Towards a System that Relieves Psychological Symptoms of Dementia by Music

Chika Oshima
Japan Society for the Promotion of Science,
Faculty of Medicine, Saga University
Saga, Japan
chika-o@ip.is.saga-u.ac.jp

Koichi Nakayama
Department of Information Science, Saga University
Saga, Japan
knakayama@is.saga-u.ac.jp

Naoki Itou
Intermedia Planning, Inc.
Tokyo, Japan
n_itou@ipi.co.jp

Kazushi Nishimoto
Research Center for Innovative Lifestyle Design,
Japan Advanced Institute of Science and Technology
Ishikawa, Japan
knishi@jaist.ac.jp

Kiyoshi Yasuda
Kyoto Institute of Technology
Chiba Rosai Hospital
Kyoto and Chiba, Japan
fwkk5911@mb.infoweb.ne.jp

Naohito Hosoi
Sodegaura Satsukidai Hospital
Chiba, Japan
hosoi@mail.satsuki-kai.or.jp

Hiroshi Okumura
Department of Information Science, Saga University
Saga, Japan
oku@is.saga-u.ac.jp

Etsuo Horikawa
Faculty of Medicine, Saga University
Saga, Japan
ethori@med.saga-u.ac.jp

***Abstract*—MusiCuddle is a system to calm the symptoms of patients with mental instability who repeat stereotypical utterances. The system presents a short musical phrase whose first note is the same as the fundamental pitch (F0) of a patient's utterances. We performed a case study to investigate how a patient's behaviors changed with MusiCuddle. The results suggested that the phrases presented by MusiCuddle may provide patients with an opportunity to stop repeating stereotypical utterances. Then, we added a vocoder function to MusiCuddle so that patients would be able to attend to the music more. We examined whether the mood of university students changed or not according to music presented with the vocoder function. We found significant differences between major harmonies and minor harmonies for the "cheerful" and "negative" moods. Namely, when a person's voice is combined with cheerful sounds, he/she can become cheerful. However, when we conducted a case study to expect a patient's repetitive utterances changed or stoped by the sound from the MusiCuddle with the vocoder, the participant's utterances did not change. We discussed reasons of the result from an aspect of characteristic of a patient according to a cause disease of dementia.**

Keywords—MusiCuddle, vocoder, FTD, Harmony in a major and minor key

I. INTRODUCTION

We are structuring a music accompaniment system to calm the symptoms of patients with mental instability who repeat stereotypical utterances. "MusiCuddle [1][2]" is a system that presents a short musical phrase. The system determines a pitch at a predetermined interval on the basis of a sound extraction technique [3]. Then, the system plays a prepared Musical Instrument Digital Interface (MIDI) sequence (a phrase) the first note of which is the same as the F0 of the patient's utterance.

The concept of MusiCuddle is derived from the "iso-principle [4]," which is a theory of music therapy, and a case of an autistic child a famous music therapist treated by extracting approximate pitches of the child's screaming and improvising based on these pitches [5]. "Iso" simply means "equal," that is, the mood or the tempo of the music must initially have an "iso" relationship with the mood or tempo of the patients. If a client is distressed or agitated, then the quality of the music should initially match his or her mood and energy [6].

In this paper, first, we introduce the MusiCuddle and results of a case study with using the system [1][2]. We performed a case study in which one of the authors used MusiCuddle to present phrases to a patient with dementia who repeated stereotypical utterances. The symptoms of dementia are divided into core symptoms and behavioral and psychological symptoms of dementia (BPSD). BPSD includes agitation, aggression, wandering behavior, hallucinations, delusions, and repetitive stereotypical utterances. However, appropriate care is thought to alleviate and slow the progression of these symptoms. Music is a method known to alleviate the symptoms of dementia.

Second, on the basis of the results of the experiment, we added a "vocoder" to MusiCuddle. The vocoder allows an individual to hear his/her voice becoming a part of the instrumental sound according to a musical phrase presented by MusiCuddle. Because our target population repeat utterances quite frequently, it will be hard for them to listen to the music presented by MusiCuddle. Therefore, the utterances should be combined with music sounds in real time, as their attention will be more likely to shift to the music than when they listen to music in parallel with their utterances.

Furthermore, if the musical phrases from MusiCuddle can manipulate the mood of patients with mental instability and make it more pleasurable, they may temporarily stop repeating utterances. There are studies showing that mood affects memory and cognitive processes [7]. For instance, Taniguchi [8] used music to manipulate subjects' mood. In [9], he considered a relationship between characteristic of music and mood induced by music. He proposed the Affective Value Scale of Music (AVSM) to indicate the property of musical pieces on the basis of 24 adjectives on five levels. Then, he conducted an experiment in which female students were rated on both the AVSM and the Multiple Mood Scale (MMS) [10], which is to evaluate subjects' mood by themselves for the five pieces, finding a significant relationship between the AVSM and the MMS. The result has shown that music can be a trigger to induce a mood.

Then, in this paper, we examined the contribution of harmonies in major and minor keys to mood induction for healthy subjects with the vocoder. They read a gloomy poem when their utterances were combined with music sounds by MusiCuddle with the vocoder. After reading the poem, they evaluated their current moods.

Finally, we performed a case study using the MusiCuddle with the vocoder for a patient with dementia who repeated stereotypical utterances. Her utterances were combined with music sounde in real time.

In the next section, we illustrate the MusiCuddle and experiments that the author presented music phrases to a patient with dementia using MusiCuddle. Section III describes the contribution of harmonies in major and minor keys to mood induction for healthy subjects by MusiCuddle with the vocoder. Section IV concludes this paper and outlines future works.

II. MusiCuddle: The First Note of a Phrase is the Same as the F0 of the Patient's Utterance

We presented a music accompaniment system, "MusiCuddle" that presents a short musical phrase. Then, we conducted a case study using MusiCuddle for a patient with dementia.

A. *Extract pitches from the utterances*

Figure 1 shows a user interface of MusiCuddle [1][2]. MusiCuddle is a system that presents music when an operator (e.g., a caregiver) pushes any of the keys of the electronic keyboard or a button on the interface of the system. Previously, we have to select a folder of musical phrases and move it into the same folder as MusiCuddle. Once the play button is pushed, the system continuously extracts pitches (F0) from sounds (the intended patient's utterances). When the operator pushed the trigger's button again, the system determines a pitch at a predetermined interval. Then, the system selects a musical phrase file in the database on the basis of that extracted pitch. The first note of the musical phrase is the same as the pitch extracted from the patient's utterances.

We employ a pitch extractor to extract pitches (i.e., C, D, E) from the patient's utterances. This is based on the technique for extraction of sounds that have unstable pitches and unclear periods, such as natural ambient sounds and the human voice, into musical notes [3].

In the original system shown by [3], if the operator gave a start trigger, the system would initiate the processing to obtain the F0 (fundamental frequency) time series from the acoustical signals (i.e., a singing voice), which were being recorded via the microphone. The short-term F0 estimation by Fast Fourier Transform (FFT) and Inverse Fast Fourier Transform (IFFT) for the power spectral is repeated until the system catches an end trigger from the operator. The system then calculates a histogram of pitches with the F0 time series between the start and end triggers. Finally, only the most frequent pitch is selected and is output as the pitch of the period.

