

Enhanced perception of various linguistic features by musicians: A cross-linguistic study

Makiko Sadakata^{a,c,*}, Kaoru Sekiyama^{b,2}

^a Max Planck Institute for Psycholinguistics, P.O. Box 310 6500 AH Nijmegen, The Netherlands

^b Division of Cognitive Psychology, Kumamoto University, 2-40-1 Kurokami, 860-8555, Japan

^c Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, P.O. Box 9104 6500 HE Nijmegen, The Netherlands

ARTICLE INFO

Article history:

Received 3 November 2010

Received in revised form 11 March 2011

Accepted 12 March 2011

Available online 2 July 2011

PsycINFO classification:

2326

2720

2340

Keywords:

Perception

Musical training

Speech

Geminate

Vowel

Nasal

ABSTRACT

Two cross-linguistic experiments comparing musicians and non-musicians were performed in order to examine whether musicians have enhanced perception of specific acoustical features of speech in a second language (L2). These discrimination and identification experiments examined the perception of various speech features; namely, the timing and quality of Japanese consonants, and the quality of Dutch vowels. We found that musical experience was more strongly associated with discrimination performance rather than identification performance. The enhanced perception was observed not only with respect to L2, but also L1. It was most pronounced when tested with Japanese consonant timing. These findings suggest the following: 1) musicians exhibit enhanced early acoustical analysis of speech, 2) musical training does not equally enhance the perception of all acoustic features automatically, and 3) musicians may enjoy an advantage in the perception of acoustical features that are important in both language and music, such as pitch and timing.

© 2011 Published by Elsevier B.V.

1. Introduction

In recent years, many studies have shown a positive transfer of learning from music to various cognitive domains. For example, musical training is associated with an increase in general IQ (Schellenberg, 2004, 2006), enhanced working memory (Tierney, Bergeson, & Pisoni, 2008) as well as enhanced (auditory) attention (Strait, Kraus, Parbery-Clark, & Ashley, 2010). Various abilities related to linguistic processing have also been shown to correlate with musical skills, such as verbal memory (Ho, Cheung, & Chan, 2003; Jakobson, Lewycky, Kilgour, & Stoesz, 2008), speech production (Slevc & Miyake, 2006) and reading (Anvari, Trainor, Woodside, & Levy, 2002), to name a few. At a more fundamental level, musical training seems to enhance the ability to perceive speech, and, in particular, pitch information (Besson, Schön, Moreno, Santos, & Magne, 2007; Marques, Moreno, Castro, & Besson, 2007; Moreno, Marques, Santos, Luis, & Besson, 2009; Musacchia, Strait, & Kraus, 2008; Wong, Skoe,

Russo, Dees, & Kraus, 2007). Longitudinal studies by Moreno et al. (2009) demonstrated that musical training, rather than visual arts training, is the underlying factor of this transfer. Thus, a transfer of learning from music to language has been shown to take place with respect to the perception of specific acoustic cues.

However, speech is realized through the complex interaction of a multitude of cues apart from pitch alone. Previous research has provided evidence indicating that musicians show enhanced sensitivity to features other than pitch (Marie, Magne, & Besson, 2011; Milovanov et al., 2009). However, the generality of this effect, and whether it extends to the perception of additional linguistic features, is not well known. Furthermore, perception of sounds can be studied at various levels requiring differing amounts of cognitive effort, e.g., from the automatic neural encoding process (Musacchia et al., 2008; Strait et al., 2010; Wong et al., 2007) to perception of speech sentences in various conditions (Parbery-Clark, Skoe, Lam, & Kraus, 2009; Slevc & Miyake, 2006). The current study addresses these issues, namely, whether the finding that musicians have a superior ability to perceive pitch information extends equally to various other aspects of speech materials and whether musical training enhances the ability to perceive sounds at different levels.

We considered two separate levels of perception. The first is an acoustical analysis that takes place early in the perception process

* Corresponding author. Tel.: +31 24 35 21414.

E-mail addresses: m.sadakata@donders.ru.nl (M. Sadakata), sekiyama@kumamoto-u.ac.jp (K. Sekiyama).

¹ Tel.: +31 24 3612689.

² Tel.: +81 96 342 2845.

enabling us to *discriminate* sounds. This discrimination can be accomplished with very little knowledge about the target sounds because a comparison can be made based purely on acoustical cues. In fact, comparing sounds and detecting odd auditory events is a highly automatic process and does not even require one's attention (Näätänen, 2001). This detection mechanism works with various acoustic features, such as pitch (Xi, Zhang, Shu, Zhang, & Li, 2010), timbre (Caclin et al., 2006), intensity (Schroeger & Winkler, 1995) and timing (Menning, Imaizumi, Zwitserlood, & Pantev, 2002). The second level includes a mechanism that establishes the internal representation of categories as well as one that further applies this categorical knowledge to an incoming sound: these allow us to *identify* sounds. This process involves far more complex steps than the mere comparison of acoustic signals. In fact, learning foreign speech categories (so-called L2 categories) is one of the major problems when acquiring new language (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999).

Previous studies have shown that musical training enhances the first step: the ability to discriminate sound differences. In particular, musicians' higher sensitivity to linguistic pitch information has been repeatedly demonstrated using various materials and methods (Besson et al., 2007; Marques et al., 2007; Moreno et al., 2009; Wong et al., 2007). Although this may very well hold for processing of other aspects of speech signals, such as timing and combined spectral information, this has not been adequately investigated up to now (but see Marie et al., 2011; Milovanov et al., 2009). We used a discrimination test to address this issue along with synthesized speech of minimal pair continuums. Among the various linguistic features, we chose timing and quality of consonants as well as spectral changes of vowels. The aim of the experiment was to investigate whether there is a positive correlation between musical training and enhanced discrimination of differences in these features.

The degree of influence of musical training on the ability to categorize and to identify speech sounds is not well known. Musicians may very well exhibit superior performance here as well because the increased ability to distinguish detailed sound features ought to have a benefit when learning new speech categories. However, abstracting the relevant information from natural speech that is required to learn a new category is more complex and difficult than discrimination alone. Thus the beneficial effect of musical training may be limited as compared to discrimination. We used an identification experiment combined with a learning procedure to address this issue. It is known to be hard to achieve learning when new categorical boundary contradicts with native categorical boundaries (Bradlow et al., 1999). For this reason, we chose L2 features which require a set of criteria that are not relevant in the L1 counterparts. Unlike the discrimination test, this test employed more complex natural utterances of multiple words spoken by multiple speakers to test the robustness of the transfer effect using more complex and variable materials resembling the types of speech encountered in every-day conversations.

1.1. Japanese speech

The mora is a perceptual temporal unit that Japanese native listeners use when segmenting incoming speech signals (Cutler, 1994; Cutler & Otake, 1994; Otake, Hatano, Cutler, & Mehler, 1993; Otake, Yoneyama, Cutler, & van der Lugt, 1996). This segmentation boundary by Japanese natives does not necessarily overlap with that by non-Japanese natives. The difference can be especially substantial when a sound contains special moraic features, such as moraic obstruents and moraic nasals. For example, Japanese natives would segment a sound *hakkaku* (meaning disclosure) into *ha-Q-ka-ku* (4 morae), while a non-native listener would segment it differently, e.g., *ha-ka-ku* (3 syllables). In this case, Japanese natives perceive a long voiceless obstruent (moraic obstruent) transcribed as /Q/ before /k/ (moraic obstruent, Vance, 1987), while non-natives would not. Similarly, native Japanese listeners would segment *konyaku* (meaning engage-

ment) into, *ko-n-ya-ku* (4 morae), again differently from non-Japanese natives, e.g., *kon-ya-ku* (3 syllables). In this case, Japanese natives perceive a temporal segment which only consists of a nasal sound (moraic nasal, Vance, 1987) in between *o* and *y* while non-Japanese natives would not.

