Influence of musical training on perception of L2 speech

Makiko Sadakata [1,2], *Lotte van der Zanden* [3] *and Kaoru Sekiyama* [4]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2] Department of Artificial Intelligence, University of Nijmegen, The Netherlands
[3] Department of Psychology, University of Nijmegen, The Netherlands
[4] Division of Cognitive Psychology, Kumamoto University, Japan

makiko.sadakata@mpi.nl, lottevanderzanden@student.ru.nl, sekiyama@kumamoto-u.ac.jp

Abstract

The current study reports specific cases in which a positive transfer of perceptual ability from the music domain to the language domain occurs. We tested whether musical training enhances discrimination and identification performance of L2 speech sounds (timing features, nasal consonants and vowels). Native Dutch and Japanese speakers with different musical training experience, matched for their estimated verbal IQ, participated in the experiments. Results indicated that musical training strongly increases one's ability to perceive timing information in speech signals. We also found a benefit of musical training on discrimination performance for a subset of the tested vowel contrasts.

Index Terms: discrimination, identification, musical training, category, learning, perception

1. Introduction

The connection between the various cognitive mechanisms responsible for processing language and music information has long attracted attention [1]. Although the traditional approach comparing cognition of language and music points to similarities between them [2], recent approaches have focused more on the transfer of ability from one domain to the other. For example, musical training has been shown to increase verbal memory [3] and IQ [4]. At a more fundamental level, musical training also seems to enhance the ability to perceive acoustical features included in linguistic signals [5-8]. The current study investigates whether musicians' superior ability to perceive linguistic signals in L2 sounds extends to timing patterns of Japanese as well as more complex linguistic features of Japanese and Dutch in which multiple elements contribute to its' realization. To this end, we carried out a series of perceptual experiments with native speakers of Dutch and Japanese having different amounts of musical training.

We tested two separate processes that are involved in the perception of speech sound categories. The first is the acoustical analysis that occurs at a very early stage of perception. This process makes it possible to *hear* differences between categories, and could be accomplished purely based on acoustical cues: one does not necessarily need to have categories for the presented sounds. Previous studies have shown that musical training enhances this process in perception of pitch patterns and metric structure in speech [5-8]. Therefore, we think that this may very well hold for processing of other aspects of speech signals. We used a discrimination test to address this issue along with synthesized speech of minimal pair continuums. A positive transfer from musical training would result in musicians demonstrating enhanced discrimination of these differences as compared to non-musicians.

The second process requires one's ability to establish internal representation of new categories and to further apply this knowledge to a presented sound: this makes it possible to *identify* categories. Obviously, this is an essential step in learning to perceive L2 speech, and is far more complex than the mere comparison of acoustical signals. However, it is not known whether musical training has a beneficial effect here. We used an identification experiment combined with a learning procedure to address this issue. Unlike the discrimination test, this test employed natural utterances of multiple words spoken by multiple speakers.

1.1. Japanese contrasts

Japanese has a distinctive syllable structure, mora [9, 10]. The previous research revealed that Japanese native speakers response differently when they are presented with moraic and non-moraic nasals, while no such difference is evident in responses by non-native speaker of Japanese [11, 12]. This difference in sensitivity to moraic structure may be able to explain why non-Japanese natives often fail to distinguish contrast such as *ko-n-ya-ku* /koN.jaku or ko~jaku/ (engagement) and *ko-n-nya-ku* /kon.njaku/ (konjac). The first one is pronounced either without alveolar closure or without any closure plus a proceeding nasalized vowel, while in the later case the first /n/ is a simple alveolar and the second /n/ is an onset nasal (Makiko Aoyagi, personal contact). When perceiver is not sensitive to the moraic structure, it may be difficult to detect difference what follows after the moraic nasals (nasalized vowel /ja/ or onset nasal /n/). Because Dutch is not based on the moraic structure, we expect this contrast to be difficult for Dutch natives to perceive.

