

# Effects of Language and Music Learning on Pitch and Duration Perception: an Experimental Investigation

Ilaria Montagni<sup>1</sup>, Andrea Peru<sup>2,\*</sup>

<sup>1</sup>Department of Public Health, University of Verona, Italy

<sup>2</sup>Department of Educational Sciences, University of Firenze, Italy

**Abstract** The ability to discriminate pitch and duration is fundamental for the processing of both music and spoken language. There is now consistent evidence that an intensive musical training may improve the ability to discriminate sound changes not only in musical but also in linguistic stimuli so that studying music may help learning foreign languages. By contrast, evidence for the reverse claim - a real transfer from linguistic experience to musical aptitude - is still lacking. This research investigates whether linguistic and musical competences can mutually influence each other especially as far as pitch and duration are concerned. Three groups of subjects (i.e. Naïves, Musicians and Bilinguals) performed a discrimination task on sound stimuli that were either linguistic (Japanese words) or musical (isolated notes), and differed in either duration (*tempi* and syllable lengths) or pitch (tones and stresses). Results showed that besides regular music practice, also an early but not a late exposure to foreign languages strongly improves pitch but not duration discrimination.

**Keywords** Pitch and Duration Processing, Music-Language Transfer

## 1. Introduction

A suggestive theory claims that language and music evolved from the same ancestor, the “musilanguage” system (Brown, 2001). Apart from any anthropological remark, there is no doubt that language and music share some structural elements and an overlapping structural processing mechanism (Fedorenko, Patel, Casasanto, Winawer and Gibson, 2009).

Among these structural elements, two features seem to be essential for music and language perception, learning and production: pitch and duration. It is quite pleonastic to discuss the role that pitch and duration play in music. By contrast, it may be worthy to note that in tone languages (i.e. Mandarin, Cantonese, Thai and many African languages) the same sequence of phonemes pronounced with subtle changes in pitch conveys very different meanings (e.g. in Mandarin: *ma* mother vs. *mā* horse). Analogously, duration is essential for the recognition and comprehension of verbal messages so that in quantity languages such as Finnish and Japanese, subtle changes in duration can imply relevant semantic differences (e.g. in Finnish: *puro* stream vs *puuro* porridge). It follows that the ability to discriminate these two features is fundamental for the analysis of linguistic stimuli as it is for musical ones (Patel, 2006).

Consistent evidence suggests that humans have an innate sensitivity to changes in duration and pitch along with other critical features of language and music like meter and timbre (see Trehub, 2003 and Levitin and Tirovolas, 2009, for reviews). For instance, it has been demonstrated that a few month-old infants can detect a small change in *tempo* of an isochronous sequence (Drake and Bertrand, 2001), as well as discriminate subtle pitch differences (Hannon and Trehub, 2005). As children mature, some musical abilities naturally improve, while others decline. In particular, while the sensitivity to harmonic structure increases (Trehub, 2003), the ability of pitch processing seems to become less accurate year after year, unless subjects are exposed to an intensive musical training (Tervaniemi, Just, Koelsch, Widmann and Schröger, 2005).

Even untrained adults (i.e. people who have not received a formal musical training), however, maintain a certain ability to process duration and pitch changes in all sound stimuli. For instance, undergraduate non-musician students presented with short extracts of Scottish music played at deviant (i.e. hastened or slackened) durations from the original tune turned out to be capable of making consistent duration judgments (Quinn and Watt, 2006). As for pitch processing, it is now accepted that, with the sole exception of amusic subjects (Stewart, 2008), also untrained adults can understand the difference between consonant and dissonant intervals (Bigand and Poulin-Charronnat, 2006; Hannon and Trainor, 2007). Quite obviously, language studies may enhance linguistic capacities and, in turn, an intensive musical training may prevent the natural decline of some abilities of

\* Corresponding author:

andrea.peru@unifi.it (Andrea Peru)

Published online at <http://journal.sapub.org/ijpbs>

Copyright © 2011 Scientific & Academic Publishing. All Rights Reserved

music processing. In this vein, linguistic expertise facilitates the processing of phonemically relevant acoustic cues in speech sounds while musical expertise facilitates the processing of subtle changes in pitch and in duration of harmonic sounds (see Marie, Kuala and Besson, 2010, for a review).

