



# Audio Engineering Society Convention Paper

Presented at the 132nd Convention  
2012 April 26–29 Budapest, Hungary

*This Convention paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This convention paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see [www.aes.org](http://www.aes.org). All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.*

## On acoustic detection of vocal modes

Eddy B. Brixen<sup>1</sup>, Cathrine Sadolin<sup>2</sup>, and Henrik Kjelin<sup>3</sup>

<sup>1</sup> EBB-consult, Smorum, Denmark  
[ebb@ebb-consult.com](mailto:ebb@ebb-consult.com)

<sup>2</sup> Complete Vocal Institute, Copenhagen, Denmark  
[cathrine@sadolin.net](mailto:cathrine@sadolin.net)

<sup>3</sup> Complete Vocal Institute, Copenhagen, Denmark  
[kjelin@shout.dk](mailto:kjelin@shout.dk)

### ABSTRACT

According to the Complete Vocal Technique four vocal modes are defined: Neutral, Curbing, Overdrive and Edge. These modes are valid for both the singing voice and the speaking voice. The modes are clearly identified both from listening and from visual laryngograph inspection of the vocal cords and the surrounding area of the vocal tract. For many reasons it would be preferred to apply a simple acoustic analysis to identify the modes. This paper looks at the characteristics of the voice modes from an acoustical perspective based on voice samples from four male and two female subjects. The paper describes frequency domain criteria for the discrimination of the various modes.

### 1. INTRODUCTION

#### 1.1. Complete Vocal Technique, CVT

Complete Vocal Technique, CVT, is developed by Cathrine Sadolin over a period of 25 years, in an attempt to encompass all sounds the human voice can produce in one singing technique model that will comply with knowledge from medical and acoustic science [1]. This has been done by: 1) Analyzing and categorizing voice sounds from all kinds of musical styles made by a vast number of singers singing live and published on LP's and CD's within the last 100 years.

2) Developing techniques on how to perform these sounds in a healthy way. 3) Testing and modifying the model and the techniques by using confirmation inquiry based learning with more than 10.000 singers and singing teachers.

#### 1.2. Vocal modes

A vital part of CVT is the categorizing of any voice sound into one of four vocal modes which have different sound characteristics, individual limitations according to pitch, volume, sound characteristics and vowel sounds produced by different settings of the voice box.

**The four vocal modes are:**

**Neutral** – the mode which is a 'non-metallic' usually softly characterized sound that might have a breathy quality to it (with or without air).

**Curbing** – the mode which is a 'half-metallic' slightly plaintive or restrained sound quality.

**Overdrive** – the mode which is a 'full-metallic' - often direct and loud - sound with a more shout like character.

**Edge** – the mode which is a 'full-metallic' light somewhat aggressive sound with a more 'screamy' character.

These modes are actually valid for both the singing voice as well as the speaking voice. They are clearly identified both from listening and from visual laryngograph inspection of the vocal cords and the surrounding area of the vocal tract (Figure 1a-1e).

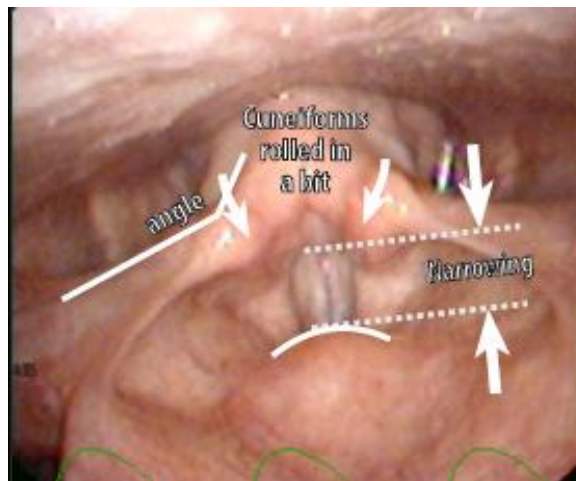


Figure 1c, above: Curbing.

