Francois Thibault
21/02/2003

Singing voice breathiness estimation using sinusoidal modeling

Introduction

The estimation of breathiness of singing voice signals have been investigated by many researchers since the early 50's because it plays an important role in the perceptual characteristic of a voice. Although largely speaker and register related, the trained singer does have an intimate control over the dimension of sound defined as going from 'pressed' to 'breathy'.

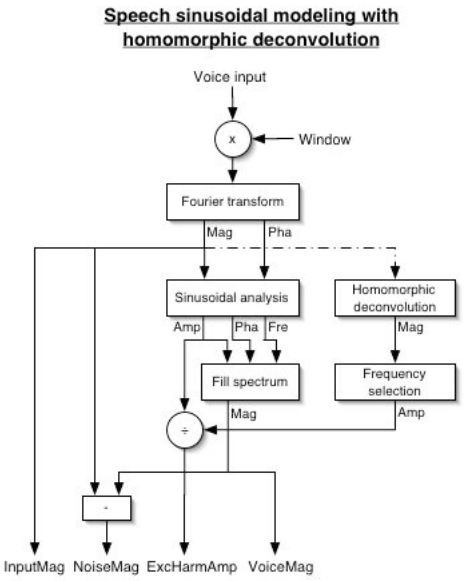
Acoustical relations

In their investigations, all researchers noticed that the breathiness of a voiced was closely related to the shape of the glottal flow pulse through the vocal cords. The glottogram, a representation of this airflow through the vocal cords has been derived according to 2 methods to analyze the breathiness of a given voice. The first method uses a sophisticated mask that measures the airflow at the mouth, and inverse filters this signal to obtain the variation of the glottal flow in time. The second method, privileged here for it's non-intrusiveness is less accurate (the dc offset cannot be measured) but more usable on a large scale. It consists of deriving the glottal waveform by inverse filtering the contribution of formants on the acoustic pressure signal (the recorded sound).

The shape of the glottal waveform (it's evolution) can be described in the spectral domain. Therefore it is possible to derive features from the spectrum of the voice excitation waveform that are linked to the breathiness of the sound. More particularly, researchers observed for a breathy voice: a stronger fundamental compared to the second harmonic (a higher H1/H2 ratio), a faster decrease of the amplitude of harmonics (a lower Harmonic Richness Factor), and a more power of the signal contained in noise (a higher noise power).

Voice sinusoidal modeling

The features mentioned above can all be obtained from a sinusoidal model adapted to voice as described by McAulay and Quatieri (1986). The model they proposed used homomorphic deconvolution to estimate the contribution in magnitude and phase due to the vocal tract filter, and were able to cancel them to obtain a sinusoidal model of the voice excitation. A simplified model of their work can be achieved here because the features we wish to extract do not require the computation of the phase contribution of the vocal tract, which might be difficult to obtain. This crucial computation step is laid out on figure 1 below.



Feature extraction

From the magnitude spectra of the input sound, the re-synthesized sinusoidal component, and the noise component, along with the inverse filtered calculated amplitudes of the voice excitation we derive the 3 features mentioned earlier to be acoustically salient to breathiness. We also derive the voiced/unvoiced estimation that will be used later to gate the estimation to contribute only the voiced part of the sound to the overall breathiness sensation estimation.

