

Zur Auswirkung der Glottis-Sprechtrakt-Kopplung auf die Stimmreinheit

B. J. Kröger

The Effect of Source-Tract Interaction on Voice Purity

By means of computer simulation of the glottal source and the vocal tract some effects of source-tract interaction on voice purity can be investigated. On the basis of given degrees of glottal frequency perturbations, the frequency and amplitude perturbations (jitter and shimmer) of the acoustic speech signal and of the excitation function are measured. It is shown that the acoustic filtering of the vocal tract leads to great differences between the perturbation values of the glottal waveform and of the acoustic speech waveform. The interaction produces extra jitter and shimmer in the waveforms of the vocal fold motion and of the excitation function.

Key words

Interaction – speech synthesis – vocal tract – voice – voicepurity

Zusammenfassung

Mittels Computersimulation von Glottis und Sprechtrakt können verschiedene Einflüsse der Glottis-Sprechtrakt-Kopplung auf die Stimmreinheit untersucht werden. Bei bekannter Stärke der glottalen Frequenzschwankungen wurden die Frequenz- und Amplitudenschwankungen (Jitter und Shimmer) des akustischen Sprachsignals und der Anregungsfunktion gemessen. Es zeigt sich, daß durch die Filterwirkung des Sprechtrakts große Differenzen zwischen den glottalen Schwankungswerten und denen des akustischen Sprachsignals auftreten. Die Kopplung erzeugt zusätzlich Jitter und Shimmer in der Stimmlippenschwingung und der Anregungsfunktion.

Schlüsselwörter

Kopplung – Sprachsynthese – Sprechtrakt – Stimme – Stimmreinheit

Einleitung

Die von verschiedenen Autoren durchgeführten Stimmreinheitsanalysen belegen eine Abhängigkeit der ermittelten Daten von Tonhöhe und Lautstärke, wie auch eine Abhängigkeit dieser von der Artikulation (Moser 1985, Milenkovic 1987, Wesselmann und Reker 1990 u. a.). Des weiteren ist bekannt, daß aufgrund der Filterwirkung des Sprechtraktes die glottalen Frequenz- oder Amplitudenschwankungen nicht in gleicher Quantität im akustischen Sprachsignal widerspiegelt werden (Milenkovic 1987, S. 535).

Zur Produktion des akustischen Sprachsignals

Die hier benutzten Sprachsynthesemodelle reproduzieren die Entstehung des Sprachsignals auf physiologisch-akustischer Ebene. Wichtige physiologische, aerodynamische und akustische Größen sind Stimmlippenöffnungsfläche, zeitliche Ableitung des glottalen Volumenstroms („glottaler Volumenstrom“ ist die durch die glottale Öffnungsfläche strömende Luftmenge pro Querschnittsfläche und pro Zeiteinheit) und das vom Mund abgestrahlte Schallsignal. Abb. 1

stellt diese Größen als Funktion der Zeit für die Produktion der Vokale /i/ und /a/ dar. Die Extrema oder Peaks (d. h. lokale Maxima und Minima) der zeitlichen Ableitung des glottalen Volumenstroms bestimmen die Intensität, mit der die Formanten des Sprechtrakts zu Schwingungen angeregt werden. Daher wird die Zeitfunktion dieser Produktionsgröße im folgenden auch als *Anregungsfunktion* bezeichnet. Wie Abb. 1 zeigt, kann an der Anregungsfunktion und am Sprachsignal des Vokals /a/ abgelesen werden, daß die stärkste Sprechtraktanregung im Moment des Stimmlippenverschlusses auftritt. Die auftretende Phasenverschiebung zwischen Extremum der Anregungsfunktion und Signalextrimum (siehe Pfeile in Abb. 1, Vokal /a/) beruht auf der endlichen Laufzeit eines Impulses von der Glottis durch den Sprechtrakt bis zum Mund. Beim Vokal /i/ ist nach der Verschlußbildung der Stimmlippen eine starke Anregung des zweiten Formanten erkennbar (starke Wellung auf der durch Grundfrequenz F₀ und Frequenz des ersten Formanten F₁ gegebenen Grundschwingung des Sprachsignals, siehe Pfeile in Abb. 1, Vokal /i/).

