

Akustische Analyse der Stimmqualität

Zwei nützliche Maße

Hier soll der Schwerpunkt auf zwei Maßen liegen, die sich als nützlich erwiesen haben, um Sprachen zu charakterisieren, die zwei oder sogar drei Stimmqualitäten entlang der Achse creaky/laryngealized - modal - breathy kontrastieren. Es gibt weitere Maße, die vor allem im Bereich der Stimmpathologie eingesetzt werden.

Die beiden Maße erfassen auf etwas unterschiedliche Weise, wie schnell die Amplitude der Harmonischen zu den hohen Frequenzen hin abnimmt (Englisch: "spectral tilt").

H1-H2

Hier ermittelt man die *Amplitude* der ersten zwei Harmonischen im Spektrum, und bildet die Differenz in dB (die *Frequenz* der ersten Harmonischen entspricht der Grundfrequenz). Ausgehend von den Modellversuchen von Klatt & Klatt (und anderen) hängt der Wert mit dem laryngealen Öffnungsquotient zusammen: Bei großem Öffnungsquotient ("breathy") stark positive Werte, d.h. Amplitude von H1 ist wesentlich größer als die Amplitude von H2. Bei kleinerem Öffnungsquotient nimmt der Wert für H1-H2 ab. Bei sehr kleinem Öffnungsquotient (creaky/gepresst) sind leicht negative Werte möglich (d.h. Amplitude von H2 ist größer als die Amplitude von H1). Vgl. Fig.1 aus Klatt & Klatt (1990). Diese Abbildung bezieht sich auf eine Modellierung des reinen Quellsignals, was uns bei normalen Mikrofonaufnahmen nicht zur Verfügung steht. Typische Verhältnisse bei einer solchen Aufnahme sieht man in der Abbildung unten. Bei Modal (Spektrum links) wäre H1-H2 leicht negativ (d.h. $H1 < H2$); bei Breathly (Spektrum rechts) wäre H1-H2 stark positiv ($> 10\text{dB}$).

Harmonics-to-Noise ratio

H1-H2 erfasst die Form des Spektrums bei den tiefsten Frequenzen. Bei Stimmtypen wie "Breathy", wo der Schließvorgang der Stimmbänder nicht sehr abrupt ist, und wo der hintere Teil der Glottis u.U. durchgehend offen bleibt, kommt es zu einer Abschwächung *aller* höheren Harmonischen (nicht nur von H2), bzw. die höheren Harmonischen sind mit Aspirationsrauschen überlagert (vgl. Klatt & Klatt. Fig. 1 rechts). Die Gesamtstärke der Harmonischen im Spektrum kann mit dem Praat-Befehl "Harmonicity" ermittelt werden.

Auszug aus der Praat-Hilfe zum Befehl 'Harmonicity':

"A Harmonicity object represents the degree of acoustic periodicity, also called Harmonics-to-Noise Ratio (HNR). Harmonicity is expressed in dB: if 99% of the energy of the signal is in the periodic part, and 1% is noise, the HNR is $10 \cdot \log_{10}(99/1) = 20 \text{ dB}$ (approx.). A HNR of 0 dB means that there is equal energy in the harmonics and in the noise.

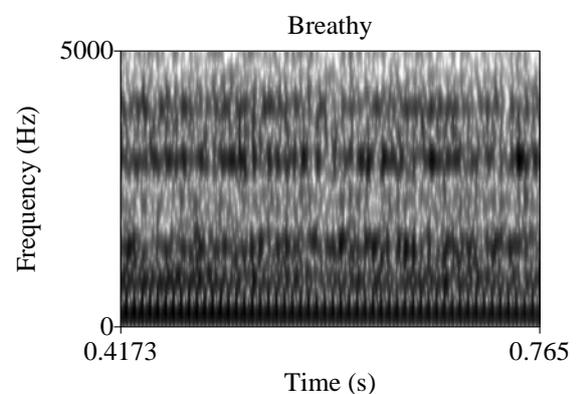
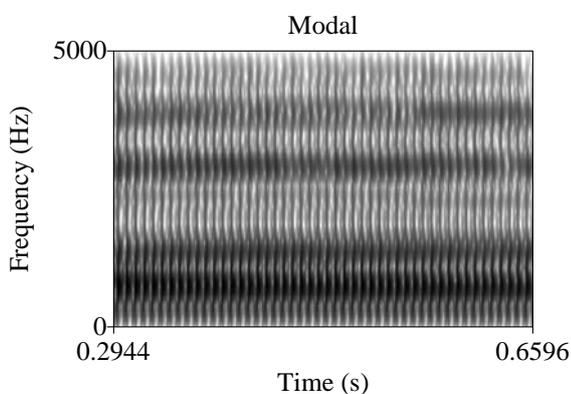
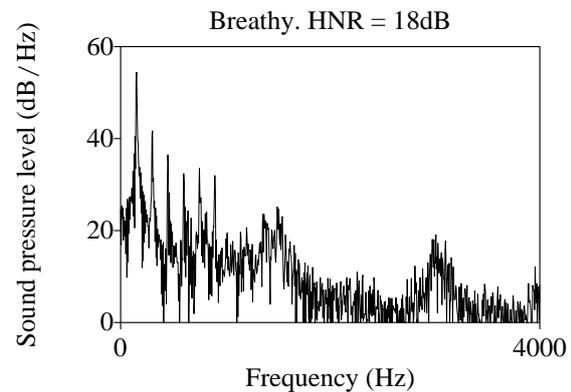
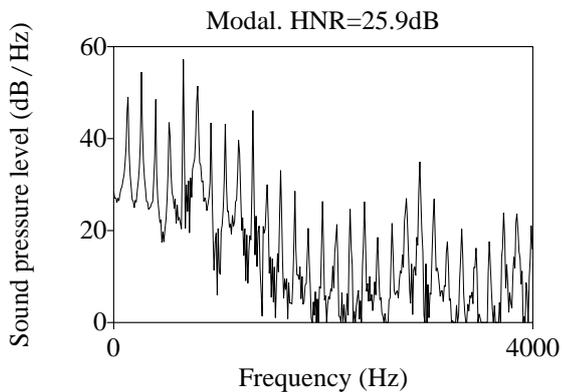
..... a healthy speaker can produce a sustained [a] or [i] with a harmonicity of around 20 dB ... Hoarse speakers will have an [a] with a harmonicity much lower than 20dB."