More interestingly, several studies in both children and adults provided convincing evidence that, besides the processing of musical stimuli, a specific musical training may improve the processing of speech sounds, too (Patel, 2008). For instance, in a behavioral and electrophysiological study, Milovanov and colleagues tested Finnish elementary school children's ability to process sound duration of both foreign (i.e. English) language sentences and melodic tunes and found that musical competence positively correlates with the linguistic one (Milovanov, Huotilainen, Esquef, Alku, Välimäki and Tervaniemi, 2009). In turn, Marques and colleagues (Marques, Moreno, Castro and Besson, 2007) presented some adult native French speakers (musicians and non-musicians) with a series of Portuguese declarative sentences ending with disyllabic words which could be prosodically congruous, weakly incongruous or strongly incongruous and asked them to judge the last word of the sentence as normal or strange. On the most difficult condition (i.e. weakly incongruous stimuli) musicians performed significantly better than non-musicians, likely because musical expertise can facilitate the detection of subtle pitch variations in word utterances (Magne, Schön and Besson, 2006). Moreover, musical expertise has proved to enhance the ability to extract prosodic information from spoken phrases (Thompson, Schellenberg and Husain, 2004), as well as the metric structure of words (Marie, Magne and Besson, 2011). In sum, musicians are supposed to detect pitch and/or duration violations better than non-musicians in both musical and linguistic stimuli (Schön, Boyer, Moreno, Besson, Peretz and Kolinsky, 2008).

While there is now a consistent corpus of literature investigating whether musical training improves non-musical (i.e. linguistic) abilities, few studies consider the reverse effect, namely whether linguistic training can improve music pitch and/or duration processing. In fact it seems that some linguistic competences can enhance music pitch and duration detection. For instance, it has been demonstrated that speaking Mandarin may help an individual acquire absolute pitch (Deutsch, Henthorn, Marvin and Xu, 2006), and that a native experience with a tone language enhances pitch encoding in both speech and non-speech contexts (Krishnan, Swaminathan and Gandour, 2009). However, supportive evidence is lacking for direct comparison of both language and music pitch and duration detection in groups of adults with either linguistic or musical skills. Moreover, in most studies subjects were tested in a foreign language that they were quite familiar with: for example Milovanov *et al.* (2009) adopted English for testing Finnish children and Marques *et al.* (2007) presented French adults with Portuguese sentences. By contrast, some recent researches have focussed on the comparison between tonal languages and tonality in music. In this vein, Japanese children have proved to be clearly

superior to their North-American peers in retaining the pitch level of familiar music (Trehub, 2003) and native Mandarin speakers have proved to be more sensitive than both English amateur musicians and English non-musicians to pitch changes in both music and language (Bidelman and Krishnan, 2009).

For the purpose of this study, we conducted an experimental investigation aimed at exploring whether linguistic and musical competences can mutually influence each other. In other words, whether subjects with linguistic or musical skills better process linguistic and/or musical stimuli, especially as far as the features of pitch and duration are concerned.

## 2. Methods

### *Participants*

Forty-eight healthy young adults (27 females and 21 males, ranging in age from 19 to 28 years) with a normal hearing ability gave their informed consent to participate in this study which was carried out in accordance with the guidelines of the Declaration of Helsinki.

All the 48 subjects were born in Italy and were Italian native speakers. According to their linguistic and musical experience and competence, they were divided into three groups of sixteen members each. The first group was labeled as "Bilinguals" and included subjects without any musical experience apart from compulsory school programs, but very proficient in one or more foreign languages. Six of them were subjects whose parents come from foreign countries and who had been exposed to their family's language (i.e. Romanian, Mandarin, American and British English) since birth. Given that none of them reported any further linguistic experience apart from compulsory school programs, following Lightbown and Spada (2006), we called them "Simultaneous Bilinguals". The remaining ten subjects studied foreign languages at University and got one or more languages certifications (i.e. Spanish, German, French and English) each corresponding to a C2 level according to the Common European Framework of Reference for Languages (CEFR). Following the definitions by Lightbown and Spada (2006), we called them "Sequential Bilinguals". The second group was labeled as "Musicians" and included subjects with a degree in Music Theory and Solfeggio, but no linguistic experience apart from compulsory school programs. All of them practiced their instruments (i.e. drums, piano, guitar and sax) at least two hours a day since the age of 6. The third group was labeled as "Naïves" and included university students with neither musical nor linguistic practice apart from compulsory school programs.

