

COMPUTATIONAL-TOOLS FOR THE ANALYSIS OF ACOUSTICAL AND PHYSIOLOGICAL CRITERIA OF THE SINGING VOICE

Ingrid Bendl

Institute of Musical Acoustics
University of Music and Performing Arts Vienna, Austria
ingrid.bendl@mdw.ac.at

ABSTRACT

Tools and methods for the calculation of acoustical, musical and physiological criteria using a sound file from the singing voice are presented, as well as tools for the automation of the time-dependant calculation processes used for analysis and statistics. The difference between recordings in an anechoic chamber and a simulated condition with reverberation is calculated using a soundfile recorded in the anechoic chamber and compared to its derivative altered by a hall simulator.

The tools presented are Praat, Wavesurfer, SNDAN (used for Sound Analysis), TAP, ProToo (an automation tool with a graphical surface) and matlab Scripts like VoiceSauce (HNR, spectral tilt, formants) and Aparat (Inverse Filter).

The acoustical criteria described are long time average spectra, α -factor, and time dependant probabilities of single vowels shown in the parameters loudness (rms), brilliancy (normalised spectral centroid), vibrato (pitch), Formants (LPC), and the spectra. Physiological criteria can be represented in spectral tilt (H1-H2, H1-A1, H1-A2, H1-A3), HNR and glottal parameters taken from inverse filtering.

1. INTRODUCTION

Good musical qualities show themselves in a complete sound that harmonises with those the singer is singing with. Singing, where the sound of the singer, the sound of the orchestra and the singer's physiology all fit together, has become an important factor in the development of classical singing tradition.

If the singer is accompanied by the piano he will adapt his interpretation to an optimal mixture of the common sound produced.

There are many aspects the singer will react to such as colour, loudness, brilliancy and interaction with his co-musicians. These affect development of the quality of the sound in the time-domain. The singer is likely to adapt his singing to feedback according to reverberation and early reflections

Therefore, time dependency plays a big role in describing musical phenomena.

2. SINGING VOICE ANALYSIS TOOLS

2.1 PRAAT [1]

This is a tool for recording and doing sound analysis, specialising in speech analysis, syntheses and manipulation. It supports the labelling and segmentation of files. It controls the management for listening tests. It also allows one to use learning algorithms and statistical tools.

You will see the speech waveform (DC level) on the top and the spectrogram (frequency in the vertical axes and energy in the intensity of gray or black colours) on the bottom. On the horizontal axis the time is displayed.

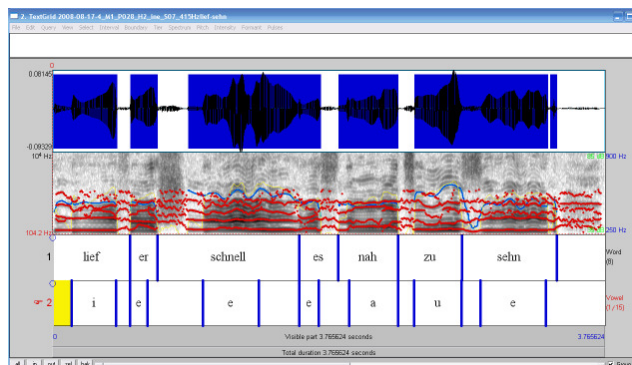


Figure 1: Spectrogram, rms, formantpeaks and pitch curve. The DC-values and the text-grids according to the sound are arranged in windows above and below the spectrogram.

If you click on a text the corresponding part of sound will be marked, and if you click on the grey area below the sound will be played. Praat allows you to use a text grid which indicates the point of beginning and ending of the words, vowels or other interesting parts of the sound files. There are two kinds of text markers, point tiers and interval tiers.

