Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Correlation between Groovy Singing and Words in Popular Music

Yuma Sakabe, Katsuya Takase and Masashi Yamada

Kanazawa Institute of Technology, Kanazawa, Japan

PACS: 43.75.Rs, 43.75.Cd, 66.Mk

ABSTRACT

Musical critics often point out that some singers start to sing slightly after the accompaniment to give a groovy feeling. In our previous studies, we revealed that a Japanese Pop diva, Namie Amuro started to sing approximately 70-90 ms after the accompaniment for the initial notes of phrases. We then conducted perceptual experiments, using blass-like tone played the melody instead of Amuro's singing. The results of the experiments showed that the 70-ms delayed singing sounded "unnatural" and "not groovy". This suggested that we have high tolerance for the delayed singing in the case that a singer sings words than in the case an instrument plays a melody. In the present study, we conducted perceptual experiments using synthesized word singing and scant singing stimuli, as well as human word singing and scat singing stimuli. The results showed that whether words are sung or not is crucial for the tolerance for the delayed singing. The scat singing stimuli showed consistent results to the case where a brass-like tone played the melody.

INTRODUCTION

It has been often discussed how popular singers sing in a groovy way. Musical critics point out that some singers start to sing slightly after the accompaniment to realize a groovy feeling. This kind of delayed singing style is used Hip-Hop or Rap music in English words mainly. Japanese words were thought not to be suitable for Hip-Hop music, because Japanese language possesses an isochronal mora structure. However, in the mid '90s a great music producer Tetsuya Komuro dramatically changed this situation, importing "Euro-beat" to Japanese popular music. He also composed, performed and arranged his music with MIDI synthesizers and computers. Komuro produced a large number of popular songs not only for his own groups, but also many singers or units. These musicians were called Komuro's family. Until 2000, top ten hit charts were filled with several musicians from Komuro's family. Since 2000, Tsunku, and then, Yasutaka Nakata, replaced the position of the number one producer instead of Komuro. However, the size of the popular music market in Japan is shrinking.

Namie Amuro is the most popular musician in Komuro's family and is recognized as the greatest diva in'90s Japanese popular music. Many of her CDs have sold more than double million copies. Generally, her singing style is recognized as "groovy." Her groovy singing style and Komuro's fascinating music correlatively becharmed people in the '90s. In '90s, it becomes to be popular that singers sing Japanese words with quite different rhythm from natural Japanese language, abandoning the isochronal mora structure. This resulted out that number of Japanese popular songs which sounded the same flavor of western popular music, increased. These songs are called J-POP. Sometimes elder Japanese listener complains that they cannot catch the words in J-POP without subtitles on the TV screen.

Music critics point out that Namie Amuro is deeply influenced by Hip-Hop or Rap music, in which the singers or rappers start to sing or rap slightly after the accompaniment. Yamada and his colleague evidenced that Amuro uses this type of delayed singing in one of her magnum opus, NEVER END [1]. They determined the asynchrony between the onsets of singing and of the accompaniment. The results showed that the asynchrony was within 50 ms. However, for the initial notes in several phrases, the onset of singing was delayed 70-90 ms from the onset of the accompaniment. On the other hand there is no notes where singing prior to the accompaniment more than 50 ms [2, 3]. It is known that if the onset of one tone is delayed from another within 50 ms, asynchrony between the two tones cannot be perceived [4,5], but asynchrony is clearly perceived for a delay over 100 ms. A delay of 70-90 ms should lead to a perception of not being just synchronous but also not clearly asynchronous. This suggested that this style of delayed singing is deeply correlated with the "groovy" impression.

Saikawa and Yamada conducted perceptual experiments to clarify the correlation between the 70-90 ms delay singing and "groovy" singing. In the first experiment, a sung melody was played by brass-like tone and mixed with the accompaniment. The results of the experiment showed that the 70-ms delayed performance was not evaluated as more groovy than the just synchronized one. In their second experiment, the singing was synthesized using the systems, HATSUNE MIKU and STRAIGHT. In this experiment, the melody was sung with lyrics like Namie Amuro does. The results showed that the 70-ms delayed singing sounded more groovy and natural than the just synchronized performance. This suggested that the delayed performance style is more effective in the case that a singer sings with words than the case that an instrument plays the melody [6]. However, the study of Sai-

kawa and Yamada did not cut two factors clearly; the factor of singing words and the factor of singing in human voice.

