MUSICIANS' AND NON-MUSICIANS' SPEAKING FUNDAMENTAL FREQUENCY RANGES IN AN ORAL READING TASK

AZ ALAPHANG TARTOMÁNYA ZENÉSZEK ÉS NEM ZENÉSZEK HANGOS OLVASÁSÁBAN

Dr. Ákos Gocsál
University of Pécs, Faculty of Music and Visual Arts, Institute of Music, Pécs, Hungary
Hungarian Academy of Sciences, Research Institute for Linguistics, Department of Phonetics, Budapest, Hungary

ABSTRACT
As a possible transfer effect of music learning, the we raise the possibility that musicians use the prosodic elements of speech with a higher degree of awareness, which, among other consequences, may result in a greater variety in the melody of speech. Ten musically untrained young adult males and ten age-matched undergraduate or graduate students of instrumental music read the same passage. The fundamental frequency (f_0) ranges did not differ significantly between the two groups, even when the f_0 ranges were narrowed by deciles around the medians. The lack of the expected difference suggests that the possible transfer effect does not exist with regard to oral reading. However, the methods applied here may not be sensitive enough to demonstrate the expected transfer effect.
Key words: transfer effect, oral reading, fundamental frequency

ABSTRACT
A zenetanulás lehetséges transzferhatásaként azt vetettük fel, hogy a zenészek tudatosabban, nagyobb változatossággal használják beszédükben a prozódiai elemeket, ami – többek között – dallamosabb beszédben jelentkezhet. A kutatásban tíz zeneileg képzetlen fiatal felnőtt férfi és tíz hangszeres zenei tanulmányokat végző egyetemi hallgató ugyanazt a szöveget olvasta fel. Az alaphang-tartományok (f_0), illetve decilisek szerint szűkített tartományok szélessége nem mutatott szignifikáns különbséget. Így a felolvasásra vonatkozó, fent említett transzferhatást nem tudtuk kimutatni, illetve elképzelhető, hogy ha létezik is ez a transzferhatás, az itt alkalmazott módszerek nem elég érzékenyek a kimutatásához.
Kulcsszavak: transzferhatás, hangos olvasás, alaphang

INTRODUCTION

Among the great variety of skills and abilities that musical training may affect, those related to spoken communication represent a special field. Researchers usually pose two kinds of questions: (1) Is musical training an advantage in speech perception? (2) Do musicians speak differently than non-musicians? Is the acoustic structure of musicians' speech more musical than that of non-musicians?

The majority of the studies report that musicians outperform non-musicians in perceptual experiments such as pitch, timbre, or temporal-interval discrimination tasks (Micheyl et al. 2006, Banai et al. 2012, Boebinger et al. 2015). In many cases, similar results were obtained when speech samples were used as stimuli: musicians were better at discriminating vocal timbres (Chartrand & Belin 2006), morphed speech sounds (Sadakata & Sekiyama 2011), and listening comprehension tasks in a foreign language learning environment (Vyspinska 2019). In other cases, such as speech-in-noise perception (Coffey et al. 2017, Madsen et al. 2019), conflicting results were found and there are experiments where no differences were demonstrated, like in speaker age estimation based on voice (Gocsál 2018). As a theoretical background, Patel's (2014) OPERA hypothesis offers a good a basis. This hypothesis suggests that (i) perceptual mechanisms for music and speech use shared brain networks, (ii) music places higher demands on the auditory perceptual process, and finally, (iii) music engages that process with emotion, repetition, and focussed attention. Musicians may therefore be expected to outperform non-musicians in auditory tasks, but this has not been proven in many cases.