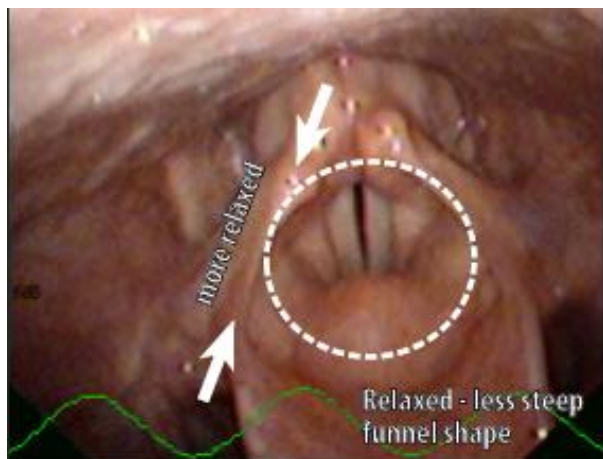


Figure 1A, above: Neutral with air

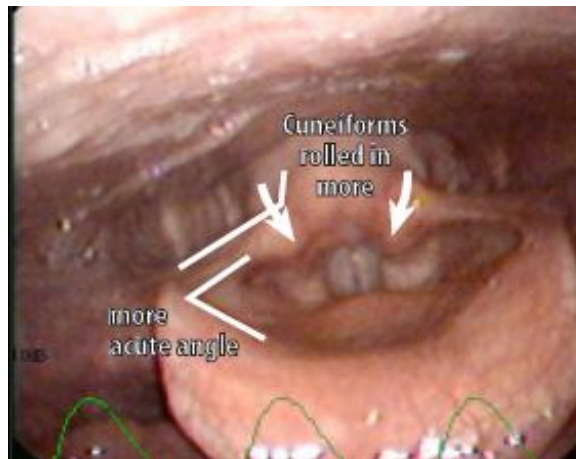


Figure 1d, above: Overdrive

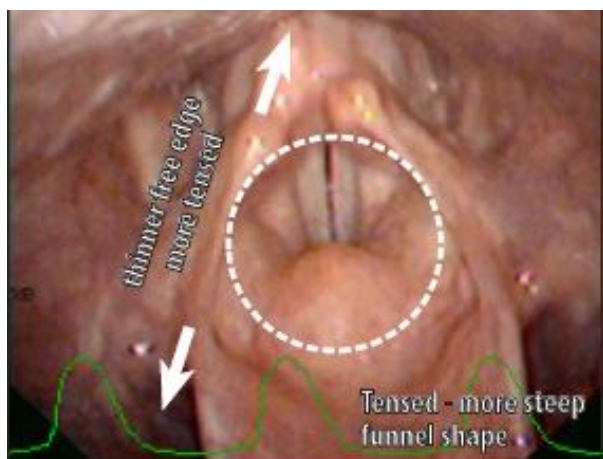


Figure 1b, above: Neutral without air.



Figure 1e, above: Edge

Also the modes to a large extent can be recognized visually from laryngograph waveform recordings [2]. However, until now there has not been presented any analysis on the frequency distribution in these modes. So for this and many other reasons (for instance as a monitoring tool while training and providing parameters for vocal mode apps) it would be preferred to establish and apply the simplest possible acoustic analysis to identify the vocal modes. This could be used as a tool in the training and as definition of mode parameters.

## 2. EXPERIMENTAL SET-UP

The experiment is based on audio recordings carried out simultaneously with laryngograph recordings. All recordings were made at the CVI in Copenhagen. Here singers are recorded to document their CVT abilities. A scheme has been worked out to ensure all students to perform the same modes, effects, etc.

### 2.1. Subjects

For this experiment the recordings from four male and two female singers have been analyzed. The focus has been to illustrate the differences between the four basic modes: Neutral, Curbing, Overdrive and Edge.

### 2.2. Recording

The recordings has been taking place in an isolation booth, a calibrated measurement microphone (DPA 4007) is used at a fixed distance of 30 cm. This microphone has a relative low sensitivity (2.5 mV/Pa) and will accept high sound pressure levels without clipping, which is necessary as some of the singers are capable of producing peak levels of more than 150 dB re 20  $\mu$ Pa at the lip plane!

