

ANALYSE NASALISierter VOKALE UNTER BERÜCKSICHTIGUNG GETRENNTER MUND- UND NASENSIGNALE

Matthias Geueke und Arild Lacroix

*Institut für Angewandte Physik, J.W. Goethe-Universität
Robert-Mayer-Straße 2-4, D-60325 Frankfurt am Main
geueke@iap.uni-frankfurt.de*

Abstract: Realisierungen nasalierter Vokale setzen sich aus einem Mund- und einem Nasensignal zusammen. Um Aussagen über die akustischen Eigenschaften dieser Teilsignale machen zu können, werden sie in einer speziellen Kabine getrennt aufgezeichnet. Dazu werden von verschiedenen Sprechern isolierte nasalisierte Vokale und Logatome gesprochen, bei denen lange Vokale in der Umgebung von Nasalen bzw. Plosiven eingebettet sind. Für die Analyse werden von ausgewählten Bereichen der getrennten Mund- und Nasensignale DFT-Spektren berechnet und BURG-Analysen durchgeführt. Hieraus werden die Formanten der Mund- und Nasensignale ermittelt. Zusätzlich zu den Aufnahmen in der Messkammer werden von denselben Sprechern auch gewöhnliche Sprachaufnahmen durchgeführt. Diese werden mit Linearkombinationen der getrennten Mund- und Nasensignale verglichen. Um Aussagen über die Formantfrequenzen der Mund- und Nasensignale machen zu können, werden aus verschiedenen Vokalrealisierungen die ersten beiden Formanten in einer Formantkarte dargestellt.

1 Einleitung

Vokale können unnasaliert oder nasalisiert gesprochen werden. Im ersten Fall wird das Anregungssignal allein durch den Vokaltrakt gefiltert, der einen unverzweigten Resonanzraum darstellt. Den Resonanzen dieses akustischen Systems entsprechen die Formanten der jeweiligen Vokale. Im zweiten Fall wird der Nasaltrakt durch Senken des Velums an den Vokaltrakt angekoppelt. Es liegt ein verzweigtes Resonatorsystem mit zwei Ausgängen vor. Dadurch kommt es zu einer Wechselwirkung zwischen Vokaltrakt- und Nasaltraktresonanzen, und man erwartet, dass sich die daraus resultierenden Effekte in den Spektren der Einzelsignale zeigen. Gelingt es also Mund- und Nasensignal zu trennen, lassen sich diese Veränderungen gezielt beobachten. STEVENS hat in [1] unter vereinfachenden Annahmen die Übertragungsfunktion von Vokal- und Nasaltrakt berechnet. In dieser Untersuchung sollen durch Messungen die Resonanzen des Sprechapparates überprüft werden. Dazu wird ein Messaufbau verwendet, der mittels einer geeigneten Dämmplatte eine Separation von Mund- und Nasensignalen ermöglicht.

2 Messaufbau

Die Aufnahmen der getrennten Mund- und Nasensignale werden in einer speziell dafür entwickelten Kabine durchgeführt, die bereits in [2] vorgestellt wurde. Im Gegensatz zu dem dort beschriebenen Messaufbau wird jedoch eine dickere Platte zur Trennung der Mund- und Nasensignale benutzt, die eine an den Kopf des jeweiligen Sprechers angepasste Aussparung enthält. Durch diesen veränderten Aufbau soll eine noch bessere Abdichtung erreicht werden. Allerdings ist jetzt für jeden Sprecher eine eigene Dämmplatte erforderlich, da diese genau auf die Größe des Kopfes des Sprechers angepasst werden muss. Der hintere Teil dieser Platte wird gegebenenfalls mit einem aufgeschnittenen Schlauch aus Schaumgummi ausgekleidet. Dies vermeidet zum einen Druckstellen, zum anderen erlaubt es durch die Elastizität eine gewisse Beweglichkeit des Kopfes, ohne dass sich die Abdichtung dadurch

verschlechtern würde. Vor Beginn der Messungen der getrennten Mund- und Nasensignale wurden Testmessungen durchgeführt, um Aussagen zur Übersprechdämpfung machen zu können. Diese Messungen wurden sowohl ohne als auch mit einer Person durchgeführt. Bei den Messungen ohne Person wurde eine Dämmplatte ohne Kopfausschnitt in die Kammer eingebaut. In der oberen Kammer wurde von einem Lautsprecher ein Sweep-Signal abgestrahlt und von beiden Mikrofonen in den getrennten Kammern aufgezeichnet. Von den aufgenommenen Signalen wurde das DFT-Spektrum berechnet. Das Übersprechen \ddot{U} ergibt sich aus dem logarithmierten Verhältnis der Betragsspektren S_1 und S_2 :

$$\ddot{U} = 20 \cdot \lg\left(\frac{S_2}{S_1}\right).$$

Als Ergebnis einer solchen Messung ergibt sich die in Bild 1 gezeigte Kurve für das Übersprechen. Man erkennt deutlich, dass bei höheren Frequenzen das Übersprechen mit -40 dB bis -30 dB gering ist. Bei den Übersprechmessungen mit Person sitzt diese – wie bei den eigentlichen Messungen – in der Messkammer, während das Sweep-Signal abgespielt und aufgezeichnet wird (vgl. Bild 2). Das Ergebnis dieser Messung ist für die Frequenzen bis 2 kHz in Bild 3 dargestellt.