### *Materials and Procedure*

Our experiment assessed subjects' performance on two domains: Language and Music. In turn, each domain included two tasks: Pitch discrimination and Duration discrimination. In the overall four tasks, subjects were pre-

sented with a series of stimuli each formed by two sounds - either words or musical notes - separated by a fixed full period of 1 second. Participants were requested to verbally report whether the two sounds were same or different. The order of the four tasks (Language Pitch and Language Duration, Music Pitch and Music Duration) was counterbalanced across the subjects according to an A-B-B-A design. All subjects were tested singularly in a separate quiet room where they were comfortably seated on a chair back to back with the examiner. Such an arrangement made verbal communication very easy, meanwhile preventing subjects to receive non-verbal feedback from the examiner. Before each task a short training session was administered. The whole experiment lasted about 30 minutes, including a short pause between tasks.

A commercial music notation software (GarageBand, Apple, Inc.) was used to generate all musical stimuli whereas linguistic stimuli were built from Japanese Grammar CDs (Banno, 1999). Both musical and linguistic stimuli were stereo sampled at 22050 Hz and stored in a wave format with 16-bit length by means of an open source software (Audacity, Inc.). Another commercial software program (E-Prime, Psychology Software Tools, Inc.) was used to implement the experimental paradigm. All the experiments were conducted on an IBM compatible notebook provided with a standard Toshiba audio card. Professional headphones set at normal volume levels served to deliver the stimuli. The examiner used the keyboard of the notebook to enter the participants' answers and to run the experiment.

Hereafter we provide further details on the stimuli used in the different tasks. The full list of the stimuli is reported in the Appendix.

### 1. Language Stimuli

All Language stimuli consisted of Japanese real words built from three mono recordings made by three (two males and a female) native Japanese speakers. More precisely, each speech stimulus contained two Japanese words which could be identical or different in consonant stress (pitch) or length (duration).

We chose Japanese because it is a pitch accent language; that is, the meaning of homographic words can change with different accent positions, in a much more restricted way than tonal languages like Chinese (Yip, 2002). Japanese is actually built on specific quantitative syllables and rhythmic flows of words. Such a rhythm is given by adequate and recurring intervals between syllables or words which attribute a melodic fluency to the sequence of each sentence (Jomori and Hoshiyama, 2009). Furthermore, Japanese is also a quantitative language, because consonant and vowel quantities carry semantic information and the meaning of a word is related also to the correct production of the phoneme duration. So, in Japanese vowel and consonant quantities are fundamental not only from a phonetic point of view, but also - and above all - from a semantic perspective since the meaning of a word completely changes if the duration of the phoneme is longer or shorter. Finally, Japanese was chosen

for being a language that all the participants were unfamiliar with.

In the Language Pitch task twelve words were employed to generate the stimuli. These words were combined for a total of twelve couples of identical words (i.e. same stimuli) and twelve stimuli made of differently stressed words (i.e. different stimuli). Different words in a pair presented deviant stress position: the stress could be on the first syllable, or on the second syllable or just absent. Considering that it is easier to recognize the difference between a word accented on the first syllable and the homograph word accented on the second syllable than the difference between stressed and unstressed words (Howell, 1993), according to the stress position in the words of a pair, different stimuli were classified as easy (e.g. *imà* vs. *ima*) or difficult (e.g. *sake* vs. *saké*). In each experimental session, the twelve same and twelve different stimuli were presented according to a completely random sequence. Then, this cycle of trials was repeated for a total of forty-eight experimental stimuli.

In the Language Duration task eight words were combined to generate a total of eight different and eight same stimuli. In the different stimuli the two words were homographic and homophonous apart from a consonant which was doubled in one part of the pair (e.g. *soto* vs. *sotto*, *sanen* vs. *sannen*), so that the two words were only slightly deviant in duration and the degree of difference was the same across the eight different stimuli. In each experimental session, same and different stimuli were presented according to a completely random sequence. In order to get forty-eight experimental stimuli, three cycles of trials were administered.

### 2. Music Stimuli

All Music stimuli were generated through an electronic grand piano tone and consisted of a sequence of two musical notes which could be identical or different in pitch or duration.

With the explicit aim to discard the culture-specific principles that govern pitch and duration relationships in melodic and harmonic progressions (Tramo, Cariani, Delgutte and Braidà, 2001), we preferred to use isolated notes instead of chords ("vertical combinations") and music lines ("horizontal combinations"). Thus, subjects were presented with pairs of isolated sounds which could differ each other for the sole parameter of pitch or duration.

