

Beziehung zwischen Mund- und Nasensignalen bei nasalisierten Vokalen

Karl Schnell, Arild Lacroix

Institut für Angewandte Physik, Goethe-Universität Frankfurt,
Max-von-Laue-Straße 1, D-60438 Frankfurt am Main
Email: {schnell, lacroix}@iap.uni-frankfurt.de

Einleitung

Bei der Artikulation von nasalisierten Vokalen werden Schallsignale von der Mundöffnung und den Nasenlöchern gleichzeitig abgestrahlt. Durch eine Superposition der beiden Schallsignale entsteht das Sprachsignal der nasalisierten Laute. Die individuellen Eigenschaften der einzelnen Signale und deren Beziehung lassen sich anhand des gemischten Sprachsignals nicht mehr eindeutig identifizieren. Daher werden in diesem Beitrag die spektralen Eigenschaften von getrennten Mund- und Nasensignalen in Bezug auf die Vokalnalisierung untersucht. Die Mund- und Nasensignale wurden dafür durch eine spezielle Vorrichtung akustisch getrennt aufgezeichnet. Mit Hilfe der getrennt aufgenommenen Signale werden Spektralanalysen und perzeptive Tests durchgeführt.

Getrennte Mund- und Nasensignale

Die Analysen der Nasallaute und des Nasaltrakteinflusses werden in der Regel mit Hilfe von akustischen Einzelaufnahmen, gewöhnlichen Sprachaufnahmen und Modellrechnungen durchgeführt [1-4]. Das Vorliegen von getrennten Mund- und Nasensignalen eröffnet darüber hinaus Möglichkeiten, das Sprachsignal der Nasalvokale explizit als Zusammensetzung von zwei Signalen zu untersuchen. Die Trennung von Mund- und Nasensignalen kann durch eine räumliche Vorrichtung akustisch vollzogen werden [5-7]. Für die hier gezeigten Ergebnisse wurde der Kopf des Sprechers in eine Dämmplatte eingelassen, die an die Wände einer gedämpften Kammer anschließen. Durch diese Maßnahme strahlen Mund und Nase ihre Schallsignale während der Artikulation in zwei akustisch getrennte Räume. In diesen Räumen ist jeweils ein Mikrofon für die Signalerfassung platziert, die größenordnungsmäßig einen Abstand von einem halben Meter zum Kopf des Sprechers aufweisen. Die Güte der Trennung kann durch das Verhältnis der Leistungen der aufgezeichneten Mund- und Nasensignale bei der Artikulation von Nasalen und unnasalisierten Vokalen abgeschätzt und kontrolliert werden, da in diesen Fällen ein Kanal eine dominante Lautstärke aufweisen muss. Hierbei ist zu beachten, dass selbst bei geschlossenem Velum der Nasaltrakt akustisch nicht vollständig abgekoppelt ist. Abb. 1 zeigt die Signale und Spektren der getrennt aufgenommenen Nasen- und Mundsignale des nasalisierten Vokals /ã/. Das Spektrum S_n des Nasensignals weist eine breitere Einkerbung bzw. Nullstelle zwischen 1 und 2 kHz auf, wie in Abb. 1(c) zu erkennen ist. Die Nasensignale der Nasalvokale klingen relativ dumpf, und aus ihnen kann in der Regel der Vokal

nicht mehr eindeutig erkannt werden. Das Zusammenwirken des Mundsignals s_m und des Nasensignals s_n kann durch unterschiedliche Linearkombinationen s_{nm} perzeptiv bewertet werden:

$$s_{nm}(k) = (1 - \lambda) \cdot s_m(k) + \lambda \cdot s_n(k - \tau) \quad (1)$$

mit $0 < \lambda < 1$.

Hörtests zeigen auf, dass sich Variationen der Zeitverschiebung τ so gut wie gar nicht auf den Höreindruck der gemischten Signale auswirken. Mit dem Mischparameter λ kann der Grad der Nasalierung eingestellt werden. Die Wahrnehmung der Nasalierung ist dabei nichtlinear mit dem Parameter λ verknüpft. Das Mund oder Nasensignal kann auch durch das Mund- bzw. Nasensignal einer anderen Realisierung derselben Äußerung ersetzt werden. Auch damit ist ein vergleichbarer Höreindruck erzielbar.