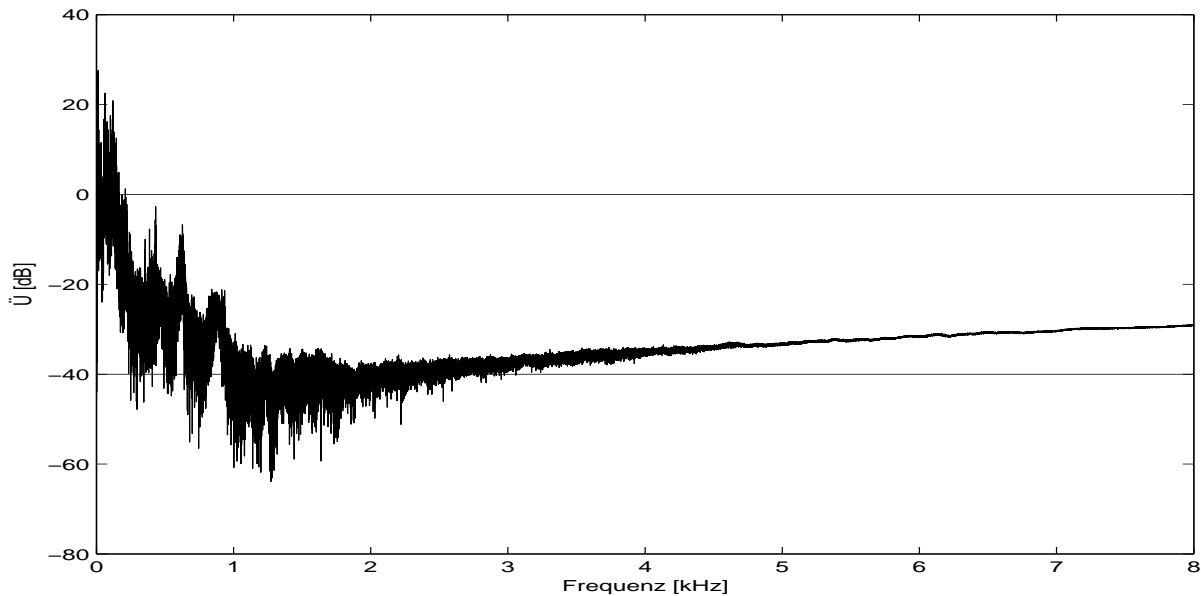


Bild 1 – Übersprechen durch die Dämmplatte ohne Person in Abhängigkeit von der Frequenz.



Bild 2 – Versuchsanordnung zur Messung des Übersprechens mit Person.

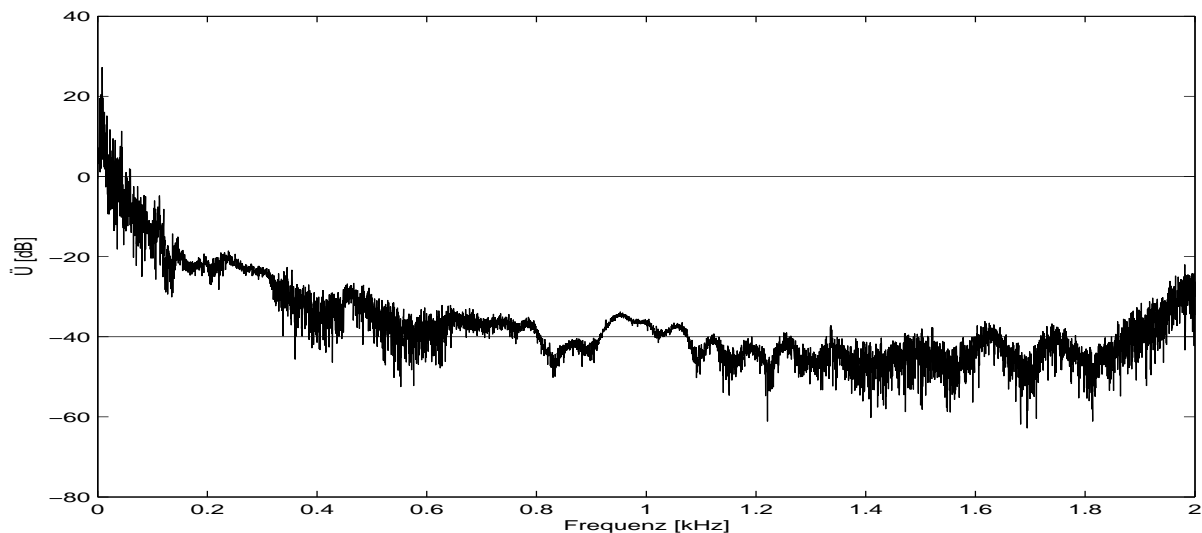


Bild 3 – Übersprechen durch die Dämmplatte mit Person in Abhängigkeit von der Frequenz

Man erkennt aus dieser Darstellung, dass das Übersprechen bei Frequenzen oberhalb von 200 Hz zwischen -40 und -30 dB liegt. Bei niedrigen Frequenzen wächst \dot{U} an, vermutlich wegen der zu geringen Abstrahlung des verwendeten Lautsprechers in diesem Frequenzbereich in Verbindung mit Undichtigkeiten der Messkammer gegenüber Außengeräuschen. Für die Formantanalysen ist vor allem der Bereich des Spektrums oberhalb von 200 Hz von Bedeutung, daher kann das Übersprechen als hinreichend gering angesehen werden.

