

ANALYSIS OF ACOUSTICAL FEATURES OF BIPHONIC SINGING VOICES – MALE AND FEMALE XÖÖMIJ AND MALE STEPPE KARGIRAA –

PACS REFERENCE: 43.75.RS

TAKEDA, Shoichi¹; MURAOKA, Teruo²

¹ Teikyo Heisei University; ² Musashi Institute of Technology

¹ 2289-23 Uruido

Ichihara-shi, Chiba 290-0193

JAPAN

Tel: +81-436-74-6098

Fax: +81-436-74-3970

E-mail: takeda@thu.ac.jp

² 1-28-1 Tamazutsumi

Setagaya-ku, Tokyo 158-8557

JAPAN

Tel: +81-3-3703-3111 (ext. 2948, 2949)

Fax: +81-3-5707-2177

E-mail: muraoka@ee.ec.musashi-tech.ac.jp

ABSTRACT

This paper clarifies spectral features of Mongolian or Tuva's biphonic singings such as Xöömij, Steppe Kargiraa, etc. Spectra of five types of Xöömij sounds sung by male singers showed that a resonance with a high Q value is necessary if a listener is to perceive two pitches, and the spectra of all the sounds were found to have second-formant peaks corresponding to the higher-pitch voice. Similar second-formant peaks were observed in Xöömij sounds sung by a female singer. In Steppe Kargiraa /a/ sounds sung by a male singer, we found that first formants have acute peaks instead.

INTRODUCTION

Traditional Asian biphonic singings, among which the Mongolian "Xöömij [1]" may be best known, are produced by a single singer articulating two voices simultaneously: a "drone," which is bass voice of almost constant low-pitch, and a "melody tone" of high-pitch. Xöömij is most popular in West Mongolia [2], and its singing technique is thought to have spread to European countries and been used in epical chants such as "two voices from a single mouth" in Yugoslavia [3]. Steppe Kargiraa is another example of biphonic singings sung in Tuva, Siberia located in the center of Asia.

The origin of Xöömij is still uncertain. It was once thought to have been a kind of conjuration, but today is most widely believed to have sprung from a vocal imitation of murmuring streams or the echoes in the Altai mountain-chain [3, 4]. It has also been suggested that Xöömij is an imitation of the sounds of the Morin Khuuh [3] and was used to pacify female animals separated from their young, a way in which it is still used [5].

This paper pursues the process of Xöömij generation by using the results of spectral analysis. Taking into account the results of previous acoustical analyses [6, 7], we formulated the following three hypotheses:

1. There actually is, in addition to a glottal source, an independent sound source (such as a whistle). (Hypothesis of Independent Sound Sources) [8]

2. Some portion of the vocal tract vibrates at a high frequency, and the product of the modulation of that high frequency vibration with a glottal source is perceived as the melody tone. (Hypothesis of Modulation)
3. A sharp resonance formed by a peculiar vocal tract shape selectively enhances some harmonics of the glottal source, and this resonance is perceived as the melody tone. (Hypothesis of Resonance)

Past soundspectrographic analyses [1], [9] did not prove any of the hypotheses because the amount of data analyzed was insufficient and the measurements were not accurate enough. We [6, 7], [17] first tested whether the "Hypothesis of Resonance" would be supported by the results of a detailed spectral analysis of a typical example of Xöömij singing and then repeated the analysis [18] using a Xöömij recording obtained under better conditions and using a state-of-the-art computer system. We then examined whether or not our results would hold for other types of Xöömij singing [11-13].

We first investigate the mechanism of Xöömij generation by using numerical speech signal analyses such as short-time FFT analysis, LPC analysis, and cepstrum analysis. Observing the harmonic structures of Xöömij sound waveforms and tracing the transitions of formant frequencies and the accompanying Q (quality of resonance) values, we obtained results consistent with the "Hypothesis of Resonance."

Adachi & Yamada recently also used FFT and LPC as part of their research on vocal tract shapes during Xöömij singing [10], [16]. They used four Xöömij samples sung by one singer (the type of Xöömij is unknown), and their results also support the Hypothesis of Resonance.

NUMERICAL SIGNAL ANALYSES [18]

We investigated the three hypotheses using Xöömij material. After careful auditory examinations, we selected a recording of unaccompanied single Xöömij singing entitled "Gooj Nanaa" (the singer is unknown) recorded on the LP "Folk Songs [Asian version]" (JVC SKX25017.25018, Japan).