There are significantly fewer authors who have dealt with the second question. Among them, Pastuszek-Lipińska (2007) tested musicians' and non-musicians' ability to imitate foreign language phrases. In a repetition task, musicians encountered fewer difficulties, they produced significantly more correct repetitions and even a few years of musical education in the past had a positive effect on the performance of the subjects. Another study examined differences in the normalized spectra of speech and song, separately in musicians and non-musicians. Using the 12-tone scale, spectral peaks were analysed and non-musicians were found to have more non-predicted peaks (i.e. occurring at intervals not predicted by the 12-tone scale) than non-musicians, especially vocalists, which suggests an increased musicality of the voice. Other parameters, including f_0 variability, were not affected by musical experience (Stegemöller et al. 2008). In a perceptual study, musicians and non-musicians were used as speakers. Musicians were judged as speaking with stronger emphasis, but the difference was not significant (Amir et al, 2020).

In the present study we also address the second question. Can we say that musicians, because of their musical experience, speak differently? Do they use more variation in the prosodic elements of speech? We hypothesize that musicians' speech is more melodic than that of non-musicians. Here, rather than analysing individual melodic patterns, we apply a statistical approach to establish which speech sample is "more melodic". If individual frequency values of speech, measured at regular intervals, spread over a wider range and their distribution is relatively flat, we can infer that this speech sample is melodic. If frequency values are limited to a narrower range, we can infer that there is a little variation of frequency around the mean value, which is a sign of more monotonous speech.

METHODS

Speech samples of ten middle-class male speakers (age range: 19-26 years) from the BEA Spontaneous Speech Database (Hungarian Academy of Sciences, Research Institute for Linguistics, Department of Phonetics, Gósy et al. 2012) were randomly selected. All of them spoke standard Hungarian and they were university students or already had university degrees. In the recordings, they read orally an article from a magazine (1804 characters). 10 male musicians (age range: 20-26 years), who were BA or MA students of the Institute of Music, Faculty of Music and Visual Arts, University of Pécs, were asked to read the same passage. They were players of piano, guitar, and string instruments with at least eight years of formal music education. Singers and players of wind instruments were excluded. The subjects read the passage in a quiet room, and the readings were recorded using a head-mounted Røde HS2 microphone (connected to a laptop computer via a Behringer UMC202 audio interface) and the Audacity software set to a sampling rate of 44 kHz and 16 bits, .wav format. None of the speakers smoked and had any speech production disorders.

The sound files were first loaded into the Praat 6.0.52 software (Boersma & Weenik 2019) where the pitch contours were visually inspected in the frequency range between 20 and 600 Hz. With all other parameters, factory settings were used. Using the *Pitch listing* command, f_0 values were listed, saved, and finally loaded into the SPSS 23 software. The amount of the f_0 values ranged between 1500 and 2400 per speaker. Slow speakers and those who occasionally repeated a word or a phrase produced more f_0 values. Then, histograms showing the distribution of the frequency values were created. The method is demonstrated in Fig. 1 and 2, using a simplified hypothetical example. The left diagram shows portion of an intonation curve. Fundamental frequency of 100 Hz is measured three times (0.01, 0.07, 0.08 s), while 115 Hz is only found once (0.05 s). Thus, the height of the bar at 100 Hz is 3 units, while that belonging to 115 Hz is only one unit (Fig. 2.).

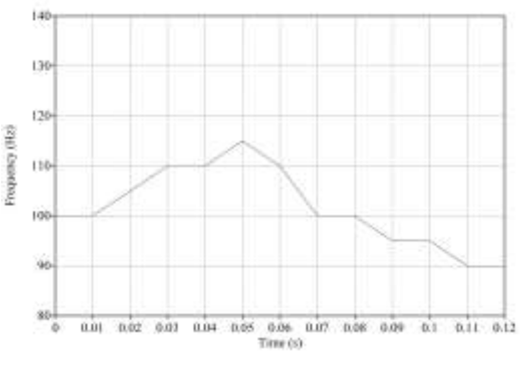

Fig. 1. A hypothetical intonation curve

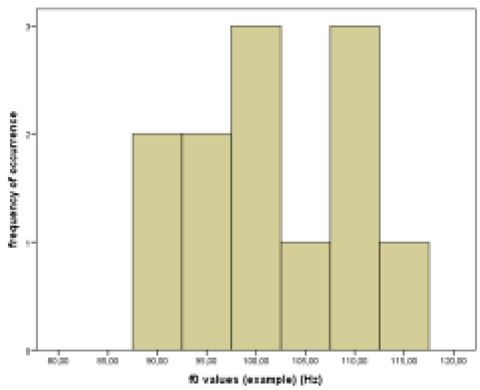

Fig. 2. Histogram of f_0 values shown in Fig. 1.