### 2.3. Tones analyzed

The singers were asked to produce sustained tones, male singers C1 and female singer B1. Also glissandos over 1.5-2.5 octaves were produced by the singers. However, initial measurement showed that the results from these were not comparable in that sense they did not cover the exact same pitch intervals. It is difficult to reach comparable results if not the glissandos have equal length and tone span. But it must be mentioned, that the results showed that the glissando of a trained singer is very consistent throughout the whole tone span.

For this experiment only the sustained single tones were compared. The duration of each tone was typically in the range of 2-6 seconds.

The Neutral mode is normally performed both with and without air added to the tone. This difference is clearly audible. However it is perceived and measured as a small amount of white noise added to the spectrum that does not affect the spectrum as such. For this experiment only data for Neutral mode without air are presented.

## 2.4. Frequency analysis

Initially FFT-based narrowband frequency analysis was performed on the recorded tones. This resulted in clear definition of the frequency contents (harmonics) making it possible to distinguish between three modes: Neutral, Edge and Overdrive/Curbing. However, this also demonstrated difficulties to determine the differences between the modes Overdrive and Curbing. And as a teaching tool it may seem too complicated to explain.

Moving from narrowband analysis to 1/3 octave band analysis improves the visual impression of the spectrum and has been proven helpful with other voice related experiments [3, 4]. Further the resolution is still high enough to display the 1<sup>st</sup> and 2<sup>nd</sup> harmonics of the singers' sustained tone as individual components.

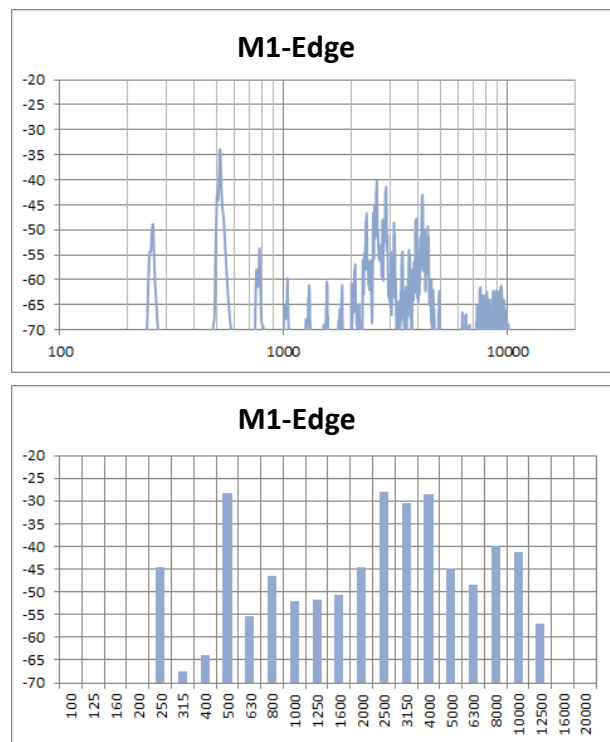


Figure 2 Upper and lower curve show the results of the same analysis. The lower curve (1/3-octave) is calculated from the upper narrowband analysis.

### 3. RESULTS

The results of the frequency analyses of four male and two female singers are shown in the next columns. From the reading of the 1/3-octave charts a number of characteristic and comparative parameters can be extracted. The horizontal axis is frequency [Hz] and the vertical axis is the level [dB]. The dynamic range of the scale is 50 dB. Three frames have been added to the plots for the identification of areas of interest.

#### 3.1. Relation between 1<sup>st</sup> and 2<sup>nd</sup> harmonic

The first frame (left frame) contains the fundamental frequency (1<sup>st</sup> harmonic/pitch/ $f_0$ ) and the 2<sup>nd</sup> harmonic ( $2 * f_0$ ). Both are marked with a circle. It can be seen that the relation between the two repeats itself through the analyses of all six subjects, male or female: Neutral and Curbing exhibit a high 1<sup>st</sup> harmonic and lower 2<sup>nd</sup> harmonic whereas it is the opposite way around regarding Overdrive and Edge.

#### 3.2. Sound energy in the 2kHz-4kHz range

To distinguish between Neutral and Curbing one can look at the raised energy level in in the second frame, ranging from 2kHz to 4 kHz. In the Neutral mode almost no sound energy is produced in this range.