For our research, some processing designs were modified. Figure 2 shows the processing of the system. Considering the attitude of the operator, we would assume that the triggers would be input after the operator catches the utterance of the patient. Therefore, we omitted the start trigger. The system starts a short-term F0 estimation just after invocation of the system and continues it thereafter. When the operator inputs a trigger that is regarded as an end trigger, the system calculates a representative pitch for a predetermined period just before the trigger based on the above-mentioned method. Then the system plays a prepared MIDI sequence (a musical phrase) that corresponds to the representative pitch. These modifications of our system improve usability by reducing the time lag between the input of the trigger and the output of phrase.

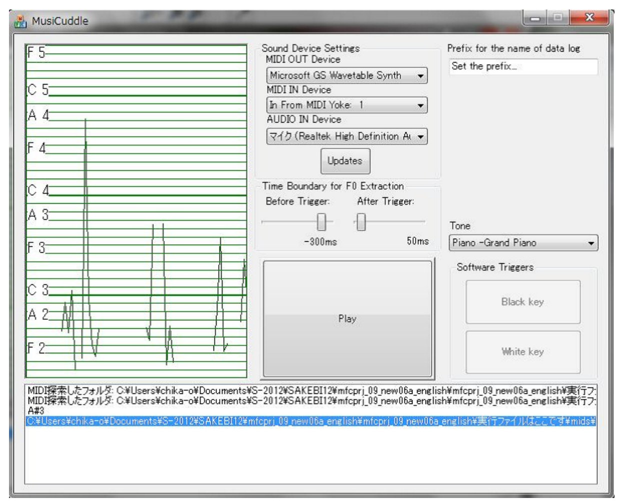

Fig. 1. User interface of MusiCuddle for a caregiver.

To extract the F0 against the mixed acoustical signal of the patient's utterance and the musical phrase output from the speaker, our system needs two of the same microphone (ideally one stereo microphone) and one speaker. Figure 3 shows the setting of the microphones. The microphones are set in front of the speaker to record the speaker's sound at the same level from both microphones. On the other hand, both microphones are displaced against the patient to record the levels of the patient's utterance that are clearly different. The system calculates the differential signals from the signals of both microphones to cancel the sounds of the MIDI sequence where they are localized in the center position. The F0 estimation is then determined with these differential signals.

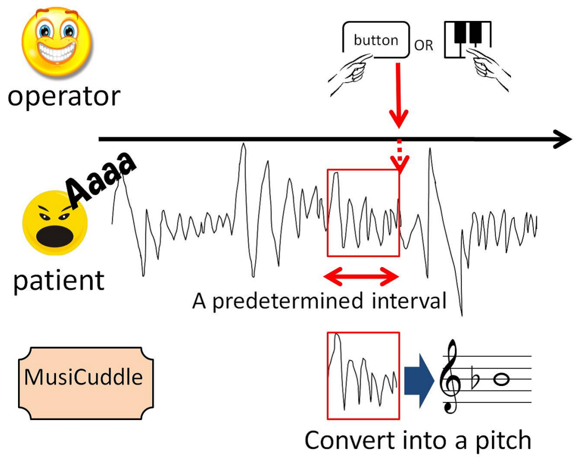

Fig. 2. How to convert an utterance into a pitch.

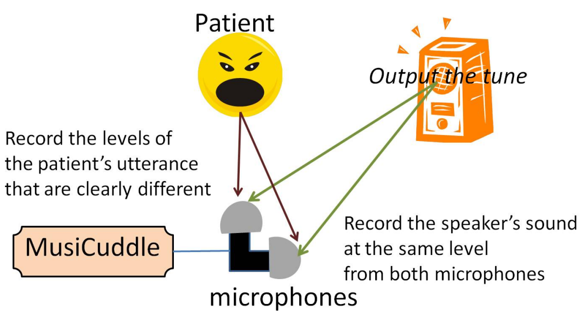

Fig. 3. Calculates the differential signals from both microphones to cancel the sounds of cadence.

B. *Case study using MusiCuddle*

We conducted a case study to investigate how a patient's behaviors changed with the application of MusiCuddle [1][2]. The symptoms of the patient targeted for this study was severe. This case study was conducted for a very short period, and we could enroll only one patient. Therefore, it is not appropriate to make any cognitive assessment [12] or to examine the patient's abilities with respect to activities of daily living [13]. Instead, we record the patient's utterances under conditions with and without the use of MusiCuddle and compare them in order to estimate the influence of MusiCuddle.

1) Ethical Considerations: This case study was approved by the Research Ethics Board of Saga University. The participant in the case study, who is a patient with dementia, her husband, and the hospital director were informed about the intentions of the case study and the treatment of personal information. Moreover, they were informed that they could withdraw from the case study at any time. Then, we obtained written consent from them.

When we conducted this case study, the hospital director and nurses worked on the same floor and could check on the condition of the participant. If the presentation of sounds from MusiCuddle had not been appropriate and the participant became more agitated, we would have had to abandon the case study immediately.

2) Participant: The participant was a 72 year-old, hospitalized patient with severe FTD (frontotemporal dementia). She repeats stereotypical utterances for many hours each day. Moreover, when she is agitated, she locks herself in a restroom for a long time while repeating stereotypical utterances. However, she is lucid enough to remember some nurses' names and greet them clearly. She can answer the date and the exact time. Her score on the HDS-R (Revised Hasegawa's dementia scale) [14] was 17 two years ago. The score shows she was mild dementia.

The following is an example of her usual utterances. This example was uttered in about thirty seconds. "P" means "Participant."

P: haittayo (repeated eight times) mashitayo imasen masen imasendesu masen (repeated three times)

"haittayo" means "have been entered" as well as "imasen" means "not being here." "mashitayo" may be fragment of "imashitayo." "imashitayo" means "being here."

Although she utters many kinds of sentences, most of them are rhythmical and fit into the same meter. Figure 4 describes some examples of her sentences. One of the authors dictated the rhythms of these sentences. These examples show that although the sentences are different, they fit into four-four time.

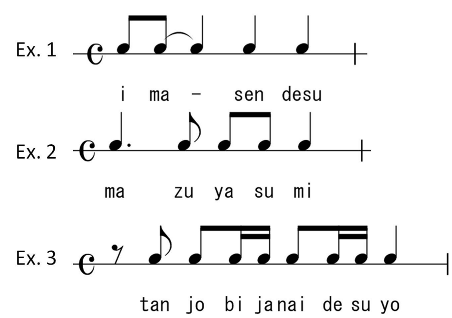

Fig. 4. The participant's sentences fit into four-four time.

She repeats stereotypical utterances especially when hungry. She often locks herself in a restroom from around eleven o'clock a.m. until lunchtime, and from around one o'clock p.m. until snack time while repeating stereotypical utterances nearly incessantly. However, she sometimes responds to nurses when they talk to her.

3) Preliminary experiment: Adopting the iso-principle, one of the authors attempted to utter according to the participant's utterances. We think that the author's utterances had the same tempo, rhythm, and pitch as those of the participant.

