Synchronization analysis of choir singing

Isao Tokuda\textsuperscript{1a)} and Akihiro Kuwahara\textsuperscript{2}

\textsuperscript{1} Department of Mechanical Engineering, Ritsumeikan University  
1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan

\textsuperscript{2} School of Information Science, Japan Advanced Institute of Science and Technology  
1-1 Asahidai, Nomi-shi, Ishikawa 923-1292, Japan

\textsuperscript{a)} isao@fc.ritsumei.ac.jp

Received August 22, 2011; Revised October 20, 2011; Published April 1, 2012

Abstract: Synchronization plays an essential role in choir singing. Fundamental frequencies of the singing voices should satisfy a rational frequency relationship to produce harmony. The present study attempts to clarify basic properties of synchronization that may influence the chorus quality. As the key features of synchronization, frequency mismatch and timing mismatch were considered. Combining the synthesis technique of choir singing, which can precisely control the detailed frequency structure of the singing voice, with psychoacoustic experiment, criterions were obtained to roughly discriminate good choruses from bad ones. To examine the reliability of the psychoacoustic experiment, acoustic analysis of the singing voices in real chorus was further carried out.

Key Words: chorus, synchronization, fundamental frequency, synthesis of singing voice

1. Introduction

Music is one of the most expressive usages of sound in human art and communication. An excellent musical performance can express a message and emotion so intensely that it moves and delights a wide audience. In particular, chorus and orchestra can sometimes give a strong impression that cannot be achieved by a solo play. Of course, each performer’s skill is important in such a cooperative music. However, even if highly skilled performers play together, a great performance cannot be always achieved. In a cooperative music, it is indispensable to find a good relationship between the performers. Each performer fits his/her rhythm and pitch to those of the other performers. In that sense, synchronization plays an essential role in a good performance of cooperative music [1, 2]. From such perspective, synchronized performance of musicians has been investigated. Rasch [3] studied timing synchronization in three small instrumental ensembles (trios). Williamson and Davidson [4] analyzed timing synchronization of piano plays, whereas Nakamura [5] studied synchronized breathing between co-performers in piano plays. Yarbrough [6] emphasized importance of eye contact and facial expression of conductor in mixed choruses. Yamamoto and Miyake [7] measured the breathing rhythm and play rhythm between co-performing pianists and found dependence of the degree of synchronization on the difficulty of the musical note.

As described above, the focus of the previous studies has been mainly on the synchronization of rhythms such as tempo, body movement, and respiration. On the other hand, synchronization of the
frequency may also play a vital role. For instance, if the frequencies of the co-performing musical instruments do not hold a good rational ratio (e.g., 4:5:6 in major triads), the musical performance instantly destroys a beautiful harmony. In most of the musical instruments such as violin, trumpet, guitar, flute, etc., frequency of the instrument is fine-tuned prior to the play. However, in a singing voice, the frequency should be precisely controlled during the whole performance. As one of the music styles, in which synchronization of frequency should be well maintained, this paper investigates the frequency relationship between two singers who sing in chorus. As the key features of synchronization that may strongly influence the quality of the chorus, both frequency mismatch and timing mismatch were considered. In real chorus, it is difficult to control such features in detail to investigate the delicate effect of synchrony on the perception of good chorus. Therefore, we synthesized the choir singing by artificially controlling the frequency and carried out the psychoacoustic experiment. Moreover, acoustic analysis of the singing voices in real chorus was also performed to examine the plausibility of the psychoacoustic experiment. On the basis of the psychoacoustic experiment and analysis of the real singing, criterions were obtained to roughly discriminate good choruses from bad ones.

The present paper is organized as follows. In Section 2, special technique for synthesizing high quality choruses is introduced. The synthesized choruses were evaluated by psychoacoustic experiment to study the key features of synchronization that affects the chorus quality. In Section 3, acoustic properties of the choir voices measured from real singers were analyzed. The final section is devoted to conclusions and discussions.