Another contrast included in the study was with or without a mora obstruent, that is "long voiceless obstruents as consisting of a phoneme that [one] transcribe[s] as /Q/ followed by a voiceless stop, affricate, or fricate [10] (P.40) ". We used a contrast with and without stop geminates, such as *ha-ka-ku* /hakaku/ (bargain) and *ha-kka-ku* /haQkaku/ (disclosure). The duration of the gap between the first /a/ and the first /k/ is the main determinant of this particular contrast (but also see [13]). This contrast is ideal for studying perception of linguistic timing when tested on individuals whose mother tongue does not use this property as a crucial factor, such as Dutch.

1.2. Dutch contrasts

The vowel qualities are mainly determined by a combination of multiple peaks in the vocal spectra (formants). The Dutch vowel inventory is large (more than 16) [14], while Japanese has only 5 vowels [10]. Obviously, many Dutch vowel categories do not completely overlap with Japanese vowels. Therefore, assimilation of categories, that is, Dutch vowel categories being jointly mapped onto a single Japanese vowel category, is likely to take place when Japanese natives perceive Dutch vowels that do not exist in the Japanese vowel inventory. The Dutch vowel *u* /ʏ/, a near-close near-front protruded vowel, may be one of these examples. The articulation of this vowel is somewhere in between *e* /ɛ/ and *oe* /u/, which are similar to Japanese vowels *e* /e/ and *u* /u/, respectively. We expect that Japanese natives will assimilate their perception of *u* /ʏ/ to either Japanese categories *e* /e/ or *u* /u/, making it difficult for them to distinguish between the Dutch vowel contrasts *u* /ʏ/ and *e* /ɛ/, as well as between *u* /ʏ/ and *oe* /u/.

2. Methods

2.1. Participants

Fifty-three native Dutch and 38 native Japanese speakers took part in the study. For both language groups, the musician group included individuals who had followed more than 5 years of formal musical training, and who were still actively playing musical instruments at the time of the experiment (NL: 5 males and 21 females, average age: 21.5 years, average musical training: 8.4 years, JP: 7 males and 12 females, average age: 21.3 years, average musical training: 12.3 years). The non-musician group included individuals who completed less than 3 years of musical training and had followed training other than music (mostly sports) for at least 5 years (NL: 6 males and 21 females, average age: 21.1 years, average training other than music: 9.6 years, JP: 12 males and 7 females, average age: 19.4 years, average training other than music: 9.1 years).

The NLV (Nederlandse Leestest voor Volwassenen [15]) and JART (Japanese Adult Reading Test [16]) were used for calculating estimation of the verbal IQ for Dutch and Japanese participants, respectively. Comparisons of NLV IQ scores and JART IQ scores revealed no significant difference between musician and non-musicians for both language groups (NL: $t_{51} = 0.393$, n.s., JP: $t_{34} = -1.76$, n.s.)

2.2. Stimuli

2.2.1. *Discrimination test*

One minimal pair per contrast was selected (see Table 1). Recordings were first low-pass filtered at 5000 Hz and average sound levels were normalized to 70 dB using Praat [17]. The Japanese nasal continuum and two Dutch vowel continuums were created in the same manner using TANDEM-STREIGHT [18]. First, target time windows which include sounds of /N.ja/-/nnja/, /ʏ/-/ɛ/, /ʏ/-/u/ were specified. For each pair, spectrum, frequency, aperiodicity and time information within the target time window were morphed in equal 19 steps. These morphed parts were embedded in context information (sound outside the target time window) of one member of the pair (konyaku, put, toet). F0 was kept constant. Amongst the 19 steps, two end points with 7 equal-distanced steps were selected after a pilot experiment. Selected steps were 1, 3, 5, 7, 9, 11, 13 for nasal and u-oe contrasts and 4, 6, 8, 10, 12, 14, 16 for u-e contrast, accordingly. The Japanese stop continuum (/hakaku/ - /haQkaku/) was made by changing the duration of the gap between the offset of the first /a/ and the onset of the first /k/ from 92 ms to 182 ms by 15 ms equal steps using Praat [17].