In the present study, perceptual experiments were conducted to clarify which factor described above is crucial for the high tolerance for the delayed singing.

EXPERIMENT 1

Experimental Method

The phrases shown in Fig. 1 were excerpted from NEVER END. For the accompaniment part, the *karaoke* track recorded in the CD was used. The singing part was synthesized using HATSUNE MIKU and STRAIGHT.

The program HATSUNE MIKU [7], developed by Crypton Future Media, Inc., is one of the character vocal synthesizer series which utilize Yamaha's Vocaloid 2 technology. Using HATSUNE MIKU program, singing with words was synthesized, easily. However, the timing of the onset of the singing voice was not able to change from the notated timing. The manipulation in the temporal feature was achieved naturally using the very high-quality speech manipulation system STRAIGHT, developed by Hideki Kawahara [8, 9].

For the two initial notes in the phrases, the perceptual onset timing of singing voice was set at -70, -50, -30, 0, +30, +50, +70, and +110 ms compared to the accompaniment (The negative values imply that the melody tone is prior to the accompaniment). Vos and Rasch conducted perceptual experiments to determine the perceptual onset timing for tones which possess various rise slopes in the time envelope. They showed that the perceptual onset timing of a tone is defined as the moment when the envelope across the level of -15 dB relative to the maximum level of the tone [10]. We adopted this definition of the perceptual onset to compare the timings of the tones we used in the experiments which possessed various types of envelope. For the notes other than the initial notes, the perceptual onset of the melody tone was set at the exact timing indicated on the score. The accompaniment started the performance four meters before the melody started to allow listeners to anticipate the timing of the initial note.

Using these nine stimuli, a perceptual experiment was conducted with Scheffe's paired comparison method. Five students from the Department of Media Informatics at Kanazawa Institute of Technology participated as listeners. They sat on a chair in a sound-proof room and diotically listened to the stimuli through STAX Lambda-pro headphones, at 81.5 dB LAeq. The participants were requested to evaluate the performances on three scales; "naturalness," "grooviness," and "synchrony between the melody and the accompaniment."

In one trial, one pair out of the nine stimuli was presented to a listener: A click indicated the start of the trial, and a 1-sec interval followed. Then a pair of the stimuli was presented. The former and latter performances were separated by a 2-sec interval. Following the latter stimulus the listeners evaluated within a 5-sec response interval using a scale of one to seven, e.g., "the former is very groovy in comparison with the latter," "the former is quite groovy in comparison with the latter," …, "the latter is very groovy in comparison with the former." The experiment was divided into three blocks. Each block corresponded to each of the three impression scales. One block consisted of 72 trials. These 72 trials were divided into three sessions of 24 trials, and the sessions were separated by a 20-minute rest period. It took approximately 25 min for one session. A listener carried out one block a day and finished the whole experiment in three days. The three blocks and the 72 trials in a block were performed in a random order for each listener.

Results

Figure 2 shows the resulting mean vales of the nine stimuli on the three scales. Figure 2 shows that the stimuli for which the singing voice delayed within 90 ms sounded "natural," and "groovy" as the just synchronized stimulus (0-ms stimulus). In contrast, the stimuli for which the singing voice precedes the accompaniment, or is delayed by 110 ms, sound "unnatural," "not groovy," and "asynchronous".

It should be noted that the 70-ms delayed singing sounded as groovy and natural as well as the 0-ms stimulus, in the case of synthesized singing with words.

EXPERIMENT 2

Experimental Method

In Experiment 2, HATSUNE MIKU sung /ne/ for all melody notes (scat style). We manipulated the onset timing using STRAIGHT. In the present experiment, the accompaniment was played by a MIDI system. For the two initial notes in the phrases, the perceptual onset timing of singing voice was set at -70, -50, -30, 0, +30, +50, +70, and +110 ms compared to the accompaniment. For the other notes, the perceptual onset of the singing voice was set at the exact timing the score indicated. Using these nine stimuli, the perceptual experiment was conducted. Five students from the Kanazawa Institute of Technology participated as listeners. The other experimental method was identical to Experiment 1.