With real sound recordings, the *x*-axis was divided into small uniform ranges (e.g. a range of 10 Hz) and the bars show the number of f_0 values measured within those ranges. Fig. 3 shows a typical histogram. It can be seen that in this case there were more than 600 f_0 values between 120 and 130 Hz, while there only a very few values between 190 and 200 Hz. Before advancing to a further analysis of the data, two important observations were made. First, histograms represented a bimodal distribution in many cases. While there is a peak around 120-130 Hz, a secondary cluster of frequencies was observed below 100 Hz, with a peak around 50-60 Hz. The reason why so low frequencies occurred is not because of the melody of reading, but it is irregular phonation, which results in unexpectedly low f_0 values in the histogram (Markó 2013: 19). Since we are only working with regular (or modal) phonation, these low f_0 values were excluded from further analysis. Similarly, we found several instances where unexpectedly high f_0 values occurred. After the inspection of these values, it was found that in most of the cases Praat had falsely identified the voiceless fricatives [s] or [ʃ] and other sounds or noises as voiced speech sounds with very high f_0. Thus, these false f_0 values were also excluded. There are a few instances of such f_0 values between 500 and 600 Hz in Fig 3. Although the number of these false f_0 values was usually low, they may have significantly biased the results of statistical calculations.

As soon as f_0 values belonging to irregular phonation and the falsely identified ones were excluded from the datasets, Kolgomorov-Smirnov normality tests were carried out which revealed that the distribution of f_0 data significantly differed from normal distribution ($p < .001$) in every speech sample, indicating a strong positive skewness. For further calculations, minimum and maximum values of f_0, f_0 ranges, medians, and deciles of f_0 distributions were established.

RESULTS

Although research questions did not include possible differences in the mean f_0 between musicians and non-musicians, Mann-Whitney *U* test was applied to examine if such a difference exists. Table 1 includes the most important statistical parameters for the medians for musicians and non-musicians separately.

Table 1. Descriptive statistics of f_0 values

	mean (Hz)	SD	min (Hz)	max (Hz)
musicians	107,1	11,92	83,23	175,14
non-musicians	112,41	14,07	74,54	187,24

The data show that the mean values of the medians slightly differed, musicians had slightly lower fundamental frequencies. However, this difference is not significant ($U = 37$, $p = .326$). The observed differences therefore cannot be attributed to musicianship. Next, frequency ranges (i.e. $f_{0max} – f_{0min}$) that the musicians and non-musicians used were compared. The boxplots in Fig. 4 represent the distribution of the range values.