2.2 WaveSurfer [2]

This is an open Source Tool for sound visualization and manipulation. Typical applications are speech and sound analysis and sound annotation and transcription. It can be used with custom plug-ins and other applications

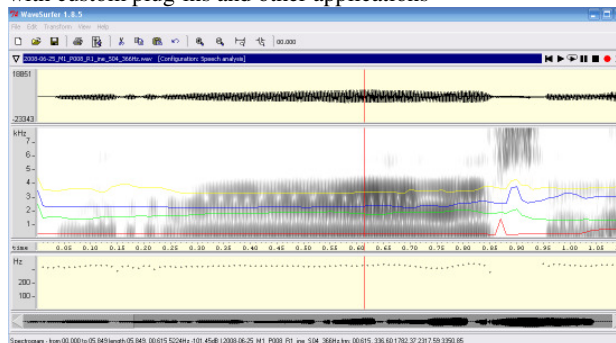


Figure 2: Graphical display from WaveSurfer: Formantpeaks, DC-curve and pitch

2.3 BIAS, VIAS and TAP [3]

BIAS and VIAS are tools for impedance measurement. The optimizer allows simulation of bore-profiles according to a

measured or simulated sound and it can be used for simulation of the vocal tract-shape. TAP is a part of VIAS. It can be used for analysis and the synthesis of sounds. It is command-line based. The script for calculating the α -Factor is a quick and simple process.

2.4 ProToo [3]

This tool allows the controlling of every script-based tool/programme. Implementation of matlab-scripts like Aparat, VoiceSauce, and of tools like TAP, PRAAT, SNDAN and others has been done. It is possible to adapt it to any kind of project because its structure is based on regular expressions.

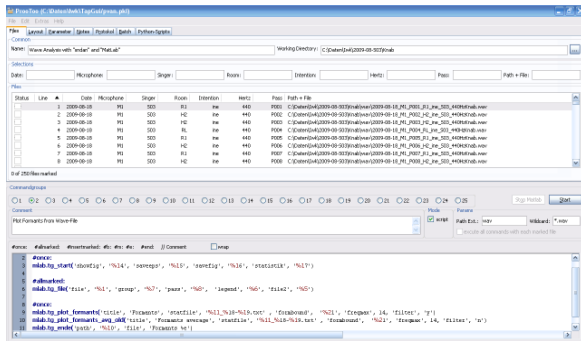


Figure 3: Graphical Display of ProToo

2.5 Aparat [4]

Aparat is a matlab-based computational tool to estimate the glottal flow by inverse filtering of microphone-recorded signals. It implements the inverse filtering methods DIF and IAIF. Aparat, like VoiceSauce, can be used with a matlab runtime-version

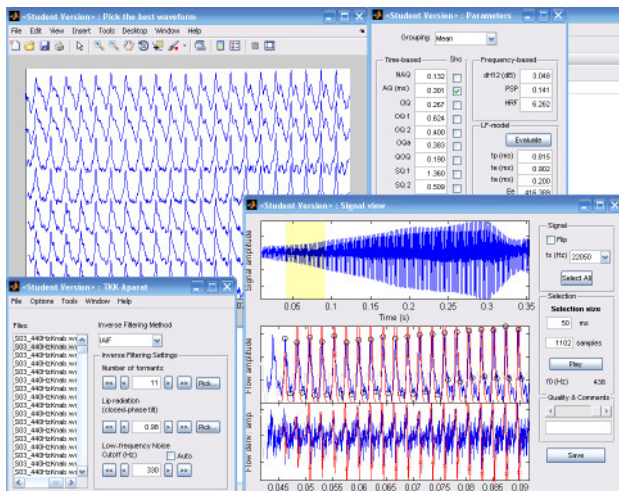


Figure 4: Aparat application for the vowel ‘a’ taken from ‘Knab’ from the first verse from Schubert’s ‘Heidenröslein’

2.6 VoiceSauce [5]

VoiceSauce matches time-dependant calculations of spectral tilt, HNR, formants and F0. It uses the Snack Sound Toolkit to find the frequencies and bandwidths of the first four formants, using the covariance method, pre-emphasis of .96, window length of 25 ms, and frame shift of 1 ms (to match STRAIGHT) as defaults. Harmonic spectra magnitudes are computed pitch-synchronously over a 3-cycle window.