Table 1. *List of minimal pairs of four contrasts, used for the discrimination test.*

Japanese		**Dutch**	
nasal	stop	u-e	u-oe
konyaku / konnyaku	hakaku / hakkaku	put / pet	tut / toet

The experiment employed a two-alternative forced choice (2AFC) task, in which participants had to judge whether the two sounds (sound A and B) were identical or not. The duration between the offset of sound A and the onset of sound B was fixed to 1500 ms. Position of the response buttons (same or different) were altered for even and odd participants.

2.2.2. *Identification test*

For each contrast, 4 native speakers (2 males, 2 females) recorded eight minimal pairs, which resulted in 32 pairs.

The test started with a learning task, which listeners had to remember two categories (e.g., with or without geminate) that were presented together with visual number *1* and *2*. Combination of categories and numbers was altered for even and odd participants. The duration between presentation of words were fixed to 2000 ms. Three minimal pairs were randomly chosen from the 32 pairs for this learning task. Participants could listen to the set multiple times until they feel confident about their understanding of categories. Furthermore we provided them with a piece of paper and a pen so that they could write down their perception of characteristic features of categories. During the identification task, a word (audio) was presented and participants pressed the button labeled 1 or 2 to indicate their identification response (to which category the presented sound belongs).

2.3. Procedure

Participants started the session with the identification test, followed by questionnaires, the IQ estimation test (NLV or JART) and the discrimination test. The whole procedure took about 80 minutes to complete. Experiments took place in sound attenuated rooms at Radboud University Nijmegen (The Netherlands) and at Kumamoto University (Japan).

2.4. Apparatus

An identical setup was used for experimentation in the Netherlands and in Japan. A DELL notebook computer with an IntellCoreDuo processer (4 GB RAM) and a Sound Blaster X-Fi sound card was used. The stimuli were presented through the Sony MDR-7506 headphones on a 15.4-inch TFT screen. Average sound pressure level (SPL) of the headphones was adjusted to around 68 dB. The responses were recorded using an USB game controller (Sanwa supply, JY-P68US). The application Presentation (version 14.3 Neurobehavioral Systems) was used for presenting instructions and stimuli, and for collecting responses. PASWStatistics18.0 (SAS) was used for the data analyses.

3. Results

Responses with reaction time longer than 5000 ms were identified as potential outliers. Consequently, 123 data points (0.6%) from the discrimination test and 191 data points (0.8%) from the identification test were discarded from the analyses.

