Voice Source Variation Between Vowels in Male Opera Singers

*,†Johan Sundberg, ‡Filipa M. B. Lã, and §Brian P. Gill, *†*Stockholm, Sweden,* ‡*Aveiro, Portugal, and* §*Bloomington, Indiana*

Summary: Objectives. The theory of nonlinear source-filter interaction predicts that the glottal voice source should be affected by the frequency relationship between formants and partials. An attempt to experimentally verify this theory is presented.

Study design. Glottal voice source and electrolaryngograph (ELG) signal differences between vowels were analyzed in vowel sequences, sung at four pitches with the same degree of vocal loudness by professional opera singers. In addition, the relationships between such differences and the frequency distance between the first formant ($F1$) and its closest partial were examined.

Methods. A digital laryngograph microprocessor was used to simultaneously record audio and ELG signals. The former was inverse filtered, and voice source parameters and formant frequencies were extracted. The amplitude quotient of the derivative of the ELG signal ($\mathrm{AQ_{dELG}}$) and the contact quotient were also compared.

Results. A one-way repeated-measures ANOVA revealed significant differences between vowels, for contact quotient at four pitches and for maximum flow declination rate (MFDR) at three pitches. For other voice source parameters, differences were found at one or two pitches only. No consistent correlation was found between MFDR and the distance between $F1$ and its closest partial.

Conclusions. The glottal voice source tends to vary between vowels, presumably because of source-filter interaction, but the variation does not seem to be dependent on the frequency distance between $F1$ and its closest partial.

Key Words: Sung vowels–Inverse-filtering–Voice source–Formant frequencies–Electrolaryngograph.

INTRODUCTION

According to classical singing pedagogy, some vowels can be produced more easily than others at a given pitch.[1] This seems to contradict the classical source-filter theory of voice production, if it is assumed to predict that the glottal airflow is independent of vocal tract resonances, that is, formants.[2] Rather, it supports the assumption that the glottal airflow is affected by the formants because of nonlinear source-filter interaction.[3,4]

The theory of nonlinear source-filter interaction in voice production has been developed over the last decades.[5] It predicts that when the first formant ($F1$) coincides with or crosses over one of the lower spectrum partials, voice instabilities may occur, for example, fundamental frequency ($F0$) jumps, subharmonic frequencies, and changes in the amplitude of the voice source fundamental.[5,6] Under certain conditions, such feedback may facilitate vocal fold oscillation, that is, elicit a more efficient conversion of aerodynamic to acoustic energy.[3] More specifically, the sound pressure level (SPL) of a vowel may increase by as much as 10 dB if one of the lowest harmonics is just below the first formant frequency. On the other hand, it may be weakened if one of those partials is located just above the first formant frequency.[5] Considering the dependence on the proximity between lower harmonics and the first formant frequency, this interaction should be milder for male speech and greater for female and child voices. In male singing, however, an interaction should be likely to occur in and above the *passaggio*, that is, E4 (±330 Hz) to G4 (±400 Hz).[7]

The theory of nonlinear source-filter interaction has been tested and confirmed in experiments using physical models, computer simulation,[8] excised larynges,[9,10] and voice source analysis in a single speaker.[11] For example, in model experiments with a simplified two-mass model connected to a straight tube, subharmonic vibrations and deterministic chaos were observed when $F0$ and $F1$ coincided.[12] The theory has also been tested in experiments. For example, Titze et al had 18 subjects, none of whom had extensive vocal training, perform vocal exercises where $F1$ was passed by a partial. In many cases, various types of $F0$ disturbances, such as pitch jumps, and bifurcations, were observed when a partial was close to F1. Moreover, using an electrolaryngograph (ELG), a noninvasive tool for documenting vocal fold contact,[13] differences have been observed in contacting and decontacting events between different spoken vowels; both the open quotient and the speed quotient were affected.[14]

In singing, control of the vocal output is crucial, so uncontrolled pitch jumps and other instabilities would be totally unacceptable. One way to circumvent them would be to avoid the situation that a partial is just above $F1$. However, the effects of the frequency relations between $F1$ and its closest partial have not been measured in singers, neither with respect to the flow glottogram of different vowels, nor with respect to the ELG waveform. Hence, it seemed worthwhile (1) to compare voice source parameters between vowels and (2) to investigate whether vowel differences between such parameters could be explained by source-filter interaction. In particular, we tested if the vocal tract excitation, that is, the maximum flow

---

Accepted for publication July 22, 2015.

This work was presented at the Annual Symposium Care of the professional Voice, Philadelphia, June 2014.

From the *Department of Speech, Music and Hearing, School of Computer Science and Communication, KTH, Stockholm, Sweden; †University College of Music Education, Stockholm, Sweden; ‡Department of Communication and Arts, INET-MD, University of Aveiro, Aveiro, Portugal; and the §Department Jacobs School of Music, Indiana University, Bloomington, Indiana.

Address correspondence and reprint requests to Johan Sundberg, Department of Speech, Music and Hearing, School of Computer Science and Communication, KTH, SE-10044 Stockholm, Sweden. E-mail: pjohan@speech.kth.se

Journal of Voice, Vol. 30, No. 5, pp. 509-517
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2015.07.010

TABLE 1.
Participant's Age, Voice Classification and Working Experience

Singer	Age (yrs.)	Classification	Experience
1	30	Tenor	Internationally touring
2	32	Baritone	Nationally touring
3	23	Baritone	Graduate student
4	25	Tenor	Graduate student
5	24	Baritone	Graduate student
6	42	Tenor	Internationally touring
7	38	Baritone	Nationally touring
8	35	Baritone	Graduate student

declination rate (MFDR), was greater when the frequency distance between $F1$ and its closest partial, henceforth Min($F1$–[n*$F0$]), was positive, that is, when $F1$ was just above the closest partial, and smaller when it was negative, that is, when $F1$ was just below its closest partial.