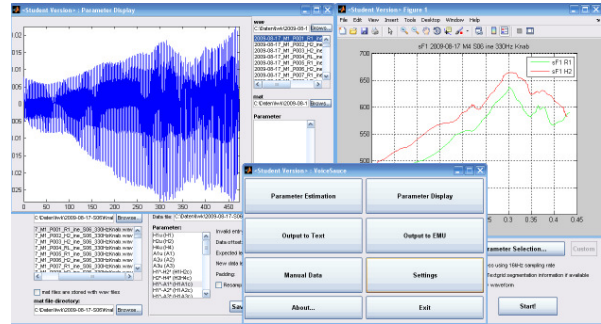


Figure 5: Graphical user surface of VoiceSauce and in the right corner results of exported data from the vowel of ‘Knab’ averaged and displayed in an automated process

2.7 R [6]

R is a tool for statistics and graphical imaging of data. It can be scripted.

3. LABOR-EXPERIMENTS VERSUS REAL CONDITIONS

3.1 An example of the effect of reverberation on the measuring of results

A female singer sang the first verse from ‘Heidenröslein’ by Franz Schubert in the unechoic chamber. She got feedback through headphones from a hall simulator which had altered her original signal. The loudness of the feedback was adjusted according to the singer’s sense of her natural volume. The recorded original signal was altered with the same hall simulator settings, rerecorded and compared to its original. The altered signal represented a realistic hearing experience. ‘Original signal’ means the signal produced by the singer recorded in the unechoic chamber at a distance of 2.0m.

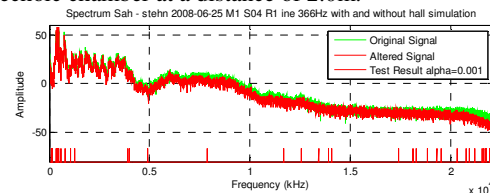


Figure 6: Comparison of an altered and an original sample of ‘Sah to stehn’ from Schubert’s ‘Heidenröslein’

In Fig. 6 it is possible to see the diagram of averaged LTAs measured in two conditions. An indicating graph marks the frequencies where a difference in loudness can be shown with a significance level of 0.001 by an alternative value which is higher than the value of the baseline. The number of samples is 5 pro condition. The delta is set to 1 dB because this represents the discrimination threshold.

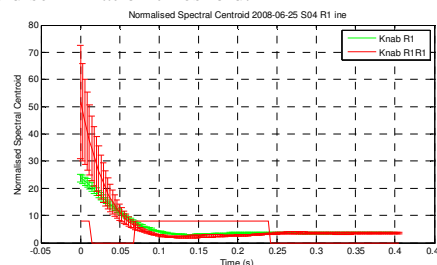


Figure 7: Normalised Spectral Centroid averages of reverberated vowel-samples from ‘Knab’ (Knab R1R1) and its original signals (Knab R1) and indicating mask for statistical results ($\alpha=0.05$)

In the example in Fig. 7 a sound recording mixed with its simulated reverberation gives the possibility to visualize differences of the spectral centroid between the sound the singer is producing and the simulation of what she was hearing at the same time. When the singer sings in a concert hall a comparable situation between the produced sound and the sound impressions coming back to the singer's ears will occur.

3.2 Statistics

For statistical comparison it is necessary to ensure that the parts to be compared are sung in the same tempo so that they are truly comparable. To avoid this problem one takes very short sections of the music such as single vowels. Another option could be stretching time by interpolation or by reducing the number of sample by averaging parts of the data to make them comparable.

When organized by automation techniques statistical evaluation can be done at each point of time. A freeware tool for this purpose is R [6]. Matlab [7] is also useful for statistics.