In Edge mode the sound energy in the 2 kHz – 4 kHz frequency band can be compare to the energy of 1<sup>st</sup> and 2<sup>nd</sup> harmonic. In most cases there is a balance between the level in frame 1 and frame 2.

#### 3.3. Sound energy in the 8kHz-10kHz range

The frequency band around 8 kHz-10 kHz can further be considered as an indicator for the Edge mode (third frame). In some cases even the two higher bands (12.5 kHz and 16 kHz) should be monitored.

The difference between Overdrive and Edge is in some cases relatively subtle. However in general it seems like the Edge mode produces the highest content in the frequency range above 8 kHz.

### 4. CONCLUSION

From the findings of this experiment it is possible to define the proportions of frequency content, that can characterise the individual vocal modes. These findings now have to be validated by non-lab experiments and the coefficient limits has to be defined. Also the influence of different voice effects must be determined.

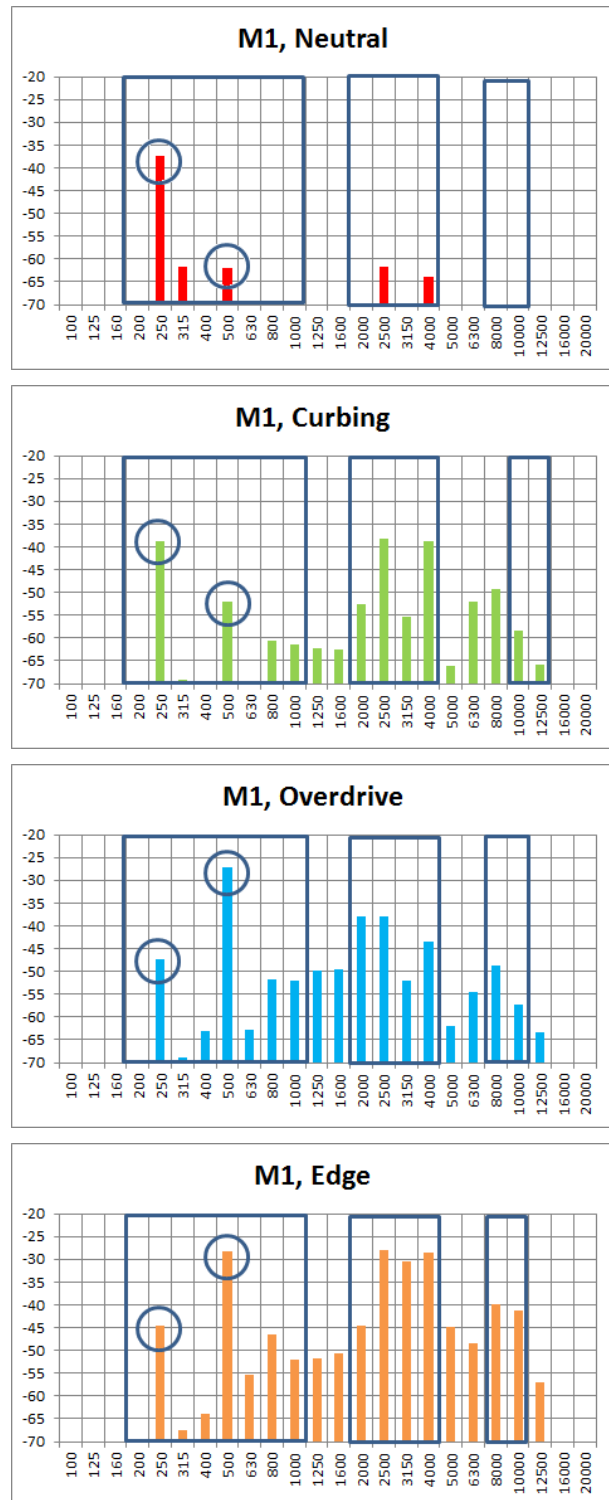


Figure 3a: Male subject no. 1, four modes.

