Overtone Singing: Productive Mechanisms and Acoustic Data

F. Klingholz

Department of Phoniatics, Ludwig-Maximilians-University, Munich, Germany

Summary: Overtone singing is where one person sings in two voices, the first voice represented by the fundamental and the second by an enhanced harmonic. Overtone singing is performed in chest register. Tuning of the first or second formant and a reduction of the formant bandwidth down to 20 Hz make harmonics prominent. Narrowing the pharynx, velar constriction, variation of the small mouth opening, and a tension of the walls of the mouth cavity are used. Changing prominent harmonics has the effect of creating an overtone melody with sustained tones, tone steps, and trillos. Key Words: Singing voice—Formant tuning—Overtone enhancement—Voice quality.

Overtone singing is where one person sings in two voices. The first voice is represented by the fundamental and the second by an enhanced harmonic. Overtone singing is performed in chest register. Overtone singing is used either in a spiritual context (e.g., in Tibet, China, Japan, Mongolia, and India) or in folkloristic songs (e.g., those of Flamenco singers, the Pygmy people, and of the indigenous people of the South American Andes and Balkan).

There are two different kinds of overtone songs. In the first case, the fundamental frequency is very low, even in the pulse register. In this way two formants coincide with the harmonics and a major third is produced. These are sustained sounds. The only frequency variation is an upward or downward shift of the fundamental frequency of one octave. This mode of chanting has been described by Smith et al. (1). In the second case, overtone melodies are created by formant tuning. This effect has been described from a perceptual point of view (2,3).

Formant tuning is a well-known singing technique (4,5). Here, the voice intensity is elevated by a shift of the first formant to the fundamental (falsetto, 500–700 Hz). In contrast, in overtone singing, tuning of the first or second formant is used to select and enhance different harmonics when the fundamental frequency is low (chest voice, 100–150 Hz).

The overtone singer prefers sound qualities between the vowels, mostly between [u] and [o]. In production, the pharynx is narrowed with a velar constriction. A rough formant tuning is performed by shaping the tongue and by using a variation of the velar constriction, whereas for the fine tuning the mouth opening is varied [Vogel (6) has already used an iris diaphragm to tune a spherical resonator]. The singer controls the tuning by ear as well as by kinesthetics.

Two problems arise in explaining the overtone effect. First, according to Fant et al. (7), the amplitude of the fundamental may gain 6–8 dB if it coincides with the first formant when the fundamental frequency is <150 Hz. However, overtone prominence may be much higher even if a harmonic coincides with the second formant. Second, it is not clear how the second formant can enhance a harmonic, although its bandwidth can exceed the distance of the harmonics. Two studies hint that a bandwidth reduction mechanism could explain selectivity and prominence. Kagen and Trendelen-
burg (8) prolonged the vocal tract by hard-walled pipes and produced prominent overtones. Smith et al. (1) declared the reduction of formant bandwidth as a condition for overtone singing, although their measured bandwidth was not less than those normally found in vowels. Therefore, in this study, apart from a demonstration of acoustic data, formant bandwidth was measured to verify the hypothesis.

**METHOD**

Good overtone singers are seldom found in Western culture. For this reason tone material for analysis was provided by Mr. Vetter, who is the best German overtone singer. Mr. Vetter has studied overtone singing in the Far East for several years. Moreover, he had developed this technique to a high standard. His songs are available on tapes, records, and CDs, although sometimes suitable recordings are hard to obtain.

The acoustic analysis was done from digital recordings. Signal segments containing only voice were carefully chosen. Recordings with superimposed accompanying musical instruments (tambura, bells, gong, etc.) were excluded. Moreover, studio recordings were made to avoid the effects of room acoustics, which are often used by overtone singers in churches and halls.

About 10 min of an overtone song was selected for study and several short segments were used. The signals were sampled at a rate of 10 kHz by an analog-to-digital converter (Data Translation 2821, Marlborough, MA). For data processing a personal computer and the following software algorithms were used. The prominent harmonics (order, amplitude, frequency) were determined from spectra [fast Fourier transform (FFT), 50-ms window length, Kaiser-Bessel window characteristics, 6 dB/octave preemphasis]. Pitch data of the fundamental were measured by the average magnitude difference function.

Methods for the determination of formant bandwidth are (a) procedures of spectral analysis-by-synthesis, (b) measurement of amplitude decay in oscillograms, and (c) linear predictive coding (LPC). The first method was excluded because source spectra needs to be known. In the second method, a 4th order Bessel bandpass filter was applied in the determination of logarithmic decrements of prominent harmonics. The decrement is calculated from

\[ d = \ln(\frac{a_{n+2}}{a_n})/(t_{n+2} - t_n) \]

where \(a_n, a_{n+2}\) are successive positive or negative maxima of the damped oscillation, and \(t_{n+2} - t_n\) are their time distance (the decrement is \(\pi\) times the bandwidth \(b\)). In the third method, the LPC analysis, the voice apparatus must conform with the source-filter model (i.e., non-nasality, neglect of source-filter coupling). Because the overtone sounds are non-nasal and because the articulatory configuration (constricted pharynx) does not suggest a reaction of the mouth-cavity resonances on the source, the LPC analysis (covariance method, filter order 14, 256 samples, 6 dB/octave preemphasis) was used.

