Design and Evaluation of Instrument Sound Identification Difficulty for the Deaf and Hard-of Hearing

Shiho Akaki [*], Rumi Hiraga[†], Keiichi Yasu[†], Keiji Tabuchi[*] and Hiroko Terasawa[*]
[*]University of Tsukuba, Tsukuba, Japan
E-mail: {akaki.shiho.ss@alumni, terasawa@slis, ktabuchi@md}.tsukuba.ac.jp
[†]Tsukuba University of Technology, Tsukuba, Japan
E-mail: {rhiraga, k-yasu}@a.tsukuba-tech.ac.jp

***Abstract*— This study investigates the level of difficulty in identifying listening to 16 different instrumental sounds in moderate to severe deaf and hard-of-hearing (DHH) subjects. Previous studies have suggested that DHH listeners' instrument identification difficulty level depends on the instrument group. Based on these findings, we hypothesized that timbre identification for similar instruments (e.g., two plucked string instruments) would be more difficult than that between dissimilar instruments (e.g., a plucked string instrument and a woodwind instrument). Based on this hypothesis, we designed the ISID, which is the difficulty of combining the sounds of four alternative instruments that comprise a group of instruments based on timbre similarity. In this study, we report the results of an instrumental timbre identification test developed using the ISID, which was administered to 20 patients with moderate to severe DHH, compare correct responses and the response times for each level of the ISID, and analyze and verify whether the identification of instrumental timbres exhibited any differences. The results show significant differences in the three levels of ISID based on the correct response rate and response time, thereby supporting the hypothesis.**

Keywords: Cochlear implant, Hearing aid, Music training, Timbre identification, Music instruments

I. INTRODUCTION

Many deaf and hard-of-hearing (DHH) people enjoy music; however, not everyone confidently distinguishes musical elements. Understanding the timbre of musical instruments helps to understand music: the distinction between melody, harmony, and rhythm becomes clear when the timbre is distinctly perceived. In this study, we analyzed the identification of timbre presented in melodic contexts using DHH subjects.

Recently, advancements in hearing aids (HAs) and cochlear implants (CIs) have enabled severely hearing-impaired people to enjoy music. However, some DHH people remain unconfident concerning whether they enjoy music in the same manner as normal hearing (NH) people or whether they can hear music correctly. The auditory education for DHH people primarily focuses on spoken language, rather than listening to or playing music. For the DHH to confidently enjoy music, providing a playful training program is necessary to gradually develop musical skills and interests [1]. However, a standard for timbre identification difficulty for DHH people has not been established, which forms the basis of musical curricula.

According to the guidebook by the Japanese Ministry of Education and Culture, Sports, Science, and Technology (MEXT) [2], DHH children can learn timbre differences across instrumental groups, e.g., to distinguish between percussion and strings, wooden and metal instruments, or percussion/wind/string instruments. However, for subjects to distinguish subtle differences within an instrument group is often difficult. This is because HAs and CIs struggle to reproduce subtle differences in timbre.

Timbre discrimination can be acquired by training. Kraus' study on NH subjects [3] determined that musicians respond better than non-musicians to stimuli in the neural representation of pitch, timing, and timbre in the human auditory brainstem. They suggested that music training might improve music processing by inducing functional and structural changes in the brain.

In this study, we investigate the instrumental sound identification difficulty (ISID), which is a tree-structured timbre stimulus set for identification tasks, that also serves as a basis for musical training curriculum design for DHH people. Instrumental sounds were classified based on their acoustic similarity. These sounds were presented as Japanese popular music (J-POP) melodies, and DHH participants performed a timbre identification task on a game-like platform developed for iOS devices. Finally, we analyzed the correct response and response time, and evaluated the properties of the ISID.

II. RELATED WORK

DHH people listening to musical instruments has been well investigated, mostly in the following two forms: test [4-9] and training [13-16].