It can be problematic for non-Japanese natives when they are asked to distinguish a word pair that contrasts these specific moraic features, for example, the distinction between singleton and geminate of stop consonants, in other words, contrast with or without a moraic obstruent /Q/. Specific examples include *hakaku/hakaku/* (meaning bargain) and *hakkaku/haQkaku/*. The duration of the gap between the first /a/ and following /k/ is the main determinant of this particular contrast (but also see Kingston, Kawahara, Chambless, Mash, & Brenner-Alsop, 2009). However, not all languages require attention to such timing features and thus it can be challenging for non-native listeners to perceive such differences (Hardison & Saigo, 2010). Such a contrast is ideal when testing the perception of linguistic timing. Based on these considerations, we used eight minimal pairs for this type of contrast (including *hakaku-hakkaku*). Discrimination test used one minimal pair among the eight pairs while identification used all eight pairs.

Another example is a distinction between singleton and geminate of nasals followed by a glide sound /j/, such as *konyaku/koNjaku* or *ko-jaku/* and *konnyaku/kon.njaku/* (*konjac*, a Japanese food). The nasal in the former case, *konyaku*, is pronounced either without alveolar closure or without any closure plus a preceding nasalized vowel /ja/, whereas the latter case, *konnyaku*, is pronounced as a simple alveolar /n/ plus an onset nasal /nja/. When one is able to perceive moraic nasals, various perceptual cues may be available to make this distinction such as duration ("moraic nasals before /n/ tend to be longer", Vance, 1987, p.36) and quality ("moraic nasals preceding vowels become a nasalized copy of the surrounding vowels", Vance, 1987, p.36). However, otherwise it is very difficult to distinguish this contrast. This contrast is useful when testing perception of combined linguistic features. Accordingly, eight minimal pairs of this type of contrast were selected for use in the experiment (including *konyaku-konnyaku*).

1.2. Dutch speech

Vowels are mainly determined by a combination of multiple peaks in the vocal spectra (formants). The Dutch vowel inventory is large – 16 vowels including monophthongs and diphthongs (Gussenhoven, 1992), whereas that of Japanese is rather limited – 5 monophthongs (Vance, 1987). Obviously, categorical boundaries between Dutch and Japanese vowels do not overlap. Because adjusting categorical boundaries is not easy, assimilation of categories may occur when Japanese natives hear Dutch vowels that do not match with the Japanese vowel categories. In other words, Japanese natives may jointly map Dutch vowel categories onto a single Japanese vowel category (Iverson & Evans, 2007). As a consequence, distinguishing two separate Dutch vowels may be extremely challenging for Japanese natives because both sounds are perceived to belong to the same category and identification of sounds belonging to the same perceptual category is known to be difficult (Harnad, 1987).

The Dutch vowel *u* /Y/, a near-close near-front protruded vowel, is one such example. This sound does not exist in the Japanese vowel inventory. Phonetically, this vowel is somewhere in between *e* /ε/ and *oe* /u/, which are similar to Japanese vowels *e* /e/ and *u* /u/, respectively. Because it is difficult for Japanese natives to develop a new category for the Dutch vowel *u* /Y/, distinctions between the Dutch vowel contrasts *u* /Y/ and *e* /ε/, as well as between *u* /Y/ and *oe* /u/ is expected to be difficult for Japanese native listeners. These contrasts are useful when testing sensitivity to perception of spectral changes in linguistic sounds. Eight minimal pairs were chosen for each type of contrast. For example, the *u* /Y/ and *e* /ε/ contrast includes the pair *put-pet* (well-cap), while the *u* /Y/ and *oe* /u/ contrast includes *tut-toet* (a girl who dresses up-toot).

Table 1

List of musical/non-musical activities. Numbers in brackets refer to number of participants who followed these activities as their primal expertise to be counted in the study.

		Training activities
JP	Musicians	Piano (15), Electone (3), Flute (2), Violin (2), Base (1), Clarinet (1), Base (1)
	Non-musicians	Volleyball (12), Calligraphy (7), Japanese fencing (2), Tennis (2), Badminton (1), Baseball (1), Basketball (1), Football (1), Martial arts (1), Swimming (1)
NL	Musicians	Piano (6), Clarinet (4), Saxophone (4), Baroque Flute (2), Flute (2), Horn (2), Violin (2), Drums (1), Guitar (1), Harp (1), Keyboard (1)
	Non-musicians	Football (6), Gymnastics (4), Hockey (3), Swimming (3), Volleyball (2), Martial arts (2), Ball Room dance (1), Basketball (1), Biking (1), Handball (1), Horse riding (1), Paragliding (1), Rowing (1)

1.3. Design

We carried out two perceptual experiments (discrimination and identification) in order to examine the benefit of musical training on perception of diverse speech sounds: the contrast of singleton and geminate of Japanese consonants (stop and nasals) and the contrast of different spectra of Dutch vowels (u–e and u–oe). The selected Dutch and Japanese speech materials were tested with Dutch and Japanese natives. In this way, both language groups received L1 and L2 stimuli. Each language group included individuals with differing levels of musical training.

2. Experiment 1

In Experiment 1, the discrimination performance of L1 and L2 contrasts were compared between musicians and non-musicians. Previous research has shown that musicians exhibit an enhanced ability to discriminate speech sound differences, especially with respect to pitch. This effect has previously been reported with regard to both L1 (Schön, Magne, & Besson, 2004) and L2 (Marques et al., 2007). Based on these findings, we expect that musicians will also exhibit enhanced discrimination performance for the speech contrasts used in this experiment. Thus musicians may perform better than non-musicians on both L1 and L2 contrasts.

2.1. Method

2.1.1. Participants

Fifty-three native Dutch (NL) and 54 native Japanese speakers (JP) participated in the study. Participants in both language groups were matched as closely as was possible with regard to educational level, age, length and type of training and language experience. Participants were recruited from student populations in the Nijmegen (Radboud University Nijmegen, The Netherlands) and Kumamoto (Kumamoto University, Japan) regions, with an age limit of 29 years. For both language groups, the musician group consisted of individuals who had received more than 5 years of formal training on any musical instrument outside of school activities (excluding singers), and who were still actively involved in the musical activities at the time of the experiment. The non-musician group included individuals who completed less than 3 years of musical training and had received training in an area other than music for at least 5 years. The number and type of musical and non-musical training activities are listed in Table 1.

A summary of the available demographic information for each group can be found in Table 2. A Kruskal–Wallis test was performed to highlight any differences among the four groups for each row. Subsequently, a multiple comparisons analysis using the Mann–Whitney *U* test was performed ($p < .05$, Bonferroni adjusted³). Results

³ Effect size *r* for all significant comparisons was greater than .4.

Table 2

Demographics of each group and summary of Kruskal–Wallis test ** $p < .001$, * $p < .05$.