When the participant was agitated and repeated the same sentences, one of the authors repeated sentences in the same melody corresponding with the participant's repetition (the same tempo, rhythm, and pitches). Figure 5 shows these sentences in musical notation. Sentence A means the participant's sentence. First, the author tries to repeat the participant's sentences in rhythm. Namely, both of them repeat Sentence A, "i-ma-se-n-yo (not being here)." Second, the author tries to repeat a different sentence using the same melody as the participant's in rhythm. Namely, the author repeats Sentence B, "go-han-de-su-yo (Time for lunch.)," although the participant is repeating Sentence A.

In the first trial, the participant turned around to pay attention to the author. However, she kept repeating the same sentence in harmony with the author's repetition. Her utterances became louder. In the second trial, the participant changed from the sentence A to the author's sentence, "go-han-de-su-yo" in the same melody. Then, the participant left the restroom and went toward a table for lunch. In a moment, however, she returned to the restroom and repeated the same sentences.

When the author repeated sentences in accordance with the participant's repetition and used the same melody, the participant kept repeating the same sentence in a loud voice. The author's utterances could have caused increased symptoms of agitation in the first trial.

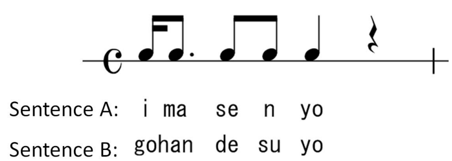

Fig. 5. The author repeated sentences in accordance with the participant's repetition.

4) Method: We stood by from ten o'clock a.m. to noon and from one o'clock p.m. to half-past two p.m. for 2 days. After this case study, we compared the participant's utterances with music with those without music (see Section II-B6) to estimate the influences of MusiCuddle. Therefore, we set two time periods, one with the use of MusiCuddle and one without MusiCuddle.

During the time with the use of MusiCuddle, we started MusiCuddle and selected a musical phrase. When the participant began to repeat stereotypical utterances, we presented the musical phrases arbitrarily by giving triggers to MusiCuddle.

The experiment was conducted in a hospital where the participant was hospitalized. Figure 6 shows the setting of the case study. The music was presented through a wireless cuboidal speaker with Bluetooth, which measured, 123×36×35 mm, and the patient's utterances were recorded through a wireless, columnar microphone with Bluetooth measuring about 75 mm in height and 24 mm in diameter. These devices were set on the door of the restroom.

Our system requires two of the same microphone (one stereo microphone). When the operator inputs the trigger, even when the previous musical phrase is being presented, the system extracts F0 against the mixed acoustical signal of the patient's utterance and the musical phrase being presented from the speaker. However, it is not so safe to use the stereo microphone in the hospital, because it is large in size and wired to its receiver. Thus, we did not use the stereo microphone in this experiment, and the operator did not input triggers when the previous musical phrase was being presented.

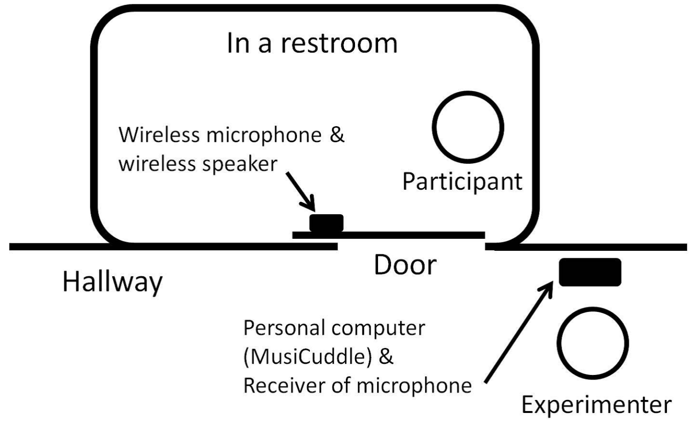

Fig. 6. A small wireless speaker and microphone are set on the door of the restroom.

5) How to use MusiCuddle: When the participant is agitated and repeats stereotypical utterances in a restroom, an operator (one of the authors) presents music by using the MusiCuddle system in front of the restroom. The operator listens to the patient's utterances outside of the restroom. When the operator finds a period during which the patient utters an almost stable pitch, she clicks the trigger button. Once the trigger button has been clicked, a musical phrase is retrieved from the database on the basis of the detected pitch, and it is automatically performed to overlap with the participant's utterances. In this case study, we prepared seven type of musical phrases: Four chords (Major seventh, Quarter note, and No volume change), Cadence, "Yuki," "Akaikutsu," "Hana," "Tsukinosabaku," and Stereotypical utterance (i-ma-se-n-yo) [1][2]. All of the musical phrases consist of very short phrases lasting 3∼30 seconds. The operator selected musical phrases considering the participant's reactions and condition. The operator clicked the trigger button again to perform the next musical phrase when the performance of the current musical phrase ended.

6) Analysis method: In this case study, we investigate how MusiCuddle influences the patient's stereotypical utterances. If the music presented from MusiCuddle distracted the patient's attention from her stereotypical utterances, her utterances would be disrupted and she would stutter. Therefore, we compare the participant's utterances while listening to music with those without music. Especially, we focus on the patient's stuttering to detect distraction of the patient's attention.

The participant's utterances are segmented into small sentences according to the method of repetition. One of the authors decided segmentation points according to the meanings of utterances by reference to the participant's breathing. For example, the following utterances (P1) are segmented like the next line (P2).

P1:imasendesuimasendesuhitoyasumimazuyasumiimasendesuyo
P2:imasendesu (not being here) / imasendesu (not being here) / hitoyasumi (taking a rest) / mazuyasumi (taking a rest) /

imasendesuyo (not being here)

Then, we analyzed the influences of the music on the patient's utterances. First, we determined whether each sentence was uttered "with music" or "without music" on the basis of the following conditions (see Fig. 7):

1) If the patient uttered a sentence while a musical phrase was being performed, the sentence was considered to be uttered with music.
2) If the patient began to utter a sentence just after a musical phrase had finished, the sentence was considered as being uttered with music.
3) If a musical phrase started after the patient had started uttering a sentence, the sentence was considered to be uttered without music.
4) Otherwise, the sentence was considered as being uttered without music.

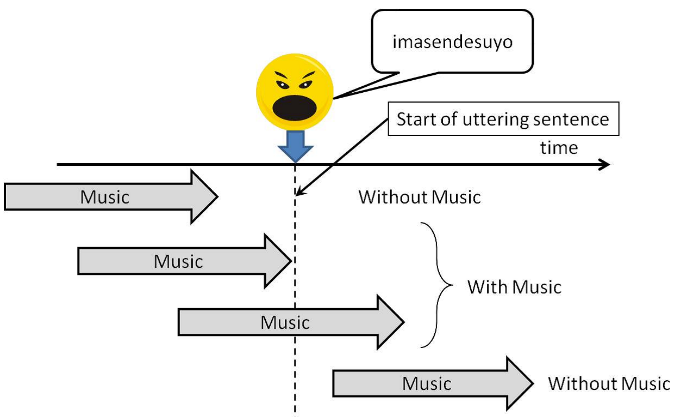

Fig. 7. Determination of whether each sentence was uttered "with or without music."

