Acoust. Sci. & Tech. **43**, 1 (2022)

PAPER

Correlation of acoustic features of pitch/rhythm/power and perceptual impressions after singing training for people with dysarthria

Maki Nanahara (Kato)[1,*], Kazumasa Yamamoto[2] and Seiichi Nakagawa[2]

[1]*Department of Computer Science and Engineering, Toyohashi University of Technology, 1–1 Hibarigaoka, Tempaku-cho, Toyohashi, 441–8580 Japan*
[2]*Department of Computer Science, Chubu University, 1200 Matsumoto-cho, Kasugai, 487–8501 Japan*

(*Received 19 January 2021, Accepted for publication 7 September 2021*)

Abstract: The aim of this fundamental study is to investigate acoustic changes and perceptual impressions by singing and vocal training in 11 dysarthric Japanese patients. We also examined to improve their speech intelligibility through this training. Dysarthria is a speech disorder caused by diseases such as stroke, intractable neurological disease, and sequelae of head trauma. In terms of vocal evaluations, pre-test, mid-test, post-test, and follow-up test were performed quarterly to investigate a correlation between the five acoustic features of the singing voice (i.e. normalized frequency (pitch) score (NFS), normalized rhythm (duration) score (NRS), normalized intensity (power) score (NIS), frequency deviation (FD) and the intensity deviation (ID)) and speech intelligibility. These factors were evaluated based on the perceptual impression of a group of music major students (MS) and a group of non-music major students (NMS). The objective acoustic features revealed that the order of NRS, NIS, NFS, FD, and ID was higher in correlation with human subjective evaluation. The pre-test results of speech intelligibility indicated a greater improvement in the low intelligibility group than in the high intelligibility group. Furthermore, no difference was found in how perceptual impressions were evaluated between the MS and NMS groups.

Keywords: Acoustic analysis, Dysarthria, Perceptual impression, Singing training, Speech intelligibility improvement

1. INTRODUCTION

Dysarthria is a common speech disorder in patients with sequelae of cerebrovascular accidents (CVAs), intractable neurological diseases, and sequelae of head trauma. Darley *et al.* [1] defined dysarthria as a neuromotor speech disorder resulting from disturbances in muscular control. The symptoms include abnormalities in the strength, speed, range, steadiness, tone, or accuracy of movements required for controlling the respiratory, phonatory, resonatory, articulatory, and prosodic aspects of speech production. Since these utterances are obscure and difficult to convey to the other person, voice communication is hindered. Consequently, the patients' social activities and roles are often limited [2]; which isolates them, leading to a lower quality of life. The estimated prevalence of dysarthria in Japan is 10% of patients with stroke and 50% of those with intractable neurological diseases. Moreover, the 2017 statistics indicated that 226,000 patients suffer from dysarthria in Japan [3]. In relation to stroke, the prevalence of dysarthria accounts for 60% when the age of onset is 70 years or older, and especially considering the current situation where the cerebral infarction rate is increasing [4] in patients with dysarthria because of aging. The proportion is also expected to increase.

Several speech therapy (ST) methods for dysarthria are currently available. According to Ramig *et al.* [5], the Lee Silverman Voice Treatment method was conceived in the United States and has been frequently used [6]. It is particularly effective for dysarthria associated with Parkinson's disease. Furthermore, it is expected to be widely applied clinically to promote the improvement of vocalization and articulation. In ST for dysarthria in Japan, tongue, face, vocalization, breathing, and nasopharyngeal closure training are performed, although the treatment method generally differs depending on the type of dysarthria and the disease and timing caused by it. It is further discussed that for patients with difficulty controlling

*e-mail: nanahara.maki.vm@tut.jp
[doi:10.1250/ast.43.22]

speech speed, using the rhythmic queuing method and pacing board as training for the adjustment method improves the speech intelligibility of patients whose native language is Japanese [7]. However, the number of speech therapists working in the long-term care insurance field is significantly smaller than the number of those working in the medical insurance field [8], and the patients in the maintenance phase are less likely to continue therapy even if desired. A music therapist can intervene as a therapist who can handle areas related to vocalization in therapies other than ST. Furthermore, it can be improved by the intervention of a music therapist, and it has been reported that music therapy improves speech.

In music therapy, several studies have conducted singing training for speech disorders and dysarthria to improve speech intelligibility. Cohen [9] found improvements in fundamental frequency variability, speech rate, and speech intelligibility. Subsequently, a comparison of the singing and rhythmic instructions of Cohen and Masse [10] indicated that the singing group improved in speech rate and intelligibility. Haneishi [11] targeted only patients with Parkinson's disease and demonstrated that singing training improved vocal intensity and speech intelligibility. In a study by Tamplin *et al.* [12], an improvement in speech naturalness was observed. Kato [13] reported an improvement in singing and speaking range, the vocal intensity of speech, and speech intelligibility.

Regardless of ST rehabilitation or music therapy, these previous studies measure speech intelligibility as a results of rehabilitation. Several attempts to predict speech intelligibility from the acoustic characteristics of speech have been reported to improve speech intelligibility in dysarthric patients. Kim *et al.* [14] created a database for people with cerebral palsy dysarthria and a resource for future research on automatic speech recognition in people with neuromotor disorders. Falk *et al.* [15] used Kim's [14] database to analyze the correlation between acoustic parameters and intelligibility and found that tempo dynamics perturbation was significantly correlated with subjective intelligibility assessment. Their results also indicated that the delta power deviation (the power contour slope) of the reading voice was highly correlated with intelligibility. Middag *et al.* [16] performed an automatic intelligibility evaluation using 55 phoneme features and 48 phonological features and found a high correlation between the subjective and objective scores. Ma *et al.* [17] studied a prosodic analysis of people with dysarthria associated with Parkinson's disease and found that they had a higher average Fo and reduced Fo variability than non-dysarthric control speakers.