Die Simulationsmodelle

Das hier benutzte Simulationsmodell des Sprechtrakts ist in der Lage, zu jedem Zeitpunkt und an jedem Ort des Sprechtrakts Luftdruck und Volumenstrom anzuge-

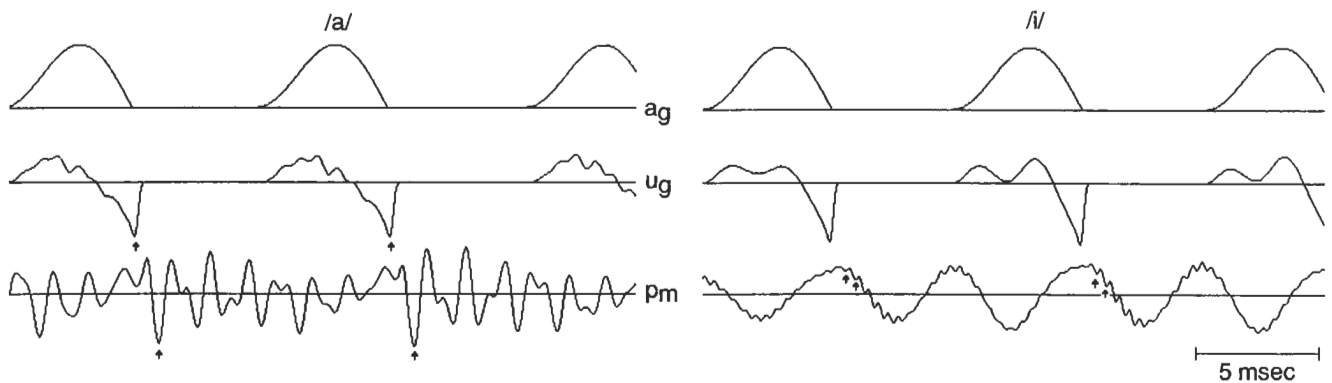


Abb. 1 Zeitfunktionen zur Sprachsignalproduktion am Beispiel der Vokale /a/ und /i/ mit $F_0 = 96$ Hz, glott. Jitter 3%: Stimmlippenöffnungsfläche a_g , zeitliche Ableitung des glottalen Volumenstroms u_g' (Anregungsfunktion) und Sprachsignal p_m . /a/: die Pfeile deuten das Maximum der Sprechtraktanregung an. /i/: die Pfeile heben die Schwingung des 2. Formanten hervor.

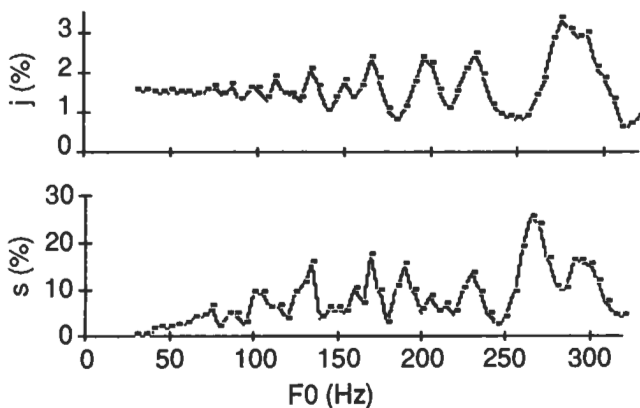


Abb. 2 Jitter j und Shimmer s des Sprachsignals als Funktion der Grundfrequenz F_0 bei Produktion des Vokals /a/ im GAI-Modell mit vorgegebenem glottalem Jitter von 1,5%.

ben. Zur Modellierung der Sprechtraktanregung wurden drei verschiedene Modelle benutzt, die verschiedene Stufen der Glottis-Sprechtrakt-Kopplung simulieren (Kröger 1990).

1. GNI-Modell: Glottismodell ohne Kopplung („no interaction“): Der glottale Volumenstrom wird am Ort der Glottis in den Sprechtrakt eingepreßt. Seine Schwingungsform ist fest vorgegeben. Glottale Anregung und Sprechtrakt sind vollständig entkoppelt.

2. GAI-Modell: Glottismodell mit aerodynamischer Kopplung („aerodynamic interaction“, auch mit „acoustic interaction“ bezeichnet (Fant und Lin 1987)): Der subglottale Druck und die Zeitfunktion der glottalen Öffnungsfläche (Zeitfunktion der Stimmlippenbewegung) sind vorgegeben. Hieraus berechnet das Modell dann nach aerodynamischen Gesetzmäßigkeiten den glottalen Volumenstrom. Die aerodynamische Kopplung beeinflusst mittels der oberhalb der Glottis vorliegenden (supraglottalen) Druckschwankungen den glottalen Volumenstrom und damit die Anregungsfunktion. Daher ist die Schwingung des ersten Formanten in der Anregungsfunktion erkennbar (Abb. 1).