Test studies [4-9] are investigated in the following manner: DHH and NH subjects identify instrumental sounds in monophonic, melodic, and ensemble forms from a closed set of stimuli. These studies compare the rate of correct responses according to subject group and instrument type. Studies from a different perspective from the test format involve examining the effect of instrumental timbre on the identification of melodic contours [10], comparing CI and NH timbre spaces

978-616-590-477-3 ©2022 APSIPA APSIPA ASC 2022

using multidimensional scaling [11], and evaluating timbre identification and comfort [12]. McDermott et. al. [4-5] tested 10 CIs to identify 16 percussion and non-percussion sounds. Instruments in the percussion category were identified more correctly than those in the non-percussion category. Participants tended to confuse instruments belonging to the same group (i.e., confusing a percussion instrument with another percussion instrument, but not with a non-percussion sound).

Training studies [13-16] are investigated in the following manner: DHH and NH participants identify instrumental sounds after learning them using audiobooks or music training. Jiam et. al. [13] investigated the effects of music training on timbre discrimination in NH and DHH participants using 16 instrumental timbres belonging to four instrument classes (woodwinds, percussion, brass, and strings). The results showed that the NH identified percussion instruments most accurately, followed by strings, brass, and woodwinds. In contrast, the CIs correctly identified percussion instruments the most, followed by brass, strings, and woodwinds. In all instrument classes, NH scored better than DHH.

These studies (both test and training forms) have not examined the relationship between identification and instrument type in more detail. In addition, no study has been conducted on the difficulty indicators for listening to musical instruments for auditory training. In most studies on listening to musical instruments, tasks involve selecting an image of the instrument after the sound of the instrument was presented. This task is significantly influenced by participants' knowledge of the instrument and their listening experience.

In this study, we performed an identification task intended to measure the hearing difficulty of the DHH, based on the similarity of instrumental sounds.

III. Instrumental Sound Identification Difficulty (ISID)

Prior research has shown that the more similar the timbre between instruments, the more difficult it is to discriminate between them. Acoustically, the similarity of timbre is caused by the similarity of the physical structure. Based on this, we designed the ISID, a tree-structured stimuli set for timbre identification inspired by musical instrument classifications.

One of the best-known instrument classifications is the Sachs-Hornbostel classification [17]. This classification is based on the principles of the structure and vibration mechanism of the instrument. The basic categories include idiophones, aerophones, membranophones, chordophones, and electrophones. However, Hayasaka's classification [18] may be more comparable with conventional musical practice: musical instruments are classified into four major types (wind, string, percussion, and electronic) based on their vibrating mechanism and shape.

In this experiment, we employed Hayasaka's classification, considered the following three conditions, and chose instruments for the stimuli set.

TABLE I Instrument classification for this study

Primary	Secondary	Tertiary	Instrument
String		Plucking String	Harp, Acoustic Guitar
		Bowed String	Violin, Cello
Wind	Woodwind	Single Reed	Clarinet, Alto Sax
		Double Reed	Oboe, Bassoon
		Air Reed	Flute, Recorder
	Brass	Lip Reed	Trumpet, Trombone
Percussion			Marimba, Vibraphone
Keyboard			Piano, Organ

1. The instrument is well known
2. The instrument has clear pitch
3. The instrument is not electronic

The first condition helps accustom students to instruments popularly used in music. The second condition enables the preparation of melodic stimuli. The last condition was to avoid confusion because our stimuli were created using sample-based synthetic sounds. While Hayasaka's classification categorized struck string instruments as a subclass of stringed instruments, few are well known. Therefore, we adopted keyboard instruments as a category and assigned piano and organ to it.

The classification of musical instruments used in this study was based on that of Hayasaka [18] and is presented in Table I.

The difficulty levels of instrumental sound identification were defined based on the musical instrument classifications listed in Table I. We hypothesized that the closer the classification, the more similar the instrumental sounds, thus increasing the difficulty of identification. Based on the classification of musical instruments, we created combinations of instrumental sounds for the four choices in the assignment, as shown in Table II. The maximum number of instruments is 16, and the higher the difficulty level, the harder the expected identification.