With regard to the rehabilitation of singing voice in patients with dysarthria, the correlation between the acoustic analysis results of singing voice and the evaluation of voice intelligibility have not been investigated. However, specific training methods for improving speech through music therapy have been developed. Furthermore, unlike the evaluation of normal utterances, it is necessary to confirm whether the evaluation of "singing voice" containing musical elements changes depending on the evaluator to propose rehabilitation by singing training.

The purpose of this study was to conduct singing training for patients with dysarthria symptoms to verify which aspect of the acoustic analysis results of the vocalization correlates with the evaluation by human hearing (perceptual impression) and also to improve speech intelligibility. In this study, we focus on the pitch (frequency), rhythm, and power (intensity) of the voice as the main elements of singing. Additionally, this study investigates any changes in speech intelligibility between evaluators who have received music training and those who have not.

2. RESEARCH DESIGN

2.1. Clinical Setting

The experiment was conducted in the rehabilitation room of a long-term care facility, which included 200 people with various levels of severity of physical disability, intellectual disability, and mental disorder from Japan.

2.2. Participant Recruitment

Given that ST was not practiced at the facility for the disabled, where the study was conducted, patients who were judged by the physical and occupational therapists to have dysarthric speech were selected. The first author and the music therapist assessed the patients who agreed to participate in the study. For assessment purposes, the therapist used Kumakura's speech intelligibility table for dysarthria [18]. In the present study, the type of dysarthria was not classified, because a speech therapist was not available. The study's inclusion criteria were as follows: (1) dysarthric speech acquired from a disease (excluding aphasia), (2) ability to understand the therapist's verbal instructions, (3) ability to read letters, (4) participation in weekly research and a music therapy session, and (5) dysarthric symptoms' severity that did not matter.

2.3. Subjects

Eleven institutional residents with dysarthria participated in this study. Nine were males, and two were females, with an age of 34–75 years old ($M = 58$, $SD = 13.5$). The causative diseases were cerebrovascular accident in six patients (CVA), spinocerebellar degeneration (SCD) in two patients, head traumatic sequelae in two patients (traumatic brain injury, TBI), and Wernicke encephalopathy (WE) in one patient. The participants had spent approximately 8–20 years since the onset of the

Table 1 Summary of patients with dysarthria.

Patient no.	Age	Gender	Diagnosis	Level of dysarthria
1	70	Female	CVA	Mild
2	49	Female	TBI	Moderate
3	52	Male	WE	Mild
4	75	Male	CVA	Moderate
5	65	Male	CVA	Moderate
6	65	Male	CVA	Severe
7	49	Male	SCD	Moderate
8	67	Male	TBI	Moderate
9	74	Male	CVA	Mild
10	34	Male	SCD	Moderate
11	39	Male	CVA	Severe

CVA, cerebrovascular accident; TBI, traumatic brain injury; SCD, spinocerebellar degeneration; WE, Wernicke encephalopathy.

disease that caused dysarthria. Their dysarthric symptoms included low vocalization; weak vocalization; fast or slow speech rate; narrow movement of the tongue, lips, and jaws; abnormal intonation and breathing; and a rattling voice. Besides vocalization, salivation, dysphagia, tremor, and accompanying movements were observed. Based on Kumakura's speech intelligibility table [18], three participants had mild speech intelligibility, six had moderate speech intelligibility, and two had severe speech intelligibility (Table 1).

2.4. Ethical Considerations

In conducting this study, we fully informed all participants, their guardians, and the welfare facility of the experiment, and we obtained their consent.

2.5. Clinical Protocol (Singing/vocal Training in Music Therapy Sessions)

The following program was devised by the therapist as a new research protocol that can be implemented in small groups with physical disabilities. A vocal and singing training research program was conducted 32 times yearly for 25 min, during a 40 min music therapy session, once a week (the remaining 15 min was spent on musical instrument activities and a conversation). Each group consisted of five or six patients.

1) Physical exercises: Upper limb and neck exercises were performed with the therapist's keyboard accompaniment to promote the relaxation of muscles around vocalization organs and blood flow. All participants were in wheelchairs and confirmed that their postures were corrected by the therapist.
2) Oral exercises: The therapist asked the patients to pull their lips sideways, squeeze their lips, and lower their chin significantly while pronouncing five vowels. They continued to practice the movement of sticking out their tongue long, moving it from side to side, inflating both cheeks, and denting inside. The therapist clapped and gave verbal instructions to keep the pace of each movement.
3) Breathing exercises: The patients exhaled slowly through the mouth for eight counts, inhaled for three counts through the nose, and held their breath for two counts during abdominal breathing. The therapist counted by clapping according to the prepared background music and repeated the above series of breaths four times.
4) Vocal exercises: The pronunciation of the time-by-count test of the diadochokinetic syllable rate [19] (sequential motion of syllables of /pa/, /ta/, /ka/, and /la/) was sung on ascending and descending melody scales. The therapist sang and led the vocalization with keyboard accompaniment.
5) Singing training: A song by Japanese Minister of Education, "*Mt. Fuji*," was used as a training and test song because all patients could sing it from memory. Furthermore, they sang a few simple and familiar songs. The singing keys and tempos of the test song were set from the lowest singing key and the slowest singing tempo in the pre-test to the highest singing key and the fastest singing tempo during the 32 sessions. The therapist set the key to rise at a semitone and the tempo to rise gradually. To control breathing, the patients were finally recommended to breathe every two bars.

2.6. Test Content, Procedures, and Schedule for Verification

The patients were individually evaluated during the pre-test, mid-test, post-test, and follow-up test. The test tasks were as follows: (1) To evaluate the prosody of the patient's speech, continuous speech was performed using the time-by-count test of the diadochokinetic syllable rate. (2) Five nonsensical sentences and five grammatically correct sentences were read out to assess speech intelligibility. (3) The "*Mt. Fuji*" song was sung by the patient in a comfortable key without accompaniment. Answers to the questionnaire about the impressions of participating in the experiment were obtained.