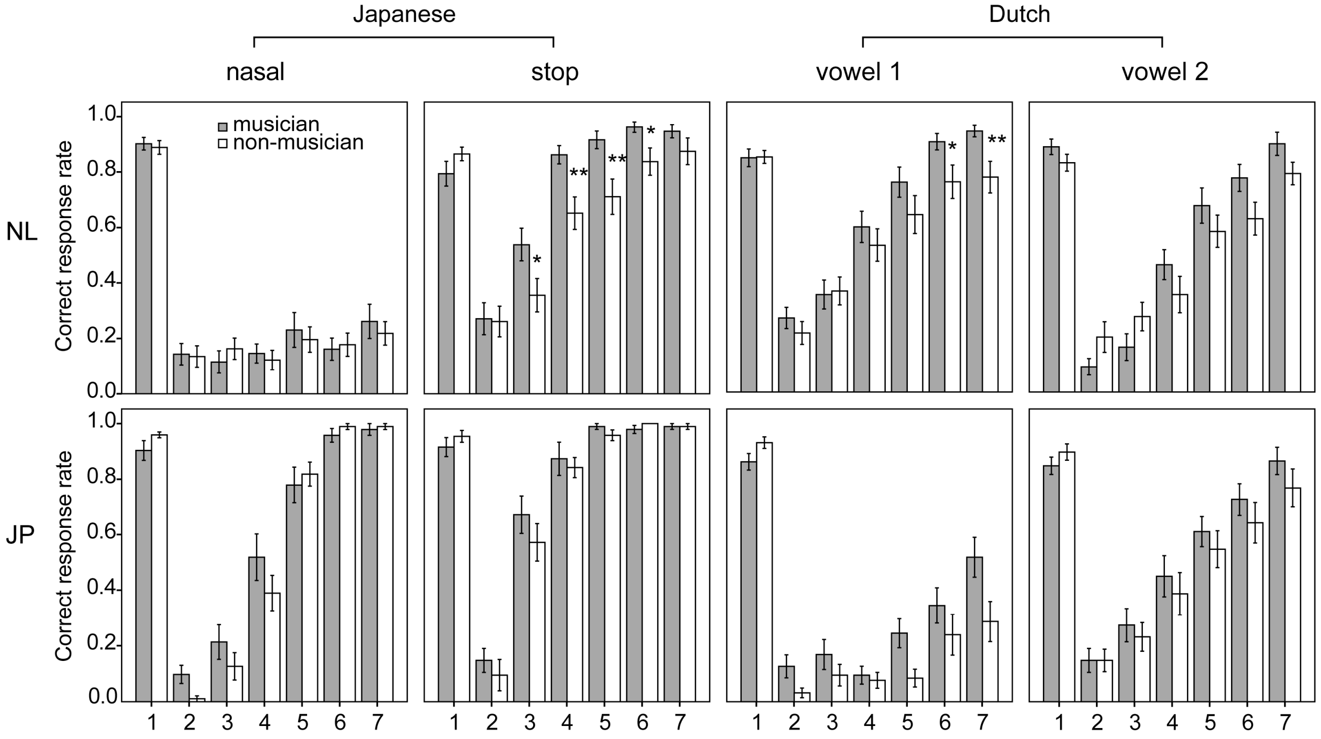

Figure 1: *Mean correct response rate and standard error of discrimination tests. One on the x-axis corresponds to the trial where sound A and B were identical, where 2 to 7 on the x-axis correspond to the trial where sound A and B were different.*

3.1. Discrimination test

Figure 1 compares correct response rate as a function of the morphing steps (1 – 7) for two groups with different levels of musical training (musician and non-musicians, respectively). One on the x axis represents results where sound A was identical to sound B while 2 to 7 on the x axis represent the results where sound A was different from sound B (the larger the number, the more different). Eight panels show the results for four contrasts (nasal, stop, and two Dutch vowels) and two groups (NL and JP). Two three-way repeated measure ANOVAs were carried out to analyze the Dutch and Japanese data separately.

The analysis of Dutch data indicated a significant main effect of musical training, morphing step and contrast ($F_{1,51}$ = 5.501, p = .023; $F_{6,306}$ = 215.369, ε_{GG} = 0.664, p < .000; $F_{3,153}$ = 92.803, p < .000; respectively). The three-way interaction among these factors was also significant ($F_{18,918}$ = 1.882, ε_{GG} = 0.670, p = .033). Significant simple main effects of musical training at different morphing steps are indicated in Figure 1 (*p < .05, **p <.001, Bonferroni adjusted). The analysis thus revealed that the performance of Dutch musicians was better than that of non-musicians for some instances of the stop contrast, as well as of the u-e contrast.

The analysis of Japanese data indicated a significant main effect of musical training, morphing step and contrast ($F_{1,36}$ = 4.147, p = .049; $F_{6,216}$ = 360.173, ε_{GG} = 0.541, p < .000; $F_{3,108}$ = 121.093, ε_{GG} = 0.727, p < .000; respectively). Although a significant interaction was found between morphing step and contrast ($F_{18,648}$ = 31.511, ε_{GG} = 0.506, p < .000), no other significant interactions were found. This means that, in general, Japanese musicians outperformed non-musicians for this test. Because no interaction reached a significant level, we did not look into simple main effects of group at different morphing steps.