3 Sprachmaterial

Die Aufnahmen werden bei einer Abtastrate von 32 kHz mit insgesamt sechs Sprechern durchgeführt, davon drei männliche und drei weibliche. Der aufgenommene Sprachdatenkorpus enthält sowohl Logatome als auch isoliert gesprochene Vokale. Die Logatome bestehen aus einem gelängten Vokal, der jeweils in eine Umgebung von zwei Plosiven bzw. zwei Nasalen eingebettet ist. Da bei Plosiven ein vollständiger Verschluss vorliegt, also auch der Nasaltrakt durch das Velum abgekoppelt ist, wird gewährleistet, dass der eingebettete Vokal unnasaliert gesprochen wird. Umgekehrt wird es ungeübten Sprechern durch die Nasal-Umgebung erleichtert, den Vokal nasaliert zu sprechen, da das Velum infolge der Koartikulation gesenkt bleibt. Zusätzlich werden auch isoliert gesprochene nasalierte Vokale aufgenommen. Die Messungen werden für alle Vokale des Deutschen durchgeführt, wobei alle Vokale lang gesprochen werden, um für die Analysen möglichst stationäre Laute zu erhalten. Dies führt bei nicht geübten Sprechern zu Problemen vor allem bei den Vokalen, die im Deutschen ausschließlich als kurzer Vokal realisiert sind. Deshalb wird das Gewicht der Analysen auf die im Deutschen naturgemäß langen Vokale und den Schwa-Laut gelegt. Das Nasen- und Mundsignal der Äußerungen $[n\tilde{i} : n]$ und $[k\tilde{i} : k]$ ist in Bild 4 dargestellt; die unterschiedlichen Pegel der Nasensignale des nasalierten und unnasalierten Vokals sind gut zu erkennen.

4 Untersuchungsergebnisse

4.1 Spektren

Aus den aufgezeichneten Vokalen wird ein Abschnitt von vier bis fünf Perioden ausgewählt und von diesen die DFT berechnet. Zusätzlich werden die Reflexionskoeffizienten nach BURG bestimmt, und daraus eine Übertragungsfunktion berechnet. Da keine Präemphase benutzt

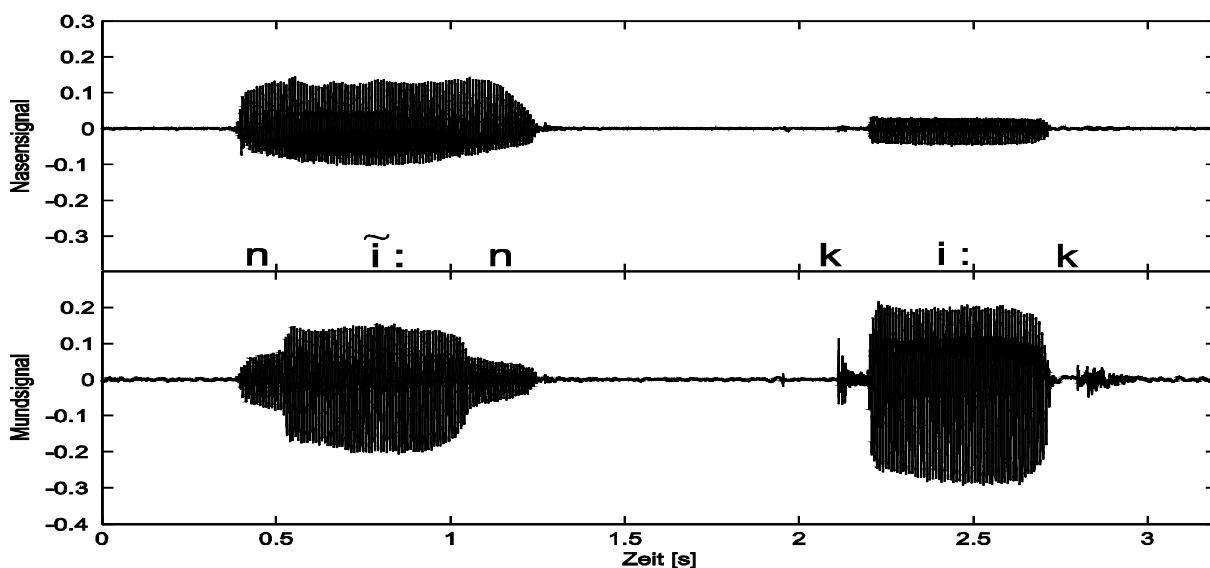


Bild 4 – Zeitsignale der Äußerungen [nĩ : n] und [k i : k] (oben Nasensignal, unten Mundsignal).

wird, enthält diese jedoch den Pegelabfall der Anregung zu höheren Frequenzen hin sowie die Charakteristik der Lippenabstrahlung. Aus dieser BURG-Übertragungsfunktion wird der Betragsgang berechnet, der in den Bildern 5 - 7 mit eingezeichnet ist, und näherungsweise die Einhüllende der Spektren darstellt. Um eine möglichst gute Nachbildung der Hüllkurve der Spektren zu erhalten, wird eine Prädiktorordnung von 60 gewählt. Die Bilder 5 - 7 zeigen die Spektren der getrennten Nasen- und Mundsignale für drei nasalierte lange Vokale des Deutschen; dargestellt ist jeweils der Bereich bis 8 kHz.

