system are activated at the same time, bimodal recognition results are recorded in relation to the musical mood. The records are later utilized in an emotion decision process along with the results of bimodal and musical mood recognition. We attempted to verify the efficiency of this approach, but could not naturally perform the evaluation since the evaluation data are obtained from participants who should imitate the emotional expression and the music clip is selected without consideration of personal musical preference. Thus, in the first experiment, the previous record of bimodal results, $R_{\text{bimodal}}(e)$, was set to zero so as to be ignored.

The second experiment was conducted in a more natural way, supported by human participants. Each participant was asked to carry out the same behavior as in the bimodal recognition experiments, looking at iRobiQ and making an emotional face and speech for respective emotions. At the same time, we played several songs categorized into a mood similar to the corresponding emotion, at a slight distance from the robot. At that instance, the robot receives three kinds of emotional data and proceeds to compute the recognition results via respective recognition modules in the perception system. It should be noted that if both musical signals and human voice signals are entered into a single microphone, two types of signals act as noise signals for each other, deteriorating the recognition accuracy. The ideal solution for this problem is to operate a process of blind source separation that divides audio signals into music and voice [37]. However, the correctness of the separation task would naturally affect the performance of the two audio recognition modules. For this reason, this study does not consider the problem caused by a single microphone in order to concentrate on the performance evaluation of respective recognition modules and their multimodality. Thus, iRobiQ receives two different types of audio signals, respectively, from two different microphones, one of which is an ear microphone which the participant wears for the voice input and the other is a general microphone equipped on the robot for the music input. The musical signals that the robot-equipped microphone receives while the ear microphone is activated are regarded as music-mixed voice signals and are excluded from the musical mood recognition task.

In this experiment, we attempted to use the previous bimodal results in the emotion decision. Once the results of bimodal and musical mood recognition were computed in respective modules, the bimodal recognition results were used to update the average value of the bimodal results on the determined musical mood. The average value was then utilized in the emotion decision process along with the bimodal and musical mood recognition results, on the basis of (4).

Figure 9 compares the performances of two multimodal experiments with bimodal results. The multimodal recognition achieved superior accuracy compared to bimodal results over all emotions. The results in this figure confirm that the proposed music-aided multimodal approach notably advanced the standard bimodal approach. The results of the first multimodal experiment (called 'Multimodal1') and the second experiment (called 'Multimodal2') respectively represented relative improvements of 15% and 46% (2.7% and 8.5% in absolute improvement) over the bimodal result. The performance improvement was more significant for emotions such as happy and angry, where the musical mood recognition indicated higher accuracy. In particular, Multimodal2 showed relative improvement of 36% over Multimodal1, which only uses the results of bimodal and musical mood recognition. This supports our idea that the record of bimodal results in relation with the musical mood provides supplementary information in the emotion decision during the current time when the relevant musical mood is entered.

From these experiments, we conclude that musical mood information can be effectively utilized as a supplementary and complementary indicator in standard emotion recognition tasks based on speech and facial expression. Nevertheless, we need to further consider that different users might enjoy different types of musical moods while being in a certain emotional state. In the human listening test conducted for the verification of mood categorization, at least 70% of the participants determined an identical mood for each clip. This result indicates that humans tend to feel similar emotions while listening to music. Even when people feel different emotions prior to listening to music, they will experience the same emotion due to the certain mood of the music. This conclusion is closely associated with the general knowledge addressed in Section 1 that music greatly influences the affective states of humans. Consequently, musical mood recognition has strong possibility of improving the reliability of affective interaction between humans and robots.

5.3. Evaluation of the Expression System. In the proposed affective interaction, the expression system provides a natural and intermediate interface between humans and service robots. As addressed in Section 4.3, the system enables service robots to appropriately react to the user emotion through visual and acoustic expressions. In particular, the proposed expression