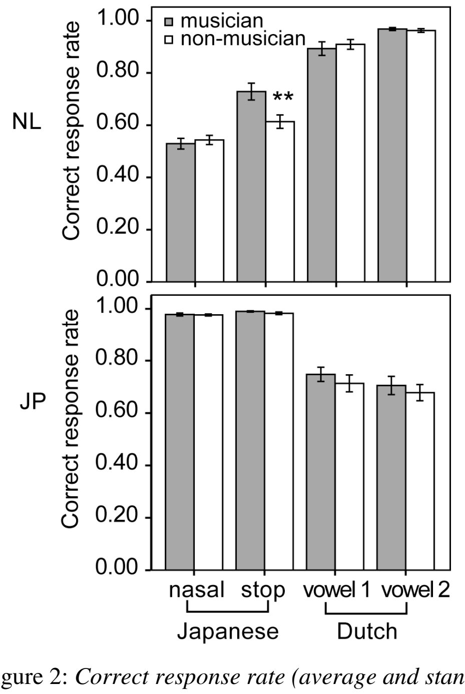

Figure 2: *Correct response rate (average and standard error) of identification test.*

3.2. Identification test

Figure 2 illustrates correct response rate for two groups per contrast. Separate two-way repeated measure ANOVAs were carried out to analyze the data of Dutch and Japanese groups. The analysis of Dutch data indicated no main effect of musical training but a significant main effect of contrast and an interaction between contrast and musical training ($F_{1,51}$ =

2.012, n.s.; $F_{3,153}$ = 193.482, ε_{GG} = 0.701, p < .000; $F_{3,153}$ = 4.168, p < .011; respectively). There was a strong simple main effect of musical training at the stop contrast condition ($F_{1,51}$ = 7.811, p < .007, Bonferroni adjusted), indicating that Dutch musicians were better at identifying stop geminates as compared to non-musicians. Further analyses revealed that Dutch listeners performed identification much better on Dutch contrasts than Japanese ones (pairwise comparisons, p < .000, Boneferroni adjusted).

The analysis of Japanese data indicated a strong main effect of contrast but no main effect of musical training and interaction between two factors ($F_{3,108}$ = 97.918, ε_{GG} = 0.621, p < .000; $F_{3,108}$ = 0.262, n.s.; $F_{1,36}$ = 1.277, n.s.; respectively). Thus, the Japanese musicians did not show any superior performance for this test. Multiple pairwise comparisons revealed that Japanese listeners' correct response rates were significantly higher for Japanese contrasts as compared to Dutch ones (p <.000, Bonferroni adjusted).

4. Discussion

The current study investigated the influence of musical training on perception of linguistic sounds by means of discrimination and identification tests. From previous studies which have demonstrated advantageous effects of musical training on perceiving pitch patterns [7-9], we predicted that musicians would also perform better at discriminating other aspects of linguistic sounds. The results confirmed this hypothesis to a certain extent. Dutch musicians detected small durational difference of a gap (30 – 75 ms) in Japanese word better as compared to non-musicians. Furthermore, Japanese musicians demonstrated overall superior performance in the discrimination task. These data strongly support the benefit of musical training for discriminating L2 sounds, not only with respect to pitch information but also to timing and more complex features of speech sounds such as vowel spectra. Moreover, the overall superior performance of Japanese musicians points to the specific case where musical training even influences L1 sound discrimination performance.

As compared to the discrimination test, the identification test required more elaborate cognitive operations, such as abstraction of category features underlying multiple examples realized in different voices, as well as using these features to analyze incoming sounds. If musicians were only better at hearing differences in speech signals, we would not see any effect of musical training for this test. Data from the Japanese participants may confirm this hypothesis, with regards to the perception of vowel qualities, as the positive effect of musical training on discrimination performances of L2 disappeared for the identification test. However, the fact that musicians perceive subtle differences more accurately may be an advantage when they are given the chance to learn these materials over longer periods of time. A study with longitudinal observations is needed for this issue.

Data from the Dutch participants showed a strong effect of musical training on both tests for Japanese stop geminates contrast while there was no effect for Japanese nasals, indicating that influence that musical training can have is not equal for perception of different aspects of speech signals. It is interesting that the most pronounced positive transfer effect was found for perception of speech timing among all the other contrasts. This may be due to the fact that timing information is crucial in both domains while the quality of vowels or nasal consonants is something more specific to the speech domain.

