Relationship between perceived vocal registers and glottal flow parameters

Gláucia Laís Salomão

Pontifical Catholical University – Applied Linguistic and Language Studies Program
Royal Institute of Technology – School of Computer Science and Communication
Department of Speech, Music and Hearing
glaucia.salomao@gmail.com

Abstract

The perception of modal and falsetto registers was analyzed in a material consisting of a total of 80 vowel sounds sung by 10 choir singers, 40 sung in modal register and 40 in falsetto register. These vowel sounds were classified by sixteen expert listeners in a force choice test and the number of votes for modal was compared with the voice source parameters (1) Closed Quotient (Q_{closed})and (2) level difference between the two lowest source spectrum partials (H1 – H2). Tones with a high value of Qclosed and low values of H1-H2 were typically associated with high number of votes for modal register, and vice versa, Q_{closed} showing the strongest correlation. Some singer subjects produced tones that could not be classified as either falsetto or modal register, suggesting that category perception of registers is not always feasible.

1. Introduction

Registration of vocal quality is observed in both speaking and singing. Voices occasionally "break" and this phenomenon can be attributed to a sudden register change. When transitions from one voice quality to another occur, most speakers and singers report some sort of non-specific, kinesthetically sensed, neuromuscular coordination adjustment in the larynx.

Register changes can be brought about by linguistic demand. For instance, abrupt changes to creaky, falsetto or fry qualities are sometimes needed for phonetic contrasts, especially in some African and Asian languages (1).

Register changes can also be brought about for expressive demand. Transition between registers can be employed, for instance, when people speak expressively within a relatively wide of fundamental frequency. Also, voluntary jumps between modal and falsetto registers can be heavily exploited for aesthetical purposes in singing, as in yodeling, country-western singing, and Hawaiian folk.

It is generally agreed that register is a phenomenon affiliated with the voice source and that it is associated with the vibrational characteristics of the vocal folds. According to authors (2–6) the vocal folds are thicker and vibrate with a greater mass in modal register, while in the falsetto they are thinner and more stretched.

The perceptual aspects of voice play a prominent role in descriptions of vocal registers as well. According to Titze (2), vocal register can be defined as "…perceptually distinct regions of vocal quality that can be maintained over some ranges of pitch and loudness". The question of how different registers are perceived along a continuum of fundamental frequencies, however, has not been completely answered.

Register investigations have often focused on register contrasts, comparing tones produced in one register with tones produced by the same subject in a different register. It appears, however, that voice source differences between registers can be sometimes quite small. Also, very clear timbral contrasts between registers are not always easily perceived aurally. Classical singers, for instance, are trained to reduce or even eliminate timbral variation between registers, performing a blended register transition instead, with no abrupt voice timbre changes.

According to Hollien (5), "the operational definition of a register must depend on supporting perceptual, acoustic, physiologic and aerodynamic evidence". Thus, analyses of the relationships between voice source characteristics and perception of register seem relevant. Experimental data investigating this relationship have been still rare in the literature, tough.

The purpose of this study was to compare the perception of modal and falsetto registers with voice source properties. A total of 80 vowel sounds sung by 10 male singers were classified as modal or falsetto register by sixteen experts. Information of the glottal voice source was derived from the analysis of electroglottographic (EGG) and inverse filtered signals.

2. Material and Method

2.1. Recording procedures and acoustical analyses

Thirteen male choir singers volunteered as subjects: two basses, five baritones and six tenors. All were members of the semi-professional State Choir in São Paulo, which frequently performs together with the São Paulo State Orchestra. Their age ranged from 21 to 30 and the duration of singing training varied from four months to fourteen years (Table 1). None of the subjects reported any voice problems at the time of the recording. All of them were giving their written consent to participate in the investigation.

Before the recording the singers were asked a set of questions regarding their ability to identify and voluntarily produce tones in both modal and falsetto registers at identical pitch. Just the subjects who answered in the affirmative to all questions (Table 1) were included in this investigation.

