"OVERDRIVE" MODE IN "COMPLETE VOCAL TECHNIQUE" A SINGLE SUBJECT STUDY OF VOICE SOURCE AND FORMANT FREQUENCIES

J. Sundberg[1, 2], M. Bitelli, [1,3], A Holmberg[4], V. Laaksonen[5]

[1]Dept of Speech Music Hearing, Scool of Computer Science Communication, KTH, Stockholm, Sweden
[2] University College of Music Education Stockholm, Sweden
[3] Studio Logopedico Voce e Linguaggio, Imola, Italy
[4] Vocal Soul Studio, Stockholm, Sweden
[5] Better Taste Music, Helsinki, Finland

Abstract

"Complete Vocal Technique" (CVT) is an internationally widespread method for teaching voice. It works with four types of voice, or "vocal modes" but their physiological correlates are unclear. This study presents an attempt to analyze voice source and formant frequency characteristics of one of these modes called "Overdrive." A male and a female expert of the Complete Vocal Technique sang a set of "Overdrive" and falsetto tones on the syllable /pæ/. The voice source could be analyzed by inverse filtering in the male subject. Results showed that subglottal pressure, measured as the oral pressure during /p/-occlusion, was low in falsetto and high in "Overdrive." Further, it was strongly correlated with each of the voice source parameters, and these correlations could be described in terms of equations. The deviations from these equations of the different voice source parameters for the various voice samples suggested that "Overdrive" phonation was produced with stronger vocal fold adduction than the falsetto tones. Further, the subject was also found to tune the first formant to the second partial in "Overdrive" tones. The results support the conclusion that the method used, to compensate for the influence of subglottal pressure on the voice source, seems promising to use for analyses of other CVT vocal modes and also for other vocal styles.

Keywords

Subglottal pressure; Voice source; Flow glottogram; Glottal adduction; Formants; Singing

INTRODUCTION

Among the numerous methods used for the teaching of singing, CVT [1] has reached a wide international scatter. The method operates with four basic vocal modes, referred to as "Neutral", "Curbing", "Overdrive" and "Edge." Brixen et al., [2] described the modes in this way:

- Neutral - 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 - a 'half-metallic' slightly plaintive or restrained sound quality.
- Overdrive - a 'full-metallic' - often direct and loud - sound with a more shout like character.
- Edge - a 'full-metallic' light somewhat aggressive sound with a more screamy character.

The modes can be produced in different artistic variants by using different vocal tract shapes, such as varying larynx height, position of the soft palate, opening of the nasal passage, tongue position, width of the mouth opening and of the supraglottal space between the petiole and the arytenoid cartilages. As vocal tract shape determines the frequencies of the vocal tract resonances, i.e., the formants, such changes affect voice quality.

Some attempts have been made to specify the acoustic and perceptual characteristics of the four modes. Brixen and associates [2] analyzed a vowel sung in the four different modes by two female and four male singers. Spectrum analysis showed that the modes differed systematically with respect to proportions of sound energy in different frequency bands. However, in a subsequent study the same authors showed that fundamental frequency and vowel quality affected the acoustic differences between the modes [3].

In a third study the possibility of identifying the four modes perceptually was analyzed in a double blind study. The result showed that to a large extent this was possible and that onset and decay features may improve the identification, especially in the modes "Edge" and "Overdrive" [4].

McGlashan and associates [5] analyzed laryngeal settings, obtained by laryngostroboscopy, EGG and audio for the vowel / ɛ / sustained in "Edge" and in "Overdrive" by ten female and ten male singers at the pitches B4 and C4 (494Hz and 262 Hz, respectively). All subjects were trained in CVT and their productions were checked by one of the authors. The fiberscope data were inspected by two of the authors, who observed that "Overdrive" was produced with less constriction of the lower part of the pharynx than "Edge." Narrow and broadband spectrum analyses, averaged over all female and over all male subjects, respectively, showed spectrum peaks (which the authors refer to as formants) at 2 kHz, 3.8 kHz, and 6 kHz. It was unclear if these differences emanated from the voice source or it they were produced by coinciding formants and partials.

Even though the different modes can be assumed to result from different voice production conditions, no further attempts have been made to describe them in quantitative voice production terms. Voice production can be characterized by the combination of respiratory, phonatory and resonatory properties, controlling subglottal pressure, voice source properties and formant frequencies. Voice source properties can be analyzed by means of inverse filtering, a classical method in voice research [see e.g., 6). It eliminates the effects of the formants, i.e., the vocal tract resonances, from the radiated signal, thus providing the glottal airflow as output. The aim of the present investigation was to find out if voice source analysis by inverse filtering could be used for analyzing phonatory and resonatory characteristics of the "Overdrive" mode in CVT.