IV. Experiments

This experiment used a memory game style: participants first listened to a melody with an instrument and then listened to four different melodies each with different instruments, where one of which used the same instrument as the first melody. Participants were asked to determine the stimulus with the same instrument as the first from four melodies presented later.

A. Sound stimuli

The sound sources used in the experiments were from a software sampler (Native Instruments' Kontakt, and the libraries GRANDEUR, FACTORY LIBRARY, and CONCERT VIBRAPHONE.) We created melodic stimuli using these software samplers as plug-ins in Studio One, a DAW software program developed by Presonus. Sixty songs were selected from the Billboard JAPAN HOT 100 2021 First Half Overall Song Chart, and four bars of the chorus of these songs were extracted to create music. The volume levels were normalized to -17.8 LUFS. All melodic stimuli were created on the same scale. Each sound stimulus was presented for approximately 6-11 s.

TABLE II ISID-instrument sound identification difficulty

	Difficulty level	Combination of instruments				Annotation
PART1	1	String	Wind	Percussion	Keyboard	
	2	String	Percussion	Wind	Wind	Woodwind & Brass
	3	String	Percussion	Wind	Percussion	Different Subcategories
	4	String	String	Wind	Percussion	Same Subcategories
	5	String	String	Woodwind	Brass	Different Subcategories
PART2	6	String	String	Woodwind	Brass	Same Subcategories
	7	String	String	Percussion	Percussion	Different Subcategories, Except Wind
	8	String	String	Percussion	Percussion	Same Subcategories, Except Wind
	9	String	String	Woodwind	Woodwind	Different Subcategories, Woodwind or Brass
	10	String	String	Woodwind	Woodwind	Same Subcategories, Woodwind or Brass
PART3	11	Woodwind	Woodwind	Brass	Brass	Different Subcategories
	12	Woodwind	Woodwind	Brass	Brass	Same Subcategories
	13	Plucking	Plucking	Bowed	Bowed	
	14	Single	Single	Double	Air	
	15	Single	Single	Double	Double	

B. Participants

All participants were hearing-impaired with moderate to severe hearing loss. To reduce the impact of age-related hearing loss, we recruited only adults in their 20s. The study included 20 participants (11 men and 9 women). The average age of the participants was 21 years (range: 20–24 years). The average hearing level was 61-130 dBHL; one with conductive hearing loss, and 19 with sensorineural hearing loss; 8 with HAs in both ears, two with bimodal HAs, six with bimodal CIs, two with neither hearing aid nor CI, and two with a bimodal CI and HA. The duration of implant use in CIs ranged from 15 to 22 years.

C. Environment

The experiment was conducted online owing to COVID-19 prevention policies, either from the participant's own rooms or from a university laboratory. The background noise level at the university laboratory was maintained at approximately 38.3 dBSPL. Participants used various playback devices such as external speakers, headphones, earphones, external inputs for HAs, and even the terminal's built-in speaker when listening to the sound stimuli. These are devices that each participant in the experiment normally used when listening to music. The experiments were conducted using the music memory application developed by the first author (Figure 1). We installed this application on an iPod Touch or iPhone and conducted experiments consisting of four main phases: (1) Volume adjustment, (2) Practice program, (3) Main program, and (4) Post-program questionnaire. In Phase (1), the volume was adjusted using sound stimuli that were not used in the main program. The volume was adjusted to the level at which the experimenter felt the most comfortable. In (2), a practice session was conducted to familiarize participants with the experimental procedure. Similar to (1), stimuli not used in the main program were presented, and a sound identification task was performed. Phase (3) consists of three difficulty levels for the stimuli combinations, as shown in Table II. Ten questions were asked at each level, resulting in a total of 30 questions. Each level and all questions within each level were presented randomly. In (4), we asked questions about the difficulty level of the experiment in a questionnaire format. Moreover, (3) consists of two tasks: timbre identification (Figure 1, left) and evaluation (Figure 1, right). In the timbre identification task, participants listened to a target melody and four other melodies and then selected a melody with the same instrument as the target melody from the four choices.