The signal was digitized (16-bit samples) at a sampling rate of 22.05 kHz for calculation of formant frequencies, bandwidth, and Q values. For spectrum display the sampling rate was only 11.025 kHz. Short-time FFT was again applied to 1024 data samples and LPC analyses were carried out with a 30-msec Hamming window weighting. The order of LPC analysis for a sectional spectrum display was 10 and that for a 3D time-varying spectrum pattern display was 12. The orders were determined empirically by observing each spectrum.

Figure 1 is an expanded view of the middle part of a Xöömij waveform, where the waveform is considered almost stationary. The melody-pitch heights that were obtained by music score transcription approximately coincided with the second formant frequency F_2 . This suggests that the movements of F_2 are perceived as melody in Xöömij singing. To trace the variation of F_2 , we calculated the successive spectrum envelopes shown in Fig. 2.

A distinctive feature of our analysis that a formant that forms the melody tone is revealed by the use of the LPC method. As shown in Fig. 2, this formant is extracted clearly and quantitatively. Notable findings are that the intensities of the second formants of Xöömij sound waveforms are quite different from those of normal speech and that the Q values of F_2 range from 6 to 98 and have an average value of 32. According to the data in the literature [14, 15], the estimated Q of formants in normal speech is at most 30.



Fig. 1. An example of portion of a Xöömij sound waveform. This part is considered almost stationary.

The spectra of a Xöömij sound signal have a harmonic structure consistent with the Hypothesis of Resonance.

SPECTRAL FEATURES OF VARIOUS XÖÖMIJ ARTICULATIONS [11-13], [17, 18]

The detailed spectral investigation described in the previous section supports the Hypothesis of Resonance but was based on the analysis of only a single Chest Xöömij sample. A stronger conclusion could be drawn from the analysis of many samples of Xöömij with different articulations.

We further investigated samples of five types of Xöömij singing in order to find out whether there are spectral differences between the different types. The samples we analyzed were (1) Nasal Xöömij, (2) Oral-Nasal Xöömij, (3) Glottal Xöömij, (4) Chest Xöömij, and (5) Throat Xöömij. This classification is based on where the singer believes the resonance point to be, and there is no proof that the resonance is actually at that place. These Xöömij samples were sung by male Mongolian singer Ganbold and were recorded on a CD entitled "Mongolian Songs" (KING RECORD, KICC-5133, Japan (1988)).

For sound pieces in which each of the present authors perceived two tones, sharp peaks could be observed in their spectra. These peaks correspond to the second formant frequencies F_2 , which thus are strikingly enhanced and are heard as the melody tone. This was commonly found for each type of Xöömij investigated in the present study, thus supporting the Hypothesis of Resonance.

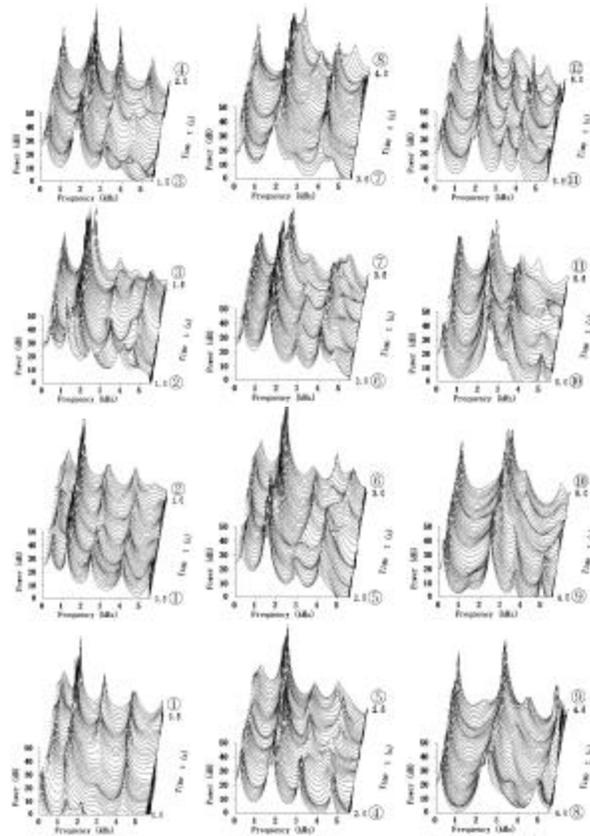


Fig. 2. Three-dimensional representation of time-varying spectra of the waveform. Acute peaks of the second formant can be clearly observed in the range between 1.2 and 2.7 kHz. These peaks correspond to the frequencies of the melody tone.

FORMANT TRANSITIONS FROM NORMAL VOWELS TO XÖÖMIJ SOUNDS [18]

We also tried to clarify the spectral features of the transition from the sounds of normal vowels to Xöömij sounds. It is widely recognized that the phonetic impressions of Xöömij sounds somehow resemble [i], [e], or [u] sounds and that Xöömij initially sounds similar to an [u] when the melody tone is not heard clearly.