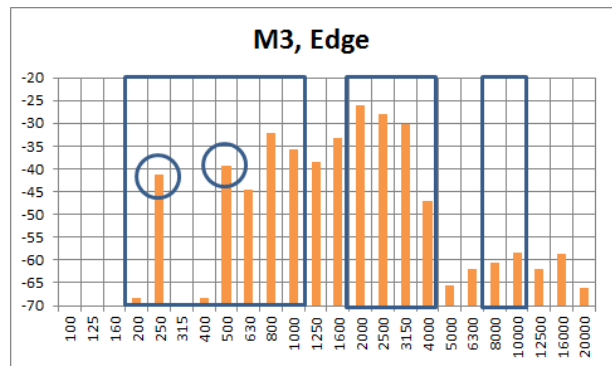
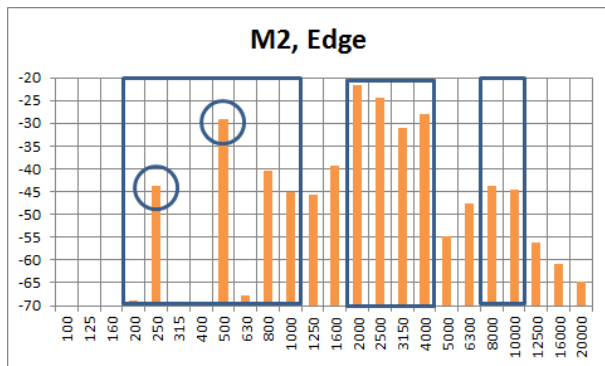
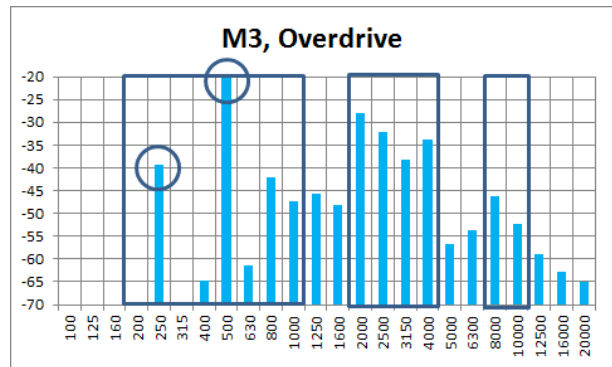
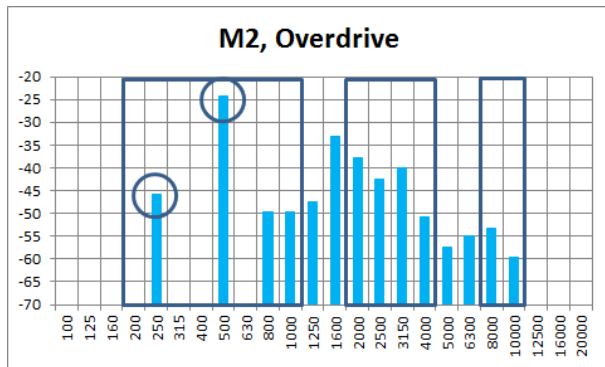
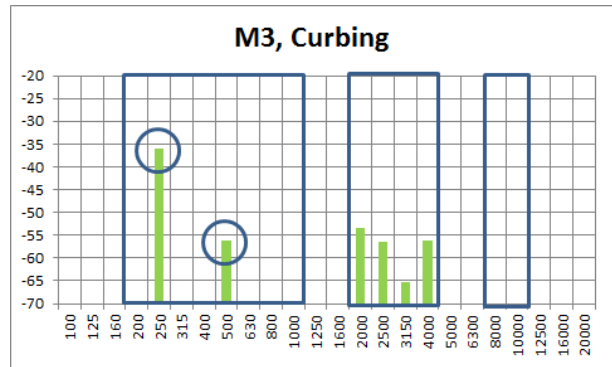
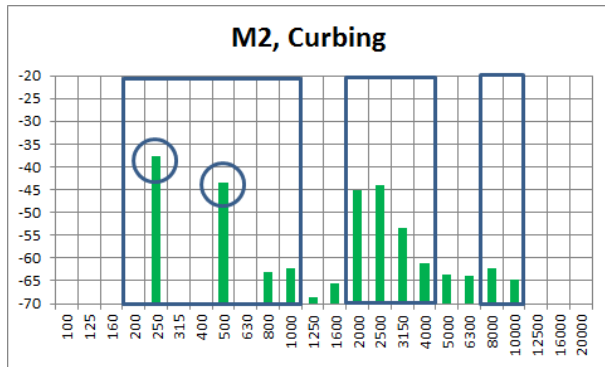
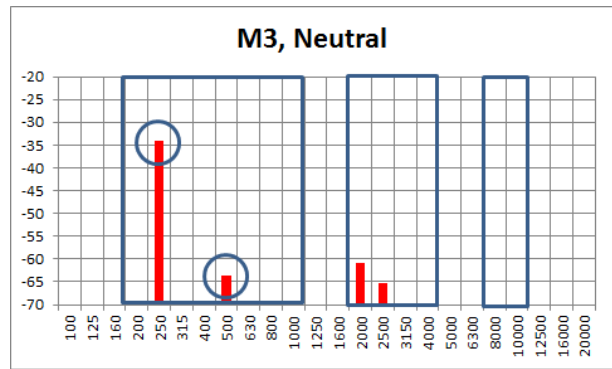
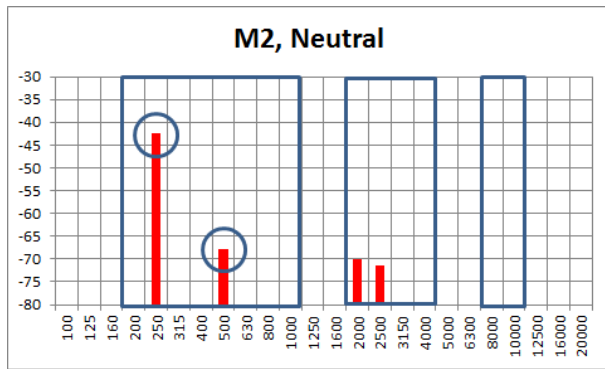


