WHY DO WE PERCEIVE TWO TONES SIMULTANEOUSLY IN MONGOLIAN TRADITIONAL SINGING, XOOMIJ?

Masashi Yamada

Department of Musicology, Osaka University of Arts
Higashiyama, Kanan, Minamikawachi
585-8555 Osaka
JAPAN
Tel: +81-93-3781 (ext.3241)
Fax: +81-93-7914
E-mail: m-yamada@osaka-geidai.ac.jp

ABSTRACT

Xoomij is a traditional Mongolian singing style in which a male sings two tones simultaneously. Acoustical analyses and psychoacoustical experiments showed that the following three factors result in two-tone perception for Xoomij: (1) A Xoomij singer use the tongue to create a vocal tract which results in a large degree of resonation for a higher component of the glottal sound. (2) The amplitude of the emphasized component is deeply modulated, whereas the other components remain stable. (3) The nature of the high tone melody is such that the component currently being emphasized was not done so previously. Such a "new comer" stands out in our auditory perceptual system.

1. INTRODUCTION

Xoomij (pronounce [Hoomii] or [Huumii]) is a type of traditional singing in Mongolia in which a male sings two pitches simultaneously. One pitch is high and varies with time (melody pitch) and the other one is low and constant (drone pitch). A similar singing style, called "Xoomei [Hoomei]," exists in Tuva, a Russian neighboring area.

Tran and Guillou proposed the "resonance" theory of Xoomij production [1]: A Xoomij singer use the tongue to divide the vocal tract into two cavities connected by a narrow opening. These two cavities create an extreme degree of resonance, emphasizing a high component of the glottal sound. Thus the emphasized component corresponds to the melody tone. On the other hand, Chernov and Maslov proposed a "double-source theory" of Xoomij production [2]: Using indirect laringoscopy, tomography, *etc.,* they observed a nozzle-like narrowing formed by ventricular vocal folds in the upper portion of the true vocal folds and suggested that the melody pitch was produces by this narrowing, whereas the true vocal folds produced the drone pitch. Several following studies by Japanese researchers presented acoustical

evidence that supported the resonance theory and refuted the double-source theory [3, 4]. Finally, Adachi and Yamada determined the three-dimensional shape of a Xoomij singer's vocal tract using MRI, and synthesized the Xoomij sound using the transfer function of the vocal tract shape. The results of this study showed that the extreme resonance is caused almost totally by the rear cavity, the area from the glottis to the narrowing of the tongue [5]. Although the series of studies described above clarified the production mechanism of Xoomij, only a few studies have been performed investigating the perceptual process.

In our environment, the sound that reaches our eardrums is usually a mixture from several sources. However, our auditory system assigns each component in the mixture to a stream (sound source) accurately with no difficulty. Consequently, the perceptual attributes (pitch, loudness, timbre, *etc*.) of events belonging to each of the streams are perceived. This process was named "auditory scene analysis" or "auditory stream segregation" by Bregman [6,7]. In the case of Xoomij, we perceive two pitches simultaneously as if two sources are producing two tones, however the sound is actually produced by a single vocal system. From the point of view of auditory scene analysis, the perception of Xoomij is one of the most interesting subjects for investigating auditory stream segregation process, *i.e.,* the question "why do we perceive two streams for the sound that is produced by a single source in Xoomij?" suggests at least a part of the answer to the general question "how does our auditory system accurately divide the mixture of sounds into streams that correspond to the individual sound sources in our environment?"