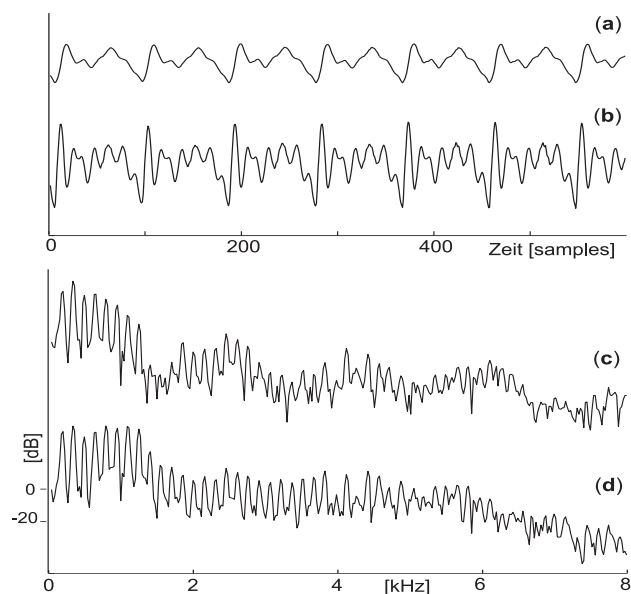


Abbildung 1: (a) Nasensignal s_n und (b) Mundsignal s_m des nasalisierten Vokals /ã/, sowie dazugehörige Spektren (c) S_n und (d) S_m .

Verhältnis von Mund- und Nasenspektrum

Da Mund- und Nasensignale an unterschiedlichen Öffnungsstellen des Sprechtrakts abgestrahlt werden, weisen sie eine bestimmte Beziehung untereinander auf, da sie vom selben Hohlraumssystem gebildet werden. Der Nasaltrakt und Mundraum sind dabei in ihrer Funktion als Seitenhöhle bzw. Hauptzweig mit Systemausgang miteinander vertauscht. Im Gegensatz zum Mundraum stellt der Nasaltrakt selbst ein

verzweigtes Rohrsystem mit Seitenhöhlen dar. Die Seitenhöhlen bestimmen bekanntlich Null- und Polstellen in der Übertragungsfunktion. Somit ist die Übertragungsfunktion Glottis zu Lippen H_M bzw. zu den Nasenlöchern H_N als Pol- Nullstellensystem anzunehmen. Nun ergibt sich das spektrale Verhältnis $R(\omega)$ zwischen Nasensignal zu Mundsignal theoretisch aus dem Verhältnis der Übertragungsfunktion zu den Nasenlöchern und dem zu den Lippen H_N/H_M . Folglich ist das Verhältnis R der Spektren S_n und S_m wiederum als Pol-Nullstellensystem H_{NdM} anzunehmen mit

$$R(\omega) = \frac{S_n(\omega)}{S_m(\omega)} \approx H_{NdM}(e^{j\omega}) = \frac{\sum_{k=0}^M b_k e^{-j\omega k}}{1 + \sum_{k=1}^N a_k e^{-j\omega k}} \quad (2)$$

Das spektrale Verhältnis $R(\omega)$ kann aus Signalabschnitten von s_n und s_m gebildet. Für die Bestimmung von H_{NdM} wird zunächst das spektrale Verhältnis $R(\omega)$ mit der inversen diskreten Fourier-Transformation in ein Zeitsignal $r(n)$ transformiert. Aus diesem Signal wird dann mit einer ARMA Schätzung das System H_{NdM} bestimmt. Dazu wird das Prony Verfahren [8] mit einer Vorverarbeitung verwendet. Das Prony Verfahren schätzt aus Impulsantworten die Zähler- und Nennerkoeffizienten der Übertragungsfunktion eines rekursiven Systems. Da anzunehmen ist, dass das Signal $r(n)$ keine Impulsantwort darstellt, wird aus diesem eine Impulsantwort $h_r(n)$ geschätzt. Dafür wird $r(n)$ mit einer linearen Prädiktion hoher Ordnung analysiert. Das resultierende Nur-Pole System wird dann mit einem Impuls angeregt, wodurch sich die geschätzte Impulsantwort $h_r(n)$ ergibt. Aus dem Signal $h_r(n)$ werden dann mittels des Prony Verfahrens die Pole und Nullstellen des Systems H_{NdM} geschätzt. Abb. 2 zeigt

