

Modellhafte Nachbildung und Analyse der Ersatzlaute eines Bauchredners

Jörg Metzner, Marcel Schmittfull, Karl Schnell*

* *Institut für Angewandte Physik, Goethe-Universität Frankfurt am Main*
 Email: {J.Metzner, M.Schmittfull}@gmx.net, Schnell@iap.uni-frankfurt.de

Einleitung

Beim Bauchreden wird das Fehlen der Lippenbewegungen durch eine geschickte Stellung des Vokaltraktes ersetzt. Besondere Schwierigkeiten ergeben sich bei den labialen Lauten, die durch Substitutionslaute ersetzt werden müssen. In diesem Beitrag werden die Substitutionslaute modellhaft simuliert und mit den Analysen von Sprachaufnahmen eines professionellen Bauchredners qualitativ verglichen. Die Untersuchungen der realen wie modellgenerierten Substitutionslaute zeigen, dass insbesondere nur der für das Sprachverstehen entscheidende untere Frequenzbereich approximiert wird.

Die Substitutionslaute

Für das Bauchreden stellen die Laute mit Lippenbewegung [b], [p], [f], [v] und [m] eine besondere Schwierigkeit dar und müssen ersetzt werden. Die Substitutionslaute müssen den Originallauten perzeptiv möglichst nahe kommen unter der Nebenbedingung der festen Lippen- und Kieferstellung.

Für die modellhafte Generierung wird das artikulatorische Syntheseprogramm Tractsyn [1] von Peter Birkholz verwendet. In Tractsyn können die Artikulatoren und ihre Bewegungen für die Substitutionslaute eingestellt werden. Mit Hilfe der berechneten Übertragungsfunktionen des Programms können die Formanten ermittelt werden. Weiterhin können synthetisierte Sprachsignale erzeugt werden, deren Ergebnisse auditiv überprüfbar sind.

Modellgestützte Untersuchungen

Die generierten Lautfolgen [ba] und [b'a] im Vergleich

Abbildung 1a zeigt die Artikulatorstellung des Plosivs [b] mit geschlossenen Lippen. Die resultierende Übertragungsfunktion des Sprechtraktes ist in Abbildung 1c zu sehen. Ausgehend von Literaturangaben [2] wurden die Artikulatoren für den Substitutionslaut [b'] mit geöffneten Lippen eingestellt, wie in Abbildung 1b dargestellt. Die Lippenschließung wird durch das Pressen der Zunge an die Vorderzähne ersetzt. Die korrespondierende Übertragungsfunktion des Substitutionslautes ist in 1d zu sehen. Es ist deutlich zu sehen, dass für die beiden Vokaltraktkonfigurationen 1a und 1b die ersten beiden Formanten fast identisch sind. Im Gegensatz dazu unterscheiden sich die höheren Formanten beträchtlich. Für eine auditive Überprüfung wurde ein Lautübergang zum Vokal [a] generiert. Die Bilder 2a und 2b zeigen die Spektrogramme der mit Tractsyn synthetisierten Sprachsignale. Die Formantabbiegungen im unteren Frequenzbereich, die für die Lauterkennung entscheidend sind, stimmen überein. Die Formantabbiegungen im oberen Frequenzbereich weichen beträchtlich voneinander ab. Eine auditive Überprüfung

zeigt, dass der Plosiv in der erzeugten Sequenz [b'a] als labialer Laut wahrgenommen wird.

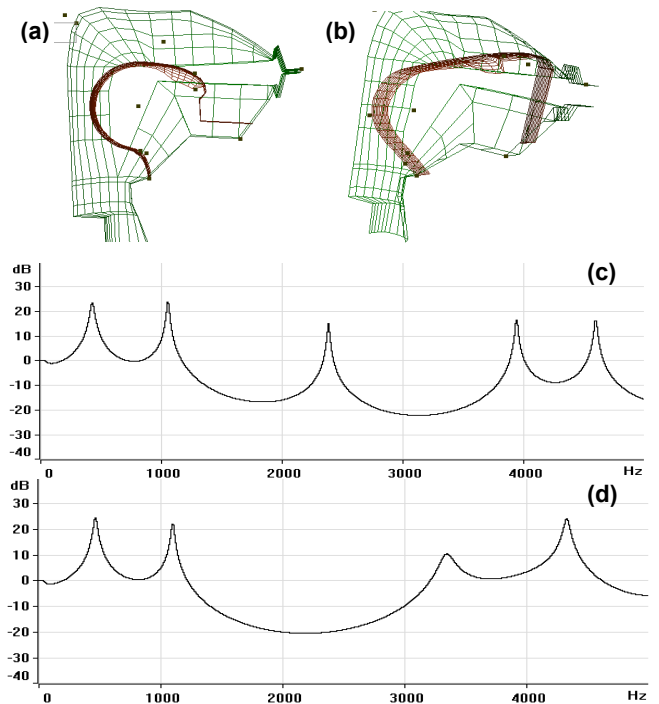


Abbildung 1: Vokaltraktstellungen von [b] in Bild (a) und [b'] in Bild (b); Betragsgänge der Vokaltraktstellungen von [b] in Bild (c) und [b'] in Bild (d).