In the Music Pitch task six quarter notes in G clef were employed. These musical notes were combined for a total of twelve different and six same stimuli. As with Language stimuli, different stimuli were ranked according to the degree of difference between notes as easy (i.e. a tone and a half, like the stimulus E-G), difficult (i.e. a tone, like the stimulus D-C) and very difficult (i.e. a semitone, like the stimulus E-F). In each cycle, the six same stimuli were doubled to obtain an equal number of same and different stimuli which were all presented according to a completely random sequence. Then, the cycle of trials was repeated for a total of forty-eight stimuli for each experimental session.

In the Music Duration task three *tempi* (i.e. 1/8 note, 1/4

note and 1/2 note) were combined to generate three same and six different stimuli, ranked according to the degree of difference between notes as easy (e.g. 1/2 note vs. 1/8 note), difficult (e.g. 1/4 note vs. 1/2 note) and very difficult (1/4 vs. 1/8). The same stimuli were doubled leading to a total of twelve stimuli presented according to a completely random sequence. Four cycles of trials were run for a total of forty-eight stimuli for each experimental session.

### 3. Results

Given that the level of difficulty was different across tasks and considering that the main question of interest was whether the performance of the three groups of participants was the same or different across tasks, we analyzed each task separately using univariate ANOVA with subjects' accuracy as the dependent variable and Group (three levels: Bilinguals, Musicians, and Naïves) as the *between-subjects* factor. Bonferroni correction for multiple comparisons was applied and a  $p$ -value of  $<.05$  was considered to indicate statistical significance. Before analyzing the data, however, further methodological considerations have to be taken into account. In each task, on each trial, subjects were requested to judge whether the two stimuli in a pair were same or different. In other words, they had to detect a difference. It follows that, qualitatively speaking, subjects' responses could belong to any one of these four categories: Hits, Misses, Correct Rejections and False Alarms. Hits occurred when subjects correctly detected differences (i.e. they responded "different" on different trials), and Misses occurred when subjects missed to detect differences (i.e. they responded "same" on different trials). Analogously, Correct Rejections occurred when subjects avoided to report a difference when it was absent (i.e. they responded "same" on same trials), and False Alarms occurred when subjects reported the presence of a difference when it was absent (i.e. they responded "different" on same trials). However, quantitatively speaking, it is noteworthy that in three out of the four tasks (i.e. Language Duration, Language Pitch, and Music Pitch) subjects were extremely accurate on same trials with a number of False Alarms absolutely negligible (i.e. less than 2%). For such a reason, in these three tests we only considered subjects' performance on different trials by using the number of Misses as index of accuracy. By contrast, in the Music Duration task, besides Misses, we also recorded a large number of False Alarms. With the aim to provide a truer representation of discrimination sensitivity of different subjects, we used  $d'$  [ $d' = z(\text{FA}) - z(\text{H})$ ] to take into account both Hit and False Alarm rates. Hereafter we describe the results of each task listed according to the number of errors.

#### 1. Language Duration

In this task, participants made very few errors, if any (overall accuracy = 97%). Although the number of Misses was too small to be entered into any quantitative analysis, some qualitative observations on the pattern of performance

are possible. First, Misses were more likely to occur the slighter was the difference between the stimuli so that the more difficult the item the more frequent the errors. Secondly - and most importantly - the pattern of performance varied across the different groups: while all Musicians performed flawless, 5 Naïves and 7 Bilinguals (2 Simultaneous- and 5 Sequential-Bilinguals) made some Misses. None of them, however, missed more than 25% of different trials so that the average of Misses was very low for both Naïves and Bilinguals (mean percentage of Misses = 3% and 6%, respectively) and no difference was observed between the two subgroups of Bilinguals.

#### 2. Music Pitch

Participants were very accurate also in this task (overall accuracy = 94%). Like in the previous task, the error rate tended to parallel the difficulty of the trials, and the pattern of performance varied across groups. Once again, all Musicians performed flawless, while 10 Naïves and 9 Bilinguals made several errors leading to a significant effect of the factor Group [ $F(2,45) = 5.151$ ;  $p = .010$ ]. However, although Naïves (mean percentage of misses = 11%) performed overall worse than Bilinguals (mean percentage of misses = 5%), only the difference between Naïves and Musicians reached statistical significance ( $p = .007$ ). Finally, it is interesting to note that the two sub-groups of Bilinguals performed quite differently from each other in this task. Indeed, all Bilinguals who made errors were Sequential-Bilinguals (mean percentage of misses = 8%). In other words, while all Simultaneous Bilinguals performed flawless like Musicians, all Sequential-Bilinguals but one made several Misses like Naïves.