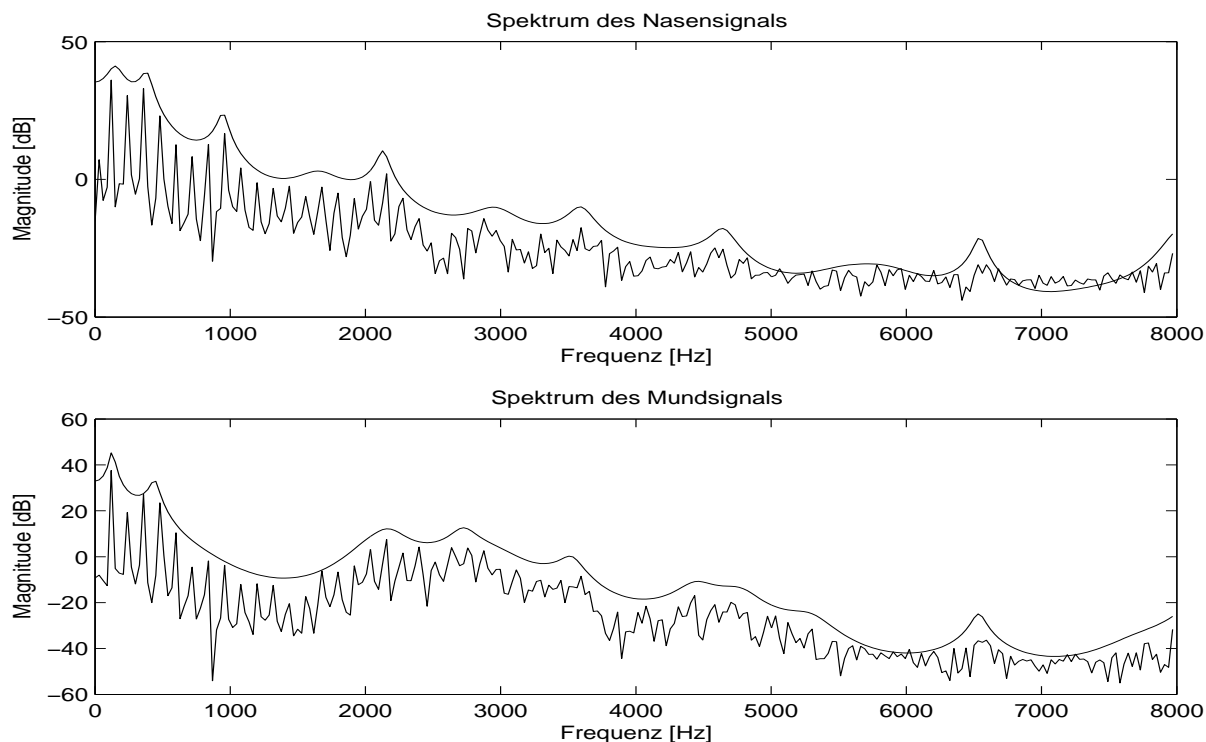


Bild 5 – Spektren der getrennten Nasen- und Mundsignale des Vokals [ĩ :] (Nasensignal oben, Mundsignal unten), die durchgezogene Hüllkurve entspricht dem BURG-Betragsgang der Ordnung 60.