3. GMI-Modell: Glottismodell mit mechanischer Kopplung („mechanical interaction“): Die Stimmlippen werden durch zwei schwingfähige Massepaare (Paar 1 = unterer, zur Trachea weisender Teil der Stimmlippen; Paar 2 = oberer, zum Sprechtrakt weisender Teil der Stimmlippen) simuliert. Die Schwingungsform der Stimmlippen ist nicht fest vorgegeben. Sie wird durch subglottalen Druck, Stimmlippenspannung und die Rückwirkung des Sprechtrakts mittels supraglottaler Druckschwankungen bestimmt.

Zur Messung der Stimmreinheitsparameter

Anhand der simulierten Schwingungsverläufe der Stimmlippenöffnungsfläche, der Anregungsfunktion und des Sprachsignals werden die Stimmreinheitsparameter relative Frequenzschwankung (Jitter, j in %) und relative Amplitudenschwankung (Shimmer, s in %) gemessen. Die Simulationsprogramme arbeiten mit einer Abtastrate von 20 kHz, so daß Zeitpunkt und Amplitude der Peaks mittels Interpolation (hier: Peakpicking-Algorithmus mit parabolischer Interpolation) bestimmt werden müssen (Titze und Horii et al. 1987). Zur Fehlerabschätzung wurden ungestörte Signale unterschiedlicher Signalform bei verschiedenen Grundfrequenzen ausgemessen. Es ergaben sich Werte von $j < 0,001\%$ und $s < 0,01\%$.

Ergebnisse

Vor allem für die geschlossenen vorderen Vokale (z. B. /i/) traten große und unsystematische Schwankungen in den Meßwerten der anhand des Sprachsignals bestimmten Stimmreinheitsparameter auf. Unproblematisch hingegen sind die Analysen beim Vokal /a/. Aus der Sichtweise der Sprachsignalproduktion erklärt sich dies leicht: Beim /a/ können die Peaks der Sprechtraktanregungsfunktion, deren Abfolge die Information über Frequenz- und Amplitudenschwankungen der Stimmlippen enthält, klar und eindeutig am Sprachsignal abgelesen werden (Abb. 1). Bei geschlossenen Vokalen (z. B. /i/ und /u/) ist dies wegen der geringen Differenz zwischen Grundfrequenz F_0 und der Frequenz des ersten Formanten F_1 nicht eindeutig möglich. Zusätzlich enthält das Signal der geschlossenen vorderen Vokale die für das Peakpicking eher störende Schwingung des zweiten Formanten (siehe Pfeile in Abb. 1, Vokal /i/). Dies erklärt

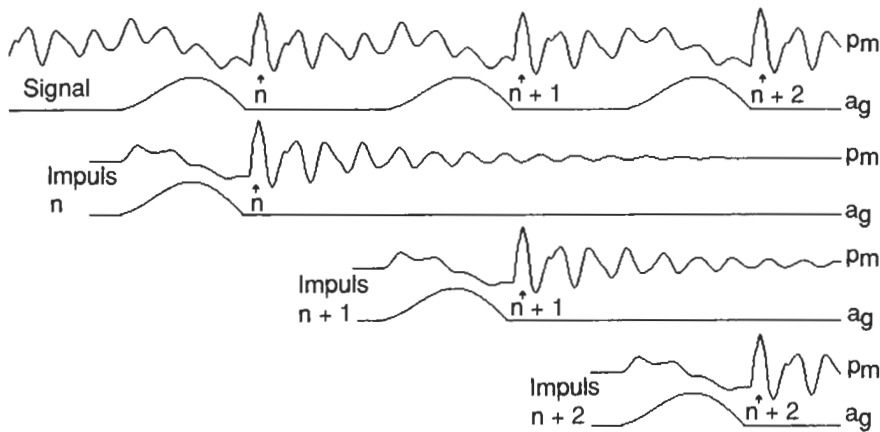


Abb. 3 Zur Entstehung von Jitterabweichungen und Shimmer durch die Überlagerung der Impulsantworten. Vokal /a/, GNI-Modell, $F_0=96$ Hz, glott. Jitter 10%.

Oben: Sprachsignal p_m und Zeitfunktion der glott. Öffnungsfläche a_g ; darunter: 3 identische Impulsantworten p_m (n , $n+1$, $n+2$) mit zugehöriger Zeitfunktion der glott. Öffnungsfläche a_g .