	NL		JP		Kruskal–Wallis test
	Musician	Non-musician	Musician	Non-musician	
Age (years)	21.5 (3.03)	21.07 (2.7)	20.6 (2.27)	19.6 (2.23)	H(3) = 11.19*
Gender	5 M, 21 F	7 M, 20 F	8 M, 17 F	12 M, 15 F	–
Training duration (years)	8.4 (2.37)	9.4 (3.88)	11.8 (4.06)	8.1 (2.43)	H(3) = 14.35*
Started year (years)	8.5 (1.84)	9.2 (3.58)	5.4 (2.89)	8 (2.25)	H(3) = 29.1128**
Level (1–7)	5.4 (0.71)	5.4 (0.91)	3.8 (1.39)	4.2 (1.07)	H(3) = 35.56**
L2 training duration (years)	6.8 (1.29)	7.9 (2.44)	8.4 (2.45)	7.3 (1.74)	H(3) = 9.6193*
L2 level (1–7)	5.1 (1.26)	5.5 (1.14)	3.2 (1.32)	3.3 (1.10)	H(3) = 48.4628**
Number of foreign language	3.23 (0.51)	3.4 (0.63)	1.8 (0.41)	1.7 (0.65)	H(3) = 78.9703**
Estimated verbal IQ	104.6 (5.96)	103.9 (5.96)	106.2 (7.32)	107.6 (8.50)	H(3) = 4.3887, n.s.

Table 3

List of minimal pairs of four contrasts used for the two experiments. Pairs with an asterisk were used for the discrimination experiment.

Japanese		Dutch	
nasal	stop	u–e	u–oe
kanyo–kannyo	shikou–shikkou	kuch–kech	cup–coup
kanyu–kannyuu	gekou–gekkou	puk–pek	kus–koes
kenyou–kennyou	hakaku–hakkaku*	pul–pel	pup–poep
konyaku–konnyaku*	haken–hakken	tul–tel	tuf–toef
senyuu–sennyuu	ritou–rittou	puls–pels	tut–toet*
shinyuu–shinnyuu	sekai–sekkai	put–pet*	puf–poef
tenyo–tennyo	shikaku–shikkaku	puts–pets	pus–poes
tenyu–tennyu	katou–kattou	pup–pep	put–poet

for the analysis of age indicated that Japanese non-musicians were significantly younger than Dutch non-musicians. Japanese musicians also had a significantly longer training duration than Japanese non-musicians and Dutch musicians. This was most likely due to the fact that Japanese musicians started their musical training at significantly younger age than the other three groups. An analysis of subjective expertise level (using a scale from 1 to 7; 1 = beginner, 7 = expert) indicated that the Dutch participants tended to estimate their expertise level in musical or non-musical activities significantly higher than the Japanese participants. Ratings of subjective fluency of L2 (reported using the scale from 1 to 7; 1 = Not fluent at all, 7 = Very fluent) also tended to be higher for the Dutch than that of the Japanese participants. Although Dutch participants tended to speak a significantly greater number of foreign languages, none of the participants were early-life bilinguals. Most importantly, none of the participants had intensive exposure to the tested L2. The NLV (Nederlandse Leestest voor Volwassenen, [Dutch Reading Test for Adults] (Schmand, Lindeboom, & van Harskamp, 1992) and JART (Japanese Adult Reading Test, Matsuoka, Uno, Kasai, Koyama, & Kim, 2006) were used to calculate the estimated verbal IQ, and no differences between the groups were observed⁴.

2.1.2. Stimuli

One minimal pair per contrast, spoken by a male native speaker of Dutch and Japanese respectively, was used (see Table 3). Recordings were first low-pass filtered at 5000 Hz and average sound levels were

⁴ NLV and JART are translated version of NART (National Adult Reading Test, Nelson & Willison 1991), which involves evaluation of ability to pronounce irregular words that is known to significantly correlate with various IQ measures (McGurn et al., 2004). Results of both tests estimate value that highly correlates to subsets of WAIS-R test (Wechsler, 1981), in particular, verbal intelligence.

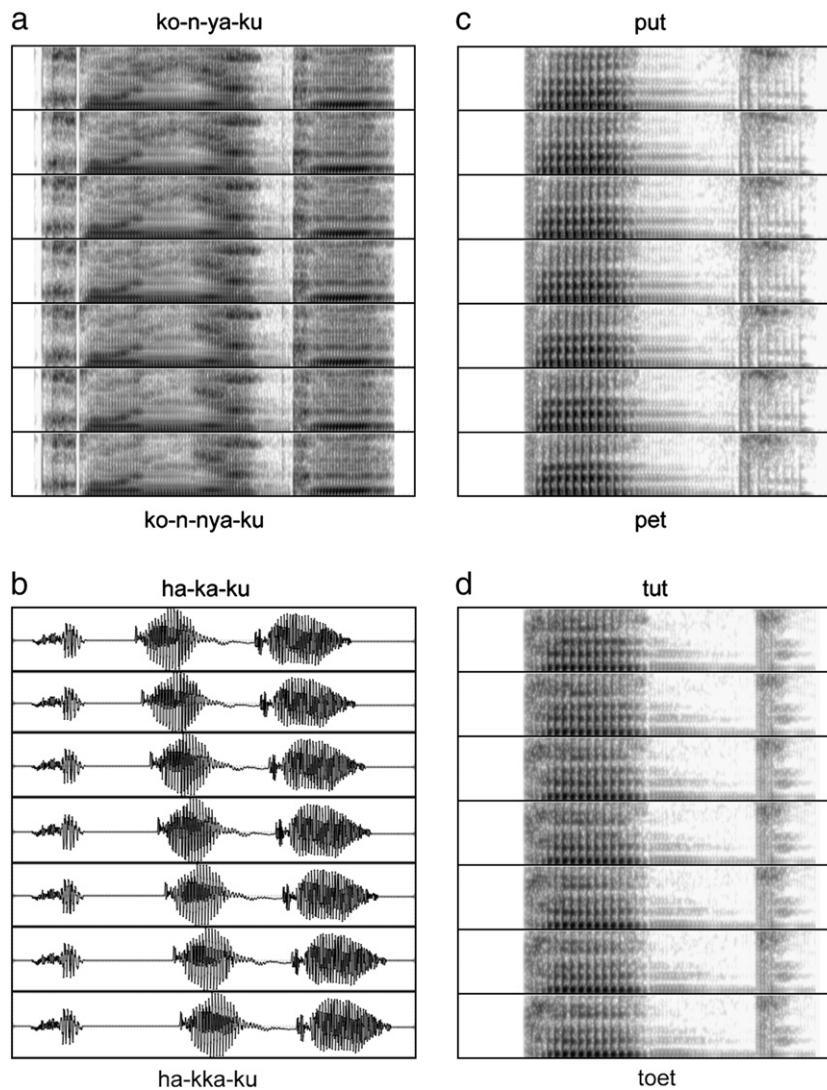


Fig. 1. Four continua used for the discrimination experiment: a) Japanese nasal from konnyaku (top) to konyaku (bottom), b) Japanese stop from hakaku (top) to hakkaku (bottom) c) Dutch vowel from put (top) to pet (bottom), and d) Dutch vowel from toet (top) to tut (bottom).