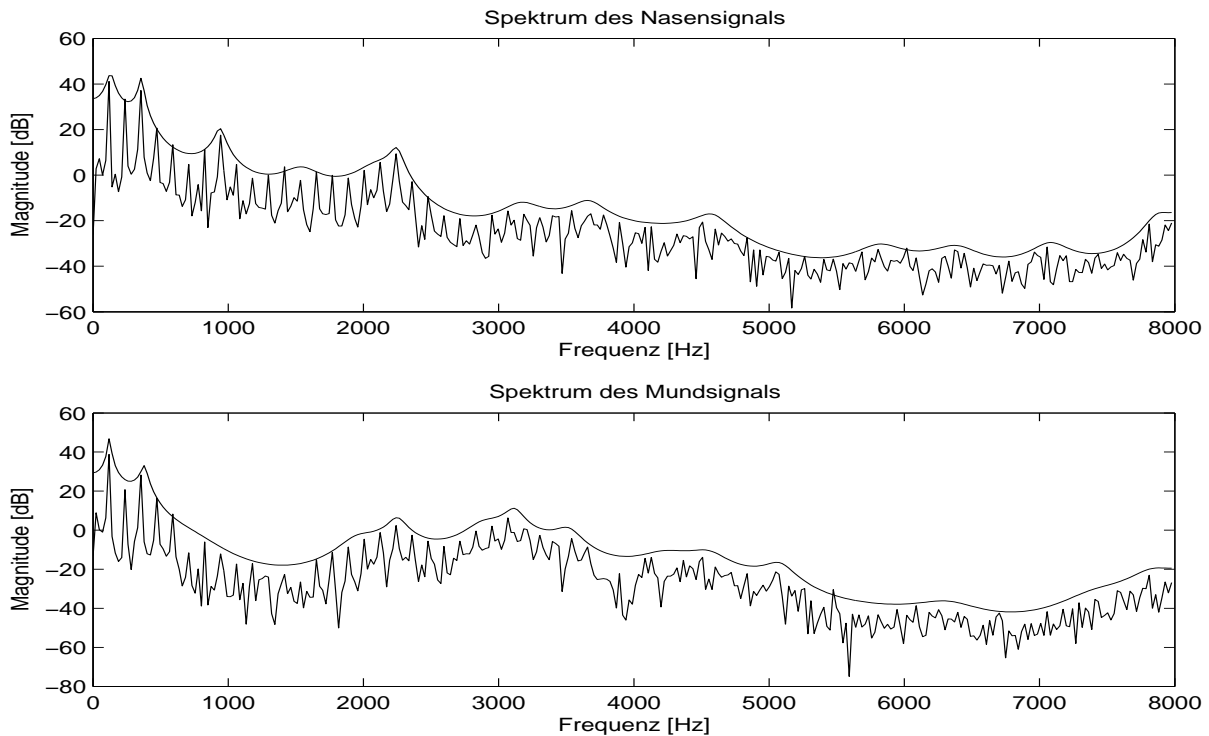


Bild 6 – Spektren der getrennten Nasen- und Mundsignale des Vokals [ĩ:] (Nasensignal oben, Mundsignal unten), die durchgezogene Hüllkurve entspricht dem BURG-Betragsgang der Ordnung 60.

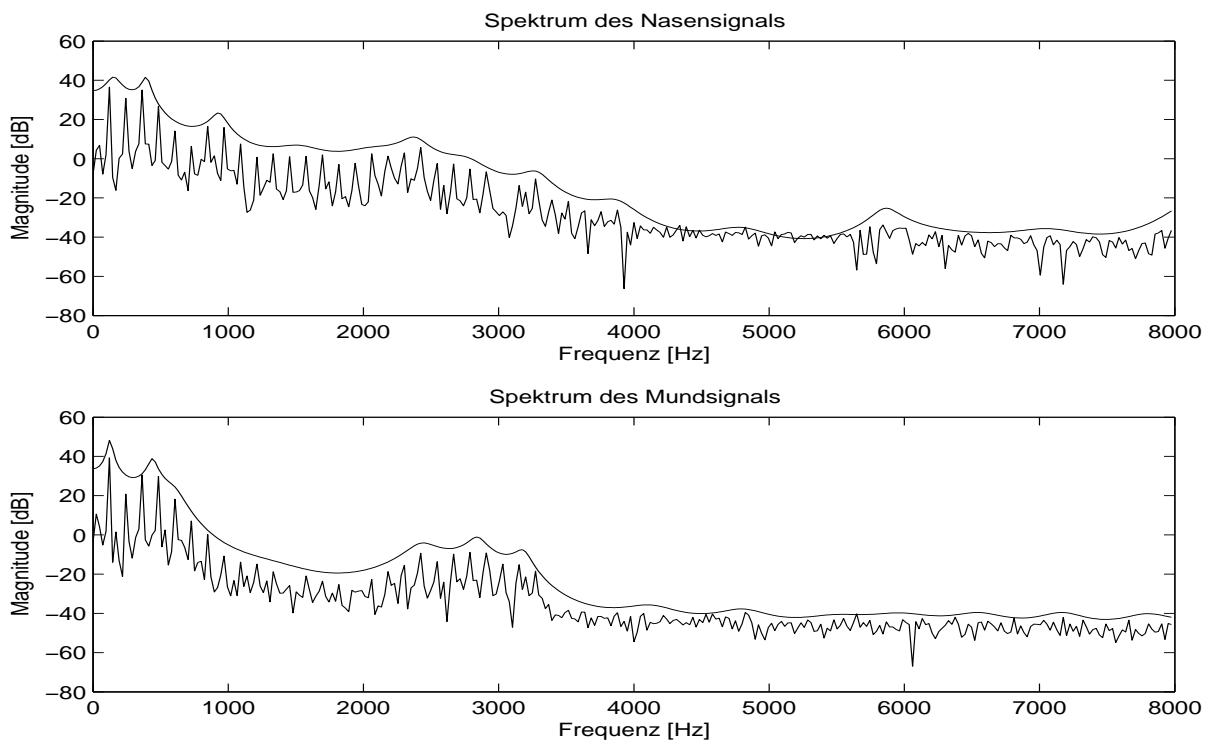


Bild 7 – Spektren der getrennten Nasen- und Mundsignale des Vokals [õ:] (Nasensignal oben, Mundsignal unten), die durchgezogene Hüllkurve entspricht dem BURG-Betragsgang der Ordnung 60.