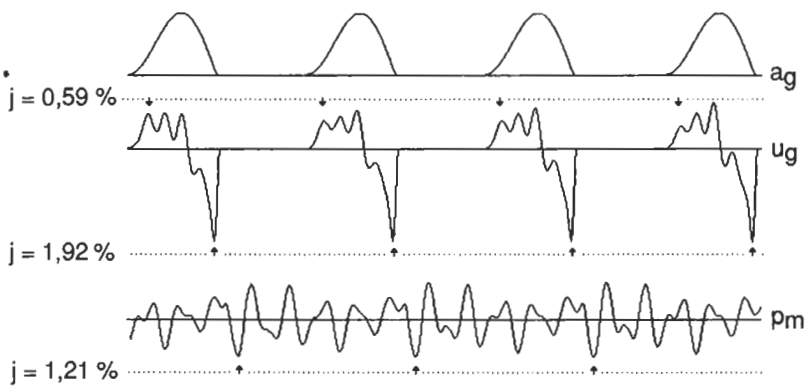


Abb. 4 Jitter verschiedener Peaks der Anregungsfunktion u_g' und des akustischen Sprachsignals p_m . Vokal /a/, GAI-Modell, $F_0 = 140$ Hz, vorgegebener Jitter in der Stimmlippenöffnungsfläche a_g von 1,9%. Die Jittermessungen wurden an den mit Pfeilen gekennzeichneten Peaks vorgenommen.

auch, daß in den für die Vokale /i/, /a/ und /u/ vorgenommenen Messungen von Milenkovic (1987, S. 535) nur beim Vokal /i/ Linearitätsabweichungen zwischen vorgegebenen und gemessenen Jitterwerten auftreten. Da der Vokal /a/ den höchsten ersten und dabei einen möglichst niedrigen zweiten Formanten besitzt, ist dieser Vokal für Reinheitsanalysen mittels Peakpicking oder Nulldurchgangsbestimmung am besten geeignet.

Zur Filterwirkung des Sprechtrakts

In der ersten Meßreihe wurden Jitter und Shimmer des akustischen Sprachsignals bei Vorgabe von konstantem Jitter in der Zeitfunktion der glottalen Öffnungsfläche im GAI-Modell für den Vokal /a/ als Funktion der Grundfrequenz F_0 untersucht (Abb. 2). Es zeigt sich, daß der Signaljitter in erheblichem Maße um den mit 1,5% vorgegebenen Wert des glottalen Jitters herum schwankt. Neben diesen Jitterabweichungen entsteht auch erheblicher Shimmer im Sprachsignal, obwohl kein glottaler Shimmer vorliegt.

Jitterabweichung und Shimmerentstehung dieser Größenordnung können, wie Messungen mit dem GNI-Modell zeigen, auch ohne Kopplung allein durch die Filterwirkung des Sprechtrakts erklärt werden. Abb. 3 zeigt die Entstehung des Sprachsignals aus der Überlagerung mehrerer Impulsantworten. Eine Impulsantwort des Sprechtrakts entsteht aufgrund eines glottalen Schwingungszyklus (siehe Darstellung der glottalen Öffnungsfläche für jede Impulsantwort in Abb. 3) und spiegelt die durch die Formanten hervorgerufenen Luftdruckschwankungen im Sprechtrakt wider. Jede neue Impulsantwort überlagert sich jeder vorhergehenden

Impulsantwort wegen der Unregelmäßigkeiten im Zeitabstand zweier aufeinanderfolgender glottaler Schwingungszyklen in unterschiedlicher Weise.

Zur aerodynamischen Kopplung

Eine erste Bewertung des Einflusses der aerodynamischen Kopplung kann gewonnen werden, indem bei vorgegebenem Jitter in der Zeitfunktion der glottalen Öffnungsfläche im GAI-Modell die Werte der Jitterabweichung und des entstehenden Shimmers am Extremum der Anregungsfunktion gemessen werden. Diese Werte sind erheblich geringer als die durch die Filterwirkung des Sprechtrakts entstehenden Werte. Es bleibt aber festzuhalten, daß sich die Stimmlippen-schwingung aufgrund der aerodynamischen Kopplung bei höheren Grundfrequenzwerten nicht ohne Entstehung zusätzlichen Jitters und Shimmers auf die Anregungsfunktion überträgt.

Bemerkenswert ist auch, daß der Sprechtrakt bei bestimmten Grundfrequenzen eine Verringerung des Signaljitters gegenüber dem eingespeisten glottalen Jitter bewirkt, da der Verlauf der Jitterwerte in Abb. 2 auch geringere Werte als 1,5% aufweisen kann. Dieser Effekt resultiert daraus, daß die Formantschwingungen des Sprechtrakts bei gleichbleibender Sprechtraktform keine Frequenzschwankungen aufweisen und somit in der Lage sind, bei bestimmten Formantfrequenz-Grundfrequenz-Verhältnissen glottalen Jitter auszugleichen. Mißt man beispielsweise bei einer Grundfrequenz von 140 Hz nicht nur den Jitter der Hauptpeaks der Anregungsfunktion (glottaler Verschuß), sondern auch den Jitter der in der Öffnungsphase der Glottis auftretenden Ne-

benpeaks, so stellt man fest, daß diese erheblich geringere Jitterwerte aufweisen (Abb. 4). Die Nebenpeaks entstehen aufgrund der aerodynamischen Kopplung und geben ein Abbild der Formantschwingungen des Sprechtrakts.

