

A PHONETIC STUDY OF OVERTONE SINGING

Gerrit Bloothoof, Eldrid Bringmann, Marieke van Cappellen,
Jolanda B. van Luipen, and Koen P. Thomassen

Research Institute for Language and Speech, University of Utrecht
Trans 10, 3512 JK Utrecht, The Netherlands

ABSTRACT

We describe the phenomenon of overtone singing in terms of the classical theory of speech production. The overtone sound stems from the second formant or a combination of both the second and third formants, as the result of careful, rounded articulation from /ɔ/, via schwa /ə/ to /y/ and /i/. Strong nasalisation provides, at least for the lower overtones, an acoustic separation between the second and first formants, and can also reduce the amplitude of the first formant. The bandwidth of the overtone peak is remarkably small and suggests a firm and relatively long closure of the glottis during overtone phonation. Perception experiments showed that listeners categorize the overtone sounds differently from normally sung vowels.

1. INTRODUCTION

Overtone singing is a special type of voice production resulting in a very pronounced, high and separate tone which can be heard over a more or less constant base sound. The technique is rarely used in Western music but in Asia (especially Mongolia and Tibet) it is more common and overtone singing can be heard during secular and religious festivities. The high tone follows a characteristic musical scale [for instance, for pitch C3 (130.8 Hz) (- and + indicate a deviation from the exact tone): C3, C4, G4, C5, E5, G5, A5+, C6, D6, E6-, F6+, G6, G#6+, A6+, B6-, C7, ...], from which it can be concluded that one really hears an overtone of the fundamental.

The literature contains only a few reports on overtone singing [1,5,7,8], which indicate both the importance of formants and register type. In this paper we

present both an acoustic analysis of overtone singing and a study to evaluate the perception of the overtone sounds, in relation to normally sung vowels.

2. MATERIAL

We have recorded series of sung overtones from a singer with many years of experience in overtone singing, both as a performer and as a teacher. In this paper we describe the results for an F_0 value of 138 Hz (C#3). In addition, 12 Dutch vowels /a/, /ɑ/, /ɔ/, /o/, /e/, /ɛ/, /i/, /i/, /œ/, /ø/, /u/, and /y/, sung in a normal way at the same F_0 , were recorded.

3. ACOUSTIC ANALYSIS

The recordings were digitized at a rate of 10 kHz and stored in a computer. From the middle, stable, part of each recording 300 ms was segmented. Average power spectra were obtained from FFT analyses (1024 points, shift 6.4 ms) over this segment. Formant frequencies were computed on the basis of appropriate LPC or ARMA analysis.

3.1. FFT-Spectra

Figure 1 shows the average FFT spectra of all overtone recordings. Despite the averaging procedure, the width of each individual harmonic is limited, indicating the stability of F_0 over the interval (standard deviation of F_0 was less than 0.1 semitone in all cases). It can be seen from the shifting peak in the spectra that overtone singing seems interpretable as a special use of a formant. Obviously, the singer tries to match a formant with the intended overtone frequency and succeeds very well.