We asked a Japanese Xöömij singer to articulate [(1) Normal vowel (2) Xöömij (3) Normal vowel] with one breath. The specific vowels used in this exercise were the four Japanese vowels [i], [u], [e], [o], and the singer was asked to pronounce them as normally as possible. It must be noted that our Japanese Xöömij singer's control of Xöömij articulation was inferior to that of expert Mongolian Xöömij singers because our singer was not as well trained as expert Mongolian Xöömij singers. The analysis results were summarized using an F_1 - F_2 diagram.

As shown in the F_1 - F_2 diagram in Fig 3, shifts of the F_1 - F_2 combinations toward the region of [i] were

always observed. This suggests that the location of the stricture during Xöömij singing is almost the same as its location during the articulation of the vowel [i]. In the transitions from vowels to Xöömij, F_1 shifted to about 250 Hz, while F_2 shifted into the range of 1.8 kHz . 2.3 kHz and its remarkable Q-increases were also observed. The frequency range of F_2 is almost the same as that of the melody tone.

ACOUSTICAL FEATURES OF FEMALE XÖÖMIJ VOICES

This section describes acoustical features of female Xöömij voices. It is known to be difficult for females to sing Xöömij songs.

Analysis was conducted using voices of Mongolian female singer Sainkho Namtchylak recorded on a CD entitled "Lost Rivers" (FMP CD 42, Germany (1992)). The signal was digitized (16-bit samples) at a sampling rate of 16 kHz for spectrum display. Short-time FFT and LPC analyses were carried out with a 30-msec Hamming window weighting.

Figures 4 (a) shows a short-time spectrum of monophonic part of a female Xöömij sound waveform, and (b) shows that of biphonic part. A sharp peak can be observed in the spectrum in Fig. 4 (b), whose sound is perceived as two pitches. This peak corresponds to the second formant frequency F_2 , which is strikingly enhanced and is heard as the higher pitch. This was commonly found for each sample of female Xöömij voices investigated in the present study, thus supporting again the Hypothesis of Resonance.

A conspicuous difference from male Xöömij voices is in that the harmonic structure of the spectrum of a female Xöömij sound waveform is coarse compare to that of a male one. This coarse harmonic structure may be the reason why it is difficult for female singers to control melody tones.

ACOUSTICAL FEATURES OF MALE STEPPE KARGIRAA VOICES

Another interesting biphonic singing is a Tuva's singing method called "Steppe Kargiraa," which is characterized by an extremely low fundamental pitch. Recently the voice-production process has been explained by Imagawa, Sakakibara, Konishi, and Niimi using a glottal source model based on a "false vocal fold [19]." In this section we describe the results of spectral analysis of Steppe Kargiraa sound waveforms that have an auditory impression near a vowel /a/.

Analysis was carried out using voices of two male singers,

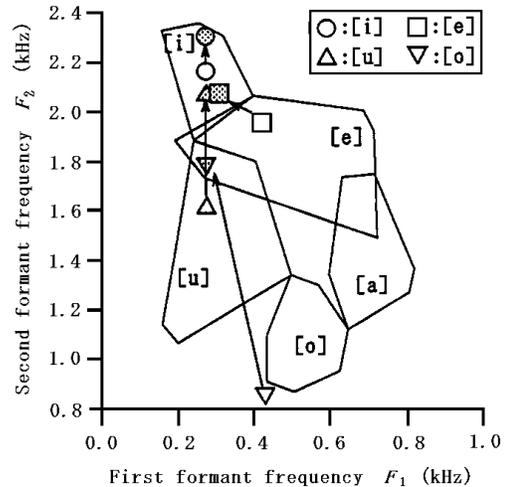
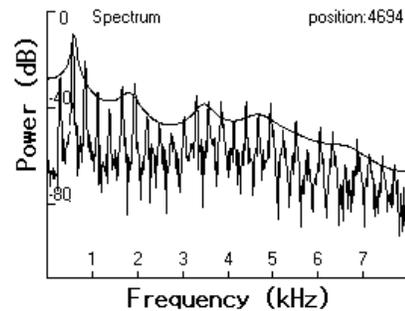
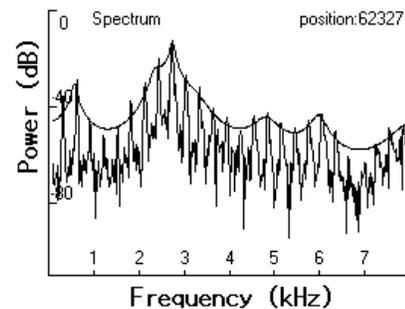


Fig. 3. Shifts of F_1 and F_2 values during transitions from normal vowel sounds to Xöömij sounds. Shifts toward the region of [i] can be commonly observed.