In the following example, the musical phrase was presented in the middle of "hirugohandeha (fragment of "It is not lunch time")." Therefore, "masende (fragment of "not being here")" and "hirugohandeha" were considered as being uttered without music, while "imasendesu (not being here)," "ima (fragment of "not being here")," and "gohanden (fragment of "It is not lunch time")" were considered as being uttered with music.

P: masende hirugohandeha imasendesu ima gohanden
[(Start music)] (Stop music)

In this case, we consider that the sentence "hirugohandeha" was unaffected by music. Moreover, we consider that "gohanden" was affected by the music, because this sentence started to be uttered immediately after the presentation of the music.

Next, we find sentences on which the participant stuttered. She often repeats several stereotypical sentences without any slight changes many times (see Section II-B2). However, if the music distracted her attention from her repetition of stereotypical utterances, she stuttered, uttering only a part of stereotypical sentence or a sentence different from a stereotypical sentence. Therefore, if we find such a sentence including words that were part of an immediately preceding sentence (but not exactly the same as the immediately preceding sentence), she was considered to have stuttered.

We determined whether each sentence included words that were part of the immediately preceding sentence. In the following example, we consider that she stuttered, as "ima" is included in the immediately previous sentence "imasendesu:"

P: imasendesu ima

7) Results: The intended records documenting our analysis of the participant's utterances constitute only three parts of the entire recorded dataset; the lengths of the three parts are 16, 8, and 3 minutes. Although we recorded for a much longer time period, we could not use the other parts because of extraneous noises that masked the patient's utterances. It seems that the restroom's iron door blocked communication between the wireless microphone and the personal computer. Moreover, we could not record when the participant moved to some unexpected rooms. Therefore, there is a large gap between the time of using MusiCuddle and the total recording time.

Of the 27 minutes of intended records, the total time of music presentation was 6 minutes and 54 seconds, approximately one-fourth of the total recording time. For the first and second recording sessions, we presented the phrases using MusiCuddle. In contrast, we did not present phrases at all during the third recording session.

The participant emitted utterances at all times during the experiment. Table I shows the kinds of sentences. Six hundred eighty sentences were segmented (84 kinds) in 27 minutes. The most uttered sentence was "imasendesu" (201 times).

In many cases, the contents of the sentences were almost the same, even when their constituent words varied slightly. For example, "imasen" and "imasendesu" have the same meaning, "I am not here." Moreover, in certain instances, only parts of sentences were uttered ("ima," "imasende").

She repeated the same words many times, uttered different words in sequence, or uttered slightly different words continuously both in the rhythm of the presented-music and not. In the following example, she repeated the same words many times:

P: mazuyasumi mazuyasumi mazuyasumi mazuyasumi...(total number of repetitions was seven)

In the following example, she uttered different words in sequence:

P: gohandashimasendesu mazu imasendesu oyatuja imasendesu mazuyasumi

In the following example, she uttered slightly different words continuously:

P: imasende imasendesu imasen imasendesu ima imasendesu

Most of the sentences were rhythmical and fitted into four-four time (see Fig. 4). However, short sentences such as, "ima" and "mazu" fitted into four-one (irregular) time.

TABLE I. SENTENCES SEGMENTED FROM THE PARTICIPANT'S UTTERANCES IN THE CASE STUDY.

Estimated meaning	Sentences
I am (not) here.	imasu (1), imasendesu (201), imasen (48), imasende (43), ima (5), deimasendesu (1), sokoniimasen (1), imasenyo (1), uruchiimasendesu (1), ryugaimasende(1)
It is (not) lunch time.	mazugohandesu (78), mazugohan (14), mazugohande (6), hirugohannarimasendesu (5) gohandesu (4), hirugohandashimasendesu (2), hirugohannarimasende (2), gohandashimasendesu (2) haisugugohandashimasendesu (2), mazugohandesuyo (2), gohan (2), gohanden (1), gohannarimasendesu (1) gohannaidesuyo (1), hirugohande (1), gohandashimasende (1), gohanninarimasu (1) gokaimenogohandashimasendesu (1), mawarinogohangoyamoyashisendesu (1)
First,	mazu (34), mazuyasumi (33), mazudesu (32), ma (6), mazuya (3), mazudesu (2), mazude (1), mazugo (1), mazuyasu (1)
Not do	masende (10), masendesu (6), masendesuyo (2)
Bath time, Break	ofurohaittadesuyo (2), ofuro (1), ofurojaimasende (1), hitoyasumi (1)
(Not) Birth day	tanjobijanaidesu (3), tanjokainaidesuyo (3), tanjobijanaidesu (1), tanjobijaimasendesuyo (1) tanjobijaarimasendesu (1), tanjobijaarimasendesuyo (1)
Time	1ji40fundesuyo (13), yoruninarimasendesu (9), 1jihandesuyo (8), 3jihanninarimasendesu (8), yoruninarimasendesuyo (5) handesuyo (4), 3jihandesuyo (4), 1jihande (3), 2jihandesuyo (2), 3jininarimasendesuyo (2), 1jihandesune (1) 1ji10fundesune (1), 1ji (1), 1jihandesuyo (1), 3jihanni (1), 3jihanninarimasendesuyo (1), mou3jininarimasendesu (1)
Snack time	oyatudesuyo (1), oyatujaimasendesu (1), keikihanaidesuyo (1), keikihanaidemasendesu (1), keikihanaitodesu (1)
Soon	suguha (1), suguhanaidesu (1)
"Yuki"	zunzuntumoru (2)
Question	imashitaka (1)
Greeting	konnichiha (1)
Others	dojoninarimasende (1), ugoninarimasende (1), sonouchimasende (1), mashi (1), bokujaarigatoarigato (1), basyohanaidesuyo (1)

Values in parentheses show the numbers of times each sentence was uttered.

Table II shows the comparison between "with music" and "without music." The numbers of different sentences uttered were 114 with music and 179 without music. The total recording time was 27 minutes, and musical phrases were presented for 6 minutes 54 seconds of that time. Changes in the sentences uttered by the participant numbered about 16 per minute with music and 9 per minute without music. Therefore, we can say that the participant changed her utterances more often with than without music.

Next, we determined whether each sentence included words that were part of the immediately preceding sentence in order to determine on which sentences the participant stuttered. The results indicated that with music, 94 out of 114 sentences (82.5%) included words from the immediately preceding sentence (see Section II-B6). On the other hand, without music, that rate was 41.3%. This result shows that the rate of sentences including words from the immediately preceding sentence was higher with than without music. If the participant stuttered, we consider that the music distracted her attention from repeating her stereotypical utterances (see Section II-B6). The results indicate that MusiCuddle may give patients an opportunity to stop repeating utterances.