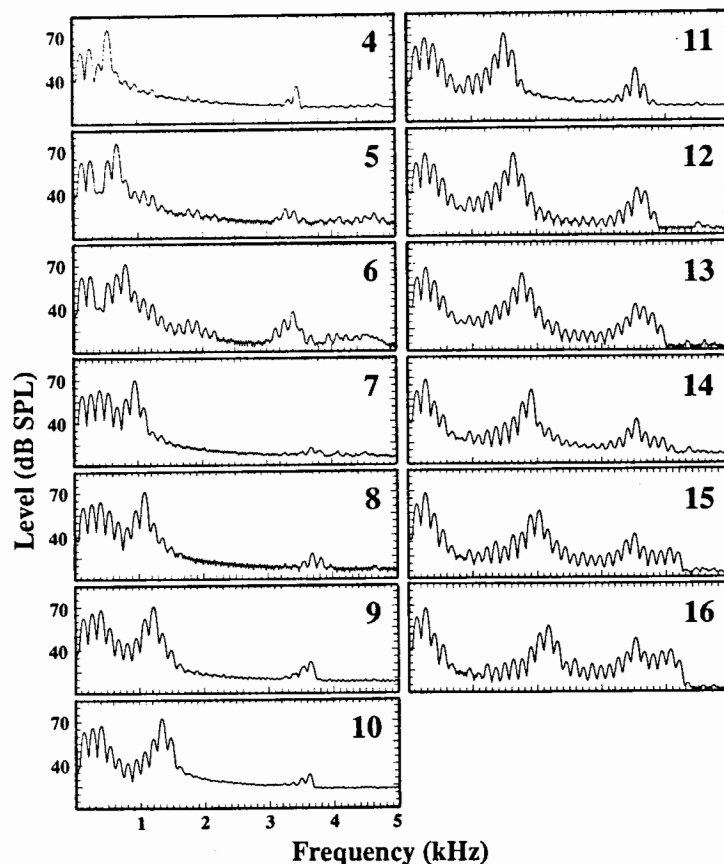


FIG. 1. Average FFT spectra for overtone sounds, sung at $F_0 = 138$ Hz (C#3). The overtone sounds are numbered according to the main partial involved.

3.2. Formant frequency analysis

In Fig. 2 we present formant frequency results for both the overtone sounds and the sung vowels in the $F_1 - F_2$ plane. The figure shows two modes in the production: firstly, the overtone sounds 4-6 around /u/, and secondly, the track from /ɔ/ to /i/.

In the first mode, it can be seen from the FFT-spectra that there is energy absorption around 400 Hz, indicating a strong nasalisation. The characteristic overtone sound resides in the second formant, as others [1,8] had already suggested. The bandwidth of the second formant is very narrow and, especially for the lower overtones, seldom exceeds

40 Hz. This indicates little acoustic damping in production: firm glottal closure and small losses in the vocal tract. All these characteristics indicate a low, rounded, nasalised, back vowel /u/ or /ɔ/ (low F_1 and F_2 , a nasal pole/zero pair, and suppressed F_3 [3]).

The second mode in the production of an overtone sound, applies for overtone frequencies higher than 800 Hz. The main peak of the spectrum still rises in tune with the intended overtone frequency and is interpreted as a combination of F_2 and F_3 . It may be of interest that the singer explains this series of overtones with the articulatory variation

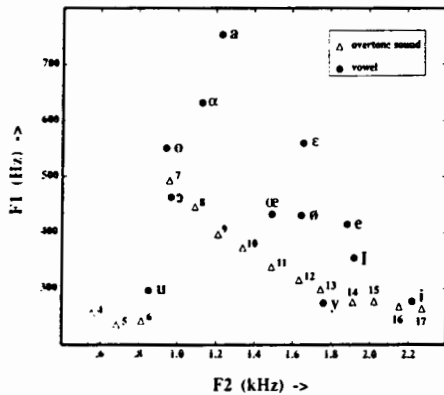


FIG. 2. $F_1 - F_2$ plane for stimuli sung at $F_0 = 138$ Hz, with positions of the vowels (IPA symbols) and overtone sounds (represented by the number of the corresponding partial).

during the word 'worry'. It is known, already from the Peterson and Barney data, that in a retroflex /r/ the F_3 frequency can be remarkably low and can approach the F_2 frequency. This has also been mentioned by Stevens (1989), especially in combination with liprounding, while Sundberg (1987) mentioned the effect as the acoustic result of a larger cavity directly behind the front teeth.

For the higher overtone sounds, the articulation comes near /y/ and /i/, where continued lip rounding makes it possible to bring F_2 and F_3 together [4], although for the highest overtones a subtle lip spread may be needed to reduce the front cavity to a minimum.

3.3. The glottal factor

The very narrow bandwidth of the "overtone formant" suggests a good and long glottal closure. We believe that the singer used modal register, with a relatively long glottal closure, originating from a firm glottal adduction. This hypothesis does not exclude that performers may use the vocal fry register as well [7]. In all cases, the long glottal closure requires a strong adduction of the vocal folds, which could easily result in general muscular hypertension in the pharyngeal region. This may relate to the prominent role of the buccal cavity, suggested by Hai (1991).

3.4. Intensity analysis

Up to an overtone frequency of 1.5 kHz, the overtone harmonic has a stable relative intensity of -10 dB relative to overall SPL, and dominates the spectrum. For higher frequencies, the relative level of the overtone harmonic sharply drops with a slope of about -18 dB/octave.