Fig. 1 Music memory application. The right side is for the identification task; the top orange button plays the target melody, and the remaining orange buttons play the choice melodies. The green buttons are used to select the

answer. The left side is for the evaluation, with questions asking about the overall difficulty, overall confidence, and ease of understanding the target melody and choice melodies on a 0-100 scale.

Participants were allowed to press the play button as many times as they wished for all melodies. There was no time limit for answering each question, but a 5-minute break was taken every 20 min. To prevent false taps during the experiment, a confirmation screen was displayed once before transitioning to the evaluation task. In the evaluation task, four questions were presented, which were rated on a 5-point scale from 0 to 100. The questions were "How difficult is this question," "Are you confident in your answer," "Is the melody of the question easy to understand," and "Is the melody of the selection sound easy to understand?"

V. Result

A. Correct Responses

We calculated the rate of correct responses by level based on the rate of correct responses obtained in the experiment. Figure 2 illustrates the results. The median value decreases as the level increases. We tested for normality and equivariance to examine significant differences among the three groups. The Shapiro-Wilk test was used to test for normal distribution, and the Levene test was used to test for equivariance. Assuming a significance level of 5%, the results of the Shapiro-Wilk test are not normally distributed because the p-values are below the significance level for all groups. The results of the Levene's test show that the three groups are not equally distributed because the p-value is below the significance level. These results indicate that the rates of correct responses by level are not normally or equally distributed. Therefore, we conducted a Kruskal-Wallis test, which is a nonparametric test with more than two groups, without these assumptions. The results of the Kruskal-Wallis test show a significant difference among the ISID, as the p-value is less than the significance level of .001 (p<.05). We conducted Dunn's test as a post-test, and the results are shown in Table III. The p-values represent those obtained after Bonferroni correction.

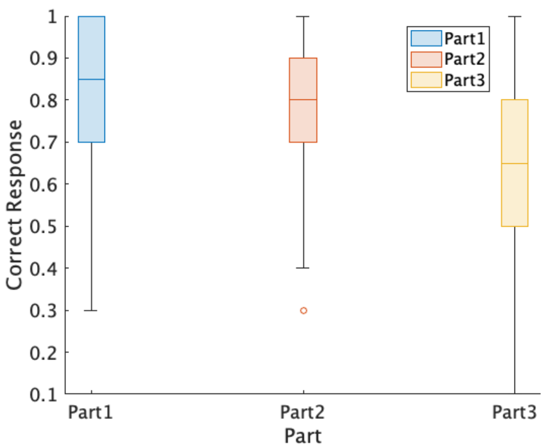

Fig. 2 Accuracy of correct responses by the parts of the ISID

TABLE III Dunn's test for rate of correct answers by the ISID

comparison	p	p_{bonf}	p_{holm}
PART1-2	.179	.537	.179
PART1-3	.001	$< .001^{**}$	.001
PART2-3	.002	$.006^{*}$	.004
	$^{*}p < .05$	$^{**}p < .001$	

The results of Dunn's test indicate a significant difference between the two groups of the ISID, as the p-values between PART 1 and PART 3 groups were below the significance level of <.001 (p<.05) and between the PART 2 and PART 3 groups at .006 (p<.05). Conversely, the p-value of .537 (p>.05) for PART 1-PART 2 shows a lack of significant difference.

B. Response time

Response time refers to the time between displaying the timbre identification task screen to the transition to the evaluation task screen. Figure 3 shows the results. We compared the response times by level based on the response times obtained in the experiment: the longer the reaction time, the more difficult the task, as the participants pressed the timbre button multiple times or had difficulty answering questions. Response time varied among participants in each experiment. Therefore, the response time used for the comparison was the median normalized value of the reaction time for each participant. We tested response time and accuracy using the Shapiro-Wilk and Levene tests. Both tests show significant differences with no evidence of normal distribution or equivariance. Therefore, we selected the Kruskal-Wallis test as a nonparametric test for more than two groups without these assumptions. The results of the Kruskal-Wallis test revealed a significant difference among the ISID, with a p-value of <.001. Table IV presents the results of Dunn's test. Comparing the Bonferroni-corrected p-values from the results of Dunn's test, the p-values of .006 and <.001 were below the significance level for the PART 1-PART 2 and PART 1-PART 3 groups, respectively.

