Audio Engineering Society

Convention Paper

Presented at the 137th Convention
2014 October 9–12 Los Angeles, USA

The importance of onset features in listeners' perception of vocal modes in singing

Eddy B. Brixen[1], Cathrine Sadolin[2], and Henrik Kjelin[3]

[1] EBB-consult, Smørum, Denmark
ebb@ebb-consult.com

[2] Complete Vocal Institute, Copenhagen, Denmark
cathrine@sadolin.net

[3] Complete Vocal Institute, Copenhagen, Denmark
henrik@shout.dk

ABSTRACT

The Complete Vocal Technique defines four vocal modes: Neutral, Curbing, Overdrive, and Edge. This paper report the result of a listening test involving 59 subjects. The goal has been to find the importance of onset and decay features when identifying the vocal modes. The conclusion is that the onset only to a minor degree is responsible for the aural detection of vocal modes.

1. INTRODUCTION

Much relevant research on singers' performance, voice quality, voice analysis, and voice characterization is available [1-21]. However, in general voice training there is still a lack of common language, which describes the physical aspects of the voice across genres and styles.

Complete Vocal Technique - CVT - is developed by Cathrine Sadolin over a period of more than 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 [22]. This has been done by: 1) Analyzing and categorizing voice sounds from all kinds of musical styles made by a vast number of singers 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.

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 and vowel sounds and different settings of the voice box. The four vocal modes are:

Neutral - which is a 'non-metallic' mode that range from a softly characterized sound that might have a breathy quality to it to a stronger clear and non-breathy sound but still without metal.

Curbing - which is a 'half-metallic' slightly plaintive or restrained sound quality.

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

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

It is basically possible to identify the individual modes by ear as the modes are described by their sound. The sound is a part of the artistic expression.

By the use of laryngograph or laryngoscope, the vocal modes can be visually identified. [23].

Acoustical analysis has been applied to samples of singers' voices in order to characterize the spectral distribution of each mode. In the initial work on this [24], it was demonstrated that identification of the individual modes is possible based on simple 1/3-octave band analysis of sustained vowels and comparison of signal energy in groups of frequency bands. In that study the vocal modes were detected from different individuals singing at the same pitch (male: C3, female: B4) and a given vowel (Neutral: EE [i], Curbing: I [ɪ], Overdrive and Edge: EH [ɛ]), providing sufficient formants to support the modes.

In [25] a more comprehensive set of voice samples was analyzed. It was investigated in which pitch-range each of the vocal modes can be defined. Also, it was investigated to which degree the individual vowels are applicable.

Further, the simple model based on spectral discrimination introduced in [24] was tested. The conclusion was that the model was not sufficient in the outer range of the pitch.

Until this point it has been stated that the voice modes were easily identified by listening. However, this was never proven by a listening test.

Further the importance of onset for the identification of a vocal mode was not clear.

So before moving on with the analysis-model, it was decided to carry out a listening test (double blind).

2. LISTENING EXPERIMENT

2.1. Scope

The scope of implementation of the listening test had two objectives:

A: To determine the importance of the tone onset (or attack) for the aural recognition of the vocal mode.

B: To provide non-biased data of the identification of the vocal modes.

2.2. Test samples

The samples for the listening test were selected from recordings of four singers (two male and two female singers). The recordings were carried out in a vocal booth, using a DPA 4007 microphone / Metric Halo 96 kHz 24 bit microphone preamp/converter. During the recording, each tone had to be sustained for at least 2.5 sec by the singer.

The recordings provided a scheme of 320 tones, in combinations of various pitch, vowels, and vocal modes (see Table 1 and 2).

Vowel	**Neutral**	**Curbing**	**Overdrive**	**Edge**
'EH' pronounced as in 'stay' [ɛ]	●		●	●
'O' pronounced as in 'woman' [o]	●	●		
'E' pronounced as in 'sit' [ɪ]	●	●		●
'UH' pronounced as in 'hungry' [ɒ]	●	●		
'OH' as in beginning of 'so' [oʊ]	●		●	
'OE' pronounced as in 'herb' [ɜr]	●			●
'A' pronounced as in 'and' [æ]	●			●

Table 1. Selection of vowels and vocal modes.

Gender	Pitch 1	Pitch 2	Pitch 3	Pitch 4	Pitch 5
Female	C4	E4	Ab4	C5	E5
Male	G3	B4	C#4	E4	G4

Table 2. Selection of pitch for the voice samples.

Each tone was performed at five pitches as listed in Table 2. The recorded tones were edited to individual files and loudness normalized, providing all tones to sound equally loud.