4. THE PERCEPTION OF OVERTONE SINGING

4.1. Material, listening experiment, and analysis

As stimuli we used the combined set of 14 overtone sounds and 12 Dutch vowels. From these stimuli we used the same segment (300 ms) as had been used for the acoustical analyses, but we shaped the first and final 25 ms sinusoidally to avoid the perception of clicks. In a computer-controlled experiment, these stimuli were judged by fifteen listeners on ten 7-point bipolar semantic scales. Further details of semantic scales will be presented in a forthcoming paper. The judgements were analyzed by means of multidimensional preference analysis MDPREF [2]. In the technique of MDPREF a stimulus space is constructed in which distance corresponds to perceptual (dis)similarity.

4.2. The perceptual stimulus space

The plane of the first two dimensions of the stimulus space is shown in Fig. 3. 41 % of the total variation in the judgements was

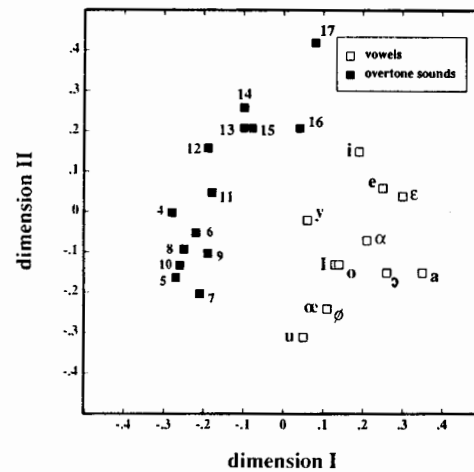


FIG. 3. The perceptual stimulus space. The overtone sounds are given by the number of their corresponding partial, the vowels by their IPA symbol.

explained in this plane, while higher dimensions each explained less than 6.3 %.

The overtone sounds and normally sung vowels are perceptually separated clusters. The vowels are situated roughly in a triangle, with the cardinal vowels /i/, /u/, and /a/ at the angles. The overtone sounds are roughly ordered according to their harmonic number, although the stimuli numbered from 4 to 10 can be described as a cluster. This probably relates to the constant relative energy of the overtone harmonic for this set. The direction of the overtone sounds is, from the lower to the higher numbers, about the same as from /u/ to /i/, as may be expected from the relation between harmonic numbers and F_2 frequency values.

4.3. A physical description of the perceptual stimulus space

We attempted to match the perceptual stimulus space with multidimensional physical descriptions of the stimuli [formant frequency space (see Fig. 2), 1/3-octave bandfilter energy space both by means of the Plomp metric and the Klatt metric [2,6]]. These attempts were not successful (low correlations between coordinate values along dimensions) because of the division into two clusters of the stimulus space, for which these metrics do not present an explanation.

Some additional perceptual sensitivity to the very small bandwidth of the "overtone formant", which clearly physically separates overtone sounds and normally sung vowels, seems necessary to explain the results.

6. REFERENCES

- [1] BARNETT, B.M. (1977), "Aspects of vocal multiphonics", *Interface* 6, 117-149.
- [2] BLOOTHOOFT, G. and PLOMP, R. (1988), "The timbre of sung vowels", *JASA* 84, 847-860.
- [3] FANT, G. (1960), "Acoustic theory of speech production" The Hague: Mouton.
- [4] FUJIMURA, O., and LINDQUIST, J. (1970), "Sweep-tone measurements of vocal tract characteristics", *JASA* 49, 541-558.
- [5] HAI, T.Q. (1991), "New experiments about the Overtone Singing Style", *Proc. Conference 'New ways of the voice'*, Becançon, 61.
- [6] KLATT, D.H. (1982), "Prediction of perceived phonetic distance from critical-band spectra: a first step", *Proc. ICASSP*, Paris, 1278-1281.
- [7] LARGE, J. and MURRY, T. (1981), "Observations on the nature of Tibetan chant", *J. of Exp. Research in Singing* 5, 22-28.
- [8] SMITH, H., STEVENS, K.N., and TOMLINSON, R.S. (1967), "On an unusual mode of chanting by certain tibetan lamas", *JASA* 41, 1262-1264.
- [9] STEVENS, K.N. (1989), "On the quantal nature of speech", *J. of Phonetics* 17, 3-45.
- [10] SUNDBERG, J. (1987), "The science of the singing voice", Dekalb: Northern Illinois University Press.