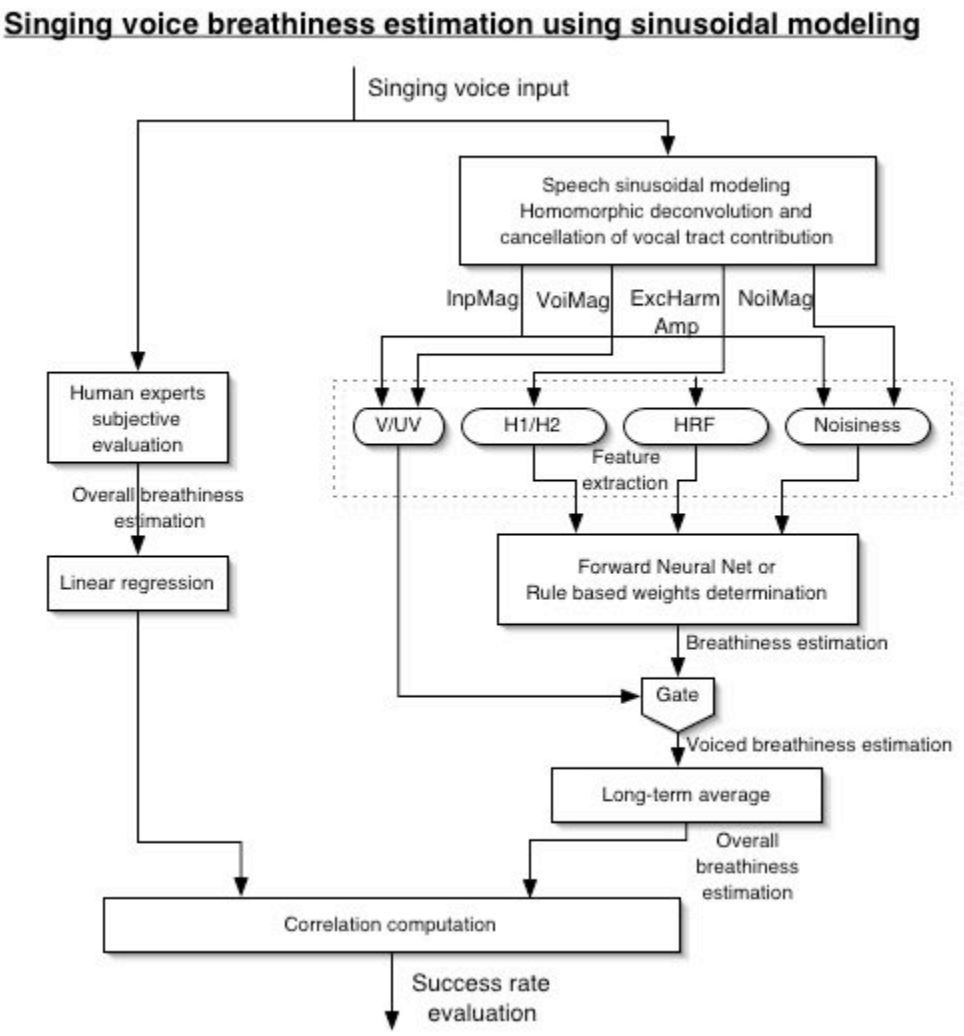
The 3 salient features for breathiness estimation are obtained as follows:

$$\text{H1/H2} = \frac{\text{Amp}_1}{\text{Amp}_2} \qquad \text{HRF} = \frac{\sum_{k=2}^{K} \text{Amp}_k}{\text{Amp}_1} \qquad \text{Noise power} = 20 \log_{10} \sum_{n=0}^{N-1} \left| X_R(k) \right|$$

Overall breathiness estimation

From the instantaneous features derived, we use a set of rules to determine the various weights that should be accorded to each of the features to provide the best measurement possible. This measure obtained for each frame is then averaged over all voiced frames for the duration of the sample to provide one overall breathiness measurement of the sample.

This measurement will be compared to the same input evaluated by many human experts for their breathiness. Using a linear regression of the human perceptual data, a correlation between these measures and the computer estimated breathiness should assess of the success of the measurement. The overall procedure is shown on the figure below.



References

1. Childers, D.G., *Vocal quality factors: Analysis, synthesis, and perception*. Journal of the Acoustical Society of America, 1991. **90**(5): p. 2394-2410.
2. Eskenazi, L., *Acoustic correlates of vocal quality*. Journal of Speech and Hearing Research, 1990. **33**: p. 298-306.
3. Serra, X. *Sound Transformations Based on the SMS High Level Attributes*. in *Digital Audio Effects Workshop*. 1998.
4. Sundberg, J., *Estimating perceived phonatory pressedness in singing from flow glottograms*. Speech, Music and Hearing, 2002. **43**: p. 89-96.