METHOD

Data Collection

A male and a female voice teacher, both formally authorized teachers of CVT (see http://completevocal.institute) volunteered as subjects. They sang in "Overdrive" mode and in two degrees of vocal loudness, loud and medium, the syllable /pɔ/ and /pɛ/ on five pitches (C4, E4, Ab4, Bb4 and C5; 262Hz, 330Hz, 415Hz, 466Hz, approximately). In addition, they sang a version of the same syllables, which they referred to as "*dark*" and "*bright*." To gain more varied material with respect to both voice source and subglottal pressure, loud and soft versions of falsetto tones, sung on the same syllables, were also recorded, even though the term "falsetto" is not used in CVT.

The recording setup is illustrated in Figure 1. The audio was picked up by an omnidirectional head-worn microphone at a distance of 7 cm from the lip opening. A flow signal was recorded by means of a Rothenberg mask, in which a thin plastic tube was inserted that fitted into the corner of the lip opening, thus allowing recording of oral pressure. Also, an EGG signal was recorded.

Insert Figure 1 here

Using the *Soundswell* workstation, the signals were recorded as wav-files on separate tracks in a PC, sampling rate 16000 Hz per channel. As the flow mask affected the singers' auditory feedback they first produced each task without the mask and then repeated the task with the flow mask in place.

Sound pressure level SPL was calibrated by recording a sustained vowel sound, the SPL of which was measured at the recording microphone; the value was announced in the recording. Oral pressure was calibrated by recording pressures that were measured by means of a manometer; also these values were announced in the recording.

Analysis

All recordings were analyzed using the custom made computer software *Sopran* (freeware developed by Svante Granqvist, KTH). It contains an inverse filtering program in which the formant frequencies and bandwidths of the inverse filters are set manually. Figure 2 shows the display of the program. In setting the filters, two criteria were used, (i) a closed phase as ripple-free as possible and (ii) a source spectrum envelope as void as possible of local maxima and minima near the formant frequencies. The program displays the resulting flow glottogram and the source spectrum in quasi real-time.

Insert Figure 2 here

The resulting voice source waveform was analyzed with respect to period, closed phase and peak-to-peak pulse amplitude (henceforth PtpAmpl) by means of the *Glottal flow parameter measurements* option of the *Sopran* software. The program then saves the filter settings and the following data in a separate file: file name, time coordinate of the measurement, period, fundamental frequency (henceforth F0), musical note, PtpAmpl, the negative peak amplitude of the differentiated flow glottogram, i.e., maximum flow declination rate (henceforth MFDR), normalized amplitude quotient, i.e. the ratio between PtpAmpl and the product of period and MFDR (henceforth NAQ), the level difference between partials 1 and 2 of the source spectrum (henceforth H1-H2), and the closed quotient defined as the ratio between the closed phase and the period (henceforth Q_{Closed}).

The oral pressure peaks during the /p/ occlusions were measured using the *Measurement / Average* option of the *Sopran* software and used as estimates of the subglottic pressure (henceforth P_{Sub}).

Inverse filtering was applied to the entire recorded material consisting of 360 utterances (2 subjects x 2 vowels, 3 "Overdrive" tones x 2 falsetto tones x 5 pitches x 3 takes). However, the flow glottogram parameters showed a limited scatter within conditions only for the two lowest pitches as produced by the male subject. The voice source data of the female singer were widely and unsystematically scattered, probably because of difficulties to perform the inverse filtering. This problem seemed to be due to a lack of a clearly demarcated closed phase in the resulting flow glottograms. Therefore, the voice source analysis was limited to the two lowest pitches as produced by the male subject.

RESULTS

The relationships between P_{Sub} and F0 are shown in Figure 3. For both singers one octave increase of F0 was accompanied by a doubling of P_{Sub}, approximately. However, the male singer used a much wider P_{sub} range than the female subject. He used about three times higher P_{Sub} for the "Overdrive" tones than for the Falsetto tones. Both singers used higher pressures in "Overdrive" than in falsetto and in the loud versions as compared with the less loud versions.

Insert Figure 3 here

The relationships between sound pressure level SPL and F0 are shown in Figure 4. For the female singer SPL tended to increase linearly with the log of F0, except for the highest pitch C5. However, for the male singer the same was true only for the falsetto tones while for the "Overdrive" tones pressure was almost constant for the lower pitches and lowered for the highest pitch.