In the following example, a sentence changed into a completely different sentence when music was not presented (without music):

P: mazu imasendesu oyatsuja mazuyasumi
(without music)

The following is an example in which a sentence included the word from the immediately preceding sentence when music was presented (with music):

P: mazugohandesu **mazu**...
[ ]
(Start music) (Stop music)

TABLE II. THE NUMBERS OF CHANGES IN REPEATED SENTENCES.

	with music	without music
changing sentence (ALL)	114	179
include the words of the immediately previous sentence.	94	74
rate (%)	82.5	41.3

8) Discussion: The participant tended to stutter when each phrase was presented from MusiCuddle. The music might shift the participant's interest to music from the repetition of stereotypical utterances. On the other hand, when one of the authors repeated the participant's sentence using the same melody and rhythmic pattern, she did also pay attention to the author (see Section II-B3). However, the participant kept repeating the same sentence.

Namely, patients might attend to the phrase according to their similarity in pitch. Meanwhile, their attention may be deflected away from their repetitive stereotypical utterances if the melody is too strikingly different from their utterances. So, the phrases presented by MusiCuddle may provide patients with an opportunity to stop repeating stereotypical utterances.

III. MOOD INDUCTION USING MUSICUDDLE WITH A VOCODER: MAJOR VERSUS MINOR HARMONIES

The result of the case study using MusiCuddle suggested that the mental instability patient's attention might shift away from her repetitive stereotypical utterances to the music (see Section II-B). We expect that the utterances should be combined with music sounds in real time, as their attention will be more likely to shift to the music than when they listen to music in parallel with their utterances. Therefore, we added a vocoder function to MusiCuddle.

A. Add a vocoder function to MusiCuddle

We added a vocoder function to MusiCuddle [1][2] so that patients would be able to attend to the music more. The vocoder is an audio processor that captures the characteristic elements of an audio signal and then uses this characteristic signal to affect other audio signals. The modulator extracts the fundamental frequencies of the voice and converts them into levels of amplitude on a series of band pass filters. Then, these band pass filter signals are passed onto the carrier wave and the final sound is created.

Fig. 8 shows the vocoder's connection to MusiCuddle. The patient's utterances are input to MusiCuddle and the vocoder (synthesizer) by two kinds of microphones. MusiCuddle extracts notes from these utterances and selects a phrase. Then, the phrase (MIDI sequence) is sent to the vocoder. The vocoder performs the MIDI sequence using the tone of the synthesizer and the patient's voice. The patient can hear his/her utterances combined with MIDI sequence.

B. Research aim

We want to examine whether the mood of the patient with mental instability changes or not according to music presented with the vocoder function. There is no research of mood induction using the vocoder function. Since it is difficult to gather the intended patients who repeat utterances continuously and it is difficult for them to express their moods in language, the subjects of this paper are healthy university students.

Moreover, we examine the difference between a mood induced by harmonies in a major key and a mood induced by harmonies in a minor key. Altshuler showed if a patient is gloomy, the quality of the music should initially be gloomier rather than happier [4]. Itoh [15] showed that individuals in a depressive state become relaxed when they listen to gloomy and calm music. After introducing these kinds of music, however, the mood of the music should gradually change to the target mood (Level attacks [4]). Takeuchi [16] conducted an experiment on university students in a state of depression and found that the group of subjects who heard music that progressed from sad to happy were put in a happier mood.

C. Pre-experiment

In this section, subjects evaluated their impressions of two musical phrases. These phrases were used in the main experiment (see Section III-D).

1) musical phrases: Hevner [17] indicated that the expressiveness of a modality, either major or minor, is more stable and more generally understood than that of any other musical element. He showed that major keys are strongly associated with happiness, gayety, playfulness and sprightliness and minor keys are deeply related to sadness, sentimental yearning, and tender effect. Moreover, consonant chords work for "delightful [18]" "cheerful [19]," and dissonant chords work for "exciting [18]" and "overcast [19] [11]."

In the main experiment, we examined the difference between a mood induced by harmonies in a major key and a mood induced by harmonies in a minor key with using the vocoder. The properties of two musical phrases should be similar although the mode (major or minor) and the harmonies are different. Therefore, we pick out these phrases from the same music piece, "Chaconne" rearranged by Busoni for a piano solo on the basis of "Chaconne from Partita No.2 for solo violin in D minor, BWV 1004" composed by Bach. Figs. 9 and 10 show two different kinds of phrases. We extracted the harmonies in the major key from bars 138-145 with one incomplete bar as well as the harmonies in the minor key from bars 1-8 with one incomplete bar. In the original score of Chaconne, there are many kinds of note values and some passing notes between the chords. However, we did not consider rhythm and passing notes. All notes in the scores were changed to whole notes due to the features of the vocoder (see Section III-A). The scores were transformed into two MIDI data files in advance. The tempo was a beat of 60 quarter notes for one minute. Namely, both phrases could be presented in about one minute.

The subjects listen to the phrases produced by the sound source of a synthesizer, "microKORG XL+ (Korg)." The synthesizer effect was made by the "ROCK" genre and "POLY SYNTH" category in microKORG XL+. In the main experiment, we used the same sound effect. However, we also used the vocoder function. Therefore, the feeling of sounds we got were different between in the pre-experiment and in the main experiment.

2) Method: The subjects were 132 engineering university students ranging from 18 to 20 years of age. Sixty-one of the subjects evaluated the harmonies in a major key (D major) first and then in a minor key (D minor). The rest of the subjects evaluated them in reverse order. We prepared AVSM [9] for the subjects to evaluate the affective value of the two phrases. The AVSM consists of 24 adjectives that can be divided into five dimensions: uplift (uplift and dysphoria), familiar, strong, lightness, and stateliness. The subjects were asked to evaluate the 24 adjectives (items) on a five-point scale: It does not apply to the adjective at all (1); It does not apply to it very much (2); I cannot say either way (3); It applies to it a little (4); It applies to it very much (5).

3) Result: We performed t-tests on the data for 24 items. Table III shows that there were significant differences between the major phrase condition and the minor phrase condition on 16 items. In particular, the evaluations of three items, "melancholy," "miserable," and "gloomy" became opposite. Their averages were more than 4-point in the evaluation for the minor phrase. Their averages were less than or equal to 3-point in the evaluation for the major phrase. These results showed that the phrases were suitable for use in the main experiment.

D. Experiment: mood induction using MusiCuddle with a vocoder

Each subject read a gloomy poem and indicated his/her current mood. Then, he/she read the same poem using MusiCuddle with the vocoder and indicated his/her mood again. The music presented from MusiCuddle included two kinds of phrases that were evaluated in Section III-C. We examined whether the mood induced in subjects using the vocoder differed according to the music from MusiCuddle.