As for the schedule of each test and the experiment (see Sect. 2.5. Clinical protocol), the first experiment was conducted immediately after the pre-test. After 16 weeks, the mid-test was conducted, and 16 weeks later, the post-test was done, and the experiment was completed. Afterward, all patients participated in regular music therapy and received a follow-up test 16 weeks later. After completing all the experiments and recordings, two groups of evaluators of music major students and non-music major students conducted perceptual evaluations.

2.7. Data Collection

Using a Linear PCM Recorder (SONY PCM-D1), the patients' and the therapist's voice data were recorded in a quiet rehabilitation room of the facility. A lapel microphone was attached to their shirts, 10 cm below their mouths. Their voices were recorded at every fourth term while they reading nonsensical sentences and grammatically correct sentences and singing the test song, "*Mt. Fuji*" song (see Sect. 2.10.1. Acoustic analysis for details of the recording settings).

2.8. Evaluators

The perceptual impression evaluation was performed using the singing voice and speech in patients with dysarthria as speech intelligibility. The evaluators were divided into two groups that consisted of seven music major students (MS) and nine non-music major students (NMS). A difference in the hearing of singing voice and speech between the two groups of students could be observed. They were paid evaluators.

2.9. Perceptual Impression Evaluations as Intelligibility

2.9.1. Speech of nonsensical sentences

The MS and NMS groups listened and evaluated the nonsensical sentences read by each patient to investigate the improvement of speech intelligibility as an effect of singing training. The task of the perceptual evaluation was to write down the nonsensical three morae embedded in one sentence, for example, this is u ga re (Korewa u ga re desu). The nonsensical morae were used so that the listener cannot guess what the patient was saying from the context of the sentence. The correct answer rate was calculated by dividing the total number of correctly written morae for each term by the total number of morae for all raters in each group. Then, the average of the correct answer rate for each group was used. The intelligibility sample size consisted of 11 patients $\times$ 4 terms $\times$ 2 evaluation groups $=$ 88 points. The number of listening samples was 11 patients $\times$ 5 sentences $\times$ 4 terms $=$ 220 sentences, and all the voices to be listened to and evaluated were randomly played.

2.9.2. Speech of grammatically correct sentences

Another way to verify the improvement in speech intelligibility was to conduct a perceptual evaluation of grammatically correct sentences read by the patients. Two relatively long sentences out of five audio recordings were used for the perceptual evaluation. The text consisted of 13–17 morae, which were often used in daily life (e.g., Watashiwa onakaga itaidesu: I have a stomach ache). This evaluation was conducted only in the NMS group because the results of the intelligibility of nonsensical sentence evaluation indicated no difference between the MS and NMS groups. The method of the evaluation compared the speech between two out of four terms (tests) for each patient reading the same sentence and choosing from three options, namely, the first, the second, or the same. We call this test the "paired comparison test." The set of sentences was a comparison of pre-test and post-test, mid-test and follow-up test, and pre-test and follow-up test. The sets of sentences in reverse order of post-test and pre-test, follow-up test and mid-test, and follow-up test and pre-test were also conducted. The order of the set of sentences and the speech of the patients to be evaluated was randomized so that the previous speech would not affect the evaluation.

2.9.3. Singing voice

The MS and NMS groups listened to and evaluated the patients' four terms of singing voices. Their 44 singing voices were randomly played and judged on a 1–5 for each of the following four items: pitch (completely out of pitch vs. perfectly in pitch), rhythm (completely out of rhythm vs. perfectly in rhythm), the intelligibility of the lyrics (very unclear vs. very clear), and overall morae (completely unrecognizable vs. perfectly recognizable as the song).

2.10. Methods of Acoustic Analysis and Acoustic Parameters

2.10.1. Acoustic analysis

All audio samples were recorded at a sampling frequency of 48 kHz, down sampled at 16 kHz, and recorded at a quantization bit rate of 24 bits per sample by a Linear PCM Recorder. Fundamental frequency (Fo) and power (dB) were used in the free software WaveSurfer 1.8.5 [20], and acoustic analysis as the intensity power (dB) was performed using the free software Praat [21]. The analysis conditions of WaveSurfer were the pitch method (ESPS), max pitch value (400 Hz), min pitch value (60 Hz), analysis window length (7.5 ms), and frame interval (10 ms).

Figure 1 illustrates an acoustic sample when the patient (upper) and the therapist (lower) sang the beginning of "*Mt. Fuji*." The dotted line marks the boundary of the mora that separated each waveform. The red line (0–300 Hz) is the frequency counter (Fo), whereas the blue line (0–60 dB) is the power counter (dB). Each mora represents a different time length that indicates the rhythm. In the beginning of the song, "atama wo" has the following morae: "a, ta, ma, wo." A preprocessing step was performed on a recorded audio file prior to the calculation of the acoustic parameters. For voice labeling, the experimenter listened to the patient's voice and performed mora segmentation manually. At that time, the voice of dysarthria started up slowly, so we labeled each mora including the sighing and a hoarse voice. After cutting these out, those with a

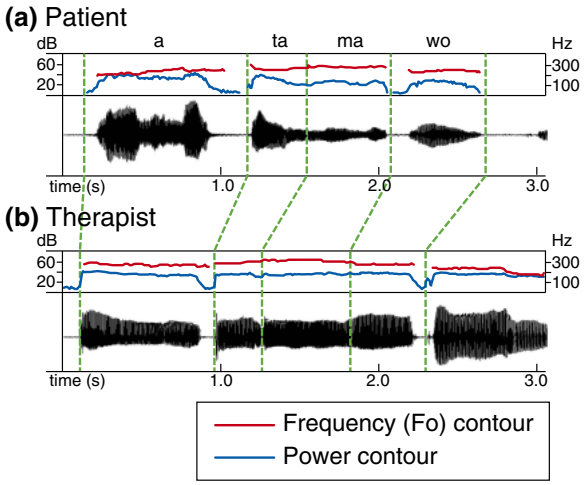

Fig. 1 An example of pitch/power/speech wave contours and time alignment for rhythm and power. The red dot lines represent the frequency contour and the blue continuous lines represent the intensity contour.