There are three possible factors in the two-tone perception of Xoomij: (1) A Xoomij singer constructs a high-Q resonator in the vocal tract, emphasizing a component of the glottal sound. This emphasis must be a factor of the segregation of the component from the other components. In fact, Xoomij sounds sung by professional singers contain extremely emphasized components and such Xoomij sounds as having very loud and dear melody tones, whereas the sounds sung by amateur singers contain poorly emphasized components and perceived as having rather soft, indistinct melody tones [8]. (2) When we listen to the performances of great Xoomij singers, we sometimes perceive deeply vibrated melody tones with stable drone tones. This may mean that only the emphasized component is deeply modulated in frequency or amplitude, while the other components remain stable. In terms of Gestalt psychology, the components that share the same motion tend to be grouped, *i.e.,* "common-fate" components tend to be grouped and the component which has a different fate tend to be segregated. This "common-fate" factor may contribute to the segregation process in two-tone perception for professionally sung Xoomij. (3) The nature of the melody is such that the component currently being emphasized is constantly changing as the pitch of the melody tone changes. Our auditory perceptual process tends to make this "new-comer" component stand out from the remaining components. Bregman called this process as "old-plus-new" heuristics [6,7]. This "old-plus-new" factor may also contribute to the segregation process.

In the present study, we use formal and informal psycho-acoustical experiments to clarify how the three factors described above contribute to the two-tone perception for Xoomij.

2. SOUND MATERIALS

One of the greatest Mongolian Xoomij singers, Ganboldt, sang three long tones for about 3.5 s, with the melody pitches of G_6, A_6 and C_7, and a consistent drone pitch F_3. He also sung several Mongolian traditional songs without instrumental accompaniment. These sounds were recorded in a soundproof room for use in the following experiments.

3. WHICH COMPONENTS CORRESPOND TO THE TWO PERCEIVED TONES?

3.1. Method

The researchers who supported resonance theory believe that the emphasized component resulting from the resonation in the vocal tract is perceived as the melody tone and the other components are combined and perceived as the drone tone. However, formal psycho-acoustical evidence has not been presented. Therefore, the first step of the present study is to conduct a formal psycho-acoustical experiment in order to determine which components correspond to the two perceived tones.

The central 3.0 s portions of the recorded three long-tone sounds were defined as original sounds and used in the present experiment. Eight students majoring in music participated as subjects. The experimental apparatus consisted of two sinusoidal wave oscillators that produced two pure tones and a MO disc player that presented one of the original sounds to the subjects. In one trial, subjects freely toggled between the original sound and the two pure tones, adjusting the frequency and attenuation of the oscillators so as to match the pitches and loudness of the two pure tones to the melody and drone tones of the original sound. This was accomplished by adjusting the frequency and attenuation of the oscillators. Both the original sounds and the pure tones were presented through headphones, and with the loudness level of the original sounds at 75 dB(A).

3.2. Results and Discussion

For all the original sounds, the intensity of the two pure tones adjusted by the subjects exceeded 50 dB SPL. This implies that the subjects perceived two tones. The mean frequency of the lower pure tone corresponded to the fundamental frequency and the high tone corresponded to the frequency of the second formant that was estimated by the LPC method. This result indicates that the drone and melody pitches perceived by the subjects consistently corresponded to the fundamental frequency and the frequency of the resonated component, respectively.

4. EFFECTS OF THE "OLD-PLUS-NEW" AND "COMMON-FATE" FACTORS

First let us consider the Xoomij singing style, where the melody pitch is changing and the drone pitch is steady. Second, consider a long-tone Xoomij singing, where both the melody and drone pitches are steady. Of the possible three factors, if the "old-plus-new" factor is significant in the two-tone perception, the melody tone in the second case would hove to be significantly softer than in the first case. Similarly, if a portion of a song where the melody and drone pitches remain constant is presented alone, the melody tone would have to sound significantly softer than when this same portion is presented in the song.

To examine the significance of the effect of this "old-plus-new" factor on the two-tone perception, several long-tone portions, where the melody and drone pitch remain constant for more than 4 s were extracted from the recorded Xoomij songs sung by Ganbold. However, the melody tone in these extracted portions still sounded very loud, and there seemed to be no significant differences in the loudness of the melody tone whether presented alone or in the songs. This conclusively shows that the "old-plus-new" effect is not a primal factor in the two-tone perception of Xoomij, but is rather a minor factor.