**RESULTS**

Apart from sustained tones, the singer uses tone steps and trillos in his melody (Fig. 1). The sound of the overtone is like that of a shawm. The small indentation in Fig. 1 occurs because of the limited spectral resolution when the frequencies of the overtones are determined from spectra.

A mean fundamental frequency of 127.5 \(\pm\) 0.73 Hz was measured over a 10-s signal segment in which the prominence of overtones varied with time. The low variance of the fundamental frequency shows high steadiness of pitch.

![FIG. 1. Signal segments from an overtone melody. Prominent harmonic (upper contour), fundamental frequency (lower contour), tone steps (top), trillo (bottom).](image-url)
In Fig. 2, the typical signal frames are shown, which the singer kept using in all his records. The frames differ in the number of oscillations per period. From Fig. 2a to d, the number of peaks per period increases from two to five. This means that the second to fifth harmonic dominate, respectively.

Figure 3 demonstrates the amplitude spectra pertinent to the signal frames in Fig. 2. As could be expected from Fig. 2, harmonics two to five are prominent, with an amplitude difference between 6 and 18 dB from their highest neighbor. Without overtone prominence, the first and second formant frequencies are about 370 and 630 Hz. In the spectra of Fig. 3a, b, and d, both formants can be identified. In the spectrum of Fig. 3c, the formants appear united because they are close together and the spectral resolution of the FFT algorithm is limited.

Table 1 lists the order of prominent harmonics, and the bandwidths of the formants coincide with these harmonics. The measurement errors in the case of the logarithmic decrement (within and between the cycles), as well as in the LPC analysis, were <10%. The bandwidth measured with both methods are very similar. Apart from the case of the prominent 2nd harmonic, the formants at the favored harmonics have lower bandwidths than are normally found in vowels (9,10). This is demonstrated in the LPC spectrum (Fig. 4), where a first formant with a normal bandwidth, and a second formant with a very small bandwidth, and hence a high amplitude, appear (5th overtone is prominent).

When the prominence of a harmonic is defined as the amplitude difference between the favored harmonic and its highest neighbor, then the prominence increases with decreasing bandwidth, as can be seen when comparing Fig. 2 and Table 1.

TABLE 1. Mean bandwidths of the formants at the prominent harmonics determined by oscillograms and LPC analysis (fundamental frequency 127.5 Hz)

| Formant bandwidth |
|-------------------|-------------------|
| Order of harmonic | From logarithmic | From LPC |
| harmonic          | decrement         | analysis |
| 2nd               | 58.3              | 56.8     |
| 3rd               | 27.8              | 29.2     |
| 4th               | 44.3              | 43.7     |
| 5th               | 15.9              | 16.3     |
DISCUSSION

Apart from the articulatory configuration in overtone singing, two results justify the application of LPC analysis (i.e., the assumption of small source-tract coupling): the correspondence of bandwidth data measured with two different methods and the steadiness of the fundamental frequency despite formant shifts. The dependence of the fundamental frequency on formant movement is usually an indicator of coupling (6,11).

Although the selectivity of the vocal tract is increased by a reduction of formant bandwidth, the more important effect for overtone singing is the prominence of a harmonic (which, of course, is linked to the bandwidth reduction), for the ear can only perceive single overtones of a harmonic sound if the tones stand out against their surroundings (12).

The mechanisms of bandwidth reduction include two essential articulatory gestures. First, the singer contracts the oral muscles and hence tenses the cheeks. In this way, he reduces the cavity wall vibration which, according to Flanagan (10), can add more than two thirds to the vocal tract damping. Second, by using a small mouth opening, formant damping due to the radiation load is lowered. The articulatory technique is described in detail by Vetter (13).

Overtone singing in the male chest voice is limited with respect to the productive as well as to the perceptive aspect. The decay of glottal spectral amplitudes is too great to favor harmonics in the region where a formant frequency is $>1.2$ kHz. For instance, if the fundamental frequency is 125 Hz and the slope in the source spectrum is $-12$ dB/oct, then the 8th harmonic amplitude is 36 dB below the first one. From an auditory point of view, one can perceive single harmonics out of a complex sound only up to the 6th or the 8th order (14). Therefore, the most beneficial frequency range of overtone prominence is limited to about 1 kHz.

According to the theory of coupled damped resonators (15), the less damped subsystem dominates the whole system. The bandwidth $b_i$ of the first and second formant normally ranges between 40 and 100 Hz, and the logarithmic decrement is about 120 Hz $\leq d_t \leq 320$ Hz. Using the damped oscillations of the vocal folds recorded with a high-speed camera by Tanabe et al. (16), one can determine the glottal damping constant $d_g$ from the decrement formula with $10$ Hz $\leq d_g \leq 20$ Hz. Hence, the glottal oscillation shows the lower damping, i.e., the glottal mechanism is the dominant part in the voice apparatus. The reduction of formant bandwidth in the case of overtone singing yields values for $d_t$ between 50 and 100 Hz. The glottal system, therefore, remains the dominant one in overtone singing. However, because the damping of a subsystem determines its autonomy, the lowered tract damping can be seen to enhance the ability of the tract to act with less influence from the glottal mechanism.

REFERENCES