Results

Figure 3 shows the resulting mean vales of the nine stimuli on the three scales. The results shows that the stimuli for which the scat singing delayed within 50 ms sounded "natural," "groovy", and "synchronous" as the just synchronized stimulus (0-ms stimulus). In contrast, the stimuli for which the scat singing precedes the accompaniment, or is delayed by 70 ms, sound "unnatural," "not groovy," and "asynchronous". The 50-ms delayed performance shows the tendency to be sounded more groovy, natural, and synchronized with the accompaniment in comparison to the 0-ms stimulus, although *t*-test showed that there were no significant differences in the significance level of $p<.05$.

It should be noted that the 70-ms delayed stimulus sounded less groovy, natural and synchronous with the accompaniment in comparison to the 0-ms stimulus, in the case of synthesized scat singing.

EXPERIMENT 3

Experimental Method

In Experiment 3, a human amateur singer sung words (Experiment 3a), or sung /ne/ for all melody notes (scat singing, Experiment 3b). This singing voice was recorded using the digital recorder, KORG 'D888' in a sound-proof room.

Then, the timing of the synthesized singing voice was manipulated using STRAIGHT. For the two initial notes in the phrases, the perceptual onset timing of singing voice was set at -70, -50, -30, 0, +30, +50, +70, and +110 ms compared to the accompaniment. For the other notes, the perceptual onset of the singing voice was set at the exact timing the score indicated.

Word singing stimuli were used in the Experiment 3a, and Scat singing stimuli were used in the experiment 3b. Five students from the Kanazawa Institute of Technology participated as listeners. The other experimental method was identical to Experiment 2.

Results

Figure 4a shows the resulting mean vales of the nine stimuli on the three scales for the human word singing. Figure 4a shows that the stimuli for which the human word singing delayed within 70 ms sounded "natural", "groovy" and "synchronous" as the just synchronized stimulus (0-ms stimulus). Figure 4b shows the results for the human scat singing. Figure 4b shows that the stimuli for which the human scat singing delayed within 70 ms sounded "unnatural", "not groovy" and "asynchronous".

It should be noted that the 70-ms delayed performance sounded "natural" and "groovy" for the human word singing but "unnatural" and "not groovy" for the human scat singing.

GENERAL DISCUSSION

The results showed that the 70-ms delayed singing sounded "natural" and "groovy", in the case that words were sung. These results are consistent with the case where a blass-like sound played the melody [6]. However, the 70-ms delayed singing sounded "unnatural" and "not groovy" in the case of scat singing. This discrepancy implies that whether words are sung or not is crucial for the tolerance for delayed singing, but whether the melody is played with an instrument or sung by human voice timbre is not crucial.

CONCLUSION

The present study showed that we have high tolerance for delayed singing when words are sung, but low tolerance when words are not sung. However, it is not clarified whether the interpretation of the meaning of the words is crucial or changing the timbre is crucial. In the next step, it has to be conduct experiments using singing with no meaning words to clarify this.

The results showed that 90-ms delayed singing was perceived as natural and groovy for the case that the words were sung. However, in the case of the scats, 90-ms delayed performance was perceived as significantly inferior to the performances with 0-70 ms delay. The results for the scats are consistent with the case of the brass tone. These results show that we have a higher tolerance for the delayed play in the case of words being sung than in the case of a single timbre playing the melody.

REFERENCES

1 N. Amuro, NEVER END, Produced, composed, arranged and written by Tetsuya Komuro AVEX TRACS, Tokyo, AVCD-30137 (2000).

2 M. Yamada and N. Masuda, "Japanese Pop diva Namie Amuro's tendency to start singing slightly after the accompaniment", *Proceedings of the 2ns International Conference of Asia-Pacific Society for the Cognitive Science of Music*, 59-64 (Seoul, 2005).

3 M. Yamada, "The correlation between a singing style where singing starts slightly after the accompaniment and a groovy, natural impression: The case of Japanese Pop diva Namie Amuro", *Proceedings of the 9th International Conference on Music Perception and Cognition*, 22-27 (Bolognia, 2006).