Singer	Age	Voice Classification	Duration of singing training		Ability to perceive and voluntarily produce tones in both registers according to the singer					
			Individual classes	Choral singing	Can you clearly identify differences in voice quality when you sing in modal and falsetto register?		Can you control your voice production so as to produce modal or falsetto registers voluntarily?		Do you think it is possible for you to sing some notes in both modal and falsetto registers?	
					yes	no	yes	no	yes	no
1	29	Tenor	9 years	14 years	X		X		X	
2	21	Baritone	2½ years	8 years	X		X		X	
3	24	Tenor	2 years	10 years	X		X		X	
4	22	Tenor	1½ year	9 years	X		X		X	
5	26	Baritone	6 years	5 years	X		X		X	
6	26	Baritone	2 years	4 years	X		X		X	
7	21	Baritone	1 year	2 years		X		X		X
8	26	Tenor	4 years	13 years		X		X		X
9	30	Tenor	2 years	4 years	X		X		X	
10	23	Bass	4 months	8 months		X		X		X
11	27	Baritone	3 years	3 years	X		X		X	
12	25	Bass	3½ years	8 years	X		X		X	
13	30	Tenor	3 years	6 yeas	X		X		X	

Table 1

Each singer was asked to sing two *chromatic scales* on the vowel /a/ in a comfortable pitch range and at comfortable vocal loudness. The subjects were instructed to sing an ascending scale in modal register, extending it as far as possible. Correspondingly, they were asked to sing a descending scale in falsetto register, starting at the highest possible pitch and extending it as far as possible.

Recording sessions took place in a soundtreated booth. The acoustic signal was picked up by a *Audiotechnica ATM 75* headset microphone fixed at 17 cm from the singer's mouth. The electroglottographic signal was recorded by a *Glottal Enterprises'* Electroglottograph EG2 with the surface electrodes fastened to each side of the thyroid cartilage. The electroglottograph was connected to a Kay Elemetrics Corporation CSL (Computer Speech Lab 4300B). Both the audio and the electroglottograph signals were recorded simultaneously on separate channels on a DAT recorder (Soundcraft 328 XD).

Sound level was calibrated by recording a 1000 Hz sine wave, the SPL of which was measured at the recording microphone by means of a sound level meter (Radio Shack Sound Level Meter, catalogue number 33-2066, Radio Shack Corporation, Forth Worth, TX). The SPL reading was announced on the tape.

The recordings were digitized and stored as sound files in the wav format. In these recordings were selected four adjacent pitches from the pitch range where the two registers overlapped, however avoiding tones close to the limits of the modal and falsetto registers. These overlapping pitches varied substantially between the subjects, the lowest being G3 (in one single subject) and the highest was B4. In most cases the lowest pitch fell within the range C4 – G#4.

As inverse filtering is difficult at high F0, it was considered risky to rely on an automatic inverse filtering program. Rather, the custom-made DeCap program (Svante Granqvist, Kungliga Tekniska Högskolan, Sweden) was used. This program requires manual tuning of the inverse filters. In the present investigation, it displayed the waveforms and spectra of the input as well as the inverse filtered signal (Figure 1). It also displayed the derivative of the EGG signal (dEGG) delayed by a time interval corresponding to the travel time of sound for a distance corresponding to the sum of vocal tract length and the microphone distance.

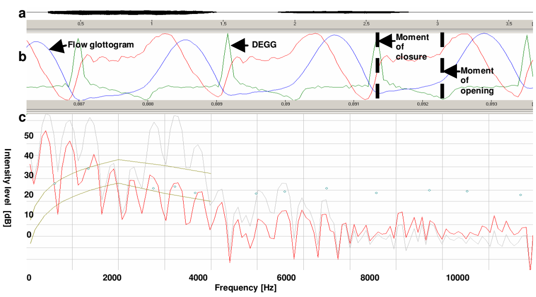

Figure 1. *DeCap inverse filtering program: (a) airflow along a compressed time scale; (b) expanded flow glottogram, its first derivative and the dEGG signal; (c) spectrum of the air flow and of the inverse filtered air flow. The x and y positions of the open circles represent the formant frequencies and bandwidths, respectively.*

For each subject and pitch a section from near the middle of the tone sung in modal and falsetto register was selected for inverse filtering. The criterion for the tuning of the filters was to obtain a ripple-free closed phase and a smooth source spectrum envelope. The identification of the closed phase was greatly facilitated by the dEGG signal; its main positive peak should be synchronized with the trailing end of the flow pulse and the negative peak with the end of the closed phase. The dEGG showed a major positive peak, also in the falsetto tones, and mostly also a much less conspicuous positive peak corresponding to the sudden decrease of glottal contact at the opening of the glottis.