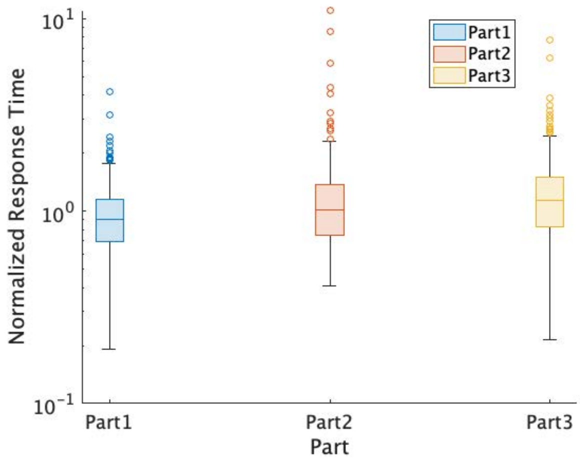

Fig. 3 Normalized response time by the parts of the ISID

TABLE IV Dunn's test for response time rate by the ISID

comparison	p	p_{bonf}	p_{holm}
PART1-2	$.002$	$.006^{*}$	$.004$
PART1-3	$< .001$	$< .001^{**}$	$< .001$
PART2-3	$.040$	$.120$	$.040$
	$^{*}p < .05$ $^{**}p < .001$		

Therefore, the difference between the PART 1-PART 2 and PART 1-PART 3 groups are significant. Conversely, no significant difference was observed between the PART 2 and PART 3 groups (p=.120).

C. *Combination of instruments*

Table IV shows the rate of correct answers by instrument. We calculated the rate of correct answers by instrument as the number of correct responses/number of stimuli presented. Table IV shows that the instrument with the highest rate of correct answers was the marimba, and the instrument with the lowest rate of correct answers is the bassoon. We calculated combinations of incorrect instruments for each instrument from the correct response rate data. Figure 4 shows the results; the vertical axis represents the presented instruments, and the horizontal axis represents the instruments answered. The diagonal line indicates the number of correct responses. Table IV lists the calculated correct response rates for the diagonals by instrument. Figure 4 shows that the most likely instrument combinations to be mistaken were violin-cello, harp-cello, and harp-vibraphone. The bassoon, which had the lowest rate of correct answers, was frequently mistaken for clarinet and oboe, which belong to the same woodwind category.

VI. Discussion

The average rate of questions answered correctly at all ISID levels was high, indicating that the questions were easy for the scale used in training. Nevertheless, we observed differences in difficulty among the three levels of ISID based on the rate of correct answers and response times. In addition, we determined that PART 2 was more difficult than PART 1, as participants repeatedly pressed the button to relisten to the target melody.

TABLE V Correct response of instrument sounds

Instrument	Correct response
Flute	0.8
Recorder	0.75
Clarinet	0.85
Alto Sax	0.825
Oboe	0.725
Bassoon	0.55
Trumpet	0.75
Trombone	0.7
Violin	0.7
Cello	0.7
Acoustic Guitar	0.7
Harp	0.7
Piano	0.775
Organ	0.75
Marimba	0.975
Vibraphone	0.65