Figure 3b: Male subject no. 2, four modes.

Figure 3c: Male subject no. 3, four modes.

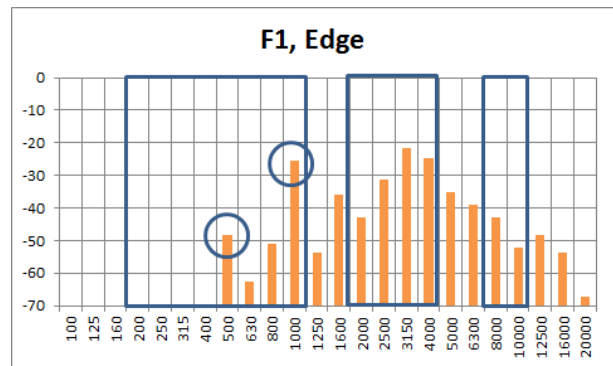
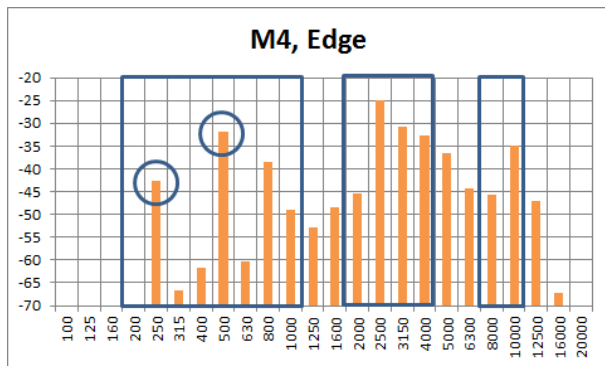
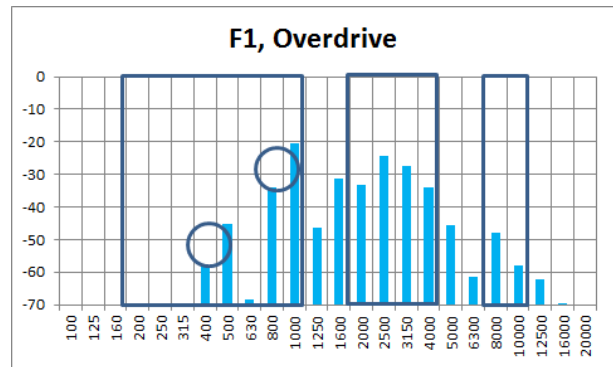