Insert Figure 4 here

The flow glottogram data for the three takes showed reasonably similar values as can be seen in Table 1 which lists the percentage variation between the three takes of the same condition for the different parameters. However, for H1-H2 the values refer to the average of the greatest level difference in dB. The values vary between 18.7% and 8.3%. For the same parameter the values for the two pitches vary by less than 5%, for H1-H2 by 0.8 dB.

Table 1. Averages of the maximum difference in the indicated flow glottogram parameters between the three takes of condition. The values are given in percent of the average across the three takes, except for H1-H2 which is the mean of the difference in dB between the highest and lowest values in the three takes.

Pitch	**ptp [%]**	**MDR [%]**	**AQ [%]**	**NAQ [%]**	**H1/H2 [dB]**	**CQ [%]**
C4	18,4	18,7	9,0	8,3	2.6	12,6
E4	*17,4*	*15,4*	*13,6*	*12,8*	*1.8*	*14,2*

The male subject's flow glottogram data for the pitches C4 and E4 are shown as function of P_{sub} in Figure 5. The values were mostly reasonably similar for the three takes of the same condition, the highest and the lowest of the three values for the same condition differing by between 10% and 20%. For H1-H2 this difference was 2 or 3 dB for the low pitch and about twice as much for the high pitch.

Insert Figure 5 here

Each of the five flow glottogram parameters showed a systematic variation with P_{Sub}. This variation could be approximated by 6 means of equations. Selecting the type of equation that yielded the highest determination constant R2 trendlines were determined for each parameter. The equations and their constants are listed in Table 2.

Table 2. Factors (k1) and constants (k2) of the equations used for describing the influence of P_{Sub} on the indicated parameters for the pitches C4 and E4 (F0≈260Hz and 330Hz, respectively).

	k1		k2		R2	
	C4	E4	C4	E4	C4	E4
SPL = k1*ln(P_{Sub}) + k2	0,7714	11,43	69,4	50,4	0,885	0,910
MFDR =k1* P_{Sub} +k2	79,985	37,13	-255	71	0,947	0,951
Ptp=k1*e^(k2* P_{Sub})	0,2327	0,254	0,038	0,025	0,873	0,699
NAQ =K1*e^(-K2* P_{Sub})	0,797	0,258	-0,679	-0,017	0,780	0,488
H1-H2 =K1* P_{Sub} ^(k2)	187,74	306,7	-1,08	-1,07	0,790	0,720
Q_{Closed}=k1*ln(P_{Sub})+k2	0,259	0,13	-0,281	0,048	0,878	0,512

R2 was mostly higher for C4 than for E4, lowest for NAQ (0.488) and highest for MFDR (0.95). In other words, about 95% of the variation in MFDR could be explained by the variation in P_{sub}, so P_{sub} was a main factor in the voice source control.

The parameter values for some conditions showed a systematic deviation from the trendlines. For example loud, dark "Overdrive" on /Ɛ/ tended to show high MFDR and high PtpAmpl, while "Overdrive" medium bright /Ɔ/ produced values below the trendline in PtpAmpl and NAQ and above the trendline in Q_{Closed}. Therefore, the mean of the three P_{sub} values and the mean of the three associated flow glottogram parameter was determined for each condition. Then the equations in Table 2 were used for calculating the trendline values for the observed average P_{sub} value. Finally these trendline values were compared with the observed average parameter values for each parameter and the deviation from the trendline value was calculated. Figure 6 shows these deviations in % of the predicted values.

Insert Figure 6 here

It can be assumed that the subject kept the same phonation type for both pitches analysed. If so and if the deviations from the equations are related to phonation mode the deviations from the trendlines should have the same sign for the two pitches.

This was the case for all parameters except the PtpAmpl; SPL, NAQ, H1-H2, and Q_{Closed} all showed positive values for the "Overdrive" tones and negative values for the falsetto tones.

Thus, for "Overdrive" the values of these parameters were higher than the trendline and for the falsetto tones they were lower. These effects are compatible with the assumption that the "Overdrive" tones were produced with firmer glottal adduction than the Falsetto tones.

The frequencies of the two lowest formants used for the inverse filtering of the three takes of each tone are shown in Figure 7. Both the first and the second formant frequencies, henceforth F1 and F2, were rather high, F1 in /Ɔ/ varying between 460 and 600 Hz and F2 between 850 and 1200 Hz, approximately. In most cases the falsetto versions had a lower F1 than the "Overdrive" versions, probably reflecting a lower larynx position in falsetto. Also, the soft versions showed a lower F1 than the loud versions and the dark version mostly showed lower F1 and/or F2 than the bright versions.