METHODS

Eight male classically trained singers, 23–42 years old (mean 31.1, SD 6.9) with varying levels of professional expertise, volunteered as subjects (Table 1). They were asked to sing a sequence of the vowels /i, e, a, o, u/ on each of the pitches E3, G3, A3, and C4, keeping vocal loudness constant. Each task was repeated once. The pitches were given to the subjects by means of the custom-made *MADDE* software (by Svante Granqvist; KTH, Stockholm, Sweden).

All recordings were made in a sound-treated studio in the Steinhardt School of Culture, Education, and Human Development at New York University. A Laryngograph microprocessor (Laryngograph Ltd, London, UK) was used to record audio and ELG simultaneously. The former was picked up by a head-mounted omnidirectional electret microphone (Knowles EK3132, Knowles Corporation, Itasca, IL) placed at a mouth-to-microphone distance of approximately 15 cm. The sound level was calibrated by means of a 1-kHz sine wave, the SPL of which was measured next to the recording microphone by means of a sound level meter. The value observed was announced in the recording. Both signals were recorded using Laryngograph Speech Studio (Laryngograph Ltd, London, UK) software and stored as wav files.

The voice source was analyzed in terms of flow glottograms derived from the audio signal after integration and inverse filtering. Inverse filtering is a classical method in voice analysis.[15,16] The strategy is to eliminate the influence of the vocal tract resonance characteristics on the radiated sound. This is realized by filtering the signal by a set of filters representing the inverse of the transfer function of the vocal tract. The method offers information on both the glottal airflow waveform (flow glottogram) and on the formant frequencies and bandwidths. The accuracy is particularly high in cases where a partial is close to a formant. This is illustrated in Figure 1, showing the effects of setting the $F1$ filter 4% above and 4% below the correct value.

Samples of the different vowels were analyzed using the custom-made *Decap* software for inverse filtering (Svante Granqvist; KTH, Stockholm, Sweden). This program can be set to display waveform and spectrum in separate windows, as described in detail elsewhere.[17] The frequencies and bandwidths of the inverse filters are set manually, and the classical equations are applied for calculating the transfer function that

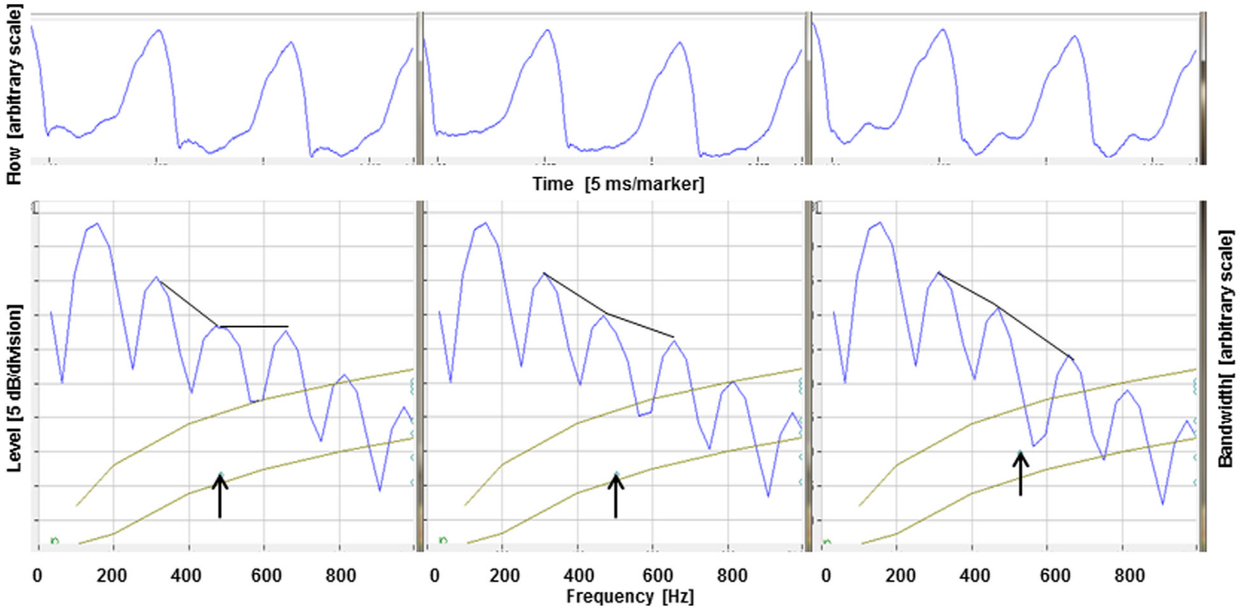

FIGURE 1. Example of the effects on the flow glottogram of mistuning inverse filter for $F1$ by -20 Hz and +20 Hz, corresponding to ±4% of $F1$ (*left and right panels*, respectively). The *middle panel* represents the result of a correct filter setting. The positions of the *arrows* along the *horizontal axis* show the frequencies of the $F1$ inverse filter. Their positions on the *vertical scale* represent their bandwidths in an arbitrary scale, and the *curves* show the range of typical bandwidth values. (Color version of figure available online.)