5. Acknowledgements

This research was supported by a Grant-in-Aid for Scientific Research (B) 21330169, the project leader Shuji Mori (Japan Society for the Promotion of Science). Authors are grateful for useful comments by James McQueen, Makiko Aoyagi, and Alex Brandmeyer, technical support by Pascal de Water and the assistance of Naomi Nakamura, Tomoko Takemura.

6. References

[1] A.D. Patel, *Music, language and the brain*, Oxford university press, 2008.

[2] F. Lerdahl, and R. Jackendoff, *A generative theory of tonal music*, MIT Press, 1983.

[3] A.S. Chan, Y.C. Ho, and M.C. Cheung, "Music training improves verbal memory," *Nature*, vol. 396, no. 6707, 1998, pp. 128-128.

[4] E.G. Schellenberg, "Music lessons enhance IQ," *Psychological Science*, vol. 15, no. 8, 2004, pp. 511-514.

[5] P.C.M. Wong, E. Skoe, N.M. Russo, T. Dees, and N. Kraus, "Musical experience shapes human brainstem encoding of linguistic pitch patterns," *Nature Neuroscience*, vol. 10, no. 4, 2007, pp. 420-422.

[6] M. Besson, D. Schön, S. Moreno, A. Santos, and C. Magne, "Influence of musical expertise and musical training on pitch processing in music and language.," *Restorative Neurology and Neuroscience*, vol. 25, 2007, pp. 399-410.

[7] C. Marques, S. Moreno, S.L. Castro, and M. Besson, "Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence," *Journal of Cognitive Neuroscience*, vol. 19, no. 9, 2007, pp. 1453-1463.

[8] C. Magne, C.Astesano, M. Aramaki, S. Ystad, R. Kronland-Martinet & M. Besson, "Influence of syllabic lengthening on semantic processing in spoken French: Behavioral and electrophysiological evidence," *Cerebral Cortex*, vol.11, no.7, 2007, pp. 2659 - 2668.

[9] A. Cutler, "The comparative perspective on spoken-language processing," *Speech communication*, vol. 21, no. 1-2, 1997, pp. 3-15.

[10] T.J. Vance, *An introduction to Japanese phonology*, State university of New York Press, 1987.

[11] T. Otake, Yoneyama, K., Cutler, A., van der Lugt, A., "The representation of Japanese moraic nasals," *Journal of the Acoustical Society of America*, vol. 100, no. 6, 1996, pp. 3831-3842.

[12] A. Cutler, Otake, T., "Mora or phoneme? Further evidence for language-specific listening," *Journal of Memory and Language*, vol. 33, 1994, pp. 824-844.

[13] J. Kingston, S. Kawahara, D. Chambless, D. Mash, and E. Brenner-Alsop, "Contextual effects on the perception of duration," *Journal of Phonetics*, vol. 37, no. 3, 2009, pp. 297-320.

[14] C. Gussenhoven, "Dutch," *Journal of the International Phonetic Association*, vol. 22, no. 1-2, 1992, pp. 45-47.

[15] B. Schmand, Lindeboom, J., Harskamp, F. van. , *NLV: Nederlandse leestest voor volwassenen: handleiding (NLV)*, Swets & Zeitlinger, 1992.

[16] Y. Matsuoka, Uno, M., Kasai, K., Koyama, K., Kim, Y., "Estimation of premorbid IQ in individuals with Alzheimer's disease using Japanese ideographic script (Kanji) compound words: Japanese version of National Adult Reading Test," *Psychiatry and Clinical Neurosciences*, vol. 60, 2006, pp. 332-339.

[17] P. Boersma, Weenink, D., "Praat, a system for doing phonetics by computer" *Grot International*, vol. 5, no. 9/10, 2001, pp. 341-345.

[18] H. Kawahara, Morise, M., Takahashi, T., Nisimura, R., Irino, T., Banno, H., "TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0 and aperiodicity estimation," *ICASSP'2008*, pp. 3933-3936.