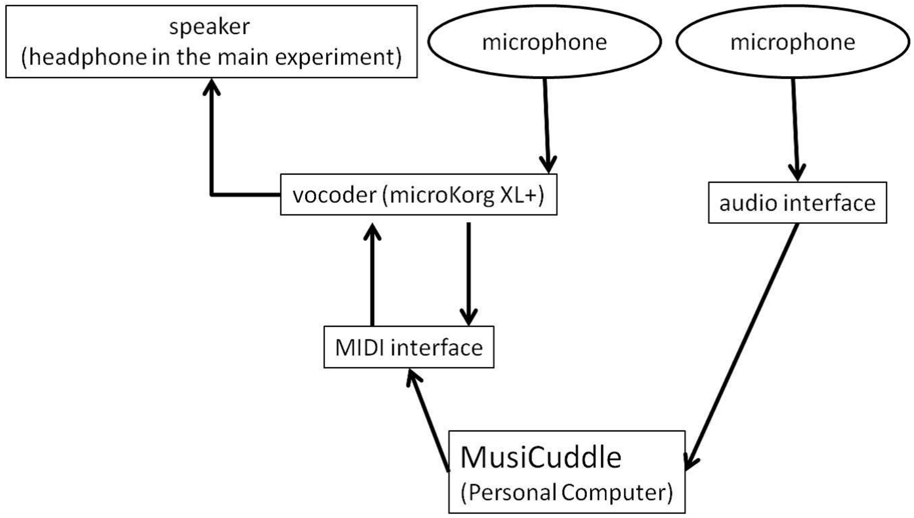

Fig. 8. Connection with a vocoder.

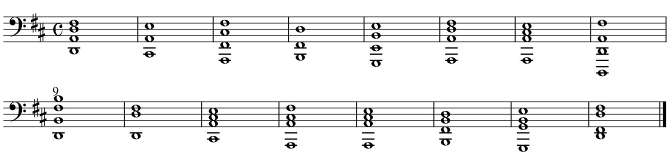

Fig. 9. Harmony in a major key.

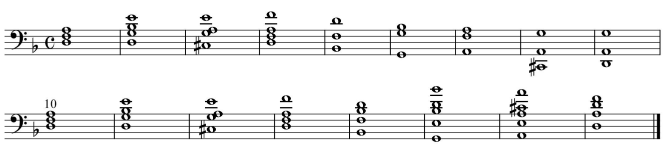

Fig. 10. Harmony in a minor key.

1) *Ethical Considerations:* This experiment was approved by the Research Ethics Board of Saga University. The subjects were informed about the purpose of the experiment and the treatment of personal information. Then, we obtained written consent from them.

2) *Method:* The subjects were 12 engineering university students between 21 and 24 years old. Two female students were included in the subjects.

Fig. 11 shows the method of the experiment. The subjects participated in the experiment one by one. First, each subject read 28 words to him/herself. These words were selected from "Personality trait words [20]." They were expressed impressions of "darkness," "stay in one's shell," and "very sensitive." Second, each subject was asked to read a poem that a 20-year-old man had composed while in a gloomy mood

TABLE III. EVALUATIONS OF TWO HARMONIES.

demension	item which has a difference	Average minor	major	t-value
dysphoria	melancholy	4.01	2.48	10.90**
dysphoria	miserable	4.14	2.54	11.59**
dysphoria	sad	4.41	3.02	8.32**
dysphoria	gloomy	4.38	2.62	8.15**
uplift	cheerful	1.38	2.38	8.02**
uplift	delightful	1.38	2.82	10.07**
uplift	joyful	1.45	2.72	7.05**
uplift	bright	1.63	3.40	7.73**
familiar	tender	2.02	3.47	11.25**
familiar	calm	2.33	3.62	5.23**
familiar	sweet	2.37	2.77	2.38*
strong	vehement	2.55	1.95	3.60**
lightness	hilarious	1.50	2.11	3.91**
lightness	eathery	1.63	2.49	4.02**
stateliness	solemn	3.62	2.79	5.13**
stateliness	ceremonious	3.29	2.62	3.61**

$^{*} < 5\%$, $^{**} < 1\%$.

and put on the Internet. We expected that the subjects would become gloomy while completing these tasks.

After reading the poem, each subject was asked to indicate his/her current mood by filling out a questionnaire. The questionnaire consisted of 40 mood-related items that were selected from the MMS [10]. The 40 items consist of 10 items on each of four dimensions: dysphoria/fatigue, active pleasure, and non-active pleasure. We lined up one set item that four items is extracted from each four dimensions. We could make 10 set items. The order of each set differed depending on the subject and on the number of times (one subject responded to the questionnaire twice). The subjects were asked to evaluate the 40 items on a four-point scale: I do not feel it at all (1); I do not feel it very much (2); I feel it a little (3); I feel it clearly (4).

Next, each subject read the same poem with headphones on. An experimenter pushed the trigger button for MusiCuddle when the subjects read the title of the poem. MusiCuddle calculated a representative pitch for a predetermined period just before the trigger to extract the pitch of each subject's

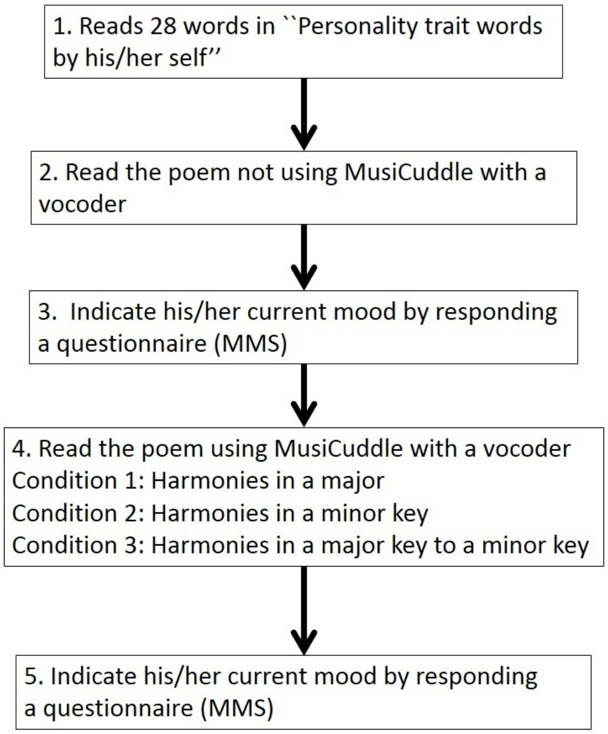

Fig. 11. Experimental method.

voice. Then, the musical phrase, the first note of which is the same as the F0 of the subject's utterance was presented. Since MusiCuddle selects an MIDI file of which the top note of the first chord is similar to the subject's F0, we transposed two phrases (Figs. 9 and 10) into other keys each, in which the top notes of the first chord are C2$\sim$C5 before the experiment. However, the male subjects' voices were quite low, making it hard to hear the selected musical phrase (harmony). Therefore, in the experiment, MusiCuddle selected an MIDI file of which the top note of the first chord was similar to the subject's F0, but one octave higher. Each subject read the poem, hearing his/her voice combined with the musical phrase according to the vocoder function.

We prepared three musical phrase conditions: (1) harmonies in a major key, (2) harmonies in a minor key, and (3) harmonies in a minor key in the early part of the poem and in a major key in the latter part of the poem. In condition (3), the experimenter pushed the trigger button again halfway through the poem. The 12 subjects were assigned to one of the three conditions (four subjects per condition). After reading the poem, each subject completed the questionnaire again.