2. Psychoacoustic experiment

2.1 Synthesis of chorus voice

As a system to synthesize a chorus voice, a vocal conversion system, which has been developed by Saitou et al. [8] to synthesize a singing voice given a speaking voice and a musical score, was utilized. This synthesizer is based on the speech manipulation system widely known as STRAIGHT [9], and controls four $F_0$ fluctuations: overshoot, vibrato, preparation, and fine fluctuation. The $F_0$ denotes fundamental frequency that determines pitch of the vocal source. The system converts speaking voices into singing voices whose quality resembles that of actual singing voices. It won the first prize at the Synthesis of Singing Challenge at INTERSPEECH 2007 (Antwerp, Belgium, 28th August, 2007). By synthesizing two singing voices corresponding to two singers who sing upper and lower parts of a chorus song and by adding the two voices together, the choir singing was created.

2.2 Listening experiment

First, a long-tone singing of vowel /a/ was recorded from two singers. Using the recorded voices as the basis sounds of the STRAIGHT, the singing voices were synthesized as described in the previous subsection. As a chorus song, Japanese traditional song, “Autumn Leave,” was chosen. Figure 1 shows upper and lower parts of the song composed of 16 notes.

![Fig. 1. Frequency portrait of the Japanese song “Autumn Leave,” which was employed to synthesize chorus. Upper and lower parts composed of 16 notes are shown with blue and red lines, respectively.](image)
As the primary factors of synchronization that may strongly influence the quality of the chorus, two features are considered. (A) Mismatch of the frequency ratio of the choir voices from the harmonic ratio (e.g., 4:5:6 in major triads) and (B) Mismatch of the timings, at which the choir voices switch from one note to the next. To see the effect of these features, two types of choir voices were synthesized. In the first synthesis, we focused on feature (A) and synthesized choir singing with various degrees of frequency mismatch. In the second synthesis, we focused on feature (B) and synthesized choir singing with various degrees of timing mismatch.

(A) Fundamental frequency of the lower part was shifted to a value smaller than the correct value by 0, 10, 20, 30, 40 cents (Cent is a measure for frequency ratio. Frequency ratio between \( F_1 \) and \( F_2 \) is given in cent by 1200 log\(_2\)(\( F_1 / F_2 \)), where \( F_1 \) stands for the correct frequency value of the lower part and \( F_2 \) stands for its shifted frequency value. Since \( F_1 \geq F_2 \), the frequency ratio takes a positive value. 100 cents and 1200 cents correspond to half note and octave, respectively). The upper part was synthesized in accordance with the original music note.

(B) Timing of the lower part was delayed from the upper part by 0 ms, 20 ms, 40 ms, 60 ms, 80 ms. The upper part was synthesized in accordance with the original music note.

It should be noted that, to synthesize the choir singing with a frequency mismatch, there exist several ways to realize the mismatch. For instance, the lower part can be shifted to an upward direction (\( F_1 \leq F_2 \)). Alternatively, the upper part can be shifted, whereas the lower part is fixed to the correct frequency value. According to our preliminary experiments, however, no significant difference was observed among the different ways of the mismatching. In the present study, the lower part was shifted to a downward direction as one of the possible realizations of the frequency mismatch.

2.3 Evaluation results
In the psychoacoustic experiment, the subjects were asked to compare a pair of two synthesized choruses and judge which chorus is better than the other with 5 evaluation scores. For all combinations of two choruses, the paired comparison was repeated. The listening subjects were composed of six males who have no professional experience of music. The Scheffé’s method of paired comparison [10] was then applied to the subjective evaluation. Figure 2 shows the results. In the synthesized chorus (A), the subjective evaluation was lowered as the frequency mismatch was increased. In particular, the evaluation was dropped sharply between 30 cents and 40 cents (significance level of \( p<0.05 \)). This implies that 30 cents provide a critical value of the frequency mismatch to distinguish good choruses from bad choruses. In the synthesized chorus (B), on the other hand, the subjective evaluation was again lowered as the timing mismatch is increased. The evaluation is dropped clearly between 40 ms and 60 ms (significance level of \( p<0.05 \)), implying that 40 ms provides the critical value. In the mixture of the synthesized choruses (A) and (B), the frequency mismatches of 30 cents and 40 cents were rated as lower quality compared to the timing mismatch of 20 ms and 40 ms. The frequency mismatch therefore had a stronger influence on the chorus quality than the timing mismatch.

3. Acoustic analysis of the singing voice
To examine the psychoacoustic experiment of the synthesized chorus, singing voices were recorded from the real singers. We focused on the difference between a group of subjects who have some experience of music and a group of subjects who have no such experience.