In the next step, we investigate whether the vibration of the melody tone in professionally sung Xoomij is caused by an amplitude modulation (AM) or a frequency modulation (FM). The recorded long tones sung by Ganbold were analyzed: Each component of the long tones was isolated using band-pass filters and then the amplitude and the period of each cycle were determined for each component. The resulting

amplitude was plotted as a function of time for each component. These plots exhibited the features of AM. Similarly, the period for one cycle was plotted as a function of time for each component. These plots exhibited the features of FM. Figure 1 and 2 show the AM and FM features for the long tone with the melody pitch of G_6 and the drone pitch of F_3. Figure 1 shows that the melody component (in this case the 9th harmonic) is deeply modulated in amplitude, while the other components remain stable. On the other hand, Fig. 2 shows that the frequency of the melody component fluctuates but the fluctuation is almost identical to that of the other components. This "different-fate" AM in the melody component may be caused by frequency fluctuation of the second formant by means of vibrating the tongue.

To examine the significance of the effect of this "common-fate" factor on the two–tone perception, sounds were synthesized with each component having an RMS amplitude matching each component in the original recorded long tones. However, there was no fluctuation in amplitude or frequency in the synthesized sounds. The melody tone for the synthesized sounds still sounded very loud, and there seemed to be no significant differences in loudness between the melody tones in synthesized sounds and original sounds This means that the "common-fate" effect is not a primal factor on the two-tone perception, but is rather also a minor factor.

The two informal psychoacoustical experiments suggest that the emphasis of a component itself acts as the primal factor in the two-tone perception. In the final step of the present study, it is quantitatively determined how the emphasized component is segregated from the other components.
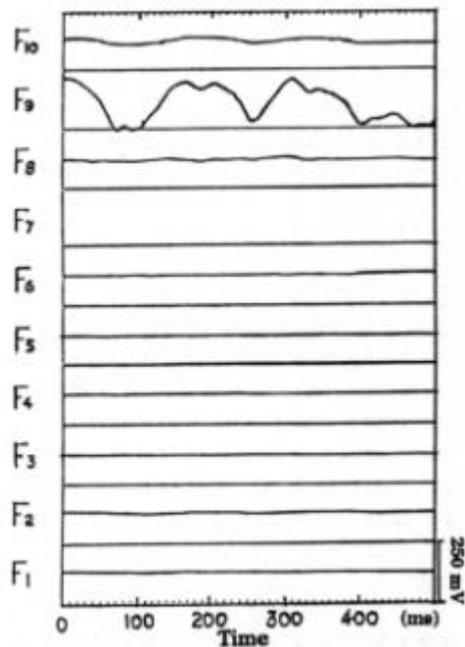
Fig. 1 The features of amplitude modulation of the first to tenth harmonics in a long-tone singing with the drone and melody pitches of F_3 and G_6, respectively, by the great Xoomij singer, Ganbold. Because the seventh harmonic was difficult to separate from the eighth harmonic, AM of the seventh harmonic is not plotted. Fn on the left side of the panels shows the n-th harmonic.
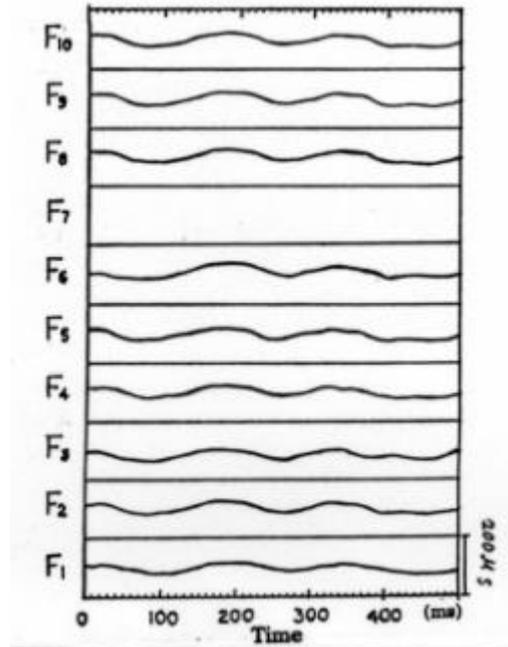
Fig. 2 The features of frequency modulation for the long-tone singing with the drone and melody pitches of F_3 and G_6, respectively.