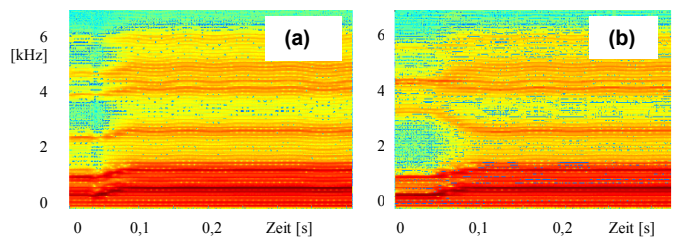


Abbildung 2: Spektrogramme der synthetisierten Lautfolgen [ba] in Bild (a) und [ba'] in Bild (b).

Die Nasale [m], [m'] und [m'']

Der Nasallaut [m] weist wie der Plosiv [b] bzw. [p] eine labiale Verengung im Mundraum auf, wie in der Vokaltrakteinstellung in Bild 3a zusehen ist. Der Nasal [m] kann auf zwei Arten substituiert werden. Die Vokaltraktstellungen der beiden Substitutionslaute [m'] und [m''] sind in den Bildern 3b und 3c zu sehen. [m'] weist einen Verschluss kurz hinter bzw. an den Zähnen auf, während der Substitutionslaut [m''] den Mundraum fast vollständig am Velum verschließt. Die Betragsgänge der Substitutionslaute (Bilder 3e,3f) sind mit dem von [m] (Bild 3d) bis ca. 1 kHz vergleichbar; was mit [m'] allerdings besser gelingt als mit [m''].

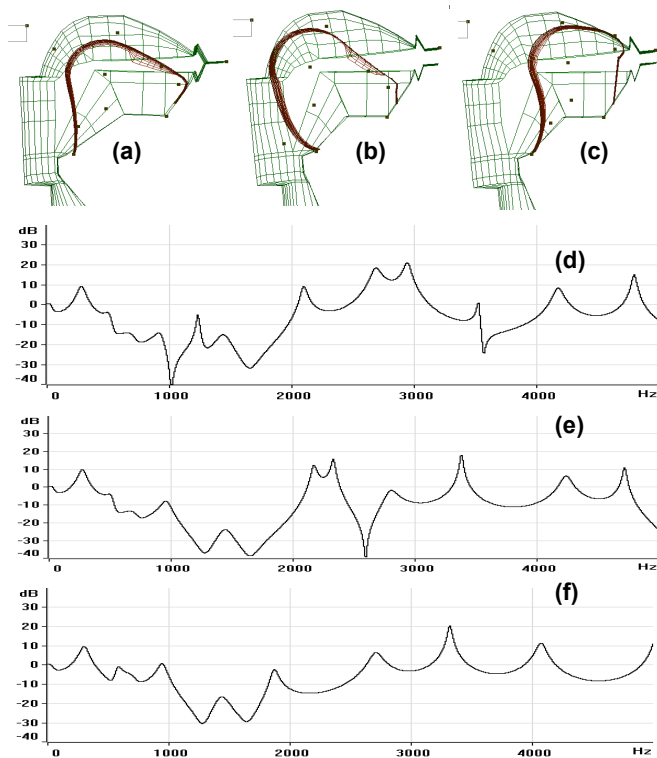


Abbildung 3: Vokaltraktstellungen von [m] in Bild (a), [m'] in Bild (b) und [m''] in Bild (c); Betragsgänge der Vokaltraktstellungen von [m] in Bild (d), [m'] in Bild (e) und [m''] in Bild (f).

Reales Modell für Nasale

Die Akustik der nasalen Substitutionslaute wird auch durch ein reales Modell untersucht. Dafür wurde ein Hohlraumgebilde aus Gips angefertigt, das die Geometrie des Sprechtraktes für die Nasale wiedergibt. Der Mundraum wird für die Modellierung der Substitutionslaute entsprechend verkürzt. Mit einer geeigneten Anregung des Gipsmodells für [m], [m'] und [m''] können Schallsignale erzeugt werden, welche perzeptiv mit dem Nasal [m] vergleichbar sind.