Fo value of 0 were regarded as silent sections (since the value is 0 in the frame where pitch estimation cannot be done with WaveSurfer) and deleted, and the rest were calculated as analysis targets. The deletion did not affect the labeled morae. The process of deleting the silent section was done manually, and the subsequent calculations were done automatically. The labeling of these patients' and the therapist's mora is indicated by the formula 2.10.2 as follows. The patient's (P) Fo (frequency) mora (i) is F_Pi, the therapist's (T) Fo mora is F_Ti, the patient's power mora is I_Pi, and the therapist's power mora is I_Ti.

2.10.2. Acoustic parameters

In this research, we applied five acoustic parameters (i.e. normalized frequency score (NFS), normalized rhythm score (NRS), normalized intensity score (NIS), frequency deviation (FD), and intensity deviation (ID)) for singing speech analysis is that the vocalization of dysarthria has characteristics of a narrow frequency (Fo), a small intensity (dB), and a weak intonation. A previous study conducted by Ma *et al.* [17] suggested that people with dysarthria have a narrow range even with high average Fo. R. Patel *et al.* [22] evaluated that Fo and intensity contours of dysarthria's production are flatter than those of healthy people. The parameters for comparison between the patient and the therapist are NFS, NRS, and NIS, and if the difference between the patient and the therapist becomes smaller, it indicates improvement. FD and ID were used because the higher the value within each individual patient, the closer his/her quality of vocalization is to that of a healthy person.

Normalized frequency score (NFS)

F_{Pi} is defined as the average Fo (frequency) of the mora (i) sung by the patient (P). Each mora was normalized with the averaged frequency of all morae:

$$\bar{F}_{Pi} = \log \frac{F_{Pi}}{\frac{1}{N}\sum_i^N F_{Pi}}, \tag{1}$$

where $\bar{F}_{Pi}$ denotes the patient's normalized frequency of mora (i) out of N sung morae in a logarithmic scale. As a reference, the therapist's normalized frequency of mora (i) was defined as

$$\bar{F}_{Ti} = \log \frac{F_{Ti}}{\frac{1}{N}\sum_i^N F_{Ti}}, \tag{2}$$

where T denotes the therapist

The *NFS* was defined as the difference between the patient and the therapist

$$NFS = \frac{1}{N}\left|\sum_i^N (\bar{F}_{Ti} - \bar{F}_{pi})\right|. \tag{3}$$

Normalized rhythm score (NRS)

dur_{pi} represents the time interval of mora (i) sung by the patient (P). To calculate the patients' singing rhythms, we divided each mora duration by the average of all morae of the song and normalized each mora deviation. Figure 1 illustrates the original sample of the rhythm. The normalized rhythm's parameter or score was similarly calculated from the differences between the therapist's normalized parameter and the patients' parameter as the normalized pitch. We calculated the absolute value from the differences and the averaged overall morae and called this result the normalized rhythm.

The patient's singing rhythm was defined as

$$\overline{dur}_{Pi} = \frac{dur_{Pi}}{\frac{1}{N}\sum_i^N dur_{Pi}}, \tag{4}$$

where $\overline{dur}_{Pi}$ denotes the patient's normalized duration of mora (i) out of N spoken morae. As a reference, the therapist normalized the mora duration (i) and was defined as

$$\overline{dur}_{Ti} = \frac{dur_{Ti}}{\frac{1}{N}\sum_i^N dur_{Ti}}, \tag{5}$$

where T denotes the therapist.

The *NRS* was defined as the difference:

$$NRS = \frac{1}{N}\left|\sum_i^N (\overline{dur}_{Ti} - \overline{dur}_{Pi})\right|. \tag{6}$$

Normalized intensity score (NIS)

I_{Pi} represents the average intensity of mora (i) sung by the patient (P). The strength of the singing voice (intensity)

of each mora, measured in decibels (Fig. 1), was subtracted from the average of all measurements of mora strengths and normalized scores. The normalized intensity parameter or score was calculated from the difference between the therapist's normalized parameter and the patients' parameters. We calculated the absolute value from this difference and averaged the results. The patient's strength of voice (intensity) of each song morae i, measured in decibels, was given as

$$\bar{I}_{Pi} = \frac{I_{Pi}}{\frac{1}{N}\sum_i^N I_{Pi}}, \tag{7}$$

where $\bar{I}_{Pi}$ denotes the patient's normalized intensity of mora (i) out of N sung morae. As a reference, the therapist's normalized mora intensity (i) was defined as

$$\bar{I}_{Ti} = \frac{I_{Ti}}{\frac{1}{N}\sum_i^N I_{Ti}}, \tag{8}$$

where T denotes the therapist.

The *NIS* was defined as the difference:

$$NIS = \frac{1}{N}\left|\sum_i^N (\bar{I}_{Ti} - \bar{I}_{Pi})\right|. \tag{9}$$

Frequency deviation (FD)

The frequency deviation, $\overline{FD}$ of the patient (P)'s singing in N sung morae was defined as

$$\overline{FD} = \frac{1}{N}\sum_i^N \left|\log F_{Pi} - \frac{1}{N}\sum_i^N \log F_{Pi}\right|. \tag{10}$$

Intensity deviation (ID)

To see the patient's P deviation of power (intensity) $\overline{ID}$ and the intensity of his/her voice in N sung morae, we defined the intensity deviation as

$$\overline{ID} = \frac{1}{N}\sum_i^N \left|I_{Pi} - \frac{1}{N}\sum_i^N I_{Pi}\right|. \tag{11}$$

3. RESULTS

3.1. Improvement of Intelligibility by Singing Training

Table 2 presents the correct answer rate of the speech of nonsensical sentences evaluated in every four terms. Patients with a "+" and "−" mark had improved intelligibility and worsened intelligibility after singing training, respectively. The intelligibility improvement evaluation was based upon sufficient verbal communication. This threshold of improvement was 60% of syllable intelligibility and 90% of sentence intelligibility [23]. Less than 60% of the patients in the pre-test are shown in bold. Six of these patients improved by 8.5% after training. Five patients who showed more than 60% in the pre-test improved by only 2%. Thus, intelligibility was improved in patients with severe dysarthria.