Auffällig ist, dass im Nasensignal bei allen Vokalen bei 1 kHz eine Resonanz auftritt. Diese ist nicht bei allen Sprechern gleichermaßen deutlich zu erkennen, wie auf den Bildern 5 – 7, tritt aber mit einer gewissen Regelmäßigkeit auf. Bild 8 zeigt das Spektrum des Vokals [ĩ:] aus der Äußerung [nĩ:n] des gleichen Sprechers. Diese Aufnahme wird außerhalb der Kabine als Monosignal aufgezeichnet, es liegt also keine Trennung von Mund- und

Nasensignal vor. Der Peak bei 1 kHz ist deutlich schwächer zu erkennen. Zum Vergleich werden die getrennt aufgezeichneten Mund- und Nasensignale derselben Äußerung desselben Sprechers linearkombiniert und daraus das Spektrum berechnet, was mit dem Spektrum in Bild 8 verglichen werden kann. Dabei wurde im einzelnen wie folgt vorgegangen: Da die absoluten Schalldrücke von Mund- und Nasensignal aufgrund der von Messung zu Messung variablen Mikrofonposition und der nicht exakt gleichen Aussteuerung der beiden Signale nicht direkt verglichen werden können, wird vor der Überlagerung der beiden Signale auditiv ein Gewichtungsfaktor für das Nasensignal bestimmt. Dazu werden verschiedene Zahlenwerte variiert, bis sich das überlagerte Signal wie ein nasaliertes Vokal anhört. Dieser Zahlenwert ist für eine bestimmte Aufnahme, während der die Messbedingungen nicht verändert werden, charakteristisch. Weiterhin ist ein möglicher Zeitversatz zwischen den beiden aufgenommenen Signalen zu berücksichtigen, der sich aus den unterschiedlichen Abständen der beiden Mikrofone von der Schallquelle ergibt. Deshalb wurde das Nasensignal vor der Überlagerung gegen das Mundsignal um einige Abtastperioden in 5er Schritten nach rechts und links verschoben, um dem Effekt unterschiedlicher Mikrofonpositionen Rechnung zu tragen. Da der Unterschied zwischen den Mikrofonpositionen maximal 15 cm beträgt, ist eine Verschiebung der Signale gegeneinander bis ca. 15 Abtastperioden realistisch. Die Bilder 9 bis 11 zeigen Linearkombinationen der Nasen- und Mundsignale des Vokals [ĩ:], die bereits in Bild 6 gezeigt wurden, für verschiedene Zeitverschiebungen. Der Gewichtungsfaktor für das Nasensignal beträgt bei dieser Messung 0,4.

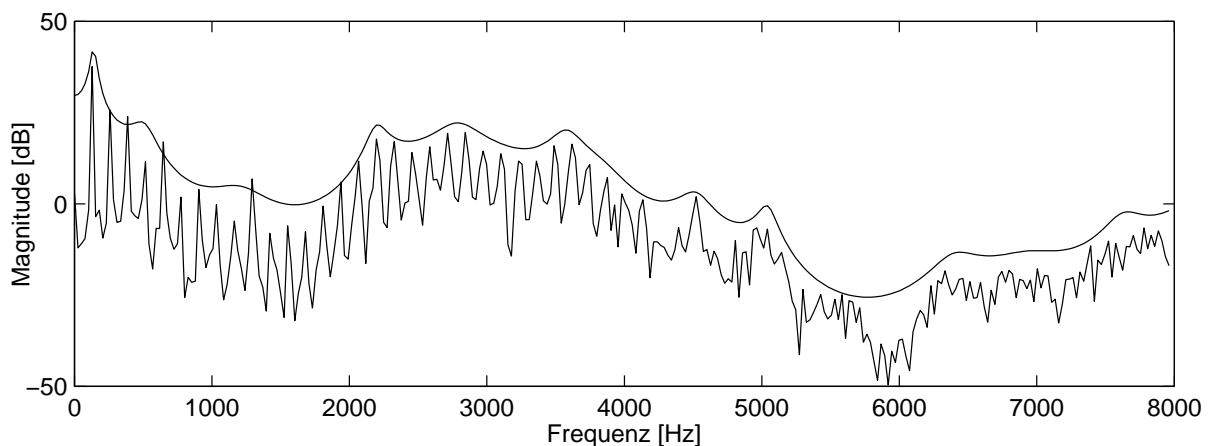


Bild 8 – Spektrum des Vokals [ĩ:] aus einer Aufnahme, die als Monosignal außerhalb der Kammer durchgeführt wurde.

