CO-MODULATION OF F_0 AND FORMANT FREQUENCIES IN SINGING

Leo-Geert van den Berg, Christel de Bruijn, Gerrit Bloothooft, Guus de Krom
Research Institute for Language and Speech
Utrecht University, The Netherlands

ABSTRACT

Vocal vibrato can be considered as the result of pulsations in muscular activity at the respiratory, laryngeal, and supra-laryngeal levels. Sometimes, the latter can be observed in singers as rhythmic movements of the entire larynx and the pharyngeal wall. This will influence formant frequencies and hence the timbre of the sound. We investigated whether modulations of formant frequencies can be observed in singing, and how they relate to modulations of the fundamental frequency.

INTRODUCTION

In singing, we often observe vibrato as a modulation of the fundamental frequency (F_0), with a rate between 5 and 7 Hz. Several researchers report a co-modulation of the activity of intrinsic laryngeal muscles with F_0, most notably perhaps the cricothyroid muscle [1]. As a second mechanism, rhythmic pulsations of the respiratory muscles resulting in modulations of the subglottal sound pressure have been mentioned [2]. In addition, a co-modulation between F_0 and the activity of several supralaryngeal muscles has been found [3]. This co-modulation, among others, influences laryngeal height, as can sometimes be seen in video-stroboscopic images of the pharyngeal cavity: the entire larynx, as well as the pharyngeal wall, can be in regular movement during singing. These findings indicate that vibrato is a complex neuromuscular phenomenon that affects all aspects of voice production. The precise coordination of the various pulsated muscular actions is still largely unknown.

In the present investigation, we concentrated on the acoustic effects of vibrato as originating from a modulation of the vibration frequency of the vocal folds, and from a modulation of the volume of the pharyngeal cavity. First, we investigated whether a regular variation of the volume of the pharyngeal cavity exists to an extent that it can be measured as a modulation of formant frequencies. Second, whether there is a co-modulation between formant frequencies and F_0.

In a stationary situation, relations between vertical larynx positioning and the singer's formant have been studied theoretically and experimentally [4]. The main effects of a raised larynx were: (1) a significant increase of F_2 in high front vowels, (2) a raise in F_1 and F_2 for open vowels, (3) a raise in F_3 and F_4. Similar effects should be found during rapid modulations of the larynx height.

METHODS

We used recordings of four professional male singers (with a classification between bass and baritone) who sang the vowels /a/, /i/, /u/, and /ə/ at F_0 = 98 Hz in three conditions: straight (none to little vibrato), normal, and exaggerated vibrato, yielding a total of 48 recordings. The recordings were digitized at 20 kHz. For these singers, we had the audio material available, but no information on larynx movements.

We chose the low F_0 value of 98 Hz to ensure a high accuracy of the F_0 measurement (peak picking algorithm in the time domain). For formant frequency measurements, we first downsampled the data from 20 kHz to 10 kHz before performing an LPC-12 analysis. This was done to increase the accuracy of the formant frequency measurement by focussing on the 0-5 kHz spectral region.

Still, LPC procedures tend to be sensitive to the distribution of harmonics (and hence to F_0). Because there are various factors that influence the value of computed formant frequencies, we give a formal description of these factors to facilitate the interpretation of the results.

MODELLING F_0 AND FORMANT FREQUENCY MODULATION

We first make the assumption that undulations in the activities of muscular structures that have an effect on sound production (respiratory, laryngeal, and supralaryngeal structures) are coordinated at a fairly central level, and have identical rates. The acoustic phenomena associated with these centrally coordinated movements may exhibit phase differences, however.

We combine the acoustic effects of the pulsated activity of respiratory muscles, influencing the subglottal pressure, and the internal laryngeal muscles, influencing the tension of the vocal folds, in a modulation of the fundamental frequency $F_0(t)$:

$$F_0(t) = F_0 + \Delta F_0 \sin(2\pi Vt) \qquad (1)$$

in which $F_0(t)$ is the time-varying value of the fundamental frequency, F_0 is its average value, ΔF_0 is the extent of the frequency variation, and V is the vibrato rate. This combination may be considered a simplification, because interactions between modulating subglottal pressure and modulating laryngeal muscular tension can be quite complex. On the other hand, $F_0(t)$ normally shows a regular pattern during vibrato.