4 R. A. Rasch, " The perception of simultaneous notes as in polyphonic music", *Acoustica* **40**, 21-33 (1978).

5 R. A. Rasch, "Synchronization in performed music", *Acoustica* **43**, 121-131 (1989).

6 M. Saikawa and M. Yamada, "A perceptual study on the groovy singing style in popular music", *Proceedings of the 10th Western Pacific Acoustics Conference,* CD-ROM, 7 pages (Beijing, 2009).

7 HATSUNE MIKU, Crypton Future Media, Inc., Sapporo, (2007).

8 H. Kawahara, "STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds," Acoustic Science and Technology 27, 349-353 (2006).

9 H. Banno, H. Hata, M. Morise, T. Takahashi, T. Irino and H. Kawahara, "Implementatioin of realtime STRAIGHT speech manipulation system: Report on its first implementation", *Acoustic Science and Technology*, **28**, 140-146 (2007).

10 J. Vos and R. A. Rasch, "The perceptual onset of musical tones," *Perception and Psychophysics,* **29**, 323-335 (1981).

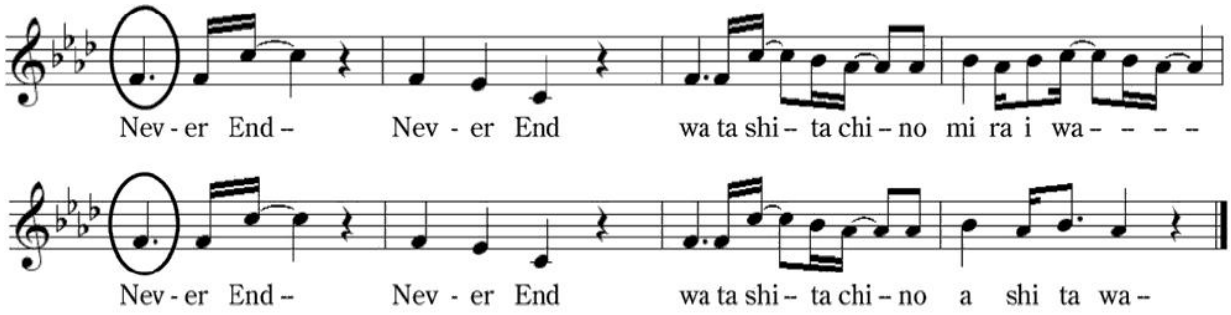

Figure 1. The excerpt from NEVER END used in experiment. For the two notes marked with circles, the melody tone was set at various timings to the accompaniment.