#### 3. Language Pitch

Participants' performance markedly worsened in the Language Pitch task with only 8 Musicians, 3 Bilinguals and 2 Naïves making no errors. Notwithstanding the considerable inter-subjects variability found in this task, ANOVA revealed a significant effect of the factor Group [ $F(2,45) = 4.515$ ;  $p = .016$ ]. Indeed, Musicians (mean percentage of Misses = 5%) turned out to be overall more accurate than Bilinguals (mean percentage of Misses = 13%) and Naïves (mean percentage of Misses = 23%), although only the differences between Musicians and Naïves reached significance ( $p = .013$ ). Also in this task, however, a strong difference in the performance of Simultaneous Bilinguals compared with Sequential Bilinguals was detected. Actually, all the Bilinguals who performed flawless were Simultaneous Bilinguals. Moreover, while the performance of Sequential Bilinguals was very similar (mean percentage of Misses = 26.5%) to that of Naïves, the performance of Simultaneous Bilinguals (mean percentage of Misses = 3.5%) was as good as that of Musicians. Finally, it is worth noting that the difference between Simultaneous and Sequential Bilinguals was statistically significant even when the performance of the six Simultaneous Bilinguals was compared with that of the six Sequential Bilinguals best scorers by means of an independent samples T-test: [ $t(10) = 2.341$ ,  $p$

= .041].

#### 4. Music Duration

As mentioned above, differently from the other three tasks, in the Music Duration task, besides Misses, we also recorded a large number of False Alarms. Because of this, we used  $d'$  to take into account both Hit and False Alarm rates [ $d' = z(\text{FA}) - z(\text{H})$ ]. According to this criterion,  $d'$  values near zero indicate chance performance, while high absolute values of  $d'$  indicate a large sensitivity to discriminate same and different pairs of stimuli. Before analyzing the data, however, a further methodological consideration must be taken into account. Computing  $d'$  is quite difficult when a respondent detects all differences giving a Hit rate = 1.0 and/or never reports the presence of a difference when absent, giving a False Alarm rate = 0. To avoid these problems, we adopted a conventional adjustment according to which subjects who made no Misses nor False Alarms were considered instead to have made 1 Miss and 1 False Alarm. Giving that there were 24 same and 24 different stimuli, for these subjects Hit and False Alarm rates were = .960 and = .040, respectively. For all the participants, but one Naïf and one Simultaneous Bilingual,  $d'$  values were positive ranging from .616 to 3.502 thus revealing a high inter-subjects variability. For such a reason the univariate ANOVA, with  $d'$  values as the dependent variable and Group as the *between-subjects* factor, was far from significance. That is the three groups of subjects performed in a quite similar way with a % of Hits ranging from 71% for Naïves to 78% for Bilinguals and 80% for Musicians and a % of False Alarms which varied from 25% for Naïves to 28% for Musicians and 40% for Bilinguals. Finally, no difference was observed between the performance of Simultaneous and Sequential Bilinguals.

## 4. Discussion

There is now a wide consensus that humans have an innate predisposition for processing linguistic and musical stimuli (e.g. Levitin and Tirovolas, 2009). In particular, along with other critical features like meter and timbre, infants can easily detect changes in duration (Drake and Bertrand, 2001) and pitch (Hannon and Trehub, 2005). These abilities naturally tend to decline although they are never entirely lost even in untrained adults (Tervaniemi *et al.*, 2005). A specific (i.e. linguistic and musical) training, however, may prevent such a decline by enhancing the processing of linguistic and musical stimuli, respectively (Marie *et al.*, 2010). More interestingly, linguistic and musical domains may influence each other. In this vein there is now increasing evidence for the proposition that music can facilitate the acquisition of a second language (Sleve and Miyake, 2006). It seems too simplistic to explain such a beneficial effect of music as only due to an increased level of arousal and attention. More likely, learning a foreign language can be positively affected by both motivational and structuring proprieties of songs (Schön, Boyer, Moreno, Besson, Peretz, and Kolinsky,

2008). Consistent with this, several studies reported an advantage for musicians compared to non-musicians in the processing of speech sounds (Patel, 2008).

One should expect that as musical expertise improves linguistic abilities, so does linguistic expertise with musical abilities. To the best of our knowledge, only few studies have addressed this issue so far. Although some reliable data suggest that speaking a tone language may enhance pitch encoding in both musical and verbal context (Deutsch *et al.*, 2006; Krishnan *et al.*, 2009; Lee and Lee, 2010), conclusive evidence is still lacking as for the effects of linguistic training on subjects' ability to process musical stimuli.