3.1 Recording condition
As an experienced group, five subjects (three males: A, B, C; two females: D, E), who belong to a chorus group and have several years of music experience, were chosen. As an inexperienced group, three subjects (three males: F, G, H), who have no particular music experience, were chosen. In each group, two subjects were paired to sing upper and lower parts of the song of “Autumn Leave.” For simplicity, the singers were asked to sing only with vowel /a/. After several practices, electroglottographic (EGG) signal (equipped with Laryngograph Ltd.) was recorded from each subject.
Fig. 2. Results of the psychoacoustic experiment based on the Scheffé’s method of paired comparison. Evaluation value close to 1 implies good chorus, whereas a value close to -1 implies bad chorus. (A) Comparison of synthesized choruses with various frequency mismatch (0 cent, 10 cents, 20 cents, 30 cents, and 40 cents). (B) Comparison of synthesized choruses with various timing mismatch (0 ms, 20 ms, 40 ms, 60 ms, 80 ms). (C) Comparison of synthesized choruses with either frequency mismatch or timing mismatch (20 cents, 30 cents, 40 cents, 20 ms, 40 ms, 60 ms).

as the vocal source. The EGG signal measures the resistance across the vocal folds in noninvasive manner and is known to provide a reliable measure for the vibratory behavior of the vocal folds [11]. The fundamental frequency was extracted from the EGG signal of each singer using the STRAIGHT-TEMPO software [9]. The phase $\phi$ was also obtained from the EGG signal by the Hilbert transform [2].

3.2 Analysis results

Figures 3 and 4 show examples of the frequency ratio and the phase difference observed in experienced singers and inexperienced singers. In the “Autumn Leave” song, the upper part ($F_u$) and lower part ($F_l$) should hold the frequency ratio of either 4:5 or 5:6. The experienced pair made clear switches between 4:5 and 5:6, whereas inexperienced pair showed no clear switches. The phase difference, $\Delta \phi = 5\phi_u - 6\phi_l$, between the upper part ($\phi_u$) and the lower part ($\phi_l$) also shows a clear difference between the experienced pair and the inexperienced pair. The experienced pair gave rise to a clear plateau from 4 s to 4.4 s, where the phase difference was bounded within a small value (Fig. 4(A), (C)). Such plateau regions were observed also between 5.3 s and 6 s and between 7 s and 9 s in Fig. 4(A) (using $\Delta \phi = 4\phi_u - 5\phi_l$, the interval between 6.1 s and 6.9 s gives a plateau too).

This implies phase synchronization of the choir singing in the experienced pair. The inexperienced pair, on the other hand, showed no clear sign of phase synchronization.

Fig. 3. Frequency ratio $F_u/F_l$ between two singers who sung the “Autumn Leave” song with upper part ($F_u$) and lower part ($F_l$). (A) corresponds to experienced pair (subjects: D, E), whereas (B) corresponds to inexperienced pair (subjects: F, G).

In Fig. 5, frequency mismatch and timing mismatch are compared between different pairs of singers.
Fig. 4. Phase difference $\Delta \phi = 5\phi_u - 6\phi_l$ between two singers who sung upper part ($\phi_u$) and lower part ($\phi_l$) of the “Autumn Leave” song. (A) and (C) correspond to experienced pair (subjects: D, E), whereas (B) and (D) correspond to inexperienced pair (subjects: F, G). (C) and (D) display enlargement of (A) and (B), respectively.

Fig. 5. Frequency mismatch and timing mismatch obtained for different pairs of singers. The mismatch value averaged over 16 notes is plotted, where the error bar indicates the standard deviation. The boxes colored by light purple indicate results of the first trial, whereas the boxes colored by blue indicate results of the second trial. (A) and (B) correspond to frequency mismatch and timing mismatch, respectively.

To compute the frequency ratio, stationary part of the singing was first extracted at each musical note, where the stationary part is defined as a duration (width: 150 ms), within which standard deviation of the fundamental frequency is less than 15 cents according to Sundberg [12]. Then, within the duration in which both upper and lower parts are stationary, frequency mismatch was computed as a frequency ratio between the upper part and the one which satisfies the harmonic ratio with the lower part (e.g., in case harmonic ratio is 4:5, frequency mismatch is given by frequency ratio between $F_u$ and $1.25F_l$). The frequency mismatch was then averaged over stationary duration. To compute the timing mismatch, the switching point of music note was detected by finding a duration, in which frequency ratio between the first half and the second half exceeds 50 cents. Then, by computing the difference of the switching points between the two singers, the timing mismatch was obtained.