If the laryngeal height modulates due to a vibratory activation of the supra-laryngeal muscles, this affects the length of the vocal tract as well as the volume of the pharyngeal cavity. As a consequence, the formant frequencies will also be modulated. There are two aspects to consider here. First, a one-tube model of the vocal tract would predict that a raised larynx will result in increased formant frequencies. However, the actual effect may be more complex. Second, there may be a phase difference between larynx height modulation and the modulation of the vibration frequency of the vocal folds. We combined both effects into one phase term, and described the modulation of the center frequency of a formant n as follows:

$$F_n(t) = F_n + \Delta F_n \sin(2\pi Vt + \varphi) \qquad (2)$$

in which F_n is the average formant frequency, and ΔF_n is the maximum deviation of F_n.

It follows from equations (1) and (2) that a co-modulation of $F_0(t)$ and $F_n(t)$ may be observed with an unknown phase difference.

Furthermore, we have to consider the possibility that the LPC procedure used to estimate the formants yielded measurement errors. Ideally, we would like to compute formant frequencies independent of the harmonic structure. However, this is not possible in practice. In an extreme situation, for instance for high-pitched vowels with $F_0 > 500$ Hz, LPC analyses will result in formant estimates that follow the frequencies of the separate harmonics. Although this artefact will be less outspoken for the low F_0 value of 98 Hz that we used, we cannot exclude the possibility that computed formant frequencies to some extent follow the fundamental frequency:

$$F_n(t) = F_n + \Delta F_n \sin(2\pi Vt) \qquad (3)$$

The resulting co-modulation of $F_0(t)$ and $F_n(t)$ should be in phase.

There is an additional interaction between F_0 and harmonics underlying a formant. That is that the *amplitude* of these harmonics will show a modulation too: in phase with F_0 modulation along the positive slope of a formant, in counter-phase along a negative slope, and with a rate that is twice the vibrato rate of F_0 if the harmonic varies symmetrically around a formant frequency [5]. If the LPC procedure does not model this variation properly, the estimated formant frequencies may show the following types of modulations:

$$\begin{aligned} F_n(t) &= F_n + \Delta F_n \sin(2\pi Vt) \\ F_n(t) &= F_n - \Delta F_n \sin(2\pi Vt) \\ F_n(t) &= F_n + \Delta F_n \sin(4\pi Vt) \end{aligned} \qquad (4)$$

The combined effects of larynx height variation and formant measurement artefacts related to F_0, the combination of Eqs. 2-4, may be rather complex. This will have to be taken into account in the interpretations of the results.

RESULTS AND DISCUSSION

Vibrato rates of the 48 recordings varied between 5 and 7 Hz. The average extent of Fo-modulation frequencies (maximal deviation from the average value) for straight, normal, and exaggerated vibrato was 1, 3, and 5.5%, respectively. These are typical values [6]. With exaggerated vibrato, however, a maximum extent up to 10% (about 2 semitones) could be measured.

In several cases modulations in formant frequencies were found, but the results were of a highly variable nature and difficult to generalize. There was a tendency, however, for modulations to be less present in the higher formants. For this reason, we present a few examples here to demonstrate several types of co-modulation between F_0 and F_1.

Figure 1 gives a clear example of co-modulation of F_0 and F_1. The extent of the modulation is 5.5% for F_0 and 6.0 % for F_1. The correlation coefficient between F_0 and F_1 is .80. The close correspondence between the two traces, with almost equal relative extent, may indicate an artefact of F_1 computation that seems to follow the fourth harmonic.

A complex pattern is shown in Figure 2. This is an example of exaggerated vibrato with an F_0-extent of 10%. Apart from modulations, there is a gradual increase in the value of F_1, which can be explained by some variation in articulation of /i/. The superimposed modulations in F_1 have and extent of 3%. First they are in counter phase, later they have a double rate. It can be seen that in the first part, for the high F_0 values (104 Hz), the F_1 values meet the frequency of the third harmonic (312 Hz), while for the low F_0 values (85 Hz), the F_1 values meet the fourth harmonic (340 Hz). This gives a strong suggestion of a computational effect. We do not have an acceptable interpretation for the last part with double rate of modulation of F_1.