Here we set out to investigate the mutual effects of linguistic and musical training by directly comparing how people with different levels of musical and linguistic expertise process musical and linguistic stimuli. For this purpose, three groups of participants with different levels of musical or linguistic expertise were tested: those with musical training (i.e. Musicians) those with linguistic training (i.e. Bilinguals, further distinguished into Simultaneous and Sequential Bilinguals), and those who were monolinguals and had no received extensive musical training (i.e. Naïves). All these subjects were asked to make same/different judgments on pairs of linguistic (Japanese words) or musical (isolated notes) stimuli which varied (or did not differ at all) in either duration or pitch.

The results were very interesting. First, subjects' performance varied across different tasks so that they were overall extremely accurate in the Language Duration task, made an increasing number of errors across Music Pitch and Language Pitch tasks, and scored markedly worse in the Music Duration task. Although a direct comparison between tasks is not possible, it seems that - at least under the conditions of our experiment - duration discrimination was easier than pitch discrimination in the linguistic domain whereas the opposite was true in the musical domain, as pitch discrimination was easier than the duration one. Another hypothesis, not mutually exclusive, can be put forward to explain these findings. Listeners rarely experience linguistic and musical stimuli as isolated sounds (Drake and Bertrand, 2001), like the single notes we used in our musical tasks. In fact, we are usually exposed to sequences of verbal or musical sounds (words and music lines, at least) and the ability to discriminate some fundamental elements allows us to catch the pauses between words. In this vein, our claim can be re-phrased by stating that under the ecological conditions in which subjects were requested to discriminate *embedded sounds* (i.e. linguistic tasks), duration discrimination turned out to be easier than pitch discrimination, while under the somehow unnatural conditions in which we investigated subjects' absolute capacity to discriminate *sounds in isolation* (i.e. musical tasks), they were much more accurate in discriminating pitch than duration. This issue could be specifically addressed in a further investigation in which musical notes are to be compared with syllables (*sounds in isolation*) and musical lines are compared with words (*embed-*

*ded sounds*).

Secondly, but most importantly, our research did reveal reliable differences between the different groups of subjects across the different tasks. That is, the three groups behaved quite differently from each other when requested to discriminate pitch either in the musical or in the linguistic domain, while they performed in a more similar way when asked to discriminate the duration of linguistic or musical stimuli. Indeed, while in the Music Duration task the percentages of Hits and False Alarms were absolutely comparable across groups leading us to the conclusion that the performances of Naïves, Bilinguals and Musicians were substantially equivalent, a subtle difference between groups was detected in the Language Duration task: Musicians performed flawless, while some Naïves and Bilinguals made some errors. With the caution that must be used in interpreting negative results (i.e. those from Music Duration task) as well as differences not supported by a statistical evidence (i.e. those from Language Duration task), these findings can be explained by arguing that even subjects with neither linguistic nor music training can distinguish syllables duration very accurately. Nonetheless, a musical, rather than a linguistic, training can be effective in further improving this spontaneous ability (Sleve and Miyake, 2006). On the other hand, neither a musical, nor a linguistic intensive training has any effect on the subjects' ability of discriminating music duration, at least when isolated notes are compared.

As said before, more consistent differences between groups were found in tasks requiring a pitch discrimination. On both Music and Language Pitch task, Musicians performed at or near ceiling, thus they turned out to be overall more accurate than Bilinguals who, in turn, were overall more accurate than Naïves. Although only the differences between Musicians and Naïves reached statistical significance, these findings seem to suggest that a musical and, even in a lesser degree, a linguistic intensive training may improve the subjects' ability of pitch discrimination in both the musical and the linguistic domains.

The conclusion that musical training is correlated with better pitch discrimination cannot be taken as especially surprising or novel. Indeed, several studies demonstrated a consistent advantage for musicians compared to non-musicians in processing musical and linguistic stimuli in general (e.g. Schön *et al.*, 2008; Patel, 2008) and in discriminating pitch in particular (e.g. Magne *et al.*, 2006; Marie *et al.*, 2010). By contrast, only few studies demonstrated the reverse effect, namely that linguistic competences can enhance pitch encoding in both speech and non-speech contexts (e.g. Krishnan *et al.*, 2009). Thus, our findings add supportive evidence to the hypothesis that a language to