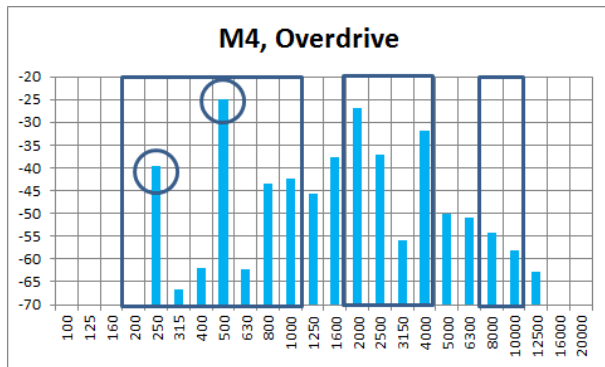
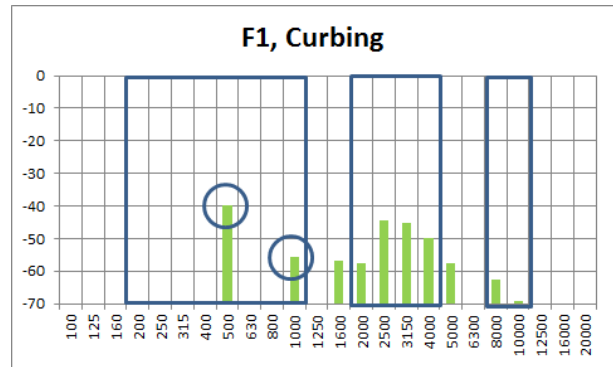
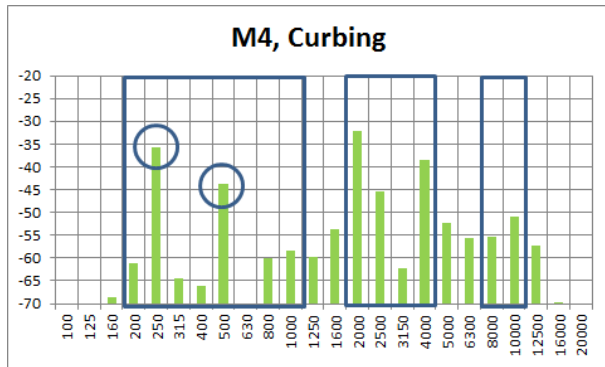
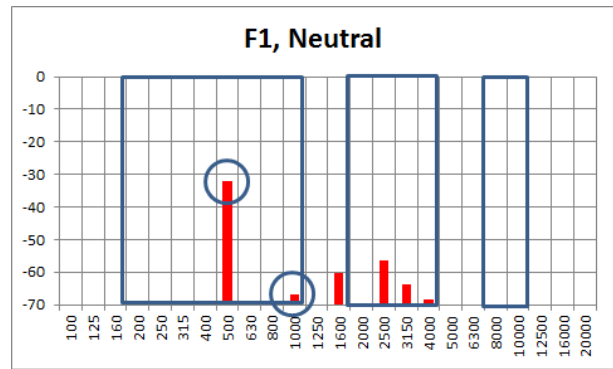
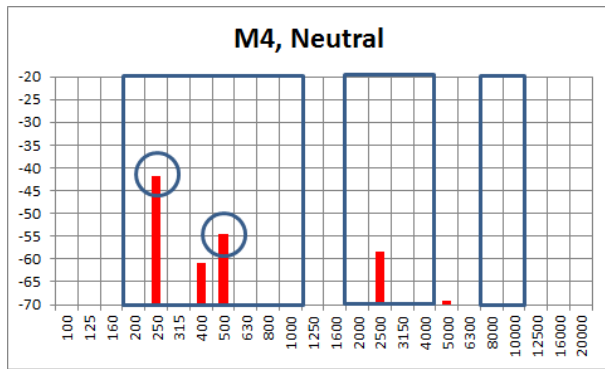


Figure 3d: Male subject no. 4, four modes.