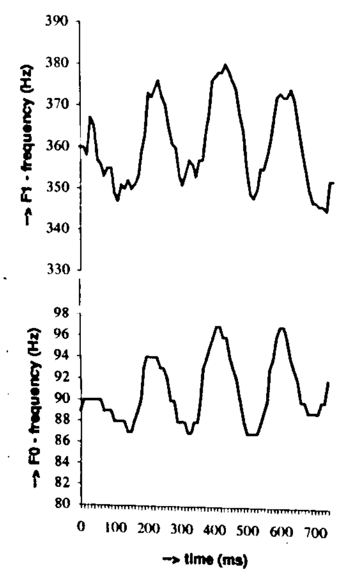

Figure 1. F_0 and F_1 for the vowel /u/, normal vibrato.

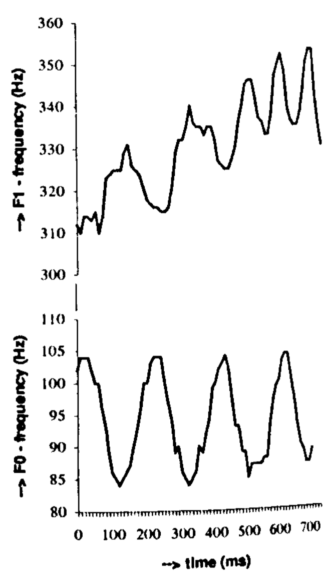

Figure 2. F_0 and F_1 for the vowel /i/, exaggerated vibrato.

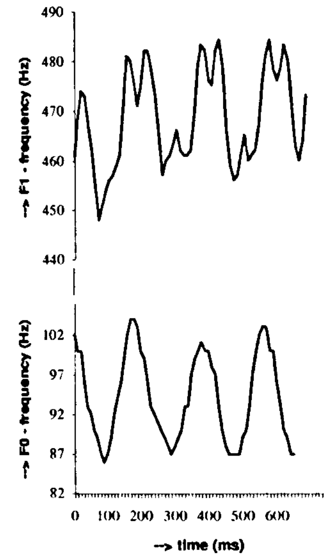

Figure 3. F_0 and F_1 for the vowel /ə/, exaggerated vibrato.

Figure 3 again shows exaggerated vibrato, with an extent of 8.5%. The F_1 trace is complex and best described as two superimposed sinusoids, one with the same rate and phase as the F_0, and one with a double rate. There is no simple harmonic relationship between F_0 and F_1 that may provide an explanation in terms of a computational effect for the entire trace. We hypothesize the following. The modulation with the same rate as F_0 may be the result of variation in the height of the larynx. To explain the modulation at the double rate, we note that the fifth harmonic varies between 440 Hz and 500 Hz and thus passes the actual formant frequency in an almost symmetrical manner. Its amplitude will therefore vary with twice the F_0 rate. As suggested earlier, the computation of F_1 may follow this rate.

CONCLUSION

Although co-modulation between F_0 and F_1 can be measured in singing, an interpretation of these results is difficult. Both a movement of the larynx and the dependency on F_0 of the LPC-based computations of F_1 can explain the results. More detailed investigations into the latter effect are needed. Measurement of larynx height with multichannel electro-glottography [7] may be of help in the interpretation of acoustic data.

REFERENCES

[1] Shipp, T., Doherty, T., and Haglund, S. (1990), "Physiologic factors in vocal vibrato production", *Journal of Voice*, vol.4, pp. 300 - 304.

[2] Sapir, S. and Larson, K. (1993), "Supra-laryngeal muscle activity during sustained vibrato in four sopranos: surface EMG findings", *Journal of Voice*, vol.7, pp. 213 - 218.

[3] Sundberg, J. and Nordstrom, P.E. (1976), "Raised and lowered larynx - the effect on vowel formant frequencies", *STL-QSPR*, vol.2, pp. 35-39.

[4] Lindblom, B.E.F. (1981), "Studies of articulation", *STL-QSPR*, vol.2, pp. 36-51.

[5] Horii, Y. (1989), "Acoustical analysis of vocal vibrato: a theoretical interpretation of data", *Journal of Voice*, vol.3, pp. 36-43.

[6] Sundberg, J. (1993), "Acoustic and Psychoacoustic Aspects of Vocal Vibrato", written version of paper for the COMETT conference, Utrecht.

[7] Rothenberg, M. (1992), "A multichannel electroglottograph", *Journal of Voice*, vol.6, pp. 36-43.