Table 2 Percentage of intelligibility of nonsensical sentences (%). The patients with marks "+" and "−" indicate that their intelligibility improved or worsened after the singing training, respectively. The patients who scored under 60% in the pre-test are shown in bold.

Patient no.	pre-test	mid-test	post-test	follow-up test
1−	78.5	87.5	84.0	77.5
2+	72.0	74.5	85.0	78.0
3+	74.5	72.0	67.0	76.0
4+	**47.5**	**62.0**	**71.5**	**68.5**
5−	74.5	42.5	50.0	62.0
6+	**26.0**	**26.0**	**39.0**	**37.5**
7+	**48.5**	**46.5**	**49.0**	**55.5**
8+	**40.0**	**44.0**	**39.0**	**43.0**
9−	89.5	74.0	87.5	87.5
10+	**42.5**	**45.5**	**41.5**	**69.0**
11+	**37.0**	**26.0**	**11.0**	**33.0**

Table 3 Correlation coefficient between the perceptual impressions for song elements and intelligibility of nonsensical speech.

Song element	NMS (44 points)	MS (44 points)	All students (88 points)	Correlation between the two groups
Pitch	0.282	0.408	0.344	0.908
Rhythm	0.292	0.235	0.264	0.898
Lyric	0.600	0.595	0.610	0.922

Table 3 presents the correlation between the measured intelligibility of each song element, such as pitch, rhythm, and lyrics, and the intelligibility of the nonsensical sentences. The two student groups rated the intelligibility for each song element on a scale of 1–5. The intelligibility between the lyrics and speech of nonsensical sentences indicated the highest correlation. Furthermore, a strong correlation was found between the two groups (rhythm: 0.898 and pitch: 0.922).

3.2. Results of Time-by Count Test of the Diadochokinetic Syllable Rate

Table 4 presents the results of the time-by count test of the diadochokinetic syllable rate of the 11 patients, who

Table 4 Average of the time-by count test of the diadokochinetic syllable rate of patients for four terms (per second).

Syllable	pre-test	mid-test	post-test	follow-up test
pa	1.9	2.6	2.9	3.0
ta	1.8	2.4	2.7	3.0
ka	1.8	2.3	2.7	3.0
pa, ta, ka	1.5	2.0	2.3	2.0

Table 5 Difference of intelligibility evaluation of grammatically correct sentences for paired comparison tests.

Patient no.	post-test/pre-test (50 points)	follow-up test/mid-test (50 points)	follow-up/pre-test (50 points)
1	22	13	27
2	27	30	21
3	5	4	11
4	−2	−7	20
5	11	40	32
6	2	7	6
7	17	17	36
8	16	25	24
9	8	20	14
10	24	14	16
11	−44	−16	−35

were each evaluated three times, showing the averages of each patient per second. As a reference, people without dysarthria can pronounce a syllable over four times per second. When the pre-test and follow-up test were compared, 90% of the patients had improved syllable repetitions, indicating that using these prosodies for oral motor function exercises is effective.

3.3. Correlation between Acoustic Parameters and Speech of Grammatically Correct Sentences

Table 5 indicates the difference between the perceptual evaluation of grammatically correct sentences and how much their intelligibility improved over each paired comparison tests. A positive number indicates intelligibility improvement (the maximum number is 50 points), and a negative number indicates worse intelligibility. From these data set, there was a significant difference [$t(10) = 1.813$, $p = 0.05$]. We determined that it improved if it was 10 points or greater. We concluded that 22 paired comparison tests of 33 tests showed improvement in intelligibility. Moreover, we concluded that less than 10 points indicated no improvement, and 11 paired comparison tests showed no improvement in intelligibility. The difference between the follow-up test and pre-test of the paired comparison tests was the largest, indicating that the effect was sustained after training. These three sets of paired comparison tests show the degree of improvement by grammatically correct sentences. In this intelligibility evaluation, most patients improved in the follow-up test and pre-test of the paired comparison tests after the training, even when the improvement was less than 10 points during the training.

Based on the results in Tables 2 and 5, which showed improvement in speech intelligibility only, we analyzed the results by disease. An increase was observed in the rate of improvement in the training effect in the order of SCD > CVA > TBI > WE.

Table 6 Average and variance of frequency (log Fo) from raw acoustic data.

Patient no.	pre-test		mid-test		post-test		follow-up test	
	average	variance	average	variance	average	variance	average	variance
1	4.845	0.046	4.843	0.053	4.738	0.054	4.853	0.093
2	5.177	0.025	5.091	0.036	5.250	0.033	5.413	0.032
3	5.031	0.069	5.028	0.045	4.991	0.109	5.077	0.036
4	5.212	0.005	5.176	0.011	5.333	0.039	5.233	0.015
5	5.134	0.014	5.178	0.007	5.189	0.123	5.104	0.614
6	5.317	0.030	5.328	0.023	5.307	0.018	5.391	0.035
7	4.729	0.004	4.717	0.001	4.795	0.064	4.800	0.072
8	4.873	0.037	4.843	0.016	5.024	0.044	4.431	1.903
9	4.972	0.007	4.904	0.008	4.981	0.019	4.983	0.561
10	5.150	0.010	5.219	0.012	5.374	0.027	5.378	1.341
11	4.743	0.011	4.854	0.012	4.977	0.119	4.849	0.523

Table 7 Average and variance of duration of mora (s) from raw acoustic data.