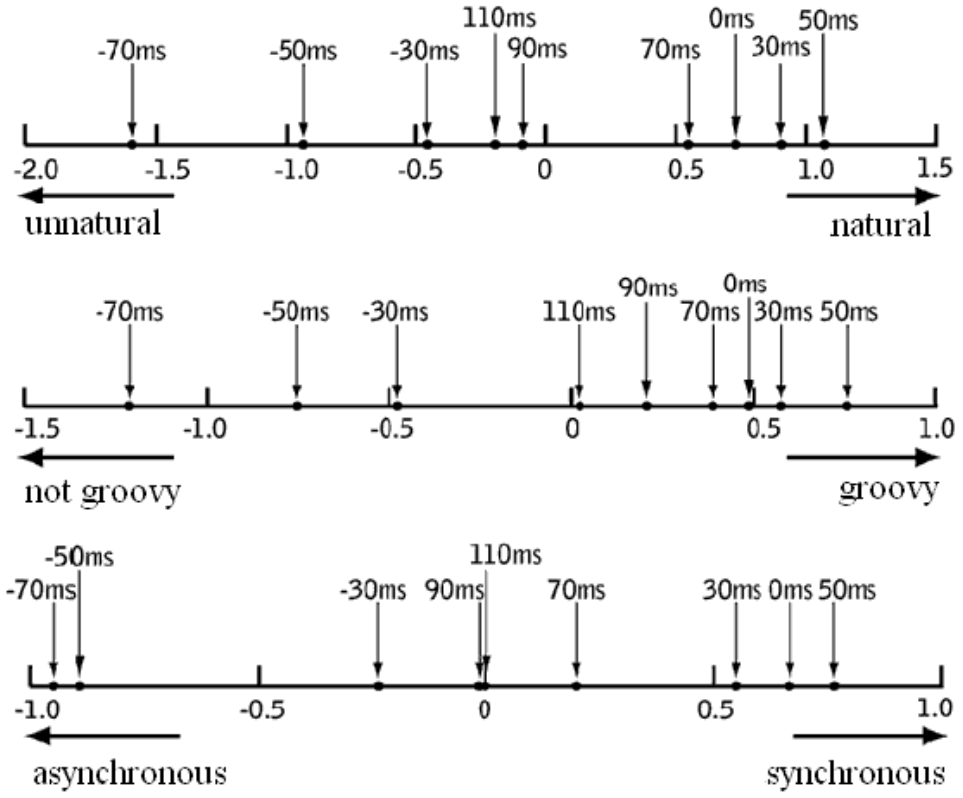

Figure 2. The results in Experiment 1 (syntrhesized word singing)

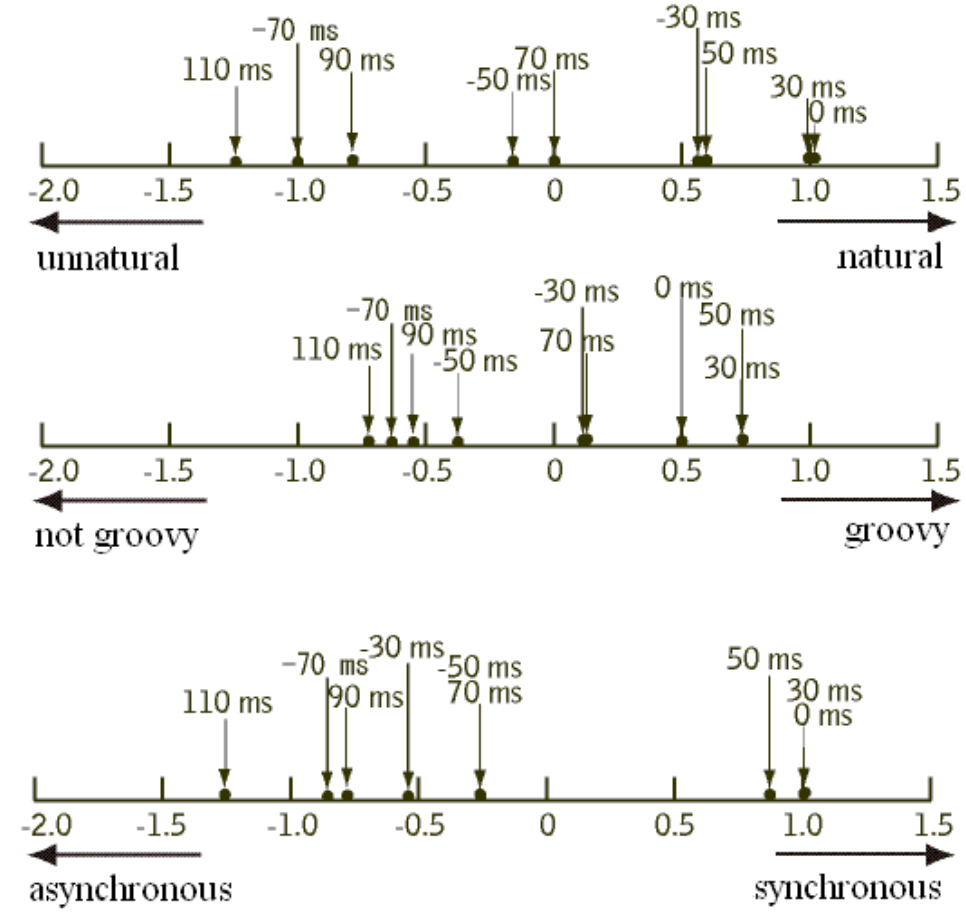

Figure 3. The results in Experiment 2 (synthesized scat singing)