Originalaufnahmen eines Bauchredners

Anhand von Originalaufnahmen lassen sich die modellhaft gewonnenen Ergebnisse überprüfen. Es ist nicht auszuschließen, dass ein Bauchredner es schafft die Ersatzlaute durch Tricks, die nicht in der Simulation dargestellt werden können, noch weiter zu perfektionieren. Bauchredner können zum Beispiel kleine Grübchen in ihren Backen bilden oder ihre Zunge schief halten. Für die Aufnahmen hat ein professioneller Bauchredner eine Sequenz von Lautfolgen gesprochen, welche abwechselnd normal und mit der Bauchredetechnik artikuliert wurden. Abbildung 4 zeigt zeitlich verschobene Frequenzanalysen der Lautfolge [ba] normal und [b'a] mit Bauchredetechnik gesprochen. Die Frequenzkurven stellen Modellbetragsgänge dar, um den Einfluss der Grundfrequenz zu eliminieren. Der Verlauf der Resonanzen ist mit vertikalen Linien hervorgehoben. Es ist zu erkennen, dass sich die Formantverläufe bis ca. 3-4 kHz zwischen den normalen und in Bauchredetechnik gesprochenen Äußerungen ähneln, während darüber eine geringe bis gar keine

Übereinstimmung zu erkennen ist. Diese Ergebnisse stimmen mit den modellhaften Untersuchungen insofern überein, dass der untere Frequenzbereich durch die Bauchredetechnik im Gegensatz zum oberen Frequenzbereich gut nachgebildet wird. Da die Approximation im relevanten Bereich der Spracherkennung vollzogen wird, stellen die Substitutionslaute perzeptiv einen Ersatz für die nachzubildenden Laute dar.

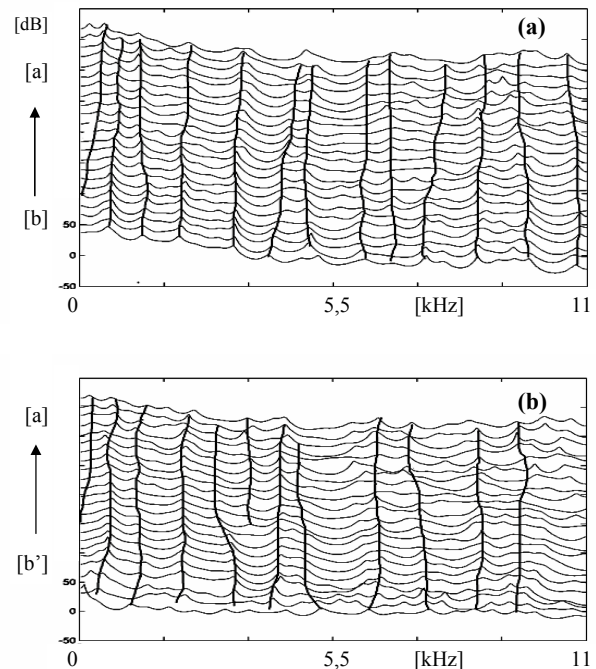


Abbildung 4: Frequenzanalysen der Äußerungen [ba] in Bild (a) und [b'a] in Bild (b) eines Bauchredners.

Zusammenfassung

Die Bauchredetechnik ist modellhaft und mittels Originalaufnahmen untersucht worden. Dabei hat sich gezeigt, dass die Einschränkungen der Lippen- und Kieferbewegungen bei den Substitutionslauten eine vernachlässigte Approximation der höheren Spektralbereiche zur Folge hat. Da die spektrale Nachbildung vorzugsweise im relevanten unteren Spektralbereich erfolgt, ist die Illusion des Bauchredens möglich.

Die Autoren danken Patrick Martin für das Sprechen der Sprachproben und Peter Birkholz für die Bereitstellung des Programms Traktsyn.

Literatur

- [1] Birkholz, Peter: „Entwicklung eines dreidimensionalen Artikulatormodells für die Sprachsynthese“, http://www.icg.informatik.uni-rostock.de/~piet/speak_main.html (2002).
- [2] Vox, Valentine: „I can see your lips moving“, Retonios Magic, Casino, Schweiz.
- [3] B. Story, I. Titze: „Vocal tract area functions from magnetic resonance imaging“, J.A.S.A. Vol. 100, pp. 537-554, (1996).