3) Result of the main experiment: The 12 subjects indicated their current mood by responding to the 40 items on the four-point scale twice. The first time, all subjects read the poem in the same condition, without MusiCuddle. We examined the null hypothesis "the medians of all conditions are equal" in the answers for each of the 40 questions using the Kruskal-Wallis one-way analysis. In the results, one of the items, "lack confidence" showed a significant difference of $p = 0.08$, although the others were $p > 0.10$. There was no evidence of differences in the remaining 39 items. Therefore, after this, we omitted "lack confidence" from the items for analysis.

In their second reading of the poem, the 12 subjects were assigned to one of the three conditions described previously. We conducted the Kruskal-Wallis one-way analysis of subjects' subsequent responses. Moreover, we calculated the differences between subjects' first and second responses to each of the 39 items. Then, we examined the Kruskal-Wallis one-way analysis for the differences in the 39 items.

The left side of Table IV shows the p values for subjects' second answers. In four of the 39 items (cheerful, well, slow-going ($p < 0.05$), and lively ($p = 0.06$)), the null hypothesis was rejected. Therefore, we performed multiple comparison analyses (Wilcoxon signed-rank test) for these items. The third, fourth, and fifth rows from the left of Table IV show the results. We can see that there was only a significant difference between the major and minor conditions for "cheerful" ($p = 0.03$).

Concerning the differences between subjects' first and second answers, we performed the Kruskal-Wallis one-way analysis and multiple comparison analyses. The right side of Table IV shows these results. In seven items, the null hypothesis was rejected. As a result of multiple comparison analyses, significant differences were observed between the major and minor conditions for "cheerful" ($p = 0.03$) and "negative" ($p = 0.06$).

Moreover, for the item "cheerful," there was a significant difference ($p = 0.03$) between four subjects' first and second answers in the major condition. Namely, we can say that the evaluations of "cheerful" for the major harmonies contributed to the result of multiple comparison analysis.

4) Discussion: We conducted an experiment in which subjects read a poem with/without using MusiCuddle with a vocoder function. When using the MusiCuddle, each subject could hear his/her voice, which was modified by harmonies in a major or minor key while reading. We examined the differences among the three conditions. The result showed that subjects' mood after reading the poem differed according to the condition. Moreover, the results of multiple comparison analyses showed that there were significant differences between subjects' cheerful mood for major harmonies and minor harmonies. In particular, it was clear that harmonies in a major key resulted in a more cheerful mood.

As another analysis method, we calculated the differences between subjects' first and second answers. Then, we examined the differences among the three conditions as well as the multiple comparison analyses. We found significant differences between major harmonies and minor harmonies for the "cheerful" ($p = 0.03$) and "negative" ($p = 0.06$) moods.

The iso-principle [4] shows that music's mood or the tempo must initially match patients' mood or tempo. If a patient is gloomy, then gloomy and/or sad music should initially be presented. However, the subjects of our experiment were induced cheerful mood by the major harmonies. The major harmonies were significantly "not melancholy (Ave.was 2.48),"

TABLE IV. CONTRIBUTION OF THE MUSIC IN THREE CONDITIONS TO MOOD INDUCTION.

items that have significant deferences	multiple comparison			
	p	second time		
		minor and major	minor and minor / major	major and minor / major
cheerful	0.02	0.03	0.43	0.17
well	0.05	0.11	1.00	0.11
lively	0.06	0.14	1.00	0.14
slowgoing	0.03	0.08	1.00	0.08
	p	differences between the first and the second time		
		minor and major	minor and minor / major	major and minor / major
cheerful	0.02	0.03	0.29	0.17
well	0.05	0.11	1.00	0.11
fresh	0.03	0.09	0.43	0.14
good mood	0.05	0.09	0.29	0.37
negative	0.03	0.06	0.09	0.40
worried	0.04	0.09	1.00	0.11
tired	0.06	0.11	0.40	0.09

The row of "p" shows the results of the Kruskal-Wallis test.

"not miserable (Ave. was 2.54)," and "not gloomy (Ave. was 2.62)" compared to the minor harmonies (see Table III).

On the other hand, traditionally, emotion was believed to stem from a physical reaction (The James-Lange theory). Then, Schachter and Singer [21] showed that emotional states may be considered a function of a state of physiological arousal and of a cognition appropriate to this state of arousal and a recognition of the factor of the emotion (Two-factor theory). These theories also support our tentative theory that a person who is gloomy can become cheerful when his/her voice is combined with cheerful sounds. Therefore, it is expected that music with the vocoder function calms the symptoms of patients with mental instability who repeat stereotypical utterances.

E. Case study using MusiCuddle with a vocoder

We performed a case study to investigate how a patient's behaviors changed with MusiCuddle using a vocoder. We expected the patient's repetitive utterances to change or stop as a result of the sound coming from MusiCuddle with the vocoder. This case study was approved by the Research Ethics Board of Saga University.

1) Method: The participant was an 81-year-old, hospitalized patient with frontotemporal dementia (FTD). She was hospitalized for depression six years ago and was discharged from the hospital. Next, she was hospitalized with a broken hip. She began to shout sometimes. Then, she moved to a dementia ward in the same hospital. Currently, she repeats stereotypical utterances for many minutes. However, she can often communicate with her care staff.

One of the authors (the MusiCuddle operator) stood by from ten o'clock a.m. to noon and again from one o'clock p.m. to half-past two p.m. In the first part of the case study, the operator played a normal electronic piano near the participant to examine the participant's interest in music.

Six months later, the second part of the case study was performed. We set two time periods, with/without the use of MusiCuddle with the vocoder. During the time MusiCuddle with the vocoder was used, the operator started MusiCuddle and presented the harmonies in either a minor key or a major key (see Section III-D). When the participant began to repeat stereotypical utterances, the operator gave triggers to MusiCuddle arbitrarily to present the harmonies.

The participant's utterances were recorded to examine the changes in her utterances. Two small wireless microphones were used to input her utterances to MusiCuddle and to the vocoder. These microphones were set behind her wheelchair. A small speaker was placed on the table near the participant.

2) Result: The participant tended to start repeating utterances only 15 minutes after going to the bathroom. She asked her care staff to take her to the bathroom, although she did not have to go to the bathroom. The following is an example of her typical utterances. "P" means "Participant."

P: Ne ne-chan ne ne-chan ne ne-chan (toots)

When the first case study was performed, other patients on the same floor were enjoying a karaoke session. When another patient sang songs, the participant temporarily transitioned from repeating utterances to singing the songs together with the other patient. After the karaoke session, the operator played the melodies of songs in which the participant was interested. Then, the participant began to sing another song making up her own lyrics.

In the second part of the case study, when the operator sang her favorite songs in front of the participant, she directed the operator to stop singing. The operator set two time periods, with/without MusiCuddle with the vocoder. The participant had to hear her utterances combined with harmonies in a minor or major key by MusiCuddle with the vocoder from a speaker. She did not push the speaker aside. So, she did not seem to hate the sound. However, there were no differences in the participant's utterances for the two time periods.

3) Discussion: In this case study, contrary to our expectations, the participant's utterances did not change. There are several possible explanations for the results.

