MUSICIANS' AND NON-MUSICIANS' VOICE-BASED IMPRESSIONS ABOUT THE BODY SIZE OF SPEAKERS

ZENÉSZEK ÉS ZENEI KÉPZETTSÉG NÉLKÜLI HALLGATÓK BENYOMÁSAI A BESZÉLŐK TESTMÉRETEIRŐL, BESZÉDÜK ALAPJÁN

Dr. Ákos Gocsál[1,2]
[1] University of Pécs, Faculty of Music and Visual Arts, Institute of Music, Pécs
[2] ELKH Hungarian Research Centre for Linguistics, Budapest

ABSTRACT
It has been demonstrated by many researchers that musicians' auditory perceptual skills are more sensitive and sophisticated than in those who have not received musical training. It is also known that speech does not only encode verbal information, but also is used to form an impression about the speaker. The objective of this study was to analyze if musicians form more accurate impressions about male speakers' body size parameters, i.e., height, weight, body mass index, than non-musicians. Some significant differences and non-significant trends were also observed: as a general impression male musicians seemed to be more accurate, while female non-musicians tended to be less accurate in the voice-based body size estimation tasks.
Key words: transfer effects, social perception, speech

ABSTRACT
Kutatások igazolták, hogy a zenészek hangészlelési mechanizmusai érzékenyebbek, kifinomultabbak, mint azoké, akik nem tanultak zenét. Ismeretes az is, hogy a beszéd nem csak nyelvi üzenetet, de benyomást is kelt a hallgatóban a beszélőről. Ebben a kutatásban azt vizsgáltuk, hogy beszéd észlelése során a zenészek pontosabb benyomást alkotnak-e a férfi beszélők testalkatát leíró paraméterekről – magasságáról, testsúlyáról, testtömeg-indexéről –, mint akik nem tanultak zenét. Néhány esetben szignifikáns különbség adódott, és nem szignifikáns tendenciák is kirajzolódtak az adatokból: a férfi zenészek inkább pontosabb, a női nem-zenészek kevésbé pontos becsléseket adtak.
Kulcsszavak: transzferhatás, társas észlelés, beszéd

INTRODUCTION
A growing body of research deals with transfer effects of music learning. Of the many authors, Hallam (2010) and Hallam and Himonides (2022) have provided an overview about the main areas of research in this field. The present study focusses on auditory perception, which has also been reported to be positively affected by music learning, as a recent meta-analysis of the literature concluded (Neves et al., 2022). This positive effect is demonstrated by experiments, in which musicians generally outperform non-musicians in a variety of tasks, such as discrimination of pure tones (Liang et al, 2016)

or temporal-interval discrimination (Banai et al., 2012). Spoken stimuli have also been used in experiments in which musical training was proven to be an advantage in discrimination of subtle pitch changes on spoken sentences (Deguchi et al., 2012), pitch violations in a foreign language (Marques et al., 2007), or understanding speech in a noisy environment (Swaminathan et al., 2015).

Speech does not only convey verbal meaning. According to Laver (2003), the acoustic structure of speech incorporates three semiotic layers: linguistic, paralinguistic, and extralinguistic. The term "paralinguistic" refers to the affective and emotional state of the speaker, while "extralinguistic" is related to (quasi-) permanent characteristics, such as gender, age, physical properties, or health status. It is therefore reasonable to extend the notion of speech perception to the non-verbal dimensions if its acoustic structure, and to examine listeners' abilities to decode speaker emotions and characteristics from it. These voice-based impressions are of particular importance in social behaviour. Sometimes listeners are just surprised by the appearance of a speaker, whose voice is familiar from phone conversations or the radio (Krauss et al., 2002), but other times they infer the age (Skoog Waller et al., 2016), body size (Pisanski et al., 2014), or dominance and attractiveness (Borkowska & Pawlowski, 2011) of the speaker, which may well establish their attitudes and even behaviour in certain situations.

Perception of the nonverbal layers of speech in the context of the listeners' musical training, however, has received little attention. Of the few studies in this field, musicians were found to be better at emotion recognition (Thompson et al., 2004; Lima & Castro, 2011), while Gocsál (2018) did not find robust differences between the age estimations by musicians and non-musicians. Other dimensions of voice-based social perception are largely unknown. Here we address this problem with an emphasis on listeners impressions about unseen speakers' height and weight, i.e. whether listeners have realistic ideas about the physique of a speaker heard on the radio or the phone. Thus, the following research questions have been formulated:

(1) Do actual and estimated body size parameters correlate?
(2) Are musicians' estimations more accurate than those of non-musicians?

In this research, "body size" refers to speaker height, weight, and Body Mass Index (BMI), subsequently calculated using the formula *weight*/*height*^{2}. We assumed that BMI might reflect listeners' general idea of "speaker size" better than individual body size parameters.

METHODS

This research is part of a larger project in which listeners were asked to judge or estimate several characteristics of unseen speakers. So far, age estimation was examined (Gocsál, 2018), but data related to body size estimations have not yet been processed.

24 male speakers (aged 20-72) of the BEA Hungarian spontaneous speech database were selected (Gósy et al., 2012). All speakers spoke standard Hungarian, none of them smoked or reported any kind of speech production disorder. All of them held at least a

B.Sc. degrees or were university students at the time of the recording. 20-30 seconds long parts were selected from the "interview" or "opinion" parts of the recordings. In the selected parts, speakers spontaneously and in an emotionally neutral manner spoke about their jobs or hobbies, or told their opinion about a theme, e.g., public transportation. Speakers' height ranged between 163 and 197 cm (mean height = 179.8 cm, SD = 8.7 cm). Weight ranged between 60 and 100 kg (mean weight = 82.08 kg, SD = 10.57 kg). BMI values ranged between 18.6 and 31.56 (mean = 25.48, SD = 3.56).