3.3 Acoustical Analysis

3.3.1 Spectra

LTAs are so-called long time average spectra. An example of two LTA averages and their statistical comparison is shown in Fig. 6. The indicating mask has a horizontal baseline and the significant statistical results are indicated by a higher value in the indicating graph. LTAs are calculated with FFT-analysis.

A spectrogram, or sonogram (for singing voice), shows intensity variations in the dimension of color and the time and frequency on the x and y axes.

If spectra are calculated with pvan or mqan algorithms, the frequency resolution is much better compared to FFT, also when the time resolution is high because of an interpolation done in the calculation process.

Time and intensity curves for every single partial can be plotted and compared statistically, as well as spectra of selected frequency bands.

3.3.2 RMS

The rms indicates the effective energy at each point of time, or when averaged the average value over time. It is possible to calculate the rms of the harmonic parts of the spectrum or of the whole sound. Rms refers to the effect of the sensation of loudness, but there are other criteria to loudness sensation which are not considered in this paper.

3.3.3 Spectral Centroid

The spectral centroid is the weighted average of the spectrum. Its dimension is frequency and the normalized spectral centroid is reached by division with f_0 and thus represents the number between which partials the spectral centre can be found. Spectral centroid is correlated to the brilliancy of a sound.

3.3.4 Pitch and Vibrato

Singing voice samples often have a vibrato and it is therefore sometimes difficult to get plausible results for spectra. Methods have been developed for pitch correction specifically when the calculations of the spectra are done pitch-dependant as in the pvan and mqan algorithms used in SNDAN or applications like PRAAT and VoiceSauce. If the

calculation of the pitch is incorrect, the result of the higher partials and derived calculations will be incorrect.

Vibrato means the time and frequency dependant behavior of the singing voice. It should be between 5 and 7 Hz and approximately in the range between one semitone above and below the centre frequency.

3.3.5 Formants

Formant calculations are always an estimation that should describe the filter characteristics of the vocal tract. There are problems with high-frequency-tones and with sounds where the first formant is identical with the f_0 . The most common method is the LPC, but there are also other algorithms used for formant-calculations, e. g. DAP – Discrete all pole modeling.

3.4 Physiological Parameters

3.4.1 Spectral Tilt

The variables describing spectral tilt and the data available from EGG-measurements and other techniques observing vocalization were combined so that a prediction about glottal closure and breathiness could be estimated by spectral data only.

3.4.2 Inverse Filtering

A microphone signal can be traced back with the inverse filtering technique to only an estimation of its original glottal flow signal. The parameters of the glottal flow are then calculated from the flow signal and some of them are robust parameters like AQ, NAQ and QOQ because they are related to amplitude characteristics.

3.4.3 HNR

This is the ratio between total energy and energy of noise (both measured in dB). If 99% of the energy is in the periodic part, then the HNR as $10 \cdot \log_{10}(99/1) = 20$ dB. HNR of 0 means that the energy is distributed to equal parts in harmonics and noise. Hoarse speakers will have a harmonicity of an $/a/$ much lower than 20 dB. HNR can for example be calculated with PRAAT and VoiceSauce

4. CONCLUSIONS

The singing voice and its behaviour can be described by any amount of acoustical and physiological criteria. It is always good to compare the results of calculations to a natural representation. Any attempt to describe the nature of singing is incomplete as it focuses only on one aspect of it. Measurements can provide information for therapy and study purposes and they may give an insight into how voice and sensation functions.

5. REFERENCES

- [1] <http://www.fon.hum.uva.nl/praat/>
- [2] <http://www.speech.kth.se/wavesurfer/>
- [3] http://193.171.88.81/ima/HTML/welcome_eng.htm
- [4] <http://www.ee.ucla.edu/~spapl/voicesauce/>
- [5] <http://dream.cs.bath.ac.uk/software/sndan/sndan.html>
- [6] <http://www.r-project.org/>
- [7] Beauchamp, J. W.: Analysis, Synthesis and Perception of Musical Sounds. New York, 2007