normalized to 70 dB using Praat (Boersma, 2001). The Japanese stop continuum (/hakaku-/hakkaku/) 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. Fig. 1a illustrates waveform of this continuum. The other three contrasts, Japanese nasal and two Dutch vowel continuums, were created using TANDEM-STRAIGHT (Kawahara et al., 2008). First, target time windows which include sounds of /N,ja-/ /nnja/, /Y/-/ε/, /Y/-/u/ were specified. For each pair, spectrum, frequency, aperiodicity and time information within the target time window were morphed in 19 equal steps. From these 19 steps, we selected the following 7 equal steps to use for this study based on the results of a pilot experiment. In the pilot experiment, two Dutch natives and two Japanese natives (none of whom participated in the main experiment) completed an identification test of their L1 categories on sound steps 1, 4, 7, 10, 13, 16, and 19, where steps 1 and 19 served as category prototypes for each contrast. The 50% crossover point was estimated and was used to define the midpoint for the stimuli continuum. A two-step interval was used to construct the continuum. The selected steps were 1, 3, 5, 7, 9, 11, and 13 for the nasal as well as the u-oe contrasts and 4, 6, 8, 10, 12, 14, and 16 for the u-e contrast. As a result, we did not use the full-range of the continuums. However, this sub-selection provides us with an opportunity to observe a potential effect of musical influence

in high-resolution. These morphed parts were embedded in context information (sound outside the target time window) of one member of the pair (/koN ja ku/, /pYt/, /tut/). F0 of these stimuli was kept constant. Spectrograms of these three continuums are illustrated in Fig. 1b, c, d respectively. The audio files can be found at <http://www.nici.ru.nl/mmm>.

2.1.3. Procedure

The experiment employed a speeded two-alternative forced choice (2AFC) task, in which participants had to judge whether the two sounds (sound A and B) were identical or not. Stimulus A was kept constant per contrast (konnyaku, hakkaku, put and tut, respectively). Stimulus B was identical to A in 50% of the trials, while one of the six morphed sounds was used in the other 50% of the trials. The inter-stimulus-interval (ISI) was 1500 ms. The four contrasts were not blocked but were randomly presented in a single block in order to keep the task challenging and avoid boredom. Each block consisted of 60 trials and lasted approximately 6 min. A total of 4 blocks were presented. Participants completed at least one practice session of 6 trials using L1 minimal pair (Dutch: *gek-rek* [crazy-rack] and Japanese: *ganen-gannen* [rock salt-January 1st]) before the real session. Participants could repeat the practice session as many times as they needed in order to understand the task. No feedback was

provided on task performance. The position of the response buttons (same or different) were altered for even and odd participants.

2.1.4. Apparatus

Experimentation in the Netherlands and in Japan employed the same setup: A DELL notebook computer with an IntelCoreDuo processor (4 GB RAM) and a Sound Blaster X-Fi. The Sony MDR-7506 headphones and a 15.4-inch TFT screen were used to present auditory stimuli and visual instructions, respectively. Average sound pressure level (SPL) of the headphones was adjusted to around 68 dB. The application *Presentation* (version 14.3, Neurobehavioral Systems) was used for presenting instructions and stimuli as well as for collecting responses. The participants responded using an USB game controller (Sanwa supply, JY-P68US). PASWStatistics18.0 (SAS) was used for the data analyses.

2.2. Results

Responses with a reaction time longer than 2.5 standard deviations from the mean were identified as potential outliers (NL: 3049 ms, JP: 3302 ms). Consequently, a minority of 2.1% of the response was discarded from the analyses.

Fig. 2 compares the correct response rate as a function of the morphing steps for two groups with different levels of musical training (musician and non-musicians, respectively). The x-axis corresponds to the characteristics of sound B, representing the degree of similarity/difference between the presented two sounds. Eight panels show the results for four contrasts (Japanese nasal, stop, and two Dutch vowels) and two language groups (NL and JP). Two-way repeated measures ANOVAs (training group*morphing step) were performed independently for the Dutch and the Japanese participants in order to analyze the data for each of the four contrasts (see Table 4 for summary statistics). Table 4 shows that a more pronounced effect of musical training in the expected direction was observed in the Japanese stop condition for Dutch participants and in the Dutch u–e condition for Japanese participants. Fig. 2 shows significant and marginally significant simple main effects of musical training at different morphing steps for two language groups. In summary, Dutch musicians detected the difference better than Dutch non-musicians

Table 4

Summary of two-way repeated measure ANOVA results on correct response rate. Significant results are highlighted with bolds.

Language group	Condition	Effects		
		Training group	Morphing step	Interaction
NL	Nasal	$F_{1,51} = .019$ $p = .890$, $\eta^2 = .000$	$F_{6,306} = 128.386$, $\epsilon_{GG} = .626$ $p < .000$, $\eta^2 = .716$	$F_{6,306} = .389$, $\epsilon_{GG} = .626$ $p = .804$, $\eta^2 = .008$
		$F_{1,51} = 5.778$ $p = .020$, $\eta^2 = .102$	$F_{6,306} = 75.231$, $\epsilon_{GG} = .637$ $p < .000$, $\eta^2 = .596$	$F_{6,306} = 3.535$, $\epsilon_{GG} = .637$ $p = .009$, $\eta^2 = .065$
	u–e	$F_{1,51} = 2.840$ $p = .098$, $\eta^2 = .053$	$F_{6,306} = 72.241$, $\epsilon_{GG} = .787$ $p < .000$, $\eta^2 = .586$	$F_{6,306} = 1.586$, $\epsilon_{GG} = .787$ $P = .804$, $\eta^2 = .030$
		$F_{1,51} = 1.054$ $p = .309$, $\eta^2 = .020$	$F_{6,306} = 92.965$, $\epsilon_{GG} = .781$ $p < .000$, $\eta^2 = .646$	$F_{6,306} = 2.960$, $\epsilon_{GG} = .781$ $p = .015$, $\eta^2 = .055$
JP	Nasal	$F_{1,52} = .597$ $p = .443$, $\eta^2 = .011$	$F_{6,312} = 273.890$, $\epsilon_{GG} = .474$ $p < .000$, $\eta^2 = .840$	$F_{6,312} = 1.764$, $\epsilon_{GG} = .474$ $p = .159$, $\eta^2 = .033$
		$F_{1,52} = 1.324$ $p = .255$, $\eta^2 = .025$	$F_{6,312} = 178.587$, $\epsilon_{GG} = .427$ $p < .000$, $\eta^2 = .774$	$F_{6,312} = .885$, $\epsilon_{GG} = .427$ $p = .437$, $\eta^2 = .017$
	u–e	$F_{1,52} = 6.699$ $p = .012$, $\eta^2 = .114$	$F_{6,312} = 113.123$, $\epsilon_{GG} = .599$ $p < .000$, $\eta^2 = .685$	$F_{6,312} = 3.668$, $\epsilon_{GG} = .599$ $p = .009$, $\eta^2 = .066$
		$F_{1,52} = .200$ $p = .656$, $\eta^2 = .004$	$F_{6,312} = 107.600$, $\epsilon_{GG} = .737$ $p < .000$, $\eta^2 = .674$	$F_{6,312} = .208$, $\epsilon_{GG} = .737$ $p = .946$, $\eta^2 = .004$

for some instances of the Japanese stop contrast, as well as for the Dutch u–e contrast. The Japanese musicians also detected the differences better than Japanese non-musicians for some instances of the Japanese nasal contrast as well as for the Dutch u–e contrast.