The filter settings were checked by synthesis using the custom-made Madde voice synthesis computer program (Svante Granqvist, Kungliga Tekniska Högskolan, Sweden). This program allows control of F0 and six formant frequencies and bandwidths, and also of the frequency and amplitude of the vibrato. The filter settings were accepted only if they generated a vowel quality similar to that produced by the subject.

Analyses of the flow glottograms obtained were carried out using the various modules in the Soundswell Signal Workstation (Hitech Development, Stockholm, Sweden). From the flow glottogram synchronized with the dEGG signal and from the spectrum of the flow glottogram the following parameters were measured and calculated:

1. Period length (T0) and Fundamental frequency (F0);
2. Duration of the closed phase (Tcl) and the closed quotient (Q_{closed}), i.e., the ratio between the duration of the closed phase and T0;
3. The level difference between the two lowest partials (H1 – H2).

For each parameter measurements were taken from three adjacent periods and the results were averaged. These averages were used in the subsequent analysis.

2.2. Perceptual Evaluation

The 80 tones selected from the recordings mentioned above (10 singers x 4 tones x 2 registers) were copied in random order onto a CD by means of the custom made Glue program (Svante Granqvist). It inserted a pause of 3 seconds between the stimuli. The program also provided a list of the stimulus order.

The 104 stimuli were evaluated by a panel of sixteen voice experts. They listened to the stimuli simultaneously at a comfortable listening level over loudspeakers in a class room. They were asked not to interact with each other during the test.

The subjects were asked to decide if the stimulus they heard was produced in modal or in falsetto register. They were given sheets containing a table with numbered lines for the 104 stimuli and two columns, one for modal and one for falsetto register votes. The subjects thus put a marking in the corresponding column just after each stimulus had been presented. The complete test took about 10 minutes.

3. Results

3.1. Acoustic Data

Typical examples of flow glottograms and of dEGG signals of modal and falsetto registers are shown in Figure 2. In the modal register, a clear closed phase in the flow glottogram can be observed together with a strong positive peak at glottal closing and a weaker peak at glottal opening in the dEGG. In the falsetto register, a smoother and often nearly sinusoidal waveform shows no abrupt change from the closed to the open phase. In contrast to the modal register, the closed phase observed in falsetto is shorter, and the dEGG peaks smaller and more similar at glottal closing and opening.

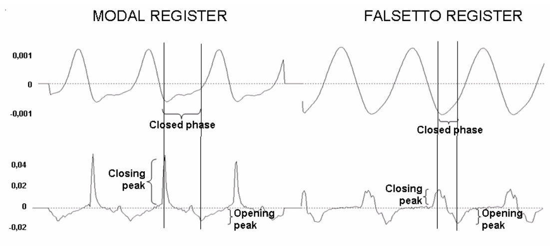

Figure 2: *Typical examples of waveforms for modal and falsetto register. The top and bottom panels show, in arbitrary scales, flow and DEGG signals***.**

The Q_{closed} values and the level difference between two lowest partials (H1 – H2) observed for the modal and falsetto register tones for each of the ten subjects are compared in the graphs shown in Figure 3. The Q_{closed} values were clearly greater in modal than in falsetto in all subjects. In modal it ranged from 0.39 to 0.67 and in falsetto from 0.13 to 0.43. On average across subjects Q_{closed} in falsetto was about half (0.48, SD = 0.15) of what it was in modal. Both in modal and falsetto the variation with pitch was mostly quite small and unsystematic, as could be expected in view of the small F0 differences between the tones analyzed. The H1-H2, on the other hand, was greater in falsetto register in all subjects except one (singer 12). It ranged from 3,1dB to 14,9dB in modal and from 10,7dB to 34,3dB in falsetto. On average across subjects the fundamental in falsetto was 12,1 dB (SD = 9,8 dB) stronger than in modal. The variation with pitch in both falsetto and modal registers was very small in some subjects, and somewhat greater though mostly unsystematic in other subjects.