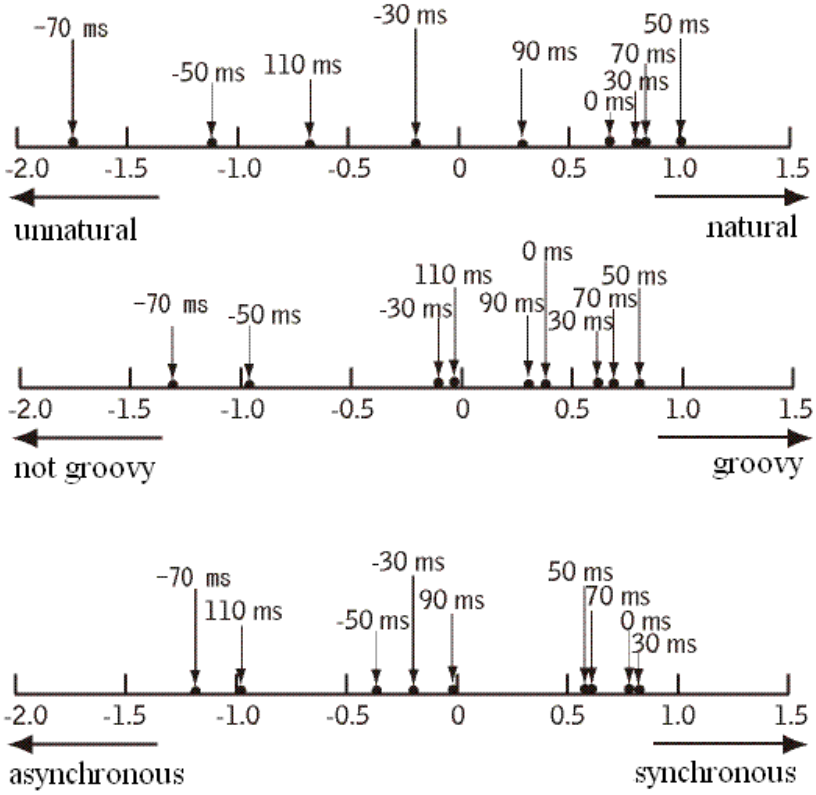

Figure 4a. The results in Experiment 3a (human word singing)

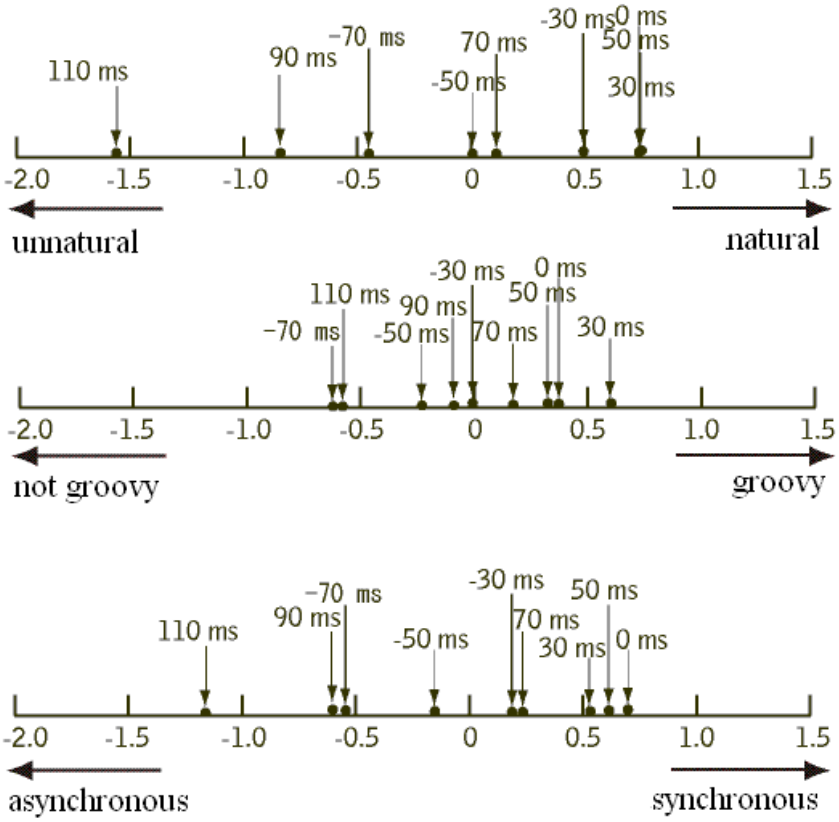

Figure 4b. The results in Experiment 3b (human scat singing)