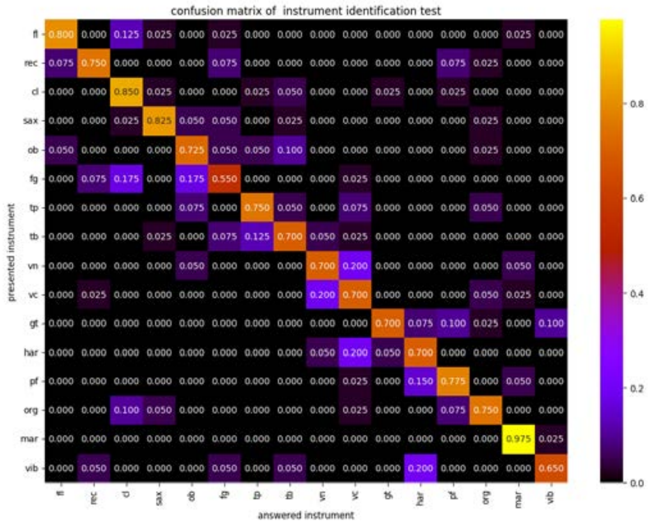

Fig. 4 Confusion matrix of instrument identification test

However, we found PART 2 was a less error-prone task. Comparing the rate of correct responses by instrument, the marimba had the highest rate of correct answers, whereas the bassoon had the lowest. Bassoons had the lowest rate of correct responses because participants were confused with oboes, which belong to the same double-reed instrument category, and clarinets, which belong to the same woodwind instrument category.

The finding that percussion instruments are easier to discriminate than other instruments is supported by results obtained by McDermott [4-5]. However, the rate of correct answers for vibraphone was low, even for percussive instruments. We consider this to be because the time envelope of the vibraphone exhibits a long release time, making it easy to mistake it for a harp.

In this study, we experimented with the synthesized sounds of various melody patterns applied to various instruments. We used melodies as we assumed a general listening environment for a musical timbre. However, depending on the problem, the melody may have disturbed the timbres. To confirm this, examining the influence of melodies on timbre is necessary.

VII. Conclusion

We conclude that the set of instrumental sounds proposed in this experimental design may be too easy to use in future music training. However, our results on correct responses and response times prove the hypothesis that instruments with the same acoustical mechanism are similar in timbre and thus difficult to identify. Future work must increase the number of instruments and create a stronger contrast between the three parts by increasing the number of stimulus options presented by more than four.

Acknowledgment

This work was supported by JSPS KAKENHI (grant numbers JP21H00884 and JP21H03759). We thank Daichi Moriyama for providing the sound sources. We also thank

Kenta Wakasa for his advice on developing the Music Memory application.

References

[1] Hiroko Terasawa and Rumi Hiraga. Support for people with disabilities by music information processing: 2. Hearing-impaired people enjoying music - Let's enjoy it more with information processing technology. IPSJ Magazine, p. 254-257, 2016, (Japanese)

[2] MEXT, Music Instructional Resources for Children with Hearing Impaired, 2021-3, https://www.mext.go.jp/content/20210317-mxt_tokubetu01-100002897-3.pdf, (reference date: 2021-11-17), (Japanese)

[3] Nina Kraus and Bharath Chandrasekaran. 2010. Music training for the development of auditory skills. Nat Rev Neurosc, 11, 8 (Aug 2010), 599-605. DOI: https://doi.org/10.1038/nrn2882

[4] Hugh J. McDermott. 2004. Music Perception with Cochlear Implants: A Review. Trends Amplif, 8, 2, 49-82. DOI: https://doi.org/10.1177/108471380400800203

[5] Hugh J. McDermott and Valerie Looi. 2004. Perception of complex signals including musical sounds, with cochlear implants. Int Congr Ser, 1273 (Nov 2004), 201-204. DOI: https://doi.org/10.1016/j.ics.2004.08.034

[6] Grace L. Nimmons, Robert S. Kang, Ward R. Drennan, Jeeff Longnion, Chad Ruffin, Tina Worman, Bevan Yueh, and Jay T. Rubinstein. 2008. Clinical Assessment of Music Perception in Cochlear, Implant Listeners, Otol Neurotol, 29, 3 (Feb 2008), 149-155. DOI: https://doi.org/10.1097/mao.0b013e31812f7244