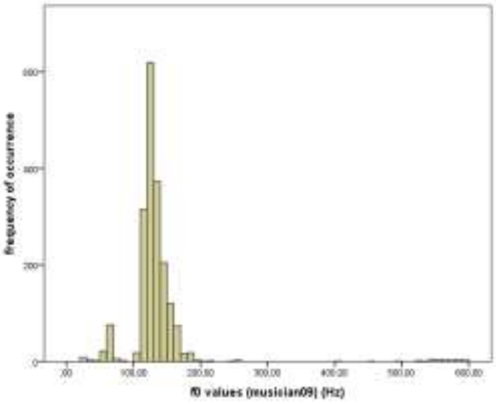

Fig. 3. f_0 distribution before excluding irregular and false f_0 values

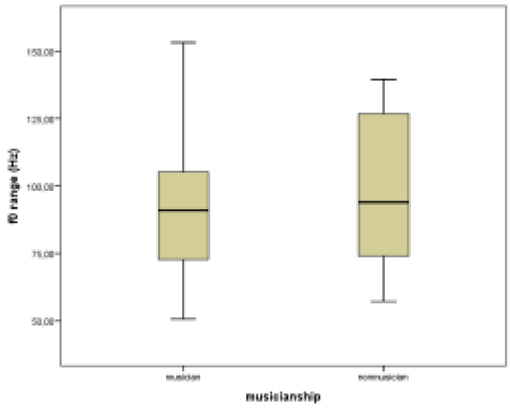

Fig. 4. Comparison of f_0 ranges in musicians and non-musicians

A Mann-Whitney *U* test revealed no significant difference between the two groups ($U = 43$, $p = .631$) which means that musicians did not use a wider range of fundamental frequency than non-musicians. However, a wider frequency range that a speaker uses does not necessarily mean that a distribution is flatter, i.e. speech is more melodic. Fig. 5 and 6 show two interesting examples. Fig. 6 shows that the f_0 distribution of musician10 has a very long right tail, which results in a wide frequency range, but higher frequency values occur only occasionally. One may then infer that despite the wide frequency range, this speaker's speech may not sound significantly "more melodic" than that of musician9 (Fig. 5), whose frequency range is narrower.

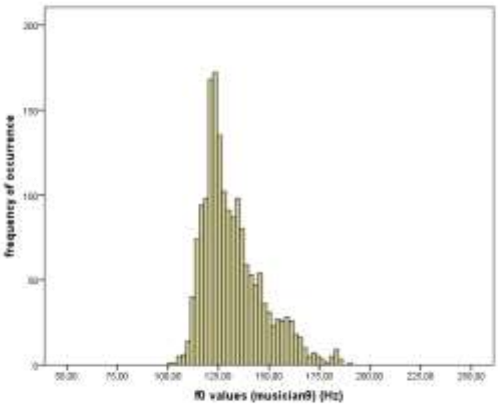

Fig. 5. Histogram of musician9

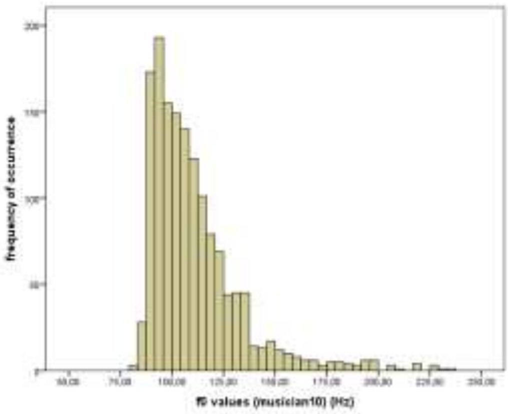

Fig. 6. Histogram of musician10

To address this problem, narrower ranges, including the peak of the distribution, were established in every case, i.e. deciles were determined and the following differences were calculated: (1) 6th decile–4th decile: this range represents 20% of the f_0 values around the median, (2) 7th decile–4th decile: this range represents 30% of the f_0 values, but the positive skewness of the distribution is considered, and (3) 8th decile–3rd decile: this range represents 50% of the f_0 values, considering the positive skewness of the distribution. Boxplot diagrams have been created to demonstrate the results. Fig. 7 shows that the ranges between the 6th and the 4th deciles do not seem to differ between musicians and non-musicians. Again, no significant difference was found between the two groups (U = 49, p = .94). Second, the ranges between the 7th and 4th deciles were compared. Fig. 8 demonstrates the differences. The diagram does not suggest a difference between musicians and non-musicians, which is confirmed by another Mann-Whitney U test (U = 48, p = .912). Third, the boxplots of Calculation (3) do not suggest a significant difference between the musician and the non-musician group. Calculations confirmed no significant difference between the two groups (U = 47, p = .853).