Two subsets of tones selected from the recordings were compiled for the listening test:

Test 1: A test sequence of 120 tones without the onset and the natural ending: A fade in of 250 ms, a sustained tone of 2 s, and a fade out of 250 ms. (See Fig. 1, upper curve). Each tone presented three times followed by a 5 sec pause for the write down of the assessment.

Test 2: A test sequence of 120 tones (including onset and decay, see Fig. 1, lower curve). Each tone presented two times and followed by a 5 sec pause for the write down of the assessment.

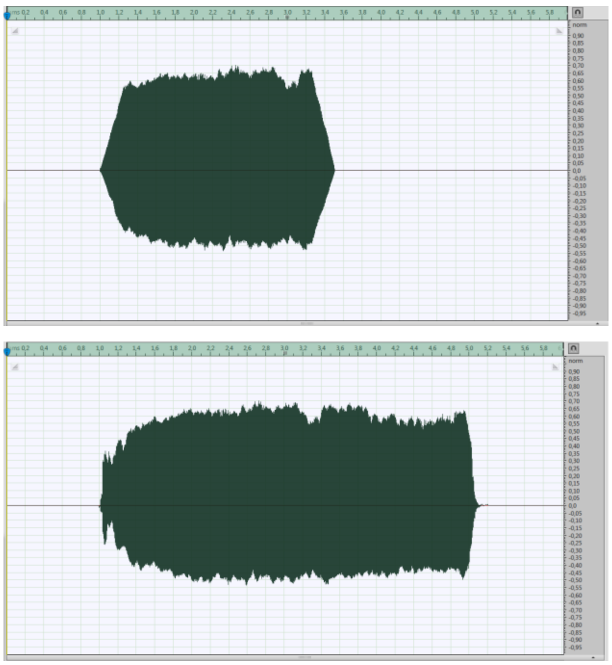

Figure 1 Two ways of presenting the same sound sample: Upper curve: A recording edited to a 2.5 sec sample without onset and decay but with 250 ms fade in / fade out. Lower curve: Full utterance as performed including onset and decay.

The source material for both tests was the same. However, the individual samples applied in random order. This providing different order of appearance in the two tests.

The test was carried out in a lecture hall of the CVI (Complete Vocal Institute, Copenhagen), all listeners present at the same time. The sound samples were reproduced by a set of Adam monitor loudspeakers.

2.3. Subjects

59 subjects attended the listening test. All subjects were graduating students or teachers of the CVI. Hence all subject were trained to perform the identification of the modes. (It should be mentioned, that this test was carried out during the very last hours of the students' graduation day.)

The subjects represent 14 countries, mainly European. The gender of the subjects was 48 female and 11 male.

2.4. Subjects' task

The subjects' task were to identify the vocal mode, the vowel of each tone, and the gender of the performer. For the present study only the vocal mode is of interest.

Each subject would answer by placing checkmarks in a printed scheme.

3. RESULTS

3.1. Correct answers

The result of this test can is expressed by the mumber of correct identification across all samples:

Test	Correct answers [%]
Without onset / decay	78.2
With onset / decay	84.6

Table 3. Test results expressed as the average percentage of correct results based on 59 subjects and 120 samples in each part of the test.

This result show that the onset of a tone improves the possibility of correctly identified vocal mode. However, only to a minor degree. This indicates that the spectrum is the most important factor for the mode identification.

Looking into the distribution of the results there is only a relatively limited spread. The boxplot in Fig. 2 has only one outlier based on a factor value of 1.5 indicating a consistency in subjects' decisions. None of the subjects has a score below 63%.

Fig. 3 show the identification rate of the individual modes, without and with onset. Here it can be seen that Curbing and Overdrive to some degree "suffer" from the absence of onset. Identification of the extremes Neutral and Edge does not exhibit the same dependence.

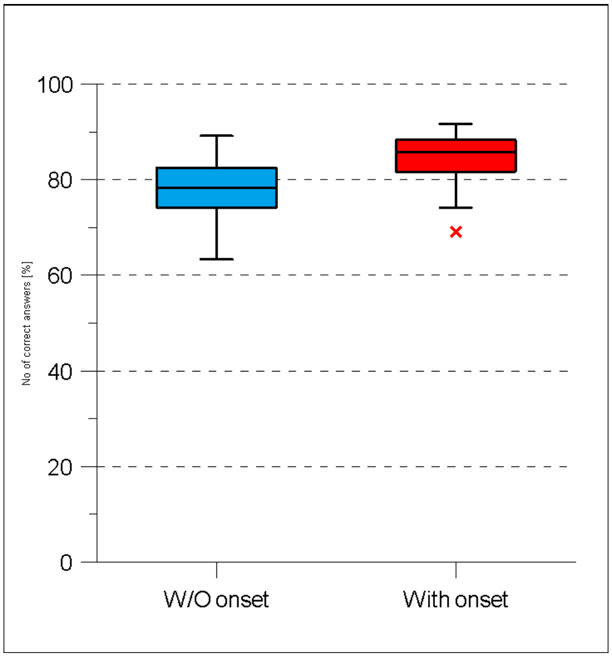

Figure 2. Box plot of results, correct answers [%].