Patient no.	pre-test		mid-test		post-test		follow-up test	
	average	variance	average	variance	average	variance	average	variance
1	0.497	0.046	0.557	0.061	0.621	0.061	0.614	0.069
2	0.693	0.062	0.597	0.082	0.572	0.079	0.706	0.070
3	0.693	0.093	0.699	0.101	0.680	0.094	0.707	0.101
4	0.316	0.022	0.514	0.052	0.478	0.043	0.484	0.070
5	0.652	0.094	0.643	0.078	0.614	0.073	0.660	0.087
6	0.526	0.039	0.516	0.043	0.557	0.051	0.563	0.050
7	0.581	0.056	0.562	0.082	0.584	0.078	0.606	0.053
8	0.648	0.132	0.710	0.131	0.704	0.136	0.623	0.111
9	0.602	0.068	0.656	0.060	0.621	0.068	0.468	0.026
10	0.363	0.034	0.787	0.273	0.894	0.175	0.774	0.120
11	0.318	0.067	0.368	0.069	0.285	0.705	0.591	0.143

Table 8 Average and variance of intensity (dB) from raw acoustic data.

Patient no.	pre-test		mid-test		post-test		follow-up test	
	average	variance	average	variance	average	variance	average	variance
1	64.284	33.537	40.889	40.503	46.601	16.351	47.851	32.051
2	46.307	43.492	43.270	63.524	42.019	49.631	51.659	34.556
3	60.697	28.119	45.320	55.168	46.302	31.754	47.620	30.173
4	61.682	47.542	46.845	72.010	54.360	57.216	53.596	70.642
5	44.477	34.096	36.812	70.310	40.190	17.384	45.737	40.826
6	63.201	38.748	50.752	75.777	52.260	53.133	54.875	66.667
7	56.619	58.231	34.339	71.467	38.536	40.877	41.800	53.415
8	46.570	50.995	39.207	94.428	41.457	47.172	50.521	98.471
9	53.740	17.217	42.213	42.137	41.285	34.671	43.714	25.067
10	38.291	45.107	43.284	86.962	49.051	36.944	58.434	41.788
11	41.320	42.469	30.183	114.136	32.222	65.716	34.088	0.872

Tables 6–8 show the average and variance values of raw acoustic data. In terms of average values, these were similar, however, in terms of the variance values, there were fluctuations in the data within the individual. In particular, frequency and duration of mora showed large

Table 9 Correlation coefficient between differences in acoustic parameters and intelligibility evaluation of grammatically correct sentences for paired comparison tests.

Parameters	Correlation between the difference in acoustic parameters and that in intelligibility
NFS	−0.593
NRS	−0.233
NIS	−0.037
FD	0.522
ID	0.289

Table 10 Correlation coefficient between the acoustic parameter and nonsensical intelligibility.

Parameters	NMS (44 points)	MS (44 points)	All students (88 points)
NFS	−0.245	−0.362	−0.288
NRS	−0.457	−0.468	−0.467
NIS	−0.434	−0.369	−0.401
FD	0.222	0.190	0.209
ID	0.240	0.256	0.248

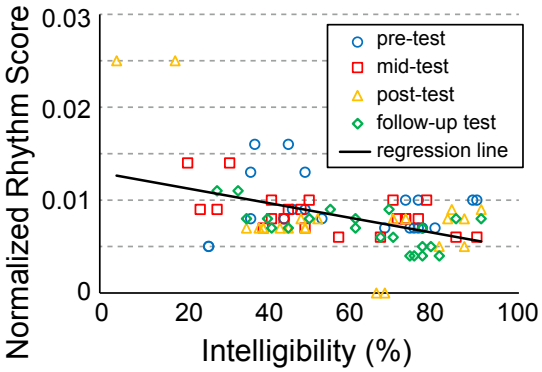

Fig. 2 Correlation of normalized rhythm scores and intelligibility of nonsensical sentences by all students.

Table 11 Correlation coefficient between the each of acoustic parameter and perceptual impression of pitch and rhythm for a song.

Parameters	NMS (44 points)	MS (44 points)	All students (88 points)
NFS	−0.470	−0.484	−0.476
NRS	−0.620	−0.597	−0.606

fluctuations in the follow-up test. In intensity, all patients had similar variance fluctuations in each test. However, the effects of training cannot be determined from these data sets. Therefore, we verified using five acoustic parameters: NFS, NRS, NIS, FD, and ID.

Table 9 shows the correlation between the paired comparison test between tests of the perceptual evaluation of grammatically correct sentences and the difference value of the acoustic parameters in the same way as the paired comparison tests. NFS has the highest correlation coefficient (slightly higher correlation) (−0.593), whereas FD was the next highest (0.522). NRS and ID had a low correlation, whereas NIS did not correlate.

3.4. Correlation between Acoustic Parameters and Speech of Nonsensical Sentences

We analyzed the correlation between five acoustic parameters (NFS, NRS, NIS, FD, and ID) and speech intelligibility by reading nonsensical sentences using Pearson's correlation coefficient (Table 10). The three normalized scores, NFS, NRS, and NIS, show the difference between the patients' and the therapist's scores, indicating that the smaller the parameter value, the better; a negative correlation was observed. The two deviations, FD and ID, indicate that the patients improved as the values increased. A statistically significant correlation exists between the NFS and the speech intelligibility test for all students at the 5% risk level. The correlation between the NRS and intelligibility test and the NIS and intelligibility test for all students was statistically significant at the 1% risk level. Thus, a correlation exists between each normalized parameter and speech intelligibility. The order of high correlation between the five acoustic parameters and speech intelligibility of the nonsensical sentence was NRS > NIS > NFS > ID > FD (Table 10). The results of the correlation between ID and intelligibility evaluation are consistent with those of Falk's study [15], which used a reading voice instead of a singing voice. Figure 2 indicates that the highest correlation (−0.467) was obtained between the NRS of 11 patients for four terms and the perceptual evaluation of nonsensical sentences of all 16 students (MS and NMS). Furthermore, the correlation coefficient of the listening evaluation of nonsensical sentences between the MS and NMS groups that were evaluated was 0.945, indicate a strong correlation between the groups. Hence, no difference was found between the two groups.