5. HOW IS THE EMPHASIZED COMPONENT SEGREGATED?

5.1. Method

The goal of this section is to quantify how a portion of the emphasized component is segregated from the other components. Therefore, sounds that contained no fluctuation in amplitude or frequency were synthesized similarly as in the previous section for use in a formal psychoacoustical experiment. Presentation of these synthesized sounds eliminated the effects of the "old-plus-new" and "common-fate" factors.

Steady sounds, which possessed the same long-term power spectra as the original sounds (3.0 s portions of the long tones recorded), were synthesized (0 dB sounds). These 0 dB sounds were presented at 75 dB(A). Other steady sounds were also synthesized, for which the emphasized component was -3, +3, +6 dB referenced to the emphasized component of the 0 dB sounds, while the other components were the same as the 0 dB sounds (-3, +3, +6 dB sounds, respectively). These four kinds of steady synthesized sounds for each of the recorded three original long-tone sounds (melody pitches of G_6, A_6 and G_7, and a consistent drone pitch of F_3) were used in the present experiment. Five musicians participated as subjects. By adjusting the attenuation of the oscillators, the subjects matched the loudness of the two pure tones to the melody and drone tones.

5.2. Results and Discussion

The power spectrum for each of the twelve synthesized sound was determined (Fig. 3 (a)), and the mean sound pressure level of the pure tone that corresponded to the melody tone was calculated (Fig. 3 (b)). The power of the melody tone was then subtracted from the emphasized component of the synthesized sound, while the power of the other components was not changed, and the resulting power spectrum was plotted (Fig. 3(c)). The resulting spectra all show a smooth and similarly shaped envelope for the −3, 0, +3, +6 dB sounds. This is true for all three original sounds.

These results suggest the following segregation process: In the perceptual process, a smooth envelope is drawn for the input Xoomij spectrum. Then the portion of the power of the resonated component that is excluded from the envelope is perceived as the melody tone, and the remaining portion of the component contributes to the drone tone along with the other components within the envelope. This envelope is rather consistent for varying degrees of emphasis of the component corresponding to the melody pitch. However, a slight, systematic difference in the envelope is also observed, *i.e.*, the spectral envelope for a more deeply resonated sound shows a steeper spectral peak. The level difference in the spectral peak was approximately 5 dB between the −3 dB and +6 dB sounds. This difference can be explained as follows: by the subjects made a more concentrated effort to "pick up" the melody tone, when the melody tone was soft.

