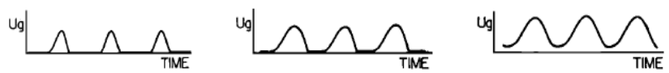
Analysis of Aspiration Noise

Risto Holopainen INF3460, 2009
http://folk.uio.no/ristoh/aspiration/index.html

An important aspect of voice quality is aspiration noise. Its source is turbulence in the vocal tract. A measure called F1F3syn has been implemented for analysis of aspiration noise.

Voice Quality

Perceptual qualities of voices, such as breathy, creaky, harsh, or rough, have their physiological correlates. Tense vocal folds produce a pressed voice, slack vocal folds produce a more breathy sound. When the vocal folds are pressed tightly together, they emit pulses of short duration, and when they are loosened, the length of pulses increase. For a vowel with an aspirated quality, the vocal folds do not close completely, and the excitation approaches a sinusoidal shape. The narrow constriction also gives rise to turbulent noise.

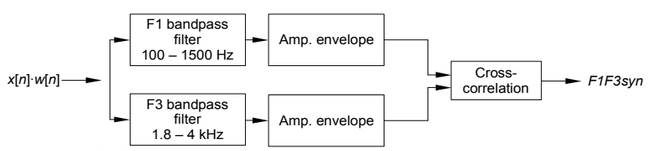


From left to right: pressure waves corresponding to (1) pressed voice (2) modal, and (3) breathy. Illustration from Klatt & Klatt [2].

Aspiration noise is thus characterized by a relatively strong fundamental frequency in the spectrum, and turbulent noise. Ishi [1] suggested two attributes to be analysed for automatic detection of aspiration noise:

• Difference in dB between the strongest partial in the first and third formant regions, called A1A3. Positive values indicate stronger first formant.

• The correlation coefficient between the amplitude envelopes of the first and third formants, called F1F3syn.



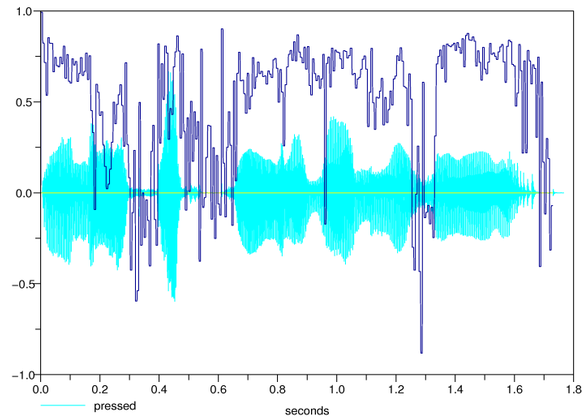
Outline of F1F3syn analysis method.

The F1F3syn measure is calculated as follows:

• The sound file is band-pass filtered with an FFT and separated into two signals corresponding to the first and third formants.

• Momentaneous amplitude envelopes are calculated with a Hilbert transformer, and smoothed with a FIR filter over 1 ms.

• The correlation coefficient of the two envelopes is calculated over each analysis window.



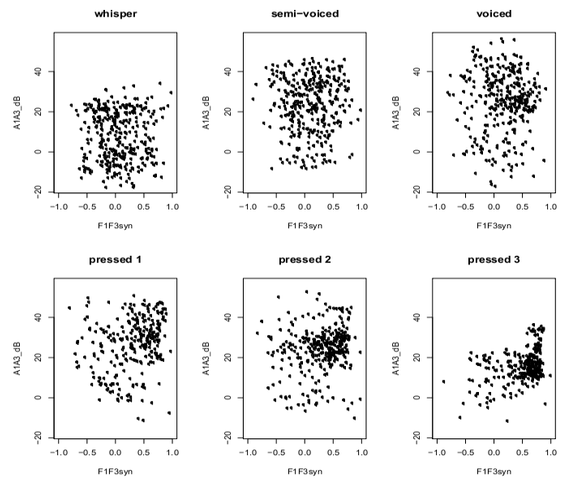
F1F3syn (dark blue) and waveform of a recording with pressed quality.

Analysis of Vocal Sounds

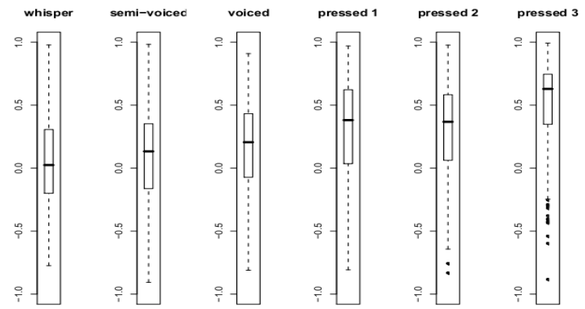
The analysis algorithm was tested with six recordings of the phrase «*Il est sorti avant le jour*», by a female speaker from the VOQUAL database [3]. These included whispered, semivoiced, voiced, and three pressed versions. It was found that:

• Fricative consonants (e.g. /f/ and /s/) are not easily distinguished from aspiration noise, and contribute to a low value of F1F3syn.

• When taking the average over the whole sound file, a correspondance can be seen with perceived degrees of aspiration noise (cf. the box plots).



Scatter plots of F1F3syn vs. A1A3 for six recordings with differing voice quality.



Box plots of the six recordings, showing a tendency of the average of F1F3syn to grow for increasingly pressed voice.

Sound Synthesis with given F1F3syn value

The analysis procedure was complemented by a sound synthesis experiment, with the aim to generate sounds that could attain arbitrary values of the F1F3syn measure. An additive synthesis model was devised as follows: The sound is split into two formant regions (F_1 and F_3), both containing harmonic partials with the same fundamental f_0.

$$F_i(n) = \sum_{k=K_i}^{L_i} a_k \cos(2\pi k f_0 n / f_s),$$

where f_s is the sample rate. All partials belonging to the F_3 formant were delayed by the same amount δ, ranging from 0 to one full period, f_s / f_0:

$$x(n) = F_1(n) + F_3(n - \delta)$$

Half a period's delay gives a minimum value of F1F3syn (–0.5), while no delay gives the maximum value (1.0). Unfortunately, these differences are imperceptible!

Another approach would be to introduce variable amounts of *shimmer* (random amplitude fluctuations) in the upper formant, which does influence the F1F3syn value, but not in such a clear-cut way as the above method. On the other hand, shimmer gives a good perceptual resemblance of turbulence noise.

References

[1] Ishi, C. T. (2004). A New Acoustic Measure for Aspiration Noise Detection. In Interspeech 2004, pp. 941-944.

[2] Klatt, D. H. & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America, 87, pp. 820-857.

[3] http://www.limsi.fr/VOQUAL