Table 11 presents the correlation between the two acoustic parameters (NFS and NRS) and the perceptual impression evaluation (pitch and rhythm) of singing. NRS shows a higher correlation (−0.606) with all students (−0.606) than NFS (−0.476).

4. DISCUSSION

In the present study, we examined any correlation between the five acoustic parameters (NFS, NRS, NIS, FD, and ID) and perceptual impression evaluation for each of the four terms of singing training. The results indicated that NRS had the highest correlation with speech of nonsensical

sentences, suggesting that rhythm is a factor in improving intelligibility. Additionally, the correlation between the acoustic parameters of NFS and NRS and the perceptual impression of singing also showed a higher correlation in NRS than that in NFS. These results also indicate that singing training makes it easier for people to hear the singing voice by practicing strengthening the rhythm. Next, since NIS had a high correlation with the speech of nonsensical sentences, the training effect of the voice volume is likely to appear during the breathing exercise and the singing exercise. Subsequently, a correlation with the nonsensical speech in the order of NFS, ID, and FD was observed, but each showed a weak correlation coefficient. With regard to the evaluation of the speech of grammatically correct sentences, the evaluators can infer the patients' words from the context, which may affect their perceptual impression. Additionally, the correlation between the perceptual impression of pitch, rhythm, and lyrics during singing and speech of nonsensical sentences was the highest in the correlation with lyrics regarding words. This is because the words in the lyrics can be guessed, so the intelligibility is likely to be highly evaluated.

Tables 2, 4, and 5 present the improvements in dysarthric speech intelligibility by singing training. Table 2 indicates that speech intelligibility in the speech of nonsensical sentences is more likely to show improvement in severely dysarthric patients. From the results of the paired comparison tests in the speech of grammatically correct sentences in Table 5, the difference between the follow-up test after training and the paired comparison test before training was large even if the improvement during training was 10 or less points. In other words, it was suggested that the effect was sustained even after the training. Additionally, in Table 4, the average of the time-by-count test of the diadochokinetic syllable rate, which is the basis of phonemes for improving speech intelligibility, 90% of the patients improved based on the follow-up test. It is probable that the effect of vocal practice was shown during singing/vocal training. Whether to use the speech of nonsensical sentences or the speech of grammatically correct sentences as the perceptual impression may differ depending on what is verified, but the validity of the analysis method should be re-examined.

When perceptually evaluating the singing and uttering voices of the dysarthric patients, a fairly high correlation was observed between the two groups. Based on the results (see Table 3), it was determined that music training experience in determining speech intelligibility did not affect the evaluation. It is considered that some estimation is possible by using only the parameters with high correlation among these data. In the future, we hope that instead of manually evaluating the patient's singing and vocal training, it could be possible to do it on a computer. As a result, rehabilitation feedback will become concrete and lead to the improvement in the patient's motivation. It will be possible for therapists to provide efficient and fulfilling singing rehabilitation.

5. CONCLUSION

In this study, it was demonstrated that singing training for dysarthric speech intelligibility correlates with the objective evaluation by the analysis of five acoustic parameters from a singing voice with three subjective perceptual impressions of singing and speech intelligibility. At the same time, it became clear that rhythm, which is an element of singing, affects speech intelligibility. In other words, singing training with enhanced rhythm shows improvement in speech intelligibility and intelligibility of lyrics during singing. The results of the perceptual impressions by the two evaluator groups also suggested that equivalent evaluations were possible regardless of whether the evaluators were trained in music. In the present study, we proposed one possibility of an objective evaluation method for singing training for dysarthria.

However, this study has some limitations. First, it was not possible to create a control group of patient subjects. The reason is that only a few dysarthric patients were admitted in the institution, and dysarthric patients other than those who participated were unable to participate in the experiment because of the combination of multiple disabilities. Since these patients were grouped, the generalization of the results was limited. Moreover, because speech therapist was not performed at the facility where this experiment was conducted, specialized assessment and evaluation of speech could not be performed and the data analysis was insufficient.

Kim, Kent, and Weismer [24] reported that speech intelligibility depends on the type of disease and severity of the medical condition. In the present study, it was found that improvement in speech intelligibility is likely to appear in SCD and CVA regardless of the severity. This is consistent with Nishio's previous research, which states that CVA becomes more severe has a large degree of improvement [7]. Regarding TBI, it was concluded that speech is expected to improve even several years after the injury and that even severe patients need a long-term follow-up treatment [2]. Therefore, rehabilitation in the maintenance phase is important.

In the future, it is desirable not only to increase the number of patients but also to compare age, disease and elapsed time after onset, degree of dysarthria type, and speech severity and to consider training plans with a speech therapist. It is necessary to study a more accurate method for acoustic analysis and evaluation of singing voices.

ACKNOWLEDGMENTS

The authors are deeply grateful to Dr. Eri Hirokawa for music therapy, Yumiko Mitsuda for speech therapy, and Dr. Ryota Nishimura and Dr. Alberto Yoshihiro Nakano for the technical supports, and for their valuable suggestions and time spent on this study. The authors also thank all the patients, who took part and the staffs for their cooperations during this experiment.