Die folgenden Abbildungen zeigen Spektrum und Sonagramm für den Vokal /a/ jeweils modal und behaucht von der gleichen Sprecherin.

Im Spektrum für Modal, sind die einzelnen Harmonischen über den ganzen Frequenzbereich bis 4kHz leicht zu erkennen. Bei Breathly dagegen ist dies aber nur bis etwa 1kHz möglich.

Dementsprechend sind die Harmonicity-Werte für Breathly deutlich niedriger als für Modal (18dB vs. knapp 26dB). Der gleiche Sachverhalt ist in den Sonagrammen daran zu erkennen, dass für

Modal die senkrechte Strukturierung durch die Glottisimpulse sich bis zu den höchsten Frequenzen erstreckt, während dies sich bei Breathy auf die niedrigeren Frequenzen beschränkt.



Weitere Anmerkungen zu den Sprachbeispielen

(1) Bei Sprachen, die mehrere Stimmtypen kontrastieren, kommt es oft vor, dass z.B Vokale mit dem Stimmtyp “breathy” nicht durchgehend behaucht sind, sondern nur am Anfang oder am Ende (ähnliches für “creaky”).

Viele Beispiele hierzu in Gordon & Ladefoged (2001), sowie Blankenship (2002)

(2) Der Stimmtyp “creaky/gespresst” weist oft sehr unregelmäßige Stimmbandschwingungen auf, was sich im Breitbandsonagramm sehr gut beobachten lässt.

Mehrere Beispiele in Gordon & Ladefoged (2001).

Weitere technische Anmerkungen

(1) Ein technisches Problem bei Maßen für *spectral tilt* besteht darin, dass sie von der Lage der Formanten beeinflusst werden. Deswegen ist es nicht ohne weiteres möglich, die Stimmqualität verschiedener Vokale (z.B. [a] und [i]) miteinander zu vergleichen. Es sind in der Literatur Korrekturverfahren vorgeschlagen worden, um dieses Problem zu umgehen (s. z.B. Garellek & Keating). Die entsprechenden Maße werden oft mit ‘*’ gekennzeichnet (z.B. H1*-H2*, statt H1-H2). Diese Möglichkeit hat das Programm VoiceSauce von Keating et al., Praat aber zur Zeit nicht. Deswegen ist es bei Praat im Moment besser, die Untersuchung auf eine Vokalkategorie (am besten [a]) zu beschränken¹.

(2) In neueren Arbeiten der UCLA-Gruppe (z.B. Blankenship, 2002; Garellek & Keating, 2011) wird ein weiteres Maß verwendet, um den Grad der Periodizität des Signals zu bestimmen: **Cepstral Peak Prominence**. Weil dieses Maß zur Zeit in Praat nicht verfügbar ist, gehen wir auf die dem Maß zugrundeliegende Signalverarbeitung hier nicht näher ein (weitere Erläuterungen hierzu in Hillenbrand & Houde, 1996, v.a. Fig. 2).

(3) In der Stimmpathologie wird oft “Jitter” als Maß verwendet, um kleinere Unregelmäßigkeiten in der Schwingungsfrequenz zu erfassen. Bei einer heiseren Stimme sind höhere Werte zu erwarten. Technische Hintergründe hierzu in Boersma (2009).

Literatur

- Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *J. Phonetics* 30, 163-191.
- Boersma, P. (2009). Should Jitter Be Measured by Peak Picking or by Waveform Matching? *Folia Phoniatrica*, 61, 305-308.
- Garallek, M. & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *J. Int. Phon. Assoc.* 41(2), 185-205
- Gordon, M. & Ladefoged (2001). Phonation types: a cross-linguistic overview. *J. Phonetics* 29, 383-406
- Hillenbrand, J. & Houde, R. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *J. Speech Hearing Res.* 39, 311-321.
- Klatt, D. & Klatt, L. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. America* 87(2), 820-857.

¹In diesem Handout verwende ich die Schreibweise von Garallek/Keating. Für H1-H2 findet man bei Gordon/Ladefoged die Notation h2-f0, also mit umgehrtem Vorzeichen! (Und etwas irreführend, weil es nicht um die *Frequenz* von f0 geht.) In der Literatur finden sich auch weitere Varianten, z.B. statt H1 mit H2 zu vergleichen, wird H1 mit der stärksten Harmonischen im Bereich von F1 oder F2 verglichen.