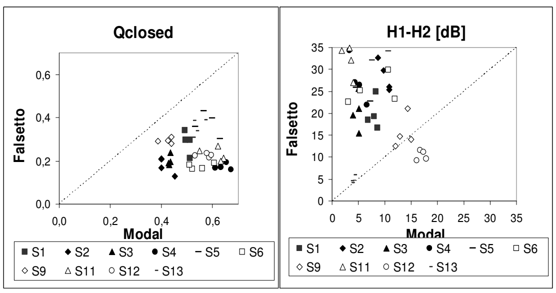

Figure 3: *Closed quotient and H1-H2 values observed in the modal and falsetto register tones for the ten singers.*

Finally, figure 4 shows the correlation between Q_{closed} and H1 – H2. For both registers, the great the Q_{closed} values, the lower the H1 – H2, and vice-versa.

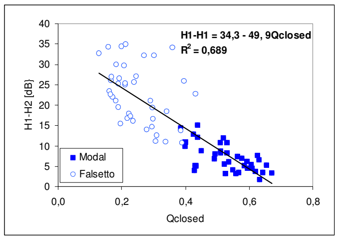

Figure 4: *Correlation between the Q_{closed} and H1* – H2

3.2. Perceptual evaluation

Figure 5 shows the number of votes for modal for each of the stimuli. In the graph, the stimuli were rank ordered according to this same number of votes for modal. Thus, the stimulus that collected the highest number of votes for modal appears at the right end of the x-axis. In the graph, 16 votes means that all 16 listeners agreed that the tone was sung in modal and zero implies that they all agreed that the tone was not sung in modal, i.e., they agreed that it was sung in falsetto.

The figure shows 16 votes in 15 cases and 0 votes in 8 cases, i.e., a total of 23 cases of complete agreement. In 19 additional cases all listeners except one agreed on the classification. In other words, the listeners were in almost perfect agreement with respect to register classification in more than half of the cases. This suggests some degree of consistency, since lack of consistency would be incompatible with complete agreement for several stimuli.

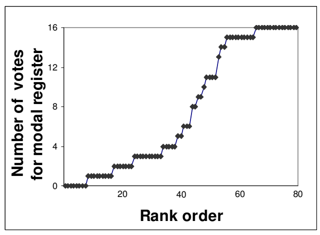

Figure 5: *Number of votes given for modal register, rank ordered according to this number*

Also, in most cases the listeners' classification also agreed with the register intended by the singers, in both modal and falsetto examples. Figure 6 shows that for both registers more than 80% of the examples were classified in accordance with the singers' intentions. Thus, most of the 80 tones were perceived as representative examples of the respective registers. Pitch did not seem a factor of relevance to the classification.

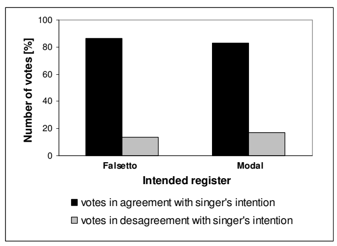

Figure 6: *Agreement (black columns) and disagreement (gray columns) between the listeners' classifications and the singers' intentions with regard to register.*

3.2. Relationships between Classification and Acoustic Data

Figure7 shows the relationship between the classification data and the Q_{closed} and H1-H2 parameters. Tones with a high value of Q_{closed} and low values of H1-H2 were typically associated with high number of votes for modal register, and vice versa. The Q_{closed} parameter showed the strongest correlation.

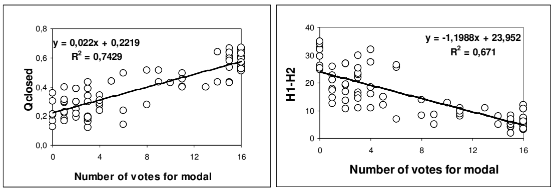

Figure 7: *Qclosed and H1-H2 as functions of the number of votes for modal register.*

4. Discussion

The results obtained have shown clear and systematic voice source differences between modal and falsetto registers. Thus, the closed phase values were predominantly greater in the modal than in the falsetto register. The H1-H2 values, on the other hand, were clearly greater in falsetto than in modal register.