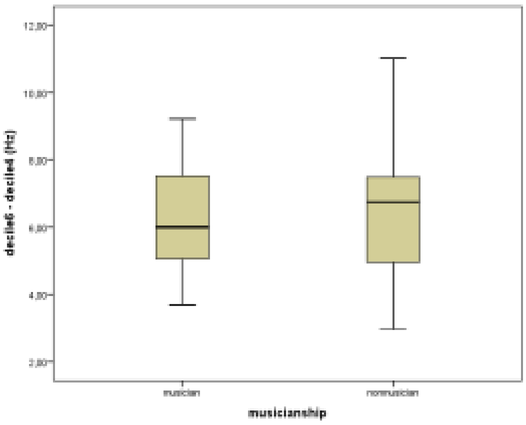

Fig. 7. Boxplots demonstrating the results of Calculation (1)

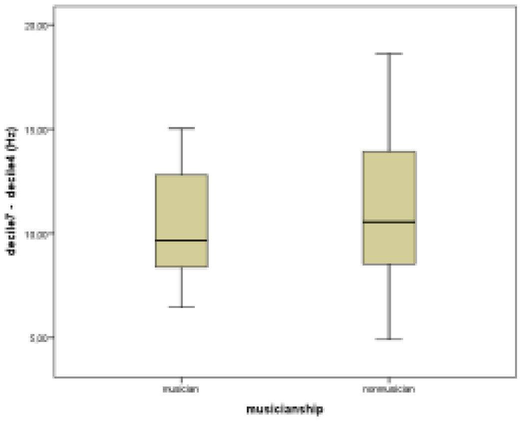

Fig. 8. Boxplots demonstrating the results of Calculation (2)

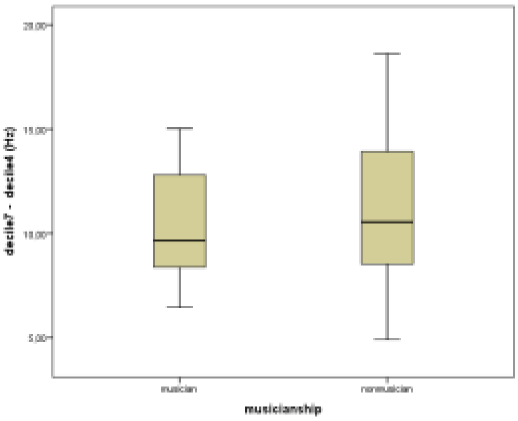

Fig. 9. Boxplots demonstrating the results of Calculation (3)

DISCUSSION

The purpose of this paper was to examine if musicians' speech in an oral reading task can be considered more melodic than that of non-musicians, based on the analysis of f_0 ranges, which did not show a significant difference. Using deciles, narrower ranges around the medians were also compared, but again, no differences were found. These findings indirectly suggest that musicians do not use more melodic intonation in reading than non-musicians, therefore they do not confirm our hypothesis. These results are in line with the findings of Stegemöller et al. (2008), who did not find difference in f_0 variation between musicians and non-musicians either.

One may infer that musical training does not influence speech melody in an oral reading task, so intonation mechanisms in reading may be at least in part independent of musicianship. Another possible explanation is that other factors, not examined here, have a more roboust effect on speech production, including intonation, than musical training. Reading skills, self-confidence, or experience in oral reading, independent of musical experience, may influence speech melody to a greater extent than musicianship. More complex studies in the future should address these questions and also if personality traits, musical aptitude, or even genetic background, as Schellenberg (2015) suggested with regard to perception, play any role in this context.

Major limitations of this study include the the number of participants, and its cross-sectional nature. More subjects may help reveal finer differences that the present study has not been able to demonstrate, and a longitudinal research, especially with children, would allow monitoring possible changes in speech in the function of years spent with musical traning, in contrast with children who do not learn music. It was also a limitation that a statistical approach to fundamental frequency was used. Thus, the existence of intonation patterns, specific to musicians, but realised within the same frequency ranges as with non-musicians, cannot be ruled out and should be analysed.