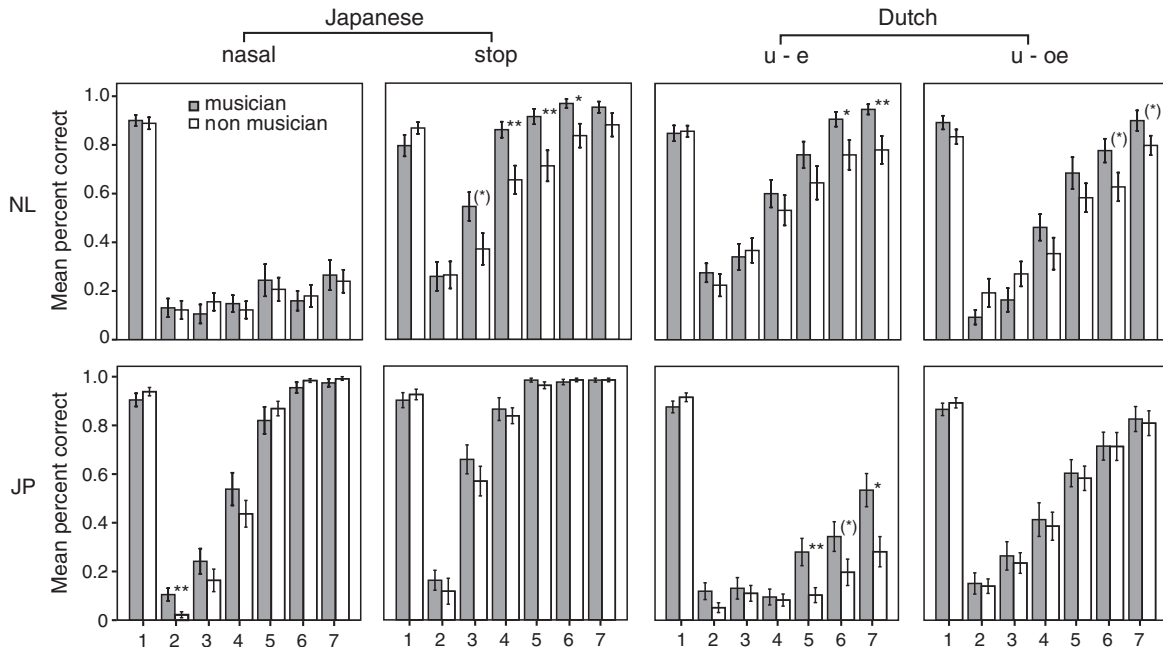


Fig. 2. Mean correct response rate and standard error of discrimination tests. The x-axis corresponds to the degree of similarity/difference between the sound A and B. The larger the number, the larger the difference between the sound A and B. The one refers to the trial where sound A and B were identical. The error bars show standard errors. **p<.01, *p<.05, (*)p<.09.

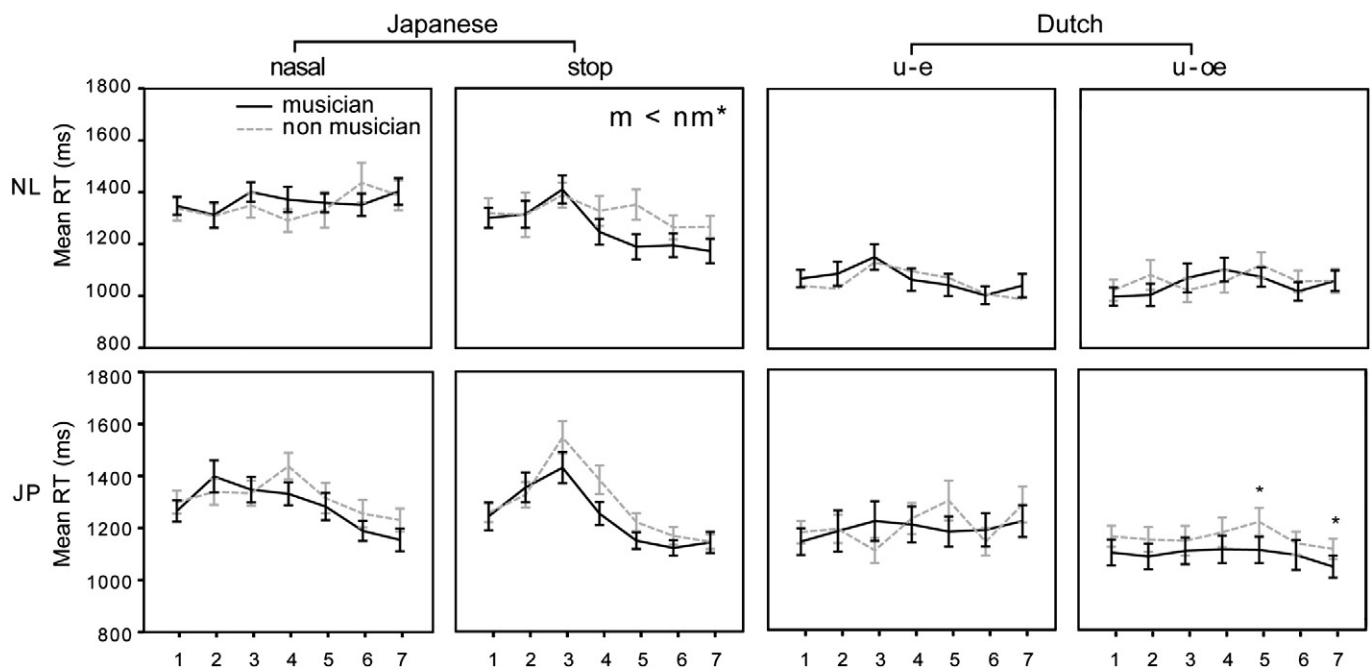


Fig. 3. Mean reaction time of discrimination test. The x-axis corresponds to the degree of similarity/difference between the sound A and B. The larger the number, the larger the difference between the presented sounds. The error bars show standard errors. * $p < .05$.

Next, we report the reaction time of correct responses in the discrimination task. Fig. 3 summarizes reaction time as a function of the morphing steps for both language groups per condition. Repeated measure ANOVAs were performed using the same factors as for correct response rate. The results for the Dutch participants in the Japanese stop contrast condition indicated a significant main effect of training without interaction ($F_{1,26} = 5.486, p = .027$; $F_{6,156} = 1.143, n.s., \eta^2 = .174$, respectively)⁵, indicating that musicians were faster at responding in this condition than non-musicians. Although no other main effect of training was observed among the eight ANOVAs, analyses of simple main effect indicated that Japanese participants also exhibited the same tendency at some steps of Dutch u-oe contrast ($p < .05$). These are indicated in Fig. 3.

2.3. Discussion

The largest effect was observed for the Japanese stop condition tested with Dutch participants: musicians tended to detect more differences than non-musicians did when timing differences were 45 ms, 60 ms and 75 ms (4, 5, and 6 of the x-axis in Fig. 2). Furthermore, musicians gave correct responses more quickly than non-musicians did. Dutch musicians also demonstrated greater discrimination accuracy to their L1 materials. Thus, the benefit of musical training for the discrimination ability was not limited to L2 but also extended to L1 materials.

On the other hand, the effect of musical training on performance was modest when Japanese stimuli were tested with Japanese participants, probably due to the ceiling effect: Japanese musicians only outperformed non-musicians in one nasal condition. However, Japanese musicians did detect more differences than Japanese non-musicians in the Dutch u-e condition and were also faster when responding correctly to some instances of u-oe contrast. Taken together, although differing in degree, the results support the idea that an enhanced sensitivity to linguistic pitch information processing

extends to other features, namely, consonant timing information and spectral information (quality of vowels and nasals).