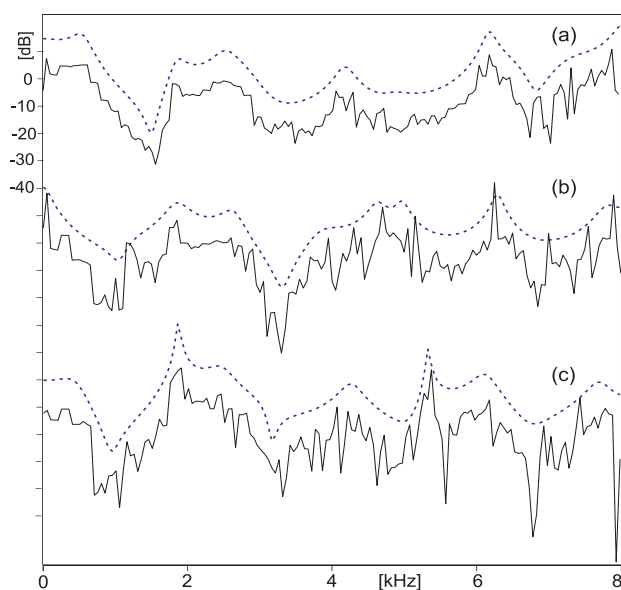


Abbildung 2: Spektrale Verhältnisse $|R(\omega)|$ (schwarze durchgezogene Linie) und Modellspektren $|H_{NdM}|$ (unterbrochene blaue Linie) für (a) /ã/, (b) /õ/ und (c) /ü/.

für die hinteren nasalierten Vokale /ã/, /õ/ und /ü/ die Beträge der spektralen Verhältnisse $|R(\omega)|$ sowie die Betragsgänge der geschätzten Modellfunktionen $|H_{NdM}|$ für die Ordnung $M = 10$ und $N = 20$. Die Modellbetragsgänge weisen eine Ähnlichkeit in ihrer groben Struktur auf. Es kann eine Nullstelle jeweils zwischen 1 und 2 kHz sowie zwischen 2 und 4 kHz erkannt werden. In [9] sind theoretische Modellanalysen von H_N/H_M eines Vokals auf Grundlage anatomischer Daten aus der Literatur präsentiert. Diese weisen ähnliche Nullstellenkonstellationen auf wie die hier aus Messungen gewonnenen Modellfunktionen. Somit ergänzen sich die hier vorgestellten experimentellen Ergebnisse mit den theoretischen Modellanalysen in [9].

Zusammenfassung

In diesem Beitrag wird das Sprachsignal von nasalierten Vokalen als Überlagerung von Mund- und Nasensignal betrachtet und dahingehend untersucht. Die getrennt aufgezeichneten Mund- und Nasensignale werden spektral analysiert und deren Zusammenwirken in Hörexperimenten beurteilt. Spektrale Zusammenhänge zwischen Mund- und Nasensignalen von nasalierten Vokalen konnten experimentell aufgezeigt werden.

Literatur

- [1] Fujimura, O., Lindqvist, J.: "Sweep-tone measurements of vocal-tract characteristics", J. Acoust. Soc. Am. 49, pp. 541-558, 1971.
- [2] Hawkins, S. W., Stevens, K. N.: "Acoustic and perceptual correlates of the non-nasal nasal distinction for vowels", J. Acoust. Soc. Am. 77, pp. 1560-1575, 1985.
- [3] Dang, J., Honda, K.: "Morphological and acoustical analysis of the nasal and paranasal cavities," J. Acoust. Soc. Am. 94 (4), pp. 2088-2100, 1994.
- [4] Feng, G., Castelli, E.: "Some acoustic features of nasal and nasalized vowels: A target for vowel nasalization", J. Acoust. Soc. Am. 99 (6), pp. 3694-3706, 1996.
- [5] Suzuki, H., Nakai, T., Dang, J., Lu, C.X.: "Speech production model involving subglottal structure and oral nasal coupling through closed velum", Proc. ICSLP'90, Kobe, pp. 437-440, 1990.
- [6] Schnell, K., Lacroix, A.: "Generation of nasalized speech sounds based on branched tube models obtained from separate mouth and nose outputs", Proc. ICASSP'03, Hong Kong, pp. 156-159, 2003.
- [7] Feng, G., Kotenkoff, C.: "New Considerations for Vowel Nasalization Based on Separate Mouth-Nose Recording", Proc. INTERSPEECH'06, Pittsburgh, pp. 2242-2245, 2006.
- [8] Kumaresan, R., Tufts, D.W.: "Estimation the Parameters of Exponentially Damped Sinusoids and Pole-Zero Modeling in Noise", IEEE Trans. ASSP-30, pp. 833-840, 1982.
- [9] Niu, X., Kain, A.B., van Santen J.P.H.: "Estimation of the acoustic properties of the nasal tract during the production of nasalized vowels", Proc. INTERSPEECH'05, Lisbon, pp. 1045-1048, 2005.