CONCLUSION

We believe that every possible dimension of transfer effects of music learning should be researched, including speech production. Even though no difference was found between musicians and non-musicians in this research, speech production of musicians and non-musicians deserves attention. Further research in a wider context, including cognitive processes or mechanisms of reading would open new perspectives, but we believe that it is also an important finding if transfer effects are found to reach a certain limit and not to excercise a certain effect in some areas.

REFERENCES

1. Amir, N., Fridman, S. B., Shakeman, O., Shuli, N., & Karni, A. (2020). Do Musicians Speak Differently? Preliminary Results of a Production Study. In *Proc. 10th International Conference on Speech Prosody 2020.* (p. 591–595). doi: 10.21437/SpeechProsody.2020-121
2. Banai, K., Fisher, S., & Ganot, R. (2012). The effects of context and musical training on auditory temporal-interval discrimination. *Hearing Research*, *284*, 59–66.
3. Boebinger, D., Evans, S., Scott, S. K., Rosen, S., Lima, C. F., & Manly, T. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *Journal of the Acoustic Society of America*, *137*(1), 378–387.
4. Boersma, P., & Weenik, D. (2019). Praat: Doing phonetics by computer [computer program] (Version 6.0.52). Retrieved from http://www.praat.org
5. Chartrand, J. P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters*, *405*, 164–167.

6. Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, *352*, 49–69.
7. Gocsál, Á. (2018). Speaker age estimation by musicians and non-musicians. In M. Gósy & T. E. Gráczi (Eds.), *Challenges in Analysis and Processing of Spontaneous Speech* (p. 185–205). Budapest: MTA Nyelvtudományi Intézet.
8. Gósy, M., Gyarmathy, D., Horváth V., Gráczi. T. E., Beke, A., Neuberger, T., & Nikléczy, P. 2012. BEA: beszélt nyelvi adatbázis [BEA: Spoken language database]. In M. Gósy (Ed.), *Beszéd, adatbázis, kutatások* [Speech, darabase, research] (pp. 9-24). Budapest: Akadémiai Kiadó.
9. Madsen, S. M. K., Marschall, M., Dau, T., & Oxenham, A. J. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Scientific Reports*, *9*(10404), 1–10. doi: 10.1038/s41598-019-46728-1
10. Markó, A. (2013). *Az irreguláris zönge funkciói a magyar beszédben*. [Functions of irregular phonation in Hungarian speech] Budapest: ELTE Eötvös Kiadó.
11. Micheyl, Ch., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*, 36–47.
12. Pastuszek-Lipińska, B. (2007). Musicians outperform non-musicians in a study with speech imitation. In *Proceedings of the International Computer Music Conference*, Kopenhagen 27-31 August 2007 (pp. 149–152). Available at https://quod.lib.umich.edu/i/icmc/bbp2372.2007.141/1
13. Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*, 98–108.
14. Sadakata, M., & Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: A cross-linguistic study. *Acta Psychologica*, *138*, 1–10.
15. Schellenberg, E. G. (2015). Music training and speech perception: A gene–environment interaction. *Annals of the New York Academy of Sciences*, *1337*, 170–177. doi: 10.1111/nyas.12627
16. Stegemöller, E. L., Skoe, E., Nicol, T., Warrier, C. M., & Kraus, N. (2008). Music Training and Vocal Production of Speech and Song. *Music Perception: An Interdisciplinary Journal*, *25*(5), 419–428.
17. Vyspinska, N. (2019). Musicians outperform non-musicians in English language vocabulary uptake and listening comprehension tasks. *Revista Românească pentru Educaţie Multidimensională*, *11*(2), 296–390. doi: 10.18662/rrem/131

Contact

Ákos Gocsál
University of Pécs, Faculty of Music and Visual Arts, Institute of Music, Hungary
gocsal.akos@pte.hu