Insert Figure 7 here

The formant data showed some application of the formant tuning principle, i.e. to tune a formant to a spectrum partial. Thus, in "Overdrive," F1 was tuned to the vicinity of the second partial. This can be seen in Figure 8, which shows the F1/F0 and F2/F0 ratios for the two pitches. The F1/F0 ratio was close to 2.0 for the "Overdrive" tones and more varied for the Falsetto tones. In some versions of /Ɔ/ the F2/F0 ratio was close to 4 for both pitches.

Insert Figure 8 here

DISCUSSION

The present study is obviously quite limited as an attempt to describe the phonatory characteristics of a particular phonation type. It may be more relevant to regard it from a methodological point of view.

Voice source characteristics are determined by the combination of laryngeal muscle adjustments and P_{Sub}. Indeed, it hardly appears meaningful to consider voice source characteristics if this pressure is not taken into account. Hence, it seemed advantageous to record voice samples produced with a large variation of P_{Sub}. This was achieved by asking the subject to produce not only "Overdrive" tones but also falsetto tones.

Our flow glottogram analysis concerned the standard measures, MFDR, PtpAmpl, H1-H2, Q_{Closed} and NAQ. MFDR represents the excitation strength of the vocal tract and is hence decisive to SPL. The PtpAmpl is strongly dependent upon P_{sub} but for a given P_{Sub} it must decrease if glottal adduction is increased. In addition, PtpAmpl must also depend on vocal fold length as for a given vibration amplitude a longer glottis will allow a greater airflow than a shorter. This may be the reason why PtpAmpl is not closely correlated with perceived degree of phonatory pressedness/hyperfunction [8]. Further, according to the same reference, H1-H2 is more strongly correlated with phonatory pressedness, firm adduction reducing it, and also Q_{Closed} and NAQ are influenced by phonatory pressedness, the former increasing and the latter decreasing for an increase of pressedness.

The availability of a wide variation range of P_{Sub} values made it possible to describe, in analytic terms, the relationship between each of the analyzed flow glottogram parameters and P_{Sub}. The emerging equations were regarded as a standard against which the parameter values of the different phonatory conditions could be compared. The result seemed systematic, suggesting that "Overdrive" tones were produced with a higher degree of phonatory pressedness/glottal adduction than the falsetto tones. It is worth noticing that NAQ, Q_{Closed} and H1-H2 unanimously supported this conclusion.

“Overdrive” and Falsetto differed systematically with regard to F1 and F2, “Overdrive” mostly having higher F1 than Falsetto. This probably reflected a higher position of the larynx, possibly combined with a wider jaw opening. Also F2 in /ɔ/ tended to be higher in “Overdrive.”

In spite of the very limited material, one subject and two pitches, it is worthwhile to note that the observed voice source differences between “Overdrive” and falsetto seem reasonable. Modal voice is produced with firmer contraction of the vocalis muscle which makes the vocal fold thick and Q_{Closed} high, while falsetto is produced with less vocalis contraction and hence thinner vocal folds and some glottal leakage [9; 10,11]. Thin vocal folds and incomplete vocal fold closure should promote a more sinusoidal flow glottogram, which should correspond to a stronger fundamental, i.e., higher H1-H2 value. The higher values of F1 and F2 for “Overdrive” suggest it was produced with an elevated position of the larynx which often is combined with an elevated degree of glottal adduction.

CONCLUSIONS

The method here used for analyzing flow glottogram data seems promising. By recording widely different types of phonation the paramount influence of P_{Sub} on the voice source could be quantitatively described. After compensating for this influence systematic voice source data could be observed for the two different types of phonation analyzed. Differences in MFDR, in Q_{Closed}, in NAQ and in H1-H2 all suggested that “Overdrive” was produced with more adduction than Falsetto. The results suggest that it would be worthwhile to apply the method to a larger material as well as to other types of phonation.

ACKNOWLEDGEMENTS

This research was carried out during co-author MB’s visit to KTH in the fall of 2015. The work was first presented at the Voce Artistica conference in Ravenna, Italy, in November 2015. The measurements and first draft of the report were made by co-author MB and the analysis jointly by co-authors MB and JS. Support for travel between Sweden and Finland was provided by Kulturfonden för Sverige och Finland.

REFERENCES

[1] C. Sadolin (2014). *Complete Vocal Technique*. Copenhagen: Shout publishing.

[2] E.B.Brixen, C. Sadolin, H. Kjelin. (2012). On acoustic detection of vocal modes. Convention Paper presented at the 132nd Convention of the Audio Engineering Society.