music transfer is possible and extensive linguistic experience may improve, to some extent, individuals' ability to discriminate pitch in music. One should claim, however, that such an evidence is quite weak, since neither in the Music, nor in the Language Pitch tasks, the difference between Bilinguals and Naïves reached statistical significance. Bilinguals, however, did not behave as a homogeneous group. On the opposite, the two subgroups of Bilinguals performed quite differently from each other: while the performance of Sequential Bilinguals was absolutely undistinguishable from that of Naïves, the performance of Simultaneous Bilinguals was as accurate as that of Musicians. Although the sample of the two subgroups was too small to allow us to draw any definite conclusion, the differences were too clear-cut to be mere coincidence. Moreover, it seems unlikely that Simultaneous Bilinguals differed from Sequential Bilinguals in motivation or other non-specific factors such as basic auditory ability etc. A more plausible explanation is that Simultaneous vs. Sequential Bilinguals differences reflect different effects of natural "early" bilingualism compared to artificial "late" bilingualism (Bialystok and Depape, 2009). In particular, these findings corroborate the idea that, analogously to musical competence, linguistic competence - provided it has been acquired in early infancy and childhood - can improve subjects' abilities to detect slight pitch variations in both the linguistic (Kaushanskaya and Viorica, 2009) and the musical (Bidelman and Krishnan, 2009) domains.

This claim, however, needs to be further substantiated by collecting data from a larger sample of Simultaneous and Sequential Bilinguals. Even more interesting, different groups of Simultaneous Bilinguals should be tested to address the issue whether linguistic competence differs depending on the spoken native languages especially with regard to tone languages. In line with recent suggestions (Schellenberg and Trehub, 2008; Pfordresher and Brown, 2009), it would be interesting to observe whether studying "just any" second language can enhance pitch or duration perception or, vice-versa, the exposure to different mother tongues where pitch and duration discrimination is crucial (i.e. tone and pitch accent languages and quantitative languages) can imply a different development of the ability to discriminate musical sounds also in non-musicians.

## ACKNOWLEDGMENTS

We are grateful to all the participants, the principals and staff of the schools for their participation. We also wish to acknowledge Giovanni Bartoli and Licia Rossi Lumachi for their help with the preparation of the stimuli.

## Appendix

**Table 1.** Language Pitch stimuli.

DIFFERENT COUPLES	DEGREE OF DIFFERENCE
hàshi (chopsticks) - hashi (edge)	Easy
hashi (bridge) - hashi (edge)	Easy
kaki (persimmon) - kàki (oyster)	Easy
kaki (persimmon) - kaki (fence)	Easy
sàke (salmon) - sake (alcohol)	Easy
sake (alcohol) - sàke (salmon)	Easy
hashi (bridge) - hàshi (chopsticks)	Difficult
hàshi (chopsticks) - hashi (bridge)	Difficult
ima (now) - imà (living room)	Difficult
imà (living room) - ima (now)	Difficult
kàki (oyster) - kaki (fence)	Difficult
nihòn (Japan) - nihon (two sticks of)	Difficult

As for the "same" couples, each word was repeated twice.  
Ex.: *ima* (now) - *ima* (now); *kàki* (oyster) - *kàki* (oyster)  
and so on.

**Table 2.** Language Duration stimuli.

DIFFERENT COUPLES	DEGREE OF DIFFERENCE
issei (all at once) - isei (opposite sex)	Easy
isei (opposite sex) - issei (all at once)	Easy
maki (scroll) - makki (last period)	Easy
makki (last period) - maki (scroll)	Easy
sannen (three years) - sanen (three yen)	Easy
sanen (three yen) - sannen (three years)	Easy
soto (outside) - sotto (softly)	Easy
sotto (softly) - soto (outside)	Easy

As for the "same" couples, each word was repeated twice. Ex.: isei (opposite sex) - isei (opposite sex); makki (last period) - makki (last period) and so on.

**Table 3.** Music Pitch stimuli.

DIFFERENT COUPLES	DEGREE OF DIFFERENCE
B-C	Easy
C-B	Easy
F-E	Easy
E-F	Easy
F-G	Difficult
G-F	Difficult
D-C	Difficult
C-D	Difficult
E-G	Very Difficult
G-E	Very Difficult
D-B	Very Difficult
B-D	Very Difficult

As for the "same" couples, each note was repeated twice. Ex.: B - B; C - C and so on.