For all pairs of singers, the computed frequency mismatch ranged between 15 cents and 55 cents, which are within the range of the previous studies (range from 10 to 50 cents in [13]; range from 10 to 16 cents in unison in [14]). The timing mismatch ranged between 25 ms and 60 ms, which are again
within the range of the previous study (range from 30 to 50 ms for trio play in [3]).

For the pairs of experienced singers (AB, AC, BC, DE), the frequency mismatch was within 30 cents (except for two instances), whereas the timing mismatch was mostly within 40 ms. This agrees well with the psychoacoustic experiment, which indicated that the quality of the chorus drops at 30 cents and 40 ms. On the other hand, the pairs of inexperienced singers (FG, FH, GH) showed frequency mismatch larger than the critical value of 30 cents, judged as low-quality chorus in the psychoacoustic experiment. The timing mismatch of the inexperienced pairs was, however, comparable to that of the experienced pairs. Many of the inexperienced pairs also showed the values of timing mismatch smaller than 40 ms (only three cases indicate the timing mismatch larger than 40 ms). According to the Welch’s t-test [15], there was no significant difference between the experienced and inexperienced pairs in the timing mismatch (significance level of $p = 0.17$). This also agrees with the psychoacoustic experiment, which suggests that the timing mismatch does not have a strong influence on auditory perception compared with the frequency mismatch.

4. Conclusions and discussions

This paper studied synchronization of singing voices in chorus. As the key features of synchronization that may strongly influence the quality of the chorus, we focused on frequency mismatch and timing mismatch. Our approach is based on the synthesis of choir singing combined with psychoacoustic experiment. The advantage of this approach is that the synchronization properties can be precisely controlled to judge which feature strongly influences the chorus quality. Such precise control is not possible in a real human singing. Acoustic analysis of the singing voices in real chorus was also performed to examine the psychoacoustic experiment.

The psychoacoustic experiment showed that the frequency mismatch of 30 cents provides a critical value to discriminate good choruses from bad ones. The critical value for the timing mismatch, on the other hand, was found to be 40 ms. The frequency mismatch had a stronger influence compared with the timing mismatch. These criterions are in good agreement with the acoustic analysis of the singing voices in real chorus, where the experienced singers showed frequency mismatch and timing mismatch smaller than the critical values.

Acoustic analysis of the real chorus also showed that the phase difference between the upper and lower parts was bounded within a small value in some time interval of the experienced pair. On the other side, such bounded regions were observed only rarely in the inexperienced pair. This implies that the phase synchronization can be observed more frequently in the chorus of experienced singers than in that of inexperienced singers. Mathematically, the phase synchronization is defined by the property that the phase difference is bounded within a finite range for an infinite duration of time [2]. Such strict definition, however, may not be applied to the choir singing, because chorus is inherently a transient process with occasional occurrence of switching among different notes. Even within a single note, the frequency ratio fluctuates to some extent. In a long span of time, the phase difference may eventually diverge to a large value. Hence, to apply the idea of phase synchronization to chorus, we need to relax the definition to adapt to a practical situation that the phase difference is bounded within a small enough range in a finite and stationary duration of musical note in a choir singing.

It is also an interesting question to address whether humans can perceive such a small time interval of phase synchronization. In the study of hearing research, however, the main quantity to characterize the auditory perception is usually based on frequency. For instance, the hearing range, which describes the range of frequencies audible to human, is widely used to measure the hearing capability. To our knowledge, there are not many studies on the auditory perception of phase difference in particular in music sounds. It should be an interesting future study to focus on the perception of phase information in music.

Finally, it should be noted that the present results are still preliminary. For a comprehensive understanding of the synchronization of choir singing, dependence of the synchronization on the type of the song such as songs with fast tempo or slow tempo should be carefully considered. The case of employing music professionals as the listening subjects should be also examined. Another interesting dynamic aspect of singing voice is vibrato, which is a musical effect consisting of a regular, pulsating
change of pitch. Synchronization of vibrato between the choir singers should be studied in a future work.

Acknowledgments
This research was supported in part by Grants-in-Aid for Scientific Research (C) (No. 23560446) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan and SCOPE (071705001) from the Ministry of Internal Affairs and Communications (MIC) of Japan.

References