1) It was necessary to use a refined speaker to ensure that the participant could hear the sound; because there were some patients with dementia on the floor, it was sometimes noisy. Since the subjects of the experiment (Section III-D) used headphones, they could hear the sound well. On the other hand, it is difficult for patients with mental instability to put headphones on. In the future, we should entertain the use of a directional loudspeaker.
2) If an FTD patient is able to hear the sound, can he/she recognize his/her voice in the sound? Is it

really necessary to recognize it? Even if he/she cannot recognize it, it may be enough to change his/her mood by the sound with the vocoder as long as the sound can shift his/her interest to music.

3) It is necessary to consider intended person for MusiCuddle with the vocoder. The participant in this case study was a patient with FTD. Generally, the following abilities in FTD patients are preserved: memory, perception, praxis, and spatial skills.
The participant repeated stereotypical utterances, telling staff members she wanted to go to the bathroom. The operator (one of the authors) presented the sound from MusiCuddle with the vocoder instead of responding to her request. However, she must have been unpleasant. She was assured she was asking her care staff to take her to the bathroom because she preserved some abilities. When an FTD patient requires something specific, it may not be appropriate to change his/her mood using a sound.

IV. Conclusion

In this paper, first, we introduced a system called "MusiCuddle" for patients with mental instability who repeat stereotypical utterances. MusiCuddle is a system that presents a short musical phrase when an operator pushes a button on the system's interface. The first note of the phrase is the same as the fundamental pitch (F0) of a patient's utterances.

We conducted a case study of a patient who repeated stereotypical utterances for many hours each day. The participant tended to stutter when each phrase was presented from MusiCuddle. The results suggest that FTD patients might attend to phrases according to their similarity in pitch, and their attention may be deflected away from their repetitive stereotypical utterances if the melody is too strikingly different from their utterances.

Then, we added a vocoder function to MusiCuddle so that patients would be able to attend to the music more. The vocoder allows a patient's utterances to combine with the phrase from MusiCuddle in real time. We examined whether the mood induced in subjects using the vocoder differed according to the music coming from MusiCuddle. Each subject read a gloomy poem, hearing his/her voice combined with the musical phrase according to the vocoder function. There are three conditions of the musical phrases: (1) harmonies in a major key, (2) harmonies in a minor key, and (3) harmonies in a minor key in the early part of the poem and in a major key in the latter part of the poem. The 12 subjects indicated their current mood by responding to the 40 items.

The results showed that subjects' mood after reading the poem differed according to the condition. We found significant differences between major harmonies and minor harmonies for the "cheerful" and "negative" moods. Namely, when a person's voice is combined with cheerful sounds, he/she can become cheerful. However, in the second case study, the participant's utterances did not change. It may not be appropriate to change his/her mood using a particular sound when an FTD patient requires something specific.

In the future, we will conduct experiments on patients with Alzheimer's disease who do not require something specific matter, but utter in the form of a monologue.

References

[1] C. Oshima, N. Itou, K. Nishimoto, K. Yasuda, N. Hosoi, H. Yamashita, K. Nakayama, and E. Horikawa, A Case Study of a Practical Use of "MusiCuddle" that is a Music Therapy System for Patients with Dementia who Repeat Stereotypical Utterances, *Proc. of Global Health 2012*, IARIA, pp. 14–20, 2012.

[2] C. Oshima, N. Itou, K. Nishimoto, K. Yasuda, N. Hosoi, H. Yamashita, K. Nakayama, and E. Horikawa, A Music Therapy System for Patients with Dementia who Repeat Stereotypical Utterances, *Journal of Information Processing*, Vol. 21, No. 2, pp. 283–294, 2013.

[3] N. Itou and K. Nishimoto, A Voice-to-MIDI System for Singing Melodies with Lyrics. *In: Proc. of the int. conf. on ACE'07*, pp. 183–189, 2007.

[4] I. M. Altshuler, The past, present and future of musical therapy, Podolsky, E. (Eds.). *Music therapy*, Philosophical Library, pp. 24–35, 1954.

[5] P. Nordoff and C. Robbins, *Creative Music Therapy*, the John Day Company, 1977.

[6] D. Grocke and T. Wigram, *Receptive Methods in Music Therapy: Techniques and Clinical Applications for Music Therapy Clinicians*, Educators and Students Jessica Kingsley Publishers, 2007.

[7] W. Aube, I. Peretz, and J.L. Armony, The effects of emotion on memory for music and vocalisations, *Memory*, 2013.

[8] T. Taniguchi, Music and Affection, Kitaooji Syobo Press, 1998 (in Japanese).

[9] T. Taniguchi, Construction of an Affective Value Scale of Music and Examination of Relations between the Scale and a Multiple Mood Scale, *The Japanese journal of psychology*, Vol. 65, No. 6, pp. 463–470, 1995.

[10] M. Terasaki, Y. Kishimoto, and A. Koga, Construction of a multiple mood scale, *The Japanese journal of psychology*, Vol. 62, pp. 350–356, 1992 (in Japanese).

[11] P. N. Juslin and J. A. Sloboda (Eds.), Music and Emotion: Theory and Research, Oxford University Press, USA, 2001.

[12] M. F. Folstein, S. E. Folstein, and P. R. McHugh, Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician, *Journal of Psychiatric Research*, Vol. 12, pp. 189–198, 1975.

[13] S. J. Sherwood, J. Morris, V. Mor, and C. Gutkin, Compendium of measures for describing and assessing long-term care populations, Boston, MA: Hebrew Rehabilitation Center for the aged, 1977.

[14] Y. Imai and K. Hasegawa, The Revised Hasegawa's Dementia Scale (HDS-R) –Evaluation of its Usefulness as a Screening Test for Dementia. Hong Kong J Psychiatr. Vol. 4, No. 2, pp. 20–24, 1994.

[15] T. Itoh and M. Iwanaga, The effect of the relation between mood and music type on positive emotions, *The Journal of Japanese Music Therapy Association*, Vol. 1, No. 2, pp. 167–173, 2001 (in Japanese).

[16] T. Takeuchi, The influence of presentation sequences of pieces of music on depressed mood reduction –An experimental study with a musical mood induction procedure, *The Journal of Japanese Music Therapy Association*, Vol. 4, No. 1, pp. 76–86, 2004 (in Japanese).

[17] K. Hevner, Experimental studies of the elements of expression in music, *American Journal Psychology*, Vol. 48, pp. 246–248, 1936.

[18] K. Hevner, The affective character of the major and minor modes in music, *American Journal of Psychology*, Vol. 47, No. 1, pp.103–118, 1935.

[19] L. Wedin, Multidimensional study of perceptual-emotional qualities in music, *Scandinavian Journal of Psychology*, Vol. 13, pp. 241–57, 1972.

[20] T. Aoki, A psycho-lexical study of personality trait words: Selection, classification and Desirability ratings of 455 words, *The Japanese journal of psychology*, Vol. 42, No. 1, pp. 1–13, 1971 (in Japanese).

[21] S. Schachter and J. E. Singer, Cognitive, social and physiological determinants of emotional state, *Psychological Review*, Vol. 69, No. 5, pp. 379–99, 1962.