[3] E.B.Brixen, C. Sadolin, H. Kjelin. (2013). Acoustical characteristics of vocal modes in singing. Paper Proceedings of the Audio Engineering Society Convention 134.

[4] E.B.Brixen, C. Sadolin, H. Kjelin. (2014). The importance of onset features in listeners’ perception of vocal modes in singing. Convention Paper presented at the 137th Convention of the Audio Engineering Society.

[5] J.McGlashan, A. Aaen Thuesen, C. Sadolin. (2016). Overdrive and Edge as refiners of “Belting”? An empirical study qualifying and categorizing “Belting” based on audio perception, laryngostroboscopic imaging, acoustics, LTAS, and EGG. Article in press, Journal of Voice.

[6] M. Rothenberg. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. J Acoust Soc Amer 53 1 (1973) 1632-1645.
[7] M. Båvegård, G.Fant, J. Gauffin, J. Liljencrants. Vocal tract sweeptone data and model simulations of vowels, laterals and nasals. STL-QPSR 34(4) (1993) 43-76.
[8] M.Millgård, J. Sundberg, T. Fors.. Flow glottogram characteristics and perceived degree of phonatory pressedness. J Voice 30 (2015) 287-292.
[9] M. Hirano, W. Vennard, J. Ohala. Regulation of register, pitch, and intensity of voice, Fol. Phoniat. 22 (1970) 1–20.
[10] M. Södersten, C. Högset, J. Sundberg. Glottal airflow and subglottal pressure in the falsetto and modal registers of countertenor singers. In A Friberg, J Iwarsson, E. Jansson, & J. Sundberg, eds. SMAC 93, Proc. of the Stockholm Music Acoustics Conference, 1993, Stockholm: Roy Sw Acad Music:, Publ Nr 79. (1994) xx-xx.
[11]J. Sundberg, C. Högset. Voice source differences between falsetto and modal registers in counter tenor, tenor, and baritone singers. Logopedics Phoniatrics Vocology 26 (2001) 26-36.

FIGURE CAPTIONS

Figure 1. Block diagram of the experimental setup.

Figure 2. Example of the display produced in inverse filtering with the *Sopran* software. The top curves show the input audio waveform and the inverse filtered flow glottogram. The lower panel shows the spectrum of the same signals. The open circles marked F1, F2, F3, and F4 show the formant frequencies and bandwidths used for the inverse filter, whereby the y-values represent the bandwidth. The two curves represent typical variation range for bandwidth values according to Båvegård et al., [1].

Figure 3. Subglottal pressure as function of pitch for the indicated vowels and tone production types observed in the female and males subjects. Black and gray symbols represent "Overdrive" and Falsetto tone production types, respectively, as indicated in the bottom of the left panel (Ov LB ɛ = "Overdrive", loud bright /ɛ/; Ov LD ɛ = "Overdrive", loud dark /ɛ/; Ov M ɛ = "Overdrive", medium loud /ɛ/; Ov LB ɔ = "Overdrive", loud bright /ɔ/; Ov LD ɔ = "Overdrive", loud dark /ɔ/; Ov M ɔ = "Overdrive", medium loud /ɔ/; Fals L ɔ = Falsetto, loud /ɔ/; Fals S ɔ = Falsetto, soft /ɔ/; Fals L ɛ = Falsetto, loud /ɛ/; Fals S ɛ = Falsetto, soft /ɛ/).

Figure 4. SPL@0.3m as function of pitch for the indicated vowels and tone production types observed in the female and males subjects (left and right panel). Black and gray symbols represent "Overdrive" and Falsetto tone production types, respectively, as listed in Figure 3.

Figure 5. Flow glottogram parameters as function of P_{Sub} observed in the male subject for the pitches E4 and C4 (upper and lower series of panels, respectively). The curves show the trendlines, the equations of which are listed in Table 2. Filled and open black symbols refer to "Overdrive" tones and gray symbols to Falsetto tones, respectively, see Figure 3.

Figure 6. Mean deviations from the trendlines shown in Figure 5 for the "Overdrive" and the Falsetto tones (black and gray columns). Filled and open columns refer to the pitches C4 and E4, respectively.

Figure 7. Frequenceis of the first and second formants F1 and F2 measured in the inverse filtering of the tone types. The left and right panels show the values for the pitches C4 and E4, respectively. Black and gray symbols represent “Overdrive” and Falsetto tone production types, respectively, as listed in Figure 3.

Figure 8. F1/F0 and F2/F0 ratios (black and gray columns) for the indicated tone types and vowels.