3. Experiment 2

Experiment 1 demonstrated some cases where musicians outperformed non-musicians with respect to perceptual discrimination of speech materials. In Experiment 2, identification performance of L1 and L2 contrasts, where learning and application of categorical information are required, was compared between musicians and non-musicians. Superior ability in discriminating sound differences may well be beneficial for performance in the identification task, especially with regard to the processing of relevant acoustic cues when learning novel phonetic categories. For example, musical training has shown to facilitate the automatic and stable representation of incoming acoustic signals (Kraus & Chandrasekaran, 2010). Such advantages would be beneficial in the context of category learning. Hence, performance was expected to be better for musicians relative to non-musicians.

3.1. Method

3.1.1. Participants

The same participants that took part in Experiment 1 also took part in this study.

3.1.2. Stimuli

For each contrast, 4 native speakers (2 males and 2 females) recorded eight minimal pairs (Table 3), which resulted in 32 word pairs per contrast. Thus, there were 128 word pairs in total. The average sound levels were normalized to 70 dB using Praat (Boersma, 2001).

3.1.3. Procedure

The experiment consisted of four blocks, each corresponding to one of the contrasts being tested (Dutch u-oe, u-e, Japanese stop and nasal). A single block lasted approximately 6 min. The presentation order of the blocks was randomized. Each block started with a learning task and was followed by an identification task. This

⁵ Degree of freedom for the ANOVAs in this section differs for conditions and language groups, because analysis only included the data of participants who showed correct responses at all morphing steps.

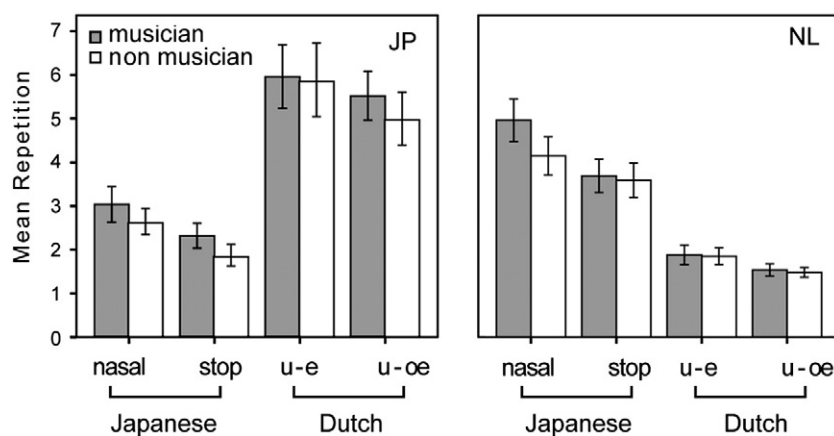


Fig. 4. Mean frequency of repetition of the memorization task. The error bars show standard errors.

procedure was repeated 4 times to test each of the contrasts. During the learning task, listeners had to learn and remember the two categories represented by a given contrast (e.g., with or without geminate). Three stimulus pairs were randomly chosen from the 32 stimulus pairs for each contrast as learning materials (8 minimal pairs recorded by 4 speakers) and used for each participant. Each category was presented along with visual number “1” or “2” with an inter-stimulus interval (ISI) of 2000 ms. The combination of visual numbers and sound categories was counterbalanced for even and odd participants. Thus half of participants learned one type of category (e.g., Japanese geminates, *hakkaku*) as category 1 and the other half learned it as category 2. A piece of paper and a pen were provided for them to write down their perception of characteristic features of categories. Participants could repeat the presentation of these three example pairs multiple times until they felt confident about their understanding of the categories.

The learning task was immediately followed by an identification task for each contrast. During the identification task, a word (audio stimulus) was presented and participants pressed the button labeled 1 or 2 to indicate their identification response (to which category the presented sound belongs). All participants but five Dutch and four Japanese participants completed a short practice session (learning task and 6 trials of identification task) using L1 minimal pair (Dutch: *g-r* comparisons such as *gek-rek* and Japanese: nasal variation contrast such as *ganen-gannen*) before the real session⁶.

3.1.4. Apparatus

The same apparatus were used as in Experiment 1.

3.2. Results

Responses with reaction time longer than 2.5 standard deviations from the mean were identified as potential outliers (NL: 3283 ms, JP: 3485 ms, respectively) and a minority of 3.0% of the identification test was discarded from the analyses. Responses to the three randomly chosen pairs that were used for the learning task were excluded from the analyses.

First we present the frequency of repetition for each contrast in the learning task. Fig. 4 shows frequency of repetition for the two groups per contrast, demonstrating that more repetitions were needed during the learning task of their L2 contrasts than the L1 contrasts for both Dutch and Japanese groups. Separate T-tests carried out per

contrast and per language group indicated that musicians and non-musicians did not differ in terms of frequency of repetition that they required during the learning task.

Fig. 5 shows the rate of correct responses for the two groups per contrast, showing that participants performed the task better in their L1 better than in their L2, as expected. Separate *t*-tests were carried out to analyze the data of Dutch and Japanese groups for each contrast. A strong effect of musical training was observed for Dutch participants in the Japanese stop contrast condition, with a mean between group difference of 12.0% ($t(51) = 2.925$, $p = .0025$, one-tailed, Cohen's $d = .80$). This indicates that among Dutch participants, musicians performed significantly better than non-musicians when identifying stop geminates. Furthermore, Japanese musicians tended to perform better at identifying the stop contrast condition than Japanese non-musicians. Although the mean difference between groups was only 1.1%, this difference was significant ($t(50) = 2.111$, $p = .020$, one-tailed, Cohen's $d = 0.59$). No other contrasts showed an influence of musical training. Likewise, analysis of reaction time did not indicate any significant differences between musicians and non-musicians.

3.3. Discussion

If musicians were only better at hearing differences in speech signals, we would not see any effect of musical training for this identification test. However, the musicians in our study demonstrated better identification performance on the Japanese geminate condition. Furthermore, this was true not only for Dutch but also for Japanese participants. Thus, there is a strong link between musical training and accuracy of the identification of timing features of linguistic signals, for example, musicians may be able to represent the incoming acoustic timing features of speech signals in an automatic and stable manner. However, the weaker effect of musical training on correct response rate observed for Experiment 2 as compared to Experiment 1 may result from increased variability in the materials used in Experiment 2, which required more complex perceptual skills.

Comparison between results with regard to correct response rate and reaction time for both experiments showed a stronger effect of musical training on correct response rate rather than on reaction time. Especially the effect on reaction time was absent for Experiment 2 while it was observed for some cases of Experiment 1. Previous results have shown an effect of musical training on the accuracy of overt response (error rate) rather than on reaction time (cf. Besson et al., 2007), which may explain the weaker effect on reaction time in this experiment.

⁶ Practice session was introduced because the first few participants took long for understanding the procedure of the experiment.