(a) At a point where only the lower pitch is perceived



(b) At a point where two pitches are perceived

Fig. 4. Spectra of monophonic and biphonic parts of a female Xöömij sound waveform.

Fedor Tau and Gundenbiliin Yavgaan. Tau's voices were recorded on a CD entitled "TUVA Voices from the Center of Asia" (Smithsonian Folkway CD SF 40017, USA (1990)), and Yavgaan's voices on a CD entitled "Mongolian Xöömij" (King KICW 1004, Japan (1999)). The signal was digitized (16-bit samples) at a sampling rate of 16 kHz for spectrum display. Short-time FFT and LPC analyses were carried out with a 30-msec Hamming window weighting.

Like Xöömij sound waveforms, the spectrum of a Steppe Kargiraa waveform in Fig. 5 (b) shows a prominent formant peak, while that of a normal vowel /a/ in Fig. 5 (a) does not. An interesting finding here is that the peaks yielding melody tones are not the second formant frequencies F_2 but the first formant frequencies F_1 .

CONCLUSIONS

We have analyzed spectral features of two types of biphonic singings: Xöömij in Mongolia and Steppe Kargiraa in Tuva. Measuring time-varying formant frequencies and Q values for a typical sample of Xöömij singing, we obtained results suggesting that resonance with an extremely large Q value is required for Xöömij generation. This is consistent with the Hypothesis of Resonance.

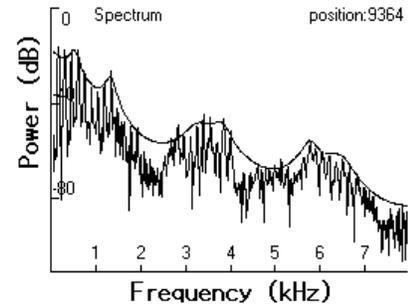
To further test this hypothesis, we evaluated samples of four types of Xöömij singing classified according to where the singer believes the resonance point to be. Sharp peaks were found in the spectra of all types of Xöömij. These results support the Hypothesis of Resonance, in which glottal waves and the sharp resonance of their higher harmonics are perceived as biphonic tones.

Another important finding in this work is that the first formant frequencies of Xöömij sound waveforms are constant. Investigating the transitions of formant frequencies from normal vowels to Xöömij sounds, we found that the F_1 - F_2 combination always shifts toward the [i] region, with the first formant frequencies shifting to about 250 Hz.

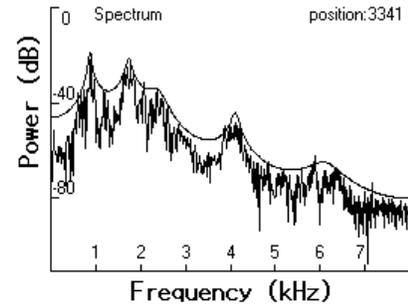
The results of analyses of spectral features of female Xöömij and male Steppe Kargiraa singings also showed sharp formant peaks in the spectra that yield perception of melody tones. A conspicuous feature of spectra of female Xöömij sound waveforms is that the harmonic structure is coarse compared to those of male Xöömij sound waveforms, which may make female singers control melody tones difficult.

ACKNOWLEDGMENTS

The authors express their sincere appreciations to Professor Kiyoko Motegi at Joetsu Kyoiku University and Mr. Masamitsu Yamakawa, previous senior engineer at JVC Company for their offer a chance to this research. And also thank with all their heart to former Professor Isao Nakamura at Teikyo Heisei University for his invaluable comments, and Messrs. Kikuji Wagatsuma, Yoshiyuki Tsuchikane, and Masato Horiuchi, the research engineers at JVC company for their cooperation to analyses, Dr. Masashi Yamada at the Osaka University of Arts for his offering useful literatures for this research, Mr. Daisuke Naganuma at Teikyo Heisei University (formerly) for his offering Xöömij sounds as a Xöömij singer, Xöömij singer Mr. G. Yavgaan, Mr. Kyoji Hoshikawa, folk music recording producer, Mr. Katsunobu Tokuda at KING RECORD Co., Ltd., President Keiko Kawashima and Ms. Hiroko Ochiai at Plankton Co. for their offering valuable information on Xöömij. Finally, the authors would like to appreciate Messrs. Masashi Itoga, Katsuhisa Tadokoro, and Masashi Miyashita, former



(a) Normal male vowel /a/



(b) Male Kargiraa /a/

Fig. 5. Spectra of normal /a/ and Kargiraa sound waveforms.

students at the Teikyo University of Technology (presently Teikyo Heisei University) for their cooperation in the experiments.