Figure 3e: Female subject no. 1, four modes.



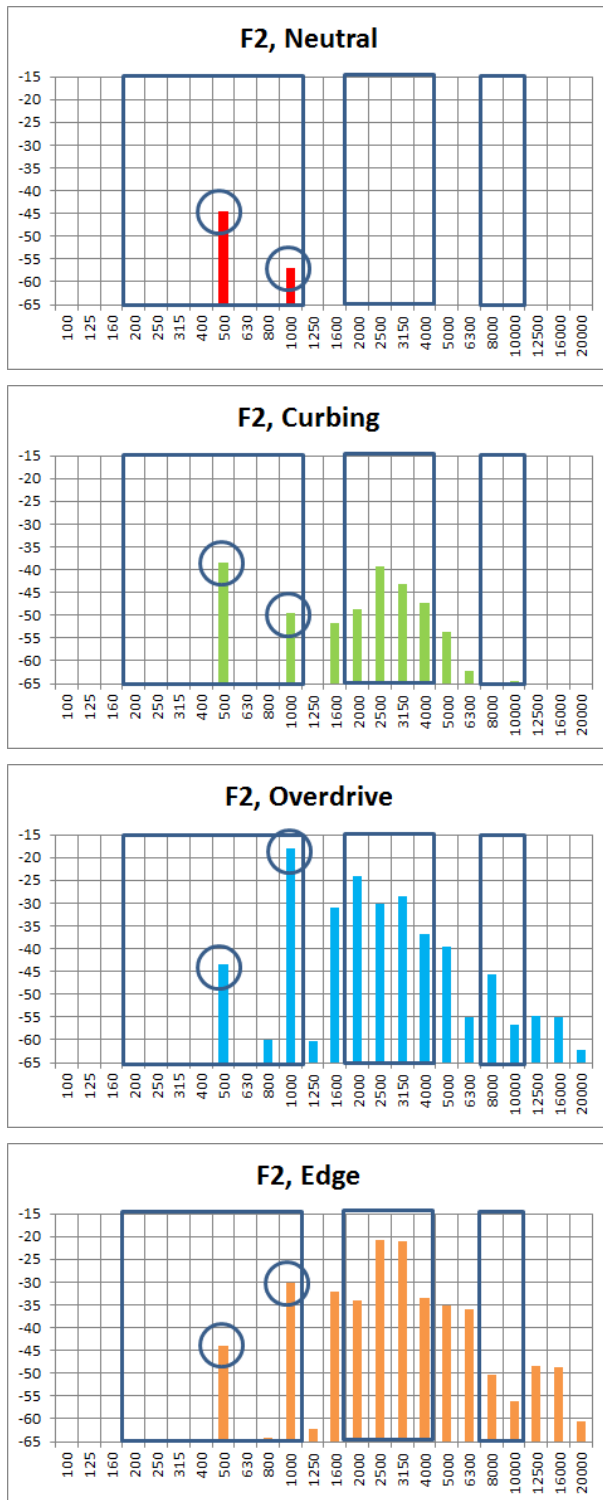


Figure 3f: Female subject no. 2, four modes.

## 5. ACKNOWLEDGEMENTS

The authors want to thank all the singers and teachers of the CVI for raising their voices.

## 6. REFERENCES

- [1] Sadolin, Cathrine: Complete Vocal Technique Copenhagen 2010 ISBN 87-986797-1-6 (English version, also available in Danish, Dutch, Finnish, French, German and Swedish).
- [2] McLashan, Julian: Extreme vocal effects: [http://www.youtube.com/watch?v=nJu\\_BQrfk3E](http://www.youtube.com/watch?v=nJu_BQrfk3E).  
Dr Julian McGlashan from Queens Medical Center in Nottingham presenting a part of a research study made in cooperation with Cathrine Sadolin and Henrik Kjelin from Complete Vocal Institute in Copenhagen.
- [3] Brixen, Eddy B.: Spectral degradation of speech captured by miniature microphones mounted on persons' head and chest. AES 100, preprint 4284.
- [4] Brixen, Eddy B.: Near field registration of the human voice: Spectral changes due to positions. AES 104, preprint 4728.