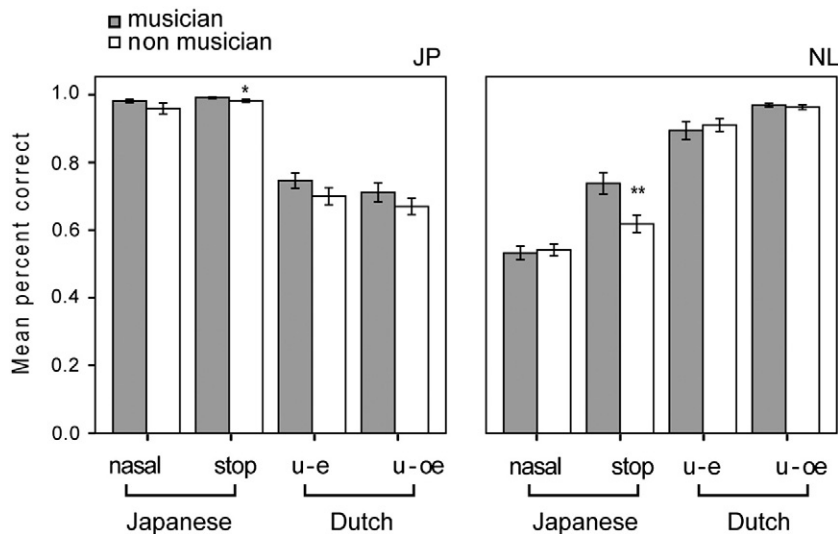


Fig. 5. Mean correct response rate (average and standard error) of identification test. The error bars show standard errors. ** $p < .01$, * $p < .05$.

4. General discussion

We studied the effect of musical training on the perception of L1 and L2 speech sounds using discrimination and identification tasks. The discrimination task tested the ability to hear subtle differences between two presented sounds, whereas the identification task included more complex cognitive tasks consisting of first learning categories and then categorizing incoming speech sounds. From previous studies that demonstrate an advantage of musical training on various linguistic skills, we predicted that musicians would perform these tasks better than non-musicians. The results of the experiments indeed showed that musicians outperformed non-musicians in these specific cases. A beneficial effect of musical training was observed not only with respect to L2 materials but also with respect to L1 materials in some cases.

The most pronounced effect of musical training was found for perception of speech timing, specifically, the duration of stop consonant. This extends previous findings with respect to musicians' enhanced ability to perceive temporal information, which has been demonstrated with diverse non-linguistic materials (Rammsayer & Altenmüller, 2006; Yee, Holleran, & Jones, 1994) as well as for some linguistic materials, such as the duration of vowels (Milovanov et al., 2009) and metric structure (Marie et al., 2011). Why was the effect largest for the temporal domain? The similarity between music and language domains with respect to prominent acoustic features may be responsible here: some features, such as pitch and timing, are equally prominent in both domains while other features, such as spectral changes of vowels, may be more specific to the language domain. Accomplishing perceptual discrimination and identification based on completely new sets of criteria (nasal, vowels) might have been much more challenging than doing so based on criteria that are somewhat more familiar (timing). Furthermore, the tasks of learning and identifying new vowel categories could be expected to be more demanding than tasks involving timing features because of the conflict with native vowel categories. Musical training may well confer a benefit for the perception of these challenging materials, however, longer training time in the context of a longitudinal experiment may be required to reveal this benefit.

Recent studies indicated that musical training shapes the way perception of sound takes place at very early stages, such as the encoding process of acoustical signals at the level of the brainstem (Musacchia et al., 2008; Strait et al., 2010; Wong et al., 2007) as well as the pre-attentive detection of various types of deviant auditory events (van Zuijen, Simoens, Paavilainen, Naatanen, & Tervaniemi, 2006;

Vuust et al., 2005). Such enhanced abilities due to musical training are likely to contribute to the enhanced discrimination performances observed in musicians. It follows that if musical training improved discrimination of any acoustical signals, then one could expect to observe superior performance in musicians in all conditions. However, there were two cases in which we did not observe enhanced performance by musicians on L2 materials: Japanese nasal tested with Dutch participants and Dutch vowel (u-oe) tested with Japanese participants. Especially in the former case, we observed no benefit of musical training and no sensitivity to differences: even when the difference between the two stimuli was maximal, Dutch participants' correct response rate was much lower (25.3%) than the chance level (50%), indicating that they did not pick up on the perceptual cues used for making distinctions in this condition very well. This is interesting because Japanese natives (both musicians and non-musicians) showed fairly good sensitivity to this condition, which means that there were enough perceptual cues available to make a distinction. This strong cross-linguistic contrast is a clear example of how the relative weighting of perceptual cues can be modulated by linguistic background (e.g., Best, McRoberts, & Goodell, 2001). Moreover, this clearly indicates that musical training does not necessarily enhance the perception of any acoustic features automatically. Rather, there appears to be an interaction between the enhancement effect and one's prior experience (i.e. linguistic background).

The positive transfer effect of musical training to speech perception observed here supports views proposing shared resources between the language and music domains during the early stages of perception, such as a shared representation of acoustic features. The differences in the degree of transfer found for the speech features investigated in this study may indicate that the degree of overlap in such representations varies among features. Such a finding is of relevance to ongoing discussions regarding the relationship between language and music processing, especially with regard to how domain specific knowledge may be structured and represented in the brain (Patel, 2008; Peretz & Coltheart, 2003).

The L1 Dutch vowel discrimination performance of Dutch non-musicians did not reach very high level. This indicates that the selected morphing steps for these conditions did not cover a wide enough range of the continuum for non-musicians. If a wider range would be covered, the performance of Dutch non-musicians might reach the level of musicians, at least around the maximum difference level. Nevertheless, the fact that Dutch musicians performed significantly better for the used range of stimuli suggests that they were more sensitive to subtle spectral changes. There were more cases in

which we observed the beneficial effect of musical training on L1 materials, specifically, Japanese nasal discrimination and stop identification by Japanese participants. Altogether, these effects on L1 conditions may support the notion that musicians are more accurate in performing auditory tasks in general due to, for example, better auditory sequence memory (Tierney et al., 2008) and better auditory attention skill (Strait et al., 2010).

Matching participant groups is a critical issue in the between-subject design. For this reason, the current study did not compare Dutch and Japanese participants because the effect of their environmental factors on their response may exceed that of their musical training. To match participants within the culture, we carefully considered age, educational level (university students), length and type of training. In fact, significant differences were found with respect to demographics when comparing the four groups (musicians, non-musicians for Dutch and Japanese), but less so when comparing musicians and non-musicians within language groups. In particular, skills related to L2 are of central importance in the current study. These skills were matched for musicians and non-musicians within each of the language groups. An equivalent level of estimated verbal IQ for musician and non-musician in both language groups suggests that the two groups had a similar vocabulary level. Schellenberg (2009) warns against the use of these IQ estimate tests because these measures may not be very accurate. Indeed, a third factor, such as general IQ, general cognitive ability *g* (but see Schellenberg & Moreno, 2009) or executive functions (Bialystok & DePape, 2009) may be responsible for enhanced skills in both music and language domains, rather than musical training being responsible for the enhancement of linguistic skills. Although our study cannot completely rule out this possibility, we think it is unlikely that our findings, that musicians showed greater ability to process temporal information rather than spectral information, could be caused by a selective influence of such factors (Tierney et al., 2008). The finding that Dutch participants were more fluent in a greater number of foreign languages than Japanese participants could possibly result in an overall better performance by Dutch participants compared to Japanese participants. For example, it has been shown that bilinguals have an advantage over monolinguals with respect to various skills in speech processing (Bialystok, 2001). However, the current experimental design does not allow us to evaluate this effect. In order to compare performance of language groups, assessing the performance of the two language groups with regard to a third language spoken by neither of the groups would be suitable.