**Table 4.** Music Duration stimuli.

DIFFERENT COUPLES	DEGREE OF DIFFERENCE
1/2 - 1/8	Easy
1/8 - 1/2	Easy
1/4 - 1/2	Difficult
1/2 - 1/4	Difficult
1/8 - 1/4	Very difficult
1/4 - 1/8	Very difficult

As for the "same" couples, each note was repeated twice. Ex.: 1/8 - 1/8 and so on.

## REFERENCES

- [1] Banno, E. (1999). *Genki: An Integrated Course in Elementary Japanese Vol. 2*. Tokyo: Japan Times
- [2] Bialystok, E., & Depape, A.M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 565-74
- [3] Bidelman, G.M., & Krishnan, A. (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *The Journal of Neuroscience*, 42, 13165-13171
- [4] Bigand, E., & Poulin-Charronnat, B. (2006). Are we 'experienced listeners'? A review of the musical capacities that do not depend on formal musical training. *Cognition*, 100, 100-130
- [5] Brown, S. (2001). Are Music and Language Homologues? *Annals of the New York Academy of Sciences*, 930, 372-374.
- [6] Deutsch, D., Henthorn, T., Marvin, E.W., & Xu, H.-S. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *Journal of the Acoustical Society of America*, 119, 719-722

- [7] Drake, C., & Bertrand, D. (2001). The Quest for Universals in Temporal Processing in Music. *Annals of the New York Academy of Sciences*, 930, 17-27
- [8] Fedorenko, E., Patel, A., Casasanto, D., Winawer, J. & Gibson, E. (2009). Structural integration in language and music: evidence for a shared system. *Memory & Cognition*, 37, 1-9
- [9] Hannon, E.E., & Trainor, L.J. (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Science*, 11, 466-472
- [10] Hannon, E.E., & Trehub, S.E. (2005). Metrical categories in infancy and adulthood. *Psychological Science*, 16, 48-55
- [11] Jomori, I., & Hoshiyama, M. (2009). Auditory brain response modified by temporal deviation of language rhythm: An auditory event-related potential study. *Neuroscience Research*, 65, 187-193
- [12] Kaushanskaya, M., & Viorica, M. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, 16, 705-710
- [13] Krishnan, A., Swaminathan, J., & Gandour, J.T. (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience*, 21, 1092-1105
- [14] Lee, C.Y., & Lee, Y.F. (2010). Perception of musical pitch and lexical tones by Mandarin-speaking musicians. *Journal of the Acoustical Society of America*, 127, 481-490
- [15] Levitin, D.J., & Tirovolas, A.K. (2009). Current Advances in the Cognitive Neuroscience of Music. *Annals of the New York Academy of Sciences*, 1156, 211-231
- [16] Lightbown, P. M., & Spada, N. (2006). *How languages are learned*. Oxford: Oxford University Press
- [17] Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, 18, 199-211
- [18] Marie, C., Kuala, T., & Besson, M. (2010). Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. *Cortex*, 27, 1-11
- [19] Marie, C., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience*, 23, 294-305
- [20] 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
- [21] 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
- [22] Patel, A.D. (2006). Language, music, syntax and the brain. *Nature Neuroscience*, 6, 674-681
- [23] Pfordresher, P.Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics*, 71, 1385-1398
- [24] Quinn, S., & Watt, R. (2006). The perception of tempo in music. *Perception*, 35, 267-280
- [25] Schellenberg, E.G., & Trehub, S.E. (2008). Is there an Asian advantage for pitch memory? *Music Perception*, 25, 241-252
- [26] Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, 106, 975-983
- [27] Sleve, L.R., & Miyake, A., (2006). Individual differences in second language proficiency: does musical ability matter? *Psychological Science*, 17, 675-681
- [28] Stewart, L. (2008). Fractionating the musical mind: Insights from congenital amusia. *Current Opinion in Neurobiology*, 18, 127-130
- [29] Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch- discrimination accuracy in musicians vs. non-musicians: an event-related potential and behavioral study. *Experimental Brain Research*, 161, 1-10
- [30] Thompson, W. F., Schellenberg, E.G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, 4, 46-64
- [31] Tramo, M.J., Cariani, P.A., Delgutte, B., & Braida, L.D. (2001). Neurobiological Foundations for the Theory of Harmony in Western Tonal Music. *Annals of the New York Academy of Sciences*, 930, 92-116
- [32] Trehub, S.E. (2003). Toward a developmental psychology of music. *Annals of the New York Academy of Sciences*, 999, 402-413
- [33] Yip, M. (2002). *Tone*. New York: Cambridge University Press