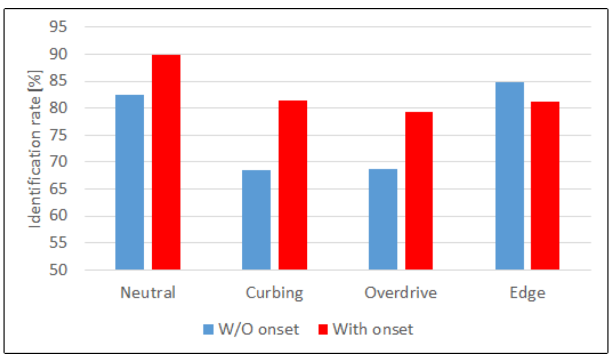

Figure 3. Identification rate [%] of the four vocal modes without onset (left/blue columns) and with onset (right/red columns).

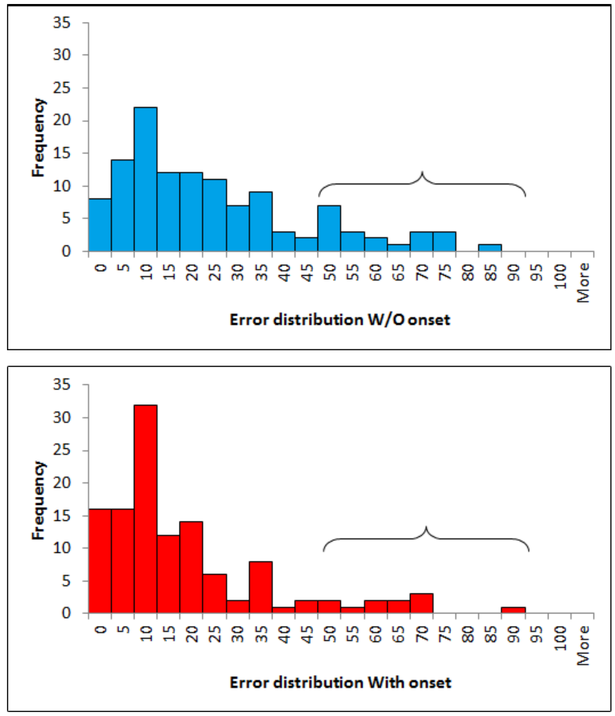

Figure 4. Histogram of failed identifications. Upper curve; Without onset/decay. Lower curve: with onset/decay. Notice that there seem to be a relatively large number of samples that are misidentified by more than 50% of the subjects.

3.2. Error analysis

Analyzing the errors provides an information that is very interesting. It rather relates to the quality of the tested samples than to the subjects' ability to identify the vocal modes.

Fig. 4 show histograms of failed answers across the 120 samples. Taking the high constituency of correct answers into account, it is striking that there are quite a number of samples that apparently are misidentified by more than 50 % of the subjects.

The most extreme example is a tone recorded and initially being performed as Edge. However, in the listening test this was wrongly assessed as Neutral and Overdrive by 51 out of 59 subjects (with onset). Without onset only 16 subjects got it wrong.

There might be several reasons for some of these misidentifications. The most likely is that the border between two modes in some cases is subtle. During the recording session the instructor has not been sharp enough to keep the singers' tone in the center of its range.

In some cases the level normalization of the test samples may cause some confusion to the listeners.

Also, another reason can be the 'anchor effect'. Here the previous stimulus always becomes a reference for the next. In that way the listener can be 'carried away', so to speak.

In listening tests, it is always a challenge to "reset" the subjects between two stimuli.

4. CONCLUSION

A test (double blind) has been carried out in order to find whether the vocal modes can be identified by listening. As a part of this work, it was studied to which degree the onset features are of any importance for the identification.

The conclusion is that it to a large extent is possible to identify the vocal modes by listening. The results also show that onset features and the decay of a tone may improve the identification process, especially in the modes Edge and Overdrive. However, overall this is only of minor importance.

The conclusion is therefore that the spectrum of the sustained tone is the most important parameter for the identification of the vocal modes in singing.

5. ACKNOWLEDGEMENTS

The authors want to thank the students and teachers of the CVI for their enthusiastic participation in the project.