Another crucial factor was the duration of expert training. Japanese musicians reported longer average training duration and earlier start of training as compared to the other groups, which may have resulted in an enhancement of their task performance. The additional analyses we performed indicated a significant correlation between training duration and the correct response rate of Japanese stop consonant identification performance by Japanese musicians ($b = 0.001$, $p = .244$, $r = .45$), a result that supports this possibility. However, no other significant correlations between training duration and task performance were found. Similarly, there was a significant difference with respect to age (Japanese non-musicians were younger than the other three groups), but we could not find an obvious influence of age on the current results. More sensitive response measures, such as neurophysiological measures, often provide more insight into such issues (e.g., Besson et al., 2007), would be a logical next step in pursuing this topic.

5. Conclusion

The findings of the current study indicate that musicians have higher sensitivity when comparing small differences in linguistic timing information and spectral information. Musicians also showed an increased ability in learning and identifying linguistic timing

information. Although there were some cases where we did not observe a benefit of musical training, the current study does not necessarily indicate null effects on these conditions. Discriminating completely new linguistic features, such as Japanese nasals, as well as learning and identifying new categories may require more exposure to the materials. A longitudinal study could clarify this issue.

Acknowledgements

This study has been supported by Grant-in-Aid for Scientific Research (B) 21330169, the project leader Shuji Mori (Japan Society for the Promotion of Science). The data collection took place in Donders Institute for Brain, Cognition and Behaviour (The Netherlands) and Kumamoto University (Japan). Authors are grateful for supports provided by Naomi Nakamura, Tomoko Takemura, Lotte van der Zanden, for comments from Makiko Aoyagi, Alex Brandmeyer, Kathleen Jenks, James McQueen, and Rebecca Schaefer as well as for constructive suggestions from Mireille Besson and a reviewer.

References

- Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology*, 83, 111–130.
- Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience*, 25, 399–410.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, 109, 775–794.
- Bialystok, E. (2001). Metalinguistic aspects of bilingual processing. *Annual Review of Applied Linguistics*, 21, 169–181.
- Bialystok, E., & DePape, A. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 565–574.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5, 341–345.
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English vertical bar /r/ vertical bar and vertical bar /l/ vertical bar: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61, 977–985.
- Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M., et al. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience*, 18, 1959–1972.
- Cutler, A. (1994). The perception of rhythm in language. *Cognition*, 50, 79–81.
- Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, 33, 824–844.
- Gussenhoven, C. (1992). Dutch. Illustrations of the IPA. *Journal of the International Phonetic Association*, 22, 45–47.
- Hardison, D. M., & Saigo, M. M. (2010). Development of perception of second language Japanese geminates: Role of duration, sonority, and segmentation strategy. *Applied Psycholinguistics*, 31, 81–99.
- Harnad, S. R. (1987). *Categorical perception: The groundwork of cognition*. Cambridge, New York: Cambridge University Press.
- Ho, Y. C., Cheung, M. C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology*, 17, 439–450.
- Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of format targets, format movement, and duration. *The Journal of the Acoustical Society of America*, 122, 2842–2854.
- Jakobson, L. S., Lewycky, S. T., Kilgour, A. R., & Stoesz, B. M. (2008). Memory for verbal and visual material in highly trained musicians. *Music Perception*, 26, 41–55.
- Kawahara, H., Morise, M., Takahashi, T., Nishimura, R., Irino, T., & Banno, H. (2008). TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0 and aperiodicity estimation. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2008* (pp. 3933–3936). (Las Vegas, US).
- Kingston, J., Kawahara, S., Chambless, D., Mash, D., & Brenner-Alsop, E. (2009). Contextual effects on the perception of duration. *Journal of Phonetics*, 37, 297–320.
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11, 599–605.
- Marie, C., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience*, 23, 294–305.
- Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, 19, 1453–1463.
- Matsuoka, Y., Uno, M., Kasai, K., Koyama, K., & Kim, Y. (2006). 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*, 60, 332–339.

- McGurn, B., Starr, J. M., Topfer, J. A., Pattie, A., Whiteman, M. C., Lemmon, H. A., et al. (2004). Pronunciation of irregular words is preserved in dementia, validating pre-morbid IQ estimation. *Neurology*, *62*, 1184–1186.
- Menning, H., Imaizumi, S., Zwitserlood, P., & Pantev, C. (2002). Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. *Learning & Memory*, *9*, 253–267.
- Milovanov, R., Huotilainen, M., Esquef, P. A. A., Alku, P., Välimäki, V., & Tervaniemi, M. (2009). The role of musical aptitude and language skills in preattentive duration processing in school-aged children. *Neuroscience Letters*, *460*, 161–165.
- Moreno, S., Marques, C., Santos, A., Luis, S., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, *19*, 712–723.
- Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research*, *241*, 34–42.
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, *38*, 1–21.
- Nelson, H. E., & Willison, J. R. (1991). *National Adult Reading Test (NART)* (2nd edn.). Windsor, UK: NFER-Nelson.
- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable – Speech segmentation in Japanese. *Journal of Memory and Language*, *32*, 258–278.
- Otake, T., Yoneyama, K., Cutler, A., & van der Lugt, A. (1996). The representation of Japanese moraic nasals. *The Journal of the Acoustical Society of America*, *100*, 3831–3842.
- Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech in noise. *Ear and Hearing*, *30*, 653–661.
- Patel, A. D. (2008). *Music, language, and the brain*. NY: Oxford University Press.
- Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*, 688–691.
- Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception*, *24*, 37–47.
- Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*, 511–514.
- Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. *Journal of Educational Psychology*, *98*, 457–468.
- Schellenberg, E. G. (2009). Music training and nonmusical abilities: Commentary on Stoesz, Jakobson, Kilgour, and Lewycky (2007) and Jakobson, Lewycky, Kilgour, and Stoesz (2008). *Music Perception*, *27*, 139–143.
- Schellenberg, E. G., & Moreno, S. (2009). Music lessons, pitch processing, and g. *Psychology of Music*, *38*, 209–221.
- Schmand, B., Lindeboom, J., & van Harskamp, F. (1992). *NLV: Nederlandse leestest voor volwassenen: handleiding (NLV)*. Lisse: Swets & Zeitlinger.
- Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music raining facilitates pitch processing in both music and language. *Psychophysiology*, *41*, 341–349.
- Schroeger, E., & Winkler, I. (1995). Presentation rate and magnitude of stimulus deviance effects on human pre-attentive change detection. *Neuroscience Letters*, *193*, 185–188.
- Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, *17*, 675–681.
- Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, *261*, 22–29.
- Tierney, A., Bergeson, T. R., & Pisoni, D. B. (2008). Effects of early musical experience on auditory sequence memory. *Empirical Musicology Review*, *3*, 178–186.
- van Zuijen, T. L., Simoens, V. L., Paavilainen, P., Naatanen, R., & Tervaniemi, M. (2006). Implicit, intuitive, and explicit knowledge of abstract regularities in a sound sequence: An event-related brain potential study. *Journal of Cognitive Neuroscience*, *18*, 1292–1303.
- Vance, T. J. (1987). *An introduction to Japanese phonology*. Albany: State University of New York Press.
- Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., et al. (2005). To musicians, the message is in the meter – Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *Neuroimage*, *24*, 560–564.
- Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale – Revised*. New York: The Psychological Corporation.
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*, 420–422.
- Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, *170*, 223–231.
- Yee, W., Holleran, S., & Jones, M. R. (1994). Sensitivity to event timing in regular and irregular sequences: Influences of musical skill. *Perception & Psychophysics*, *56*, 461–471.