hz (Minimalwert) und 17.500 Hz (Maximalwert). Wobei beim

Vergleich analoger Signale zwischen den einzelnen Sprechern nur völlig unerhebliche minimale Differenzen festgestellt werden konnten. Die allgemein übliche Annahme, dass die dritten und vierten Formanten die Anatomie des Sprechers und sein Timbre charakterisieren, fanden wir somit in der usbekischen Sprache nicht bestätigt.

Bei der Produktion der Studioaufzeichnungen waren die einzelnen Sprecher angehalten, mit normaler Lautstärke und normalem Tempo zu sprechen. Damit sollte der Einfluss der allgemein bekannten Pegelanhebung bei hoher Intensität und Pegelsenkung bei niedriger Intensität weitestgehend in der Analyse unberücksichtigt bleiben. Bei der Analyse der Sprachsignale wurden bei jedem der Sprecher jedoch bei der Artikulation der Signale Schwankungen der Intensität mit Werten zwischen 39 und 90 dB festgestellt (das Signal wurde für die Analyse geglättet und verstärkt). Eine entsprechende Absenkung der Intensität, begleitet von einem sprunghaften Anstieg der Grundfrequenz war typisch für die Artikulation von Konsonanten und Zischlauten.

Im Rahmen der Untersuchungen wurde den Zischlauten besondere Aufmerksamkeit geschenkt. Dazu wurden aus den aufgezeichneten Signalen Die einzelnen Zischlaute als Phon extrahiert.

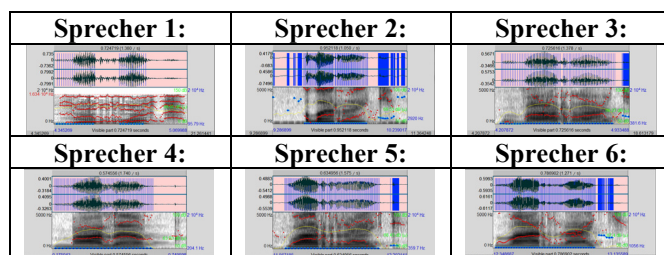
Zischlaute werden in der Regel bei der Identifikation und Erkennung von Sprachsignalen meistens als Störgeräusch wahrgenommen und dementsprechend eingeordnet. Im Rahmen der durchgeführten Untersuchungen wurde jedoch einwandfrei festgestellt, dass es sich bei den Zischlauten um ein Signal mit hohen Frequenzen handelt. So sind Werte der Grundfrequenz zwischen 5.000 und 6.000 Hz für Zischlaute charakteristisch, wobei die ersten beiden Formanten mit Frequenzen stets unter der Grundfrequenz ( $f_1 = 1.000 - 4.500$  Hz;  $f_2 = 4.700 - 5.2000$  Hz) liegen. Die dritten und vierten Formanten lagen weit über der Grundfrequenz in den bereits oben genannten Frequenzbereichen. Charakteristisch für Zischlaute war ebenfalls der Umstand, dass die Grundfrequenz bei der Artikulation von Zischlauten nahezu unverändert stabil blieb, während alle Formanten während der Signaldauer dynamisch durch phasenversetzte gedämpfte Schwingungen beschrieben wurden.

### Praktische Anwendungen

Die oben genannten Analyseergebnisse wurden im Weiteren dazu verwendet, um für die usbekische Sprache entsprechend nach Altersgruppen differenzierte Sprachtests für die Pädiatrie zu entwickeln. Die praktische Approbation der entwickelten Sprachtest ist derzeit bereits abgeschlossen und hat die Zweckmäßigkeit zur Reanimation der audiometrischer Untersuchungen nachgewiesen.

Im Verlauf der Untersuchungen der Sprachsignale wurden ebenfalls Vergleiche der Messdaten gleicher Sprachsignale zwischen verschiedenen Sprechern vorgenommen. Dabei wurde festgestellt, dass im Wesentlichen, unabhängig von Alter und Geschlecht der Sprecher, wie am Beispiel anhand der grafischen Visualisierung in Abbildung 1 dargestellt, im wesentlichen ein so gut wie identischer Verlauf des Signals mit fast übereinstimmender Strukturierung der einzelnen Komponenten des Signals vorliegt. Die Dehnung und

Stauchung der Signale auf der Zeitachse entstand durch nicht ganz exakte Synchronisation der Sprechgeschwindigkeit der einzelnen Sprecher, kann aber in diesem Fall vernachlässigt werden.



**Abbildung 1:** Grafische Visualisierung eines Signals mit unterschiedlichen Sprechern (Sprecher1 – männlicher Erwachsener, Sprecher 2 – weibliche Erwachsene, Sprecher 3 – männlicher Teen, Sprecher 4 – weiblicher Teen, Sprecher 5 – männliches Kind, Sprecher 6 – weibliches Kind).

Ausgehend von der einheitlichen Strukturierung der Signale werden derzeit die Untersuchungen bezüglich der Nutzung von Methoden der Mustererkennung zur Identifizierung und Erkennung von Sprachsignalen fortgesetzt. Obwohl die Untersuchungen sich noch in einem sehr frühen Stadium befinden, ist es nicht auszuschließen, dass diese Verfahren für bestimmte Anwendungsgebiete durchaus geeignet sind.

### Literatur

- [1] Jörg Mayer: Linguistische Phonetik. Universität Stuttgart, 2010.
- [2] Jörg Mayer: Phonetische Analysen mit Praat. Ein Handbuch für Ein- und Umsteiger. Universität Stuttgart, 2009.
- [3] Praat: doing phonetics by computer, URL: <http://www.fon.hum.uva.nl/praat/>
- [4] Nina Berend, Stefan Kleiner, Ralf Knöbl: Sprachliche Variabilität des Deutschen und ihre Erfassung mit Methoden der automatischen Spracherkennung. Institut für Deutsche Sprache Mannheim, 2005.
- [5] Б. Видеке, В. Хамраева, М. Паязов: Анализ компонентов речевых аудио-сигналов в расширенном диапазоне частот. – Ташкент: Вестник ТУИТ, № 1, 2010.
- [6] Б. Видеке, М. Паязов, В. Хамраева: Шипящие сигналы в процессе идентификации и распознавания речи. – Ташкент: Вестник ТУИТ, № 3, 2010.