REFERENCES

[1] F. Darley, A. Aronson and J. Brown, *Motor Speech Disorders* (W. B. Saunders, Philadelphia, London, Toronto, 1975), p. 2.
[2] K. M. Yorkston, "Treatment efficacy: Dysarthria," *J. Speech Hear. Res.*, **39**, 46–57 (1996).
[3] M. Kariyasu, T. Matsuhira and M. Toyama, "Communication disorder and estimated number of persons with disability," *Dep. Bull. Pap., Kyoto Gakuen Daigaku Sogo Kenkyujo Shohou*, **18**, 55–60 (2017) (in Japanese).
[4] A. Toyota, "Actual status of stroke in the working generation based on patient statistics of rosai general hospitals in Japan," *J. Occup. Med. Traumatol.*, **58**, 89–93 (2010) (in Japanese).
[5] L. Ramig, C. Mead, R. Scherer, Y. Horii, K. Larson and D. Kohler, "Voice therapy and Parkinson's disease: A longitudinal study of efficacy," *Paper presented at the Clinical Dysarthria Conf.*, San Diego, (1988).
[6] K. Nakayama, T. Yamamoto, C. Oda, M. Sato, T. Murakami and S. Horiguchi, "Effectiveness of Lee Silverman voice treatment LOUD on Japanese-speaking patients with Parkinson's disease," *Rehabil. Res. Pract.*, Article ID 6585264 (2020).
[7] M. Nishio, Y. Tanaka and N. Abe, "Efficacy of speech therapy for dysarthria," *J. Jpn. Soc. Logop. Phoniatr.*, **48**, 215–224 (2007).
[8] M. Ozono, "Current status and future tasks of speech therapist education," *J. Health Sci.*, **9**, 1–6 (2012) (in Japanese).
[9] N. Cohen, "The effect of singing instruction on the speech production of neurologically impaired persons," *J. Music Ther.*, **29**, 87–102 (1992).
[10] N. Cohen and R. Masse, "The application of singing rhythmic instruction as a therapeutic intervention for persons with neurogenic communication disorders," *J. Music Ther.*, **30**, 81–99 (1993).
[11] E. Haneishi, "Effect of music therapy voice protocol on speech intelligibility, vocal acoustic measures, and mood of individuals with Parkinson's disease," *J. Music Ther.*, **38**, 273–290 (2001).
[12] J. Tamplin and D. Grocke, "A music therapy treatment protocol for acquired dysarthria rehabilitation," *Music Ther. Perspect.*, **26**, 23–29 (2008).
[13] M. Kato, "The effect of music therapy for improvement of vocal range and vocal intensity on central neurological disease patients with dysarthria: Based on the prosodic analysis," *Jpn. J. Music Ther.*, **8**, 67–75 (2008) (in Japanese).
[14] H. Kim, M. Hasegawa-J., A. Perlman, J. Gunderson, T. Huang, K. Watkin and S. Frame, "Dysarthric speech database for university access research," *Proc. ICSLP*, pp. 1741–1744 (2008).
[15] T. H. Falk, R. Hummel and W.-Y. Chan, "Quantifying perturbations in temporal dynamics for automated assessment of spastic dysarthric speech intelligibility," *Proc. ICASSP*, pp. 4480–4483 (2011).
[16] C. Middag, J.-P. Martens, G. V. Nuffelen and M. D. Bodt, "Automated intelligibility assessment of pathological speech using phonological features," *EURASIP J. Adv. Signal Process.*, ID 629030, pp. 1–9 (2009).
[17] J. K. Y. Ma and R. Hoffmann, "Acoustic analysis of intonation in Parkinson's disease," *Proc. Interspeech*, pp. 2586–2589 (2010).
[18] I. Kumakura, *Dysarthria* (Kenkousya, Tokyo, 2005), p. 62 (in Japanese).
[19] S. G. Fletcher, "Time-by-count measurement of diadochokinetic syllable rate," *J. Speech Hear. Disord.*, **15**, 763–770 (1972).
[20] WaveSurfer, https://www.speech.kth.se/wavesurfer/ (accessed April 14, 2008).
[21] P. Boersma and D. Weenink, Praat program, version 5. 0. 22, https:www.fon.hum.uva.nl/praat/ (2008) (accessed April 14, 2008).
[22] R. Patel and P. Campellone, "Acoustic and perceptual cues to contrastive stress in dysarthria," *J. Speech Lang. Hear. Res.*, **52**, 206–222 (2009).
[23] S. Imai, Y. Yamashita, N. Suzuki and K. Michi, "Development of sentence intelligibility assessment method for patients with articulation disorders," *Jpn. J. Logop. Phoniatr.*, **38**, 357–365 (1997) (in Japanese).
[24] Y. Kim, R. D. Kent and G. Weismer, "An acoustic study of the relationship among neurologic disease, dysarthria type, and severity of dysarthria," *J. Speech Lang. Hear. Res.*, **54**, 417–429 (2011).

Maki Nanahara (Kato) received her B.M. degree (Music Performance) from University of Alaska, Anchorage in 1995 and M.M. degree (Music Education/Music Therapy course) from Nagoya College of Music in 2007. She has been working at welfare facilities and a hospital as a Music Therapist certified by Japanese Music Therapy Association and Gifu-prefecture and is currently a student toward her Ph.D. in Computer Science and Engineering, Toyohashi University of Technology, Japan.

Kazumasa Yamamoto received his B.E., M.E. and Dr. Eng. degrees in Information and Computer Sciences from Toyohashi University of Technology in 1995, 1997, and 2000. He joined Shinshu University in 2000, and he moved to Toyohashi University of Technology in 2007. In 2017, he moved to Chubu University and he is currently serving as a professor in the Department of Computer Science. His current research interests include automatic speech recognition, spoken dialogue system and sound signal processing.

Seiichi Nakagawa graduated a doctor course of Kyoto University in 1976. He joined the Faculty of Kyoto University in 1976, as a Research Associate in the Department of Information Sciences. In 1980, he moved to Toyohashi University of Technology. From 1990 to 2014, he was a Professor in the Department of Information and Computer Science and from 2014 to 2018, he was a Special Appointment Professor. From 2017 to 2021, he joined the Department of Computer Science, Chubu University as a Professor. He is now a Visiting Professor of Chubu University.