Zur mechanischen Kopplung

Des weiteren wurden Jitter und Shimmer der freien Stimmlippenbewegung im GMI-Modell gemessen. Bemerkenswert ist zunächst, daß das selbstschwingende Glottismodell überhaupt Jitter und Shimmer produziert. Bei der Produktion des Vokals /a/ mit einer Grundfrequenz von 120 Hz ergaben sich für die Zeitfunktion der Stimmlippenöffnungsfläche 0,15% Jitter und 0,5% Shimmer. Betreibt man das Glottismodell mit abgekoppeltem Sprechtrakt, indem man den zeitlich konstanten Umgebungsdruck als supraglottalen Druck ansetzt, so treten keine meßbaren Schwingungsunregelmäßigkeiten der Stimmlippenbewegung auf. Diese entstehen bei ideal schwingenden Stimmlippen also erst aufgrund der Glottis-Sprechtrakt-Kopplung: Der glottale Volumenstrom beeinflußt den Luftdruck am Ort der Stimmlippen und damit die auf die Stimmlippen wirkenden äußeren Kräfte (mechanische Kopplung).

Diskussion

Zunächst kann anhand einfacher Betrachtungen zur Sprachproduktion verdeutlicht werden, daß der Vokal /a/ wegen des hohen ersten und niedrigen zweiten Formanten zur Stimmreinheitsanalyse mittels Meßmethoden im Zeitbereich am besten geeignet ist.

Die hier anhand von recht realitätsnahen Simulationsmodellen der Sprachsignalproduktion durchgeführten Messungen zeigen aber auch beim Vokal /a/ die starken Differenzen zwischen glottalen Schwankungswerten und denen des akustischen Sprachsignals. Als Grund ist in erster Linie die Filterwirkung des Sprechtrakts zu nennen. Um eine genauere Meßmöglichkeit glottaler Schwankungswerte an-

hand des Sprachsignals zu erhalten, könnte inverses Filtern des Sprachsignals durchgeführt werden (Rothenberg 1973).

Der Einfluß der aerodynamischen Kopplung ist dem Einfluß des obigen Effekts zwar quantitativ untergeordnet, bewirkt aber mit steigender Grundfrequenz eine zunehmende Erhöhung der Schwingungsunregelmäßigkeiten in der Anregungsfunktion. Bei ideal schwingenden Stimmlippen ist die mechanische Kopplung überhaupt erst der Verursacher von glottalen Schwingungsunregelmäßigkeiten. Die Abhängigkeit dieser Unregelmäßigkeiten von der Grundfrequenz wird Gegenstand weiterer Untersuchungen sein.

Literatur

- Fant, G., Q. Lin: Glottal source-vocal tract acoustic interaction. *Speech Transm. Lab.-Quart. Prog. Stat. Rep.* 1 (1987) 13–27. Royal Inst. Techn., Stockholm
- Kröger, B. J.: Die Synthese der weiblichen Stimme unter besonderer Berücksichtigung der Phonation. Dissertation, Köln 1989
- Kröger, B. J.: Three glottal models with different degrees of glottal source-vocal tract interaction. *IPKöln-Berichte* 16 (1990) 59–67
- Milenkovic, P.: Least mean square measures of voice perturbation. *J. Speech Res.* 30 (1987) 529–538
- Moser, M.: Artikulatorische Einflüsse auf die Stimmreinheit. *Sprache – Stimme – Gehör* 9 (1985) 117–120
- Rothenberg, M.: A new inverse filtering technique for deriving the glottal air flow waveform during voicing. *Journ. Acoust. Soc. Am.* 53 (1973) 1632–1645
- Titze, I. R., Y. Horii, R. C. Scherer: Some technical considerations in voice perturbation measurements. *J. Speech Res.* 30 (1987) 252–260
- Wesselmann, U., U. Reker: Abhängigkeit der Stimmreinheit von der Artikulation. *Sprache – Stimme – Gehör* 14 (1990) 148–152

Dr. phil. Dipl.-Phys. Bernd J. Kröger

Institut für Phonetik
der Universität zu Köln
Greinstr. 2
5000 Köln 41