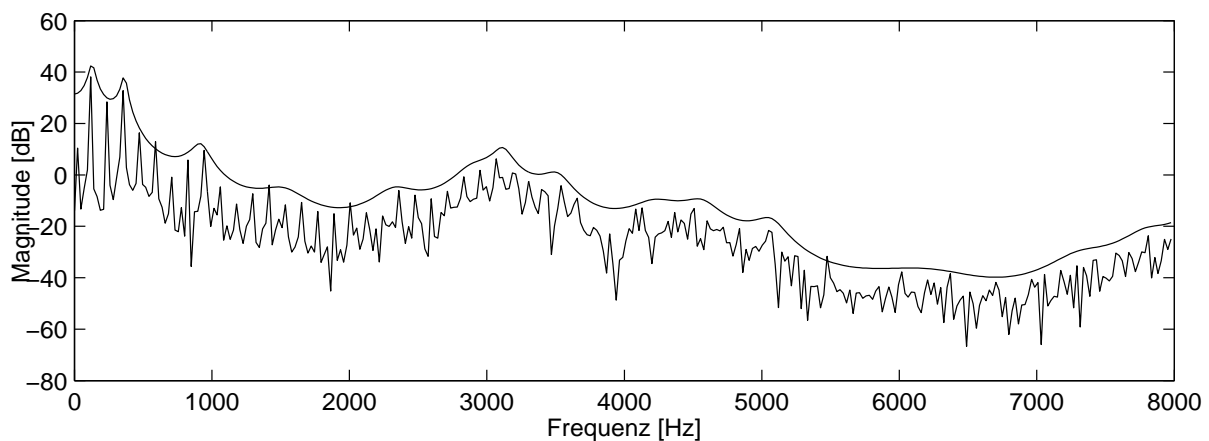


Bild 9 – Spektrum der Linearkombination der Mund- und Nasensignale für den Vokal [ĩ:] ohne Zeitverschiebung.

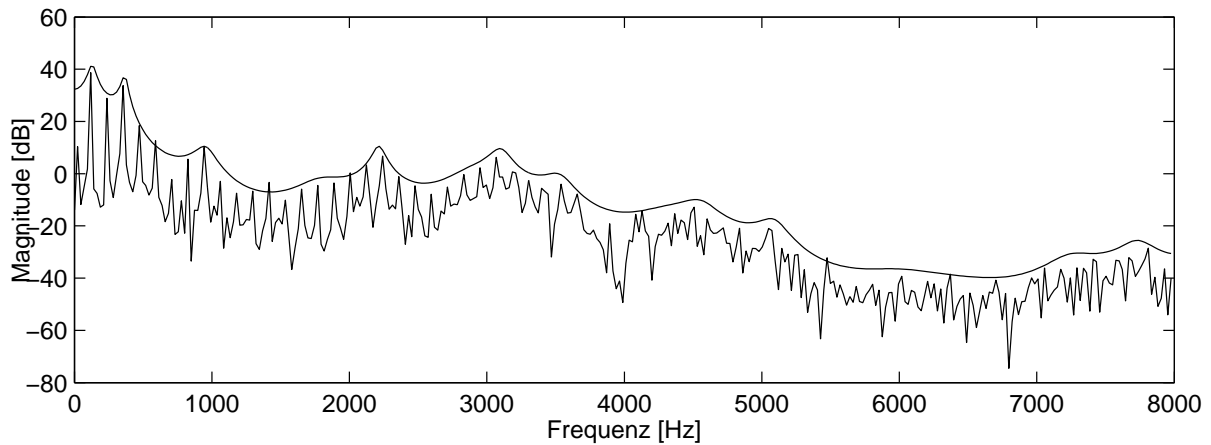


Bild 10 – Spektrum der Linearkombination der Mund- und Nasensignale für den Vokal [ĩ:] mit einer Zeitverschiebung von 5 Abtastperioden.

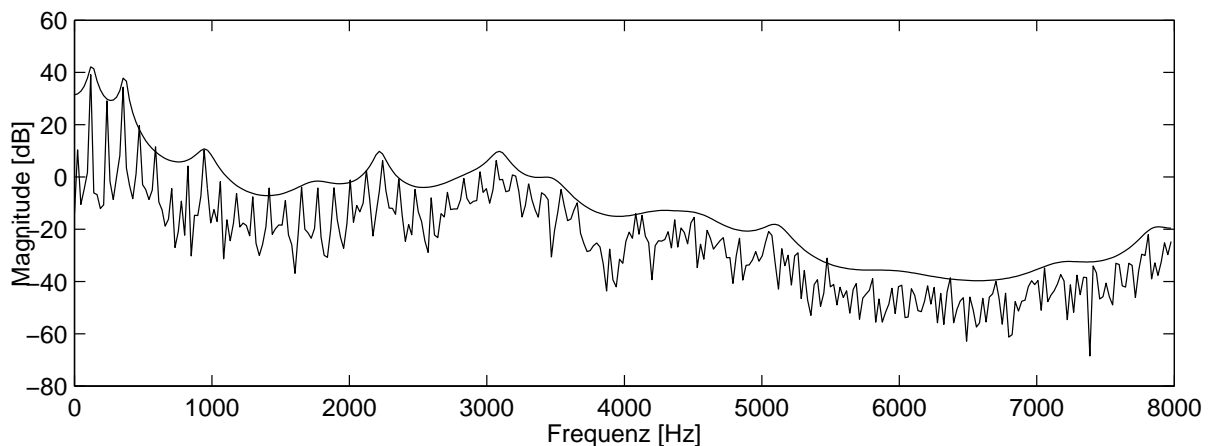


Bild 11 – Spektrum der Linearkombination der Mund- und Nasensignale für den Vokal [ĩ:] mit einer Zeitverschiebung von 10 Abtastperioden.

Die Modellspektren lassen erkennen, dass bei der Überlagerung die 1 kHz Resonanz nahezu unabhängig von der Zeitverschiebung ist. Zusätzlich kann beobachtet werden, dass die Resonanz bei ca. 2,2 kHz, die in Bild 6 sowohl im Mund- als auch im Nasensignal vorhanden ist, bei der Überlagerung in Abhängigkeit von der Zeitverschiebung unterschiedlich stark ausgeprägt ist. So sieht man beispielsweise, dass in Bild 9 diese Resonanz stark abgeschwächt ist, während sie in den Bildern 10 und 11 deutlich erkennbar ist.