[7] Valerie Looi, Hugh J. McDermott, Colette McKay, and Louise Hickson. 2008. Music Perception of Cochlear Implant Users Compared with that of Hearing Aid Users. Ear Hear, 29, 3 (Jun 2008), 421-434. DOI: https://doi.org/10.1097/AUD.0b013e31816a0d0b

[8] Ward R. Drennan, Jacob J. Oleson, Kate Gfeller, Jillian Crosson, Virginia D.Driscoll, Jong Ho Won, Elizabeth S. Anderson, and Jay T. Rubinstein. 2015. Clinical evaluation of music perception, appraisal and experience in cochlear implant users. Int J Audiol, 54, 2 (Feb 2015), 114-123. DOI: https://doi.org/10.3109/14992027.2014.948219

[9] Robert Kang, Grace L. Nimmons, Ward Drennan, Jeff Longnion, Chad Ruffin, Kaibao Nie, Jong H. Won, Tina Worman, Bevan Yueh, and Jay Rubinstein. 2009. Development and Validation of the University of Washington Clinical Assessment of Music Perception Test. Ear Hear, 30, 4 (Aug 2009), 411-418. DOI: https://doi.org/10.1097/AUD.0b013e3181a61bc0

[10] John J. Galvin 3rd, Qian-Jie Fu, and Sandra Oba. 2008. Effect of instrument timbre on melodic contour identification by cochlear implant users. J Acoust Soc Am, 124, 4 (Oct 2008) 189-195. DOI: https://doi.org/10.1121/1.2961171

[11] Oliver Macherey and Alexia Delpierre. 2013. Perception of Musical Timbre by Cochlear Implant Listeners: A Multidimensional Scaling Study. Ear Hear, 34, 4 (Jul-Aug 2013), 426-436. DOI: https://doi.org/10.1097/AUD.0b013e31827535f8

[12] Kate Gfeller, Aaron Christ, John Knutson, Shelley Witt, Kathleen T. Murray, and Richard S. Tyler. 2000. Musical backgrounds, listening habits and aesthetic enjoyment of adult cochlear implant recipients. J Am Acad Audiol, 11, 7 (Jul-Aug 2000), 390–406. DOI: https://doi.org/ 10.1055/s-0042-1748126

[13] Nicole T. Jiam, Mickael L. Deroche, Patpong Jiradejvong, and Charles J. Limb. 2019. A Randomized Controlled Crossover Study of the Impact of Online Music Training on Pitch and Timbre Perception in Cochlear Implant Users, J Assoc Res Otolaryngol, 20, 3 (Jun 2019), 247-262. DOI: https://doi.org/10.1007/s10162-018-00704-0

[14] Virginia D. Driscoll. 2012. The Effects of Training on Recognition of Musical Instruments by Adults with Cochlear Implants. Semin Hear, 33, 4 (Nov 2012), 410-418. DOI: https://doi.org/10.1055/s-0032-1329230

[15] Virginia D.Driscoll, Jacob Oleson, Dingfeng Jiang, and Kate Gfeller. 2009. Effects of Training on Recognition of Musical Instruments Presented through Cochlear Implant Simulations. J Am Acad Audiol, 20, 1 (Jan 2009), 71-82. DOI: https://doi.org/10.3766/jaaa.20.1.7

[16] Bjorn Peterson, Malene Vejby Mortensen, Mads Hansen, and Peter Vuust. 2012. Singing in the Key of Life: A Study on Effects of Musical Ear Training After Cochlear Implantation. Psychomusicology Music Mind Brain 22, 2 (Dec 2012), 134-151. DOI: https://doi.org/10.1037/a0031140

[17] Erich M. von Hornbostel, Curt Sachs, Classification of Musical Instruments: Translated from the Original German by Anthony Baines and Klaus P. Wachsmann, The Galpin Society Journal, 14, (Mar 1961), 3-29. DOI: https://doi.org/10.2307/842168

[18] Toshio Hayasaka. 1992. The Science of Musical Instruments, Journal of the Institute of Electronics, Information and Communication Engineers. (Japanese)