This research was partly supported by Grant-in-Aid from Teikyo Heisei University as well as Grant-in-Aid for Scientific Research on Priority Areas (2) "Diversity of Prosody and its Quantitative Description" from the Ministry of Education, Culture, Sports, Science and Technology, Japan (No. 12132206).

REFERENCES

- [1] Trần Q. H. and D. Guillou, "Original research and acoustical analysis in connection with the Xöömij style of biphonic singing," *Musical Voices of Asia, Individual research reports | Mongolia*, pp.162-173 (1980).
- [2] M. Yamada, "Mongolian biphonic singing Xöömij," *Journal of the Acoustical Society of Japan Vol. 54-9*, pp.680-685 (1998).
- [3] "A general survey of Mongolian music," *Asian traditional performing arts 1978*," The Japan Foundation, pp.5-9 (1978.11).
- [4] Batzengel, "Urtin duu, Xöömij, and Morin xuur," *Musical Voices of Asia, Seminar information and documentation | Mongolia*, pp.52-53 (1980).
- [5] H. Hasumi, "Understanding Mongolian music," *Musical Voices of Asia, Seminar information and documentation | Mongolia*, pp.142-148 (1980).
- [6] T. Muraoka, K. Wagatsuma, and M. Horiuchi, "Acoustic Analysis of the Mongolian singing Xöömij," *Preprint of the Acoustical Society of Japan 2-3-9*, pp.385-386 (1983.10).
- [7] T. Muraoka, K. Wagatsuma, Y. Tsuchikane, and M. Horiuchi, "On a Consideration of Mongolian Singing Xöömij and its Specialities," *Preprint of the seminar on Musical acoustics, The Acoustical Society of Japan MA84-1*, pp.1-6 (1984).
- [8] B. Chernov, and V. Maslov, "Larynx -double sound generator," *Proc. 11th Int'l. Cong. Phonetic Sci.*, pp.40-43 (Tallin, Estonia, 1987).
- [9] S. Gunji, "An acoustical consideration of Xöömij," *Musical Voices of Asia, Individual research reports | Mongolia*, pp.135-141 (1980).
- [10] S. Adachi, and M. Yamada, "An Acoustical Study of Sound Production in Biphonic Singing, Xöömij," *Proceedings of 1997 Japan - China Joint Meeting on Musical Acoustics*, pp.21-26 (Tokyo, 1997).
- [11] S. Takeda, M. Itoga, Y. Sato and Y. Ueda, "Analysis of Acoustical Features of Mongolian Singing "Khöömij"," *Proc. Acoust. Soc. Jap. 2-7-15*, pp605-606 (Oct, 1992).
- [12] S. Takeda, M. Itoga, "On the differences in Spectra in Accordance with the Phonemic and Tone-height Differences in Mongolian Singing "Khöömij"," *Proc. Acoust. Soc. Jap. 2-3-3*, pp.499-500 (March, 1993).
- [13] S. Takeda, M. Itoga, "Analysis of Acoustic Features of Mongolian Singing "Khöömij"," *Technical Report on Musical Information Sci.1-4*, pp.1-4 (April, 1993).
- [14] J. Ohizumi, and Y. Fujimura, *Onseikagaku (Science of Human Voices)*, Tokyo University Publishing (1972).
- [15] K. Nakata, *Onsei (Human voices)*, Acoustic Engineering Series by the Acoustical Society of Japan (Corona Publishing Co., Ltd., Tokyo, 1977).
- [16] S. Adachi, and M. Yamada, "An Acoustical Study of Sound Production in Biphonic Singing, Xöömij," *Journal of the Acoustical Society of America*, 105, pp.2920-2932 (May, 1999).
- [17] T. Muraoka, S. Takeda, and M. Itoga, "Analysis of Acoustic Features of Mongolian Xöömij Singing," *Journal of the Acoustical Society of Japan Vol. 56-5*, pp.308-317 (May, 2000).
- [18] T. Muraoka, S. Takeda, and M. Itoga, "An Acoustical Analysis of Mongolian Xöömij Singing," *Journal of the Acoustical Society of America* (in submission).
- [19] H. Imagawa, K. Sakakibara, T. Konishi, and S. Niimi, "Glottal Source Model for Throat Singing Based on Vocal Fold and False Vocal Fold Vibrations," *Proc. Acoust. Soc. Jap. 1-6-14*, pp.255-256 (March 2001).