6. REFERENCES

[1] Bartholomew, Wilmer T.: A physical definition of 'good voice-quality' in the male voice. J. Acoust. Soc. Am. 1934;6:25–33.

[2] Sundberg Johan: Vocal tract resonance. In: Staloff RT, editor. Professional Voice: The Science and Art of Clinical Care. 2nd ed. San Diego: Singular Publishing Group, Inc.; 1997. pp. 167–184.

[3] Sundberg Johan: Articulatory interpretation of the "singing formant". J. Acoust. Soc. Am. 1974 Apr; 55 (4):838–844.

[4] Sundberg Johan: What's so special about singers? J. Voice. 1990 Jun; 4(2):107–119.

[5] Sundberg Johan: Level and center frequency of the singer's formant. J. Voice. 2001 Jun; 15(2):176–186.

[6] Sundberg J., Lindblom B., Liljencrants J.: Formant frequency estimates for abruptly changing area functions: a comparison between calculations and measurements. J. Acoust. Soc. Am. 1992 Jun; 91(6):3478-82.

[7] Bloothooft, G.; Plomp, R.: The sound level of the singer's formant in professional singing. J. Acoust. Soc. Am. 1986, vol. 79: 2028-2033

[8] Oliveira Barrichelo V.M.; Heuer R.J.; Dean CM, Sataloff, R.T.: Comparison of singer's formant, speaker's ring, and LTA spectrum among classical singers and untrained normal speakers. J. Voice. 2001, Sep; 15 (3):344–350.

[9] Stone R. E.; Jr.; Cleveland T.F.; Sundberg J.: Formant frequencies in country singers' speech and singing. J. Voice, 1999 Jun; 13(2):161–167.

[10] Cleveland T. F.; Sundberg J.; Stone R. E.: Long-term-average spectrum characteristics of country singers during speaking and singing. J. Voice. 2001, Mar; 15(1):54–60.

[11] Boersma, Paul; Kovacic, Gordana: Spectral characteristics of three styles of Croatian folk singing. J. Acoust. Soc. Am. 119(3), Mar. 2006: 1805-1816

[12] Lundy, D.S.; Roy, S.; Casiano, R.R.; Xue, J.W.; Evans, J..: Acoustic analysis of the singing and speaking voice in singing students. J. Voice. 2000 Dec; 14(4):490–493.

[13] Titze, I.R.; Jin S.M.: Is there evidence of a second singer's formant? J. Singing. 2003; 59(4):329–331.

[14] Nawka T.; Anders, L.C.; Cebulla, M.; Zurakowski, D.: The speaker's formant in male voices. J. Voice. 1997 Dec; 11 (4):422–428.

[15] Howard, D.M.; Brereton, J.; Welch, G.F.; Himonides, E.; DeCosta, M.; Williams, J., and Howard, A.W.: Are Real-Time Displays of Benefit in the Singing Studio? An Exploratory Study, J. Voice, 21(1):20-34, 0892-1997, January, 2007

[16] Winckel, Fritz: A quick test method for the diagnosis of speaker's and singer's voices under stress. AES 53rd Convention, preprint A-4 (1976).

[17] Zwan, Pawel: Expert system for automatic classification and quality assessment of singing voices. AES 121st Convention, preprint 6898 (2006).

[18] Zwan, Pawel; Kostek, Bozena: System for automatic singing voice recognition. J. Audio Eng. Soc. Vol 56, No. 9 (2008).

[19] Presentation of singing voice classification system, http://www.multimed.org/singing

[20] Mayor, Oscar; Bondana, Jordi; Loscos, Alex: Performance analysis and scoring of the singers voice. Proceedings of the AES 35th International Conference, London, UK, 2009.

[21] Sundberg, Johan: My research on the singing voice from a rear-view-mirror perspective. Voice Research

Centre, Department of Speech Music Hearing, KTH, Stockholm. (Invited presentation).

[22] Sadolin, Cathrine: Complete Vocal Technique, 3rd edition, Copenhagen 2012, ISBN 978-87-992436-7-9 (English version, also available in Danish, Dutch, Finnish, French, German and Swedish).

[23] McGlashan, Julian: Extreme vocal effects: 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.

[24] Brixen, Eddy B.; Sadolin, Cathrine, and Kjelin, Henrik: On acoustic detection of vocal modes. AES 132nd Convention, preprint 8620 (2012).

[25] Brixen, Eddy B.; Sadolin, Cathrine, and Kjelin, Henrik: Acoustical characteristics of vocal modes in singing. AES 134nd Convention, preprint 8897 (2013).