The physiological origin of the differences observed could be related to the shape and mode of vibration of the vocal folds. According to Van den Berg (3) and Hirano (4) the vocal folds are elongated and thin in falsetto register while in modal register the folds are thicker and shorter, due to activation of the thyroarytenoid muscle.

It seems reasonable to assume that thin vocal folds should be associated with a smaller time lag between the upper and the lower layers of the folds, resulting in a more symmetrical waveform, typically found in falsetto. By contrast, thick vocal folds with lengthening of the closed phase should be associated with longer time lag between the vocal fold layers, making the waveform with the same period as the fundamental less similar to a sine wave, as commonly found in modal register. Thus, the more sinusoidal falsetto waveform would explain why the H1-H2 level difference was much greater in falsetto than in modal and also the correlation between Q_{closed} and H1-H2, illustrated in Figure 4.

Consistent differences between voice source properties in the different registers were found. However, they could vary substantially in quantity among singers. For example in some subjects the mean Q_{closed} in modal and falsetto differed by no less than 0,469, while in other subjects, this difference was no more than 0,1.

This variation is not surprising, considering that our subjects differed considerably with regard to vocal training and experience of singing. On the other hand, quite small voice source differences between registers could result in timbral differences difficult to be perceived.

Indeed, the listening test showed that our singers produced a number of tones that were not unanimously classified as either falsetto or modal registers. In about 20% of the cases nearly half of the judges came up with different classifications. Thus, some voices produced tones that could not be unanimously classified as either falsetto or modal register, suggesting that the timbral contrast between the two registers was quite substantial in some subjects but rather slight in other subjects.

The cases of disagreement could result from poor skill in voice control. Also, it could be possible that the listeners found it easier to hear that a tone was produced in modal than in falsetto register. A third possibility is that some singers actually used a voice source that combined characteristics of the modal and falsetto registers, varying each of flow glottogram parameters continuously and performing timbral differences between registers by a gradual rather than an abrupt change. This ability would cause the listeners difficulties in their classification.

In any event, our results appear to challenge the idea that modal and falsetto registers are necessarily associated with clear timbral differences, suggesting that category perception of registers is not always feasible.

In view of these results, and considering that lack of perceptual evidence of timbral differences would imply in lack of registers differences, it is tempting to ask whether registers exist in all voices.

5. Conclusion

Our results show that: 1. the tones with a high value of Q_{closed} and low values of H1-H2 were typically associated with high number of votes for modal register, and vice versa; 2. the Q_{closed} parameter presented the strongest correlation with the results of the register classification test; and 3. some tones were not unanimously classified as neither as modal nor as falsetto, suggesting that classification of registers in terms of perceptual category is not always feasible.

Acknowledgements
The author would like to thank Professor Johan Sundberg for the invaluable contributions in writing of this paper. This work was carried out at the Speech, Music and Hearing Department (TMH) at Royal Institute of Technology (KTH) and it is part of the author's thesis work to be presented at the Applied Linguistic and Language Studies Program (LAEL) at Pontifical Catholic University, São Paulo (PUC-SP), Brazil, under supervision of Professor Sandra Madureira. Her research at KTH was financially supported by the Brazilian Ministry of Education (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior - CAPES).

6. References

[1] Ladefoged, P., 1983. The linguistic use of different phonation types. In: *Vocal fold physiology: Contemporary research and clinical issues.* Bless & J. Abbs (eds.). San Diego: College-Hill Press, 351-360.
[2] Berg, Jw van den., 1968. Mechanisms of the larynx and the laryngeal vibrations. In: Malberg, B. *Manual of Phonetics*. Amsterdam: North Holland Publishing Company, 278-308.
[3] Hirano, M., 1974. Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniat.* 26: 89-94.
[4] Hiroto, I., 1981. Introductory Remarks. In: Stevens, K. N., Hirano, M. *Vocal Fold Physiology.* Tokyo: University of Tokyo Press, 3–9.
[5] Titze, I., 1994. *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice Hall, 252-278.
[6] Hollien, H., 1994. On vocal registers. *J. Phonetics*. 2: 125–143.