6. CONCLUSIONS

The present study showed that a consistent spectral envelope segregated the melody and the drone tones, and that additionally the "common-fate" factor in AM features and the "old-plus-new" factor also may contribute to the two-tone perception. In the next stage, the "common-fate" and "old-plus-new" factors will be accounted for in the formal psycho-acoustical experiments, and the overall perceptual process of Xoomij will be holistically quantified.
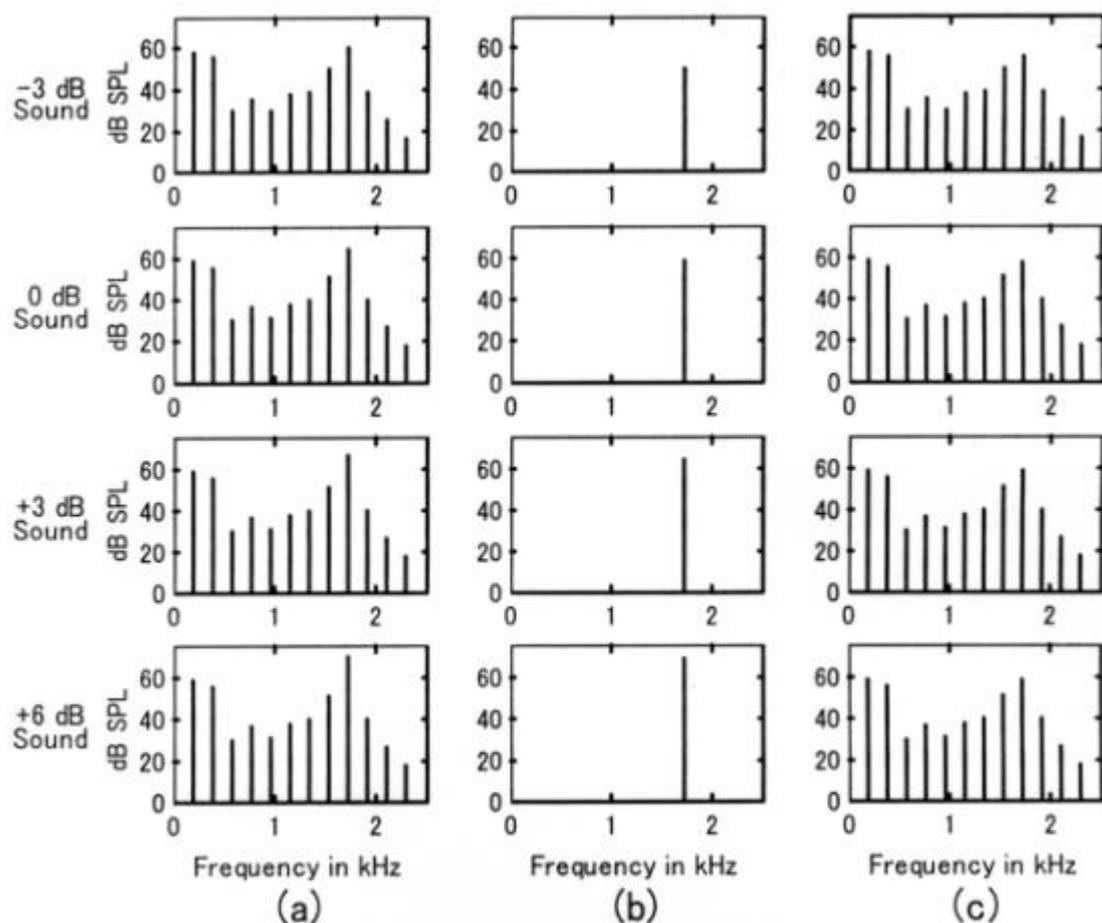
Fig. 3 Panels (a) and (b) show examples of the power spectra of the synthesized Xoomij sounds and the perceived melody tone, respectively. Panel (c) shows the spectra after the power of the melody component is subtracted from the spectra of the synthesized sounds.

BIBLIOGRAPHYCAL REFERENCES

[1] Tran, Q. H. and D. Guillou, "Original research and acoustical analysis in connection with the Xoomij style of biphonic singing," In *Musical voices of Asia* (Heibonsya, Tokyo, 1980), pp.162-173.

[2] B. Chernov and V. Maslov, "Larynx—double-sound generator," Proc. 11th Int'l Cong. Phonetic Science, (Tallinn, Estonia) 40-43 (1987).

[3] T. Muraoka, S. Takeda and M. Itoga, "Analysis of acoustic features of Mongolian xoomij singing," *J. Acoust Soc. Jpn.* **56**, 308-317 (2000) (in Japanese).

[4] S. Gunji, An acoustical consideration of xoomij," In *Musical voices of Asia* (Heibonsya, Tokyo, 1980), pp.135-141.

[5] S. Adachi and M. Yamada, "An acoustical study of sound production in biphonic singing, Xoomij," *J. Acoust. Soc. Am.* **105**, 2920-2932 (1999).

[6] A. S. Bregman, *Auditory scene analysis* (The MIT Press, Cambridge, 1990).

[7] A. S. Bregman, "Auditory scene analysis," In *Thinking in sound*, S. McAdams and E. Bigand Eds., Chap. 2, pp. 10-36.

[8] M. Yamada, "Stream segregation in Mongolian traditional singing, Xoomij," Proc. Int. Symp. Music Acoust., Soc. Franciase d'Acoust, 539-545 (Dourdan, 1995).