4.2 Formantkarte

Mit Formantkarten kann die Lage der Formanten von Vokalen graphisch zu veranschaulicht werden. Daher ist es interessant, die Formanten, die aus den Mundsignalen der unnasalierten und nasalierten Vokale ermittelt werden, in einer solchen Formantkarte gegenüberzustellen. So können eventuell auftretende Formantverschiebungen sichtbar gemacht werden. Bild 12 zeigt die Formanten aus den Mundsignalen für die unnasalierten und nasalierten Vokale /a:/, /i:/, /u:/ und für den Schwa-Laut. Ihre genaue Lage wird aus den Polstellen der bereits oben erwähnten Übertragungsfunktion nach BURG ermittelt. Anhand dieser Daten können außer beim Vokal /u:/ keine Verschiebungen der Formantlagen beobachtet werden. Die Veränderungen der Formanten im Mundsignal, die durch die Nasalierung zustande kommen, sind offenbar viel geringer als die natürliche Streuung der Formanten über verschiedene Realisierungen. Beim /u:/ erkennt man, dass der F1 der nasalierten Realisierungen signifikant höher liegt als bei den unnasalierten. Dies kann man durch den in der Literatur [3]

beschriebenen Effekt erklären, dass bei nasalierten Vokalen der Öffnungsgrad des Mundes im allgemeinen größer ist, als bei ihren unnasalierten Realisierungen.

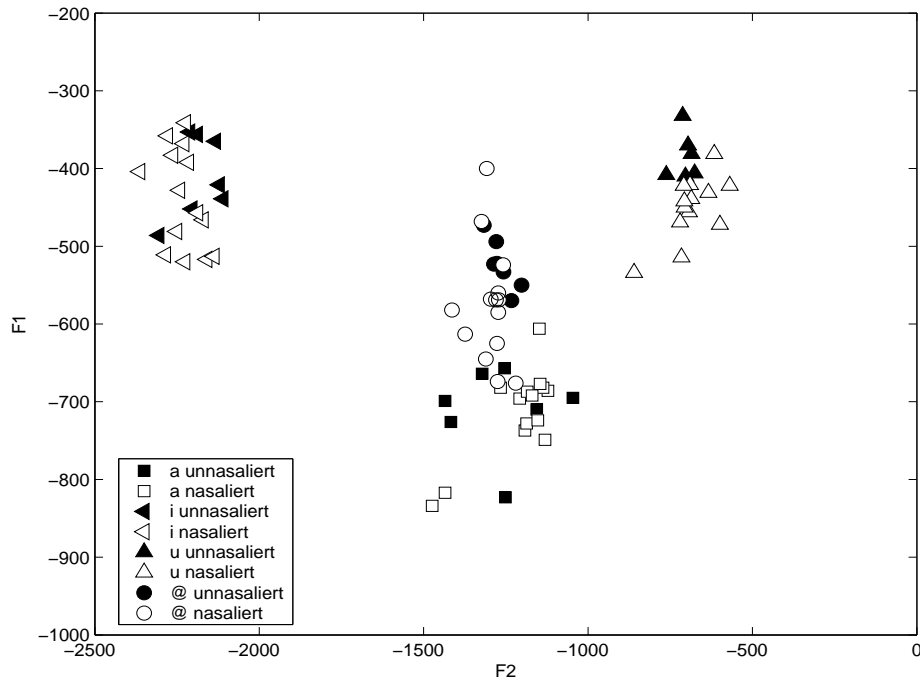


Bild 12 – Formantkarte zur Darstellung der aus den Mundsignalen unnasalierter und nasalierter Laute ermittelten Werte für die ersten beiden Formantfrequenzen.

5 Zusammenfassung

Zur Untersuchung der spektralen Eigenschaften von Mund- und Nasensignalen nasalierte Vokale werden diese in einer Kabine getrennt aufgezeichnet; die Separation wird durch eine an die jeweilige Versuchsperson angepasste Dämmplatte erreicht. Aus diesen Signalen werden DFT-Spektren und BURG-Modellspektren berechnet. Die aus den überlagerten Mund- und Nasensignalen berechneten Spektren werden mit gewöhnlichen Sprachaufnahmen verglichen. Dabei wird beobachtet, dass der Zeitversatz zwischen Mund- und Nasensignal das Gesamtspektrum beeinflusst. Die Formantfrequenzen von Mundsignalen nasalierte und unnasalierter Vokale werden in einer Formantkarten gegenübergestellt und verglichen.

Literatur

- [1] Stevens K. N.: „Acoustic Phonetics“, MIT Press, Cambridge - London 1998, Chapter 3
- [2] Bettinelli M.; Schnell K., Lacroix A.: „Separate Messung und Analyse von Mund- und Nasensignalen bei natürlicher Sprache“, Studententexte zur Sprachkommunikation: Band 24, 13. Konferenz ESSV 2002, Dresden, S. 237-244.
- [3] Neppert J.; Pétursson M.: „Elemente einer akustischen Phonetik“, Helmut Buske Verlag, Hamburg, 3. durchgesehene Auflage 1992, S. 130.