Listeners included 85 university students (age: 19-37, median = 22 years). 42 of them were students at the Institute of Music (14 males, 28 females), with formal music education of at least 8 years. They were all instrumental musicians. 43 listeners were students of other faculties of the university (14 males, 29 females), who never received any music education apart from the compulsory singing lessons at school, and never played any musical instruments. No listener had any kind of hearing impairment.

Speech samples were played in several sessions in groups, in a silent seminar room, using a multimedia computer and professional Genlec speakers. Before the experiment, the researcher played three samples to the participants to familiarize them with the task and to make sure that all of them can clearly hear the speech samples. Listeners were given printed sheets on which they indicated their own demographic data (age, gender, major, musical training if any), and their height (in cm) and weight (in kg) estimations after listening to the individual speech samples.

RESULTS

We first examined if listeners were able to estimate speakers' height. First, mean estimated height values for each speaker were calculated. Table 1 shows the descriptive statistics for all listeners and by the individual listener groups.

Table 1. Descriptive statistics for estimated heights (in centimetres)

listener group	$height_{min}$	$height_{max}$	mean	SD
all listeners	170.65	182.09	176.92	3.68
male musicians	170.86	183.21	177.04	3.84
female musicians	170.64	184.21	177.87	4.05
male non-musicians	169.00	183.86	177.73	4.64
female non-musicians	170.14	181.48	175.56	3.39

Pearson's correlation coefficients were determined to assess if there is an association between the actual and mean values of estimated height. Table 2 shows the correlation coefficients with all participants and broken down to subgroups by gender and musicianship. Three out of the four subgroups demonstrated significant correlation, which means that members of those groups systematically judged taller speakers taller, and shorter ones shorter. This association with female non-musicians did not reach statistical significance but approached it. We have also calculated correlation coefficients of height estimations between the subgroups. Strong and statistically significant correlation was found (*r* > .774, *p* < .001 in every case), which indicated a strong agreement among the groups.

Table 2. Pearson's correlation coefficients between actual and estimated speaker heights. * indicates statistical significance (*p* < .05)

listener group	*r*	*p*
all listeners	.480	.018*
male musicians	.446	.029*
female musicians	.503	.012*
male non-musicians	.458	.024*
female non-musicians	.401	.052

Next, the accuracy of the estimations was analysed. The accuracy of estimation was defined as the difference between the estimated and the actual height, thus, smaller differences indicate more accurate estimations. Every possible difference value (85 listeners × 24 speakers = 2040 difference values) was calculated. Data were submitted to a mixed ANOVA with musicianship (musician and non-musician) and listener gender (male and female) as between-subject factors, and speakers as the repeated measures factor. While no significant model emerged, and main effects were not significant either, listener gender×musicianship interaction approached significance ($F(1,81) = 3.105$, $p = .082$). An examination of estimated marginal means revealed that means were relatively close to each other when those of the musician and non-musician males' groups were compared (EMM=2.777 and 2.060, std. error: .911 and .911, 95% CI: .931–4.558 and .246–3.873 respectively), somewhat larger differences between the estimated marginal means in the musician and non-musician female groups were found (EMM=1.917 and 4.006, std. error: .644 and .633, 95% CI: .634 –3.199 and 2.746–5.266 respectively). So, it cannot be stated that any of the subgroups defined by listener gender and musicianship performed worse or better than any other, but still certain trends were observed, musician females seemed to be non-significantly better estimators.

Next, the mean of the estimated weight values was computed for each speaker. Table 3 shows the descriptive statistics for all listeners and subgroups.

Table 3. Descriptive statistics for estimated weights (in kilograms)

listener group	$weight_{min}$	$weight_{max}$	mean	SD
all listeners	72.69	83.71	77.88	2.96
male musicians	68.50	88.36	76.18	4.42
female musicians	72.00	84.61	78.98	3.28
male non-musicians	72.43	82.86	77.67	3.13
female non-musicians	72.00	83.03	77.73	2.73

Again, Pearson's correlation coefficients were computed if association exists between actual and mean estimated weight. Table 4 presents the results of these calculations.

Table 4. Pearson's correlation coefficients between actual and estimated speaker weight. * indicates statistical significance ($p < .05$)

listener group	*r*	*p*
all listeners	.140	.515
male musicians	.411	.046*
female musicians	.054	.802
male non-musicians	.118	.582
female non-musicians	-.003	.989

The *r* values suggest that except the male musicians, listener subgroups were unable to differentiate speakers by weight. However, speaker weight estimations between the subgroups strongly and significantly correlated (in every case $r > .693$, $p < .001$)
Since actual and estimated weight did not correlate in three subgroups, no significant ANOVA model was expected for accuracy. Calculations confirmed this expectation. Neither main effects, nor their interaction was significant (gender: $F(1,81) = 1.320$, $p = .254$, musicianship: $F(1, 81) = .010$, $p = .921$, and gender×musicianship: $F(1, 81) = 1.204$, $p = .256$). None of the *p* values approached statistical significance ($\alpha = .05$) so we cannot infer nonsignificant trends.
The same methodology was applied to estimated BMI. Table 4 shows the descriptive statistics for all listeners and by the individual listener groups.

Table 5. Descriptive statistics for estimated BMI (in kg/m^2)

listener group	BMI_{min}	BMI_{max}	mean	SD
all listeners	22,92	27,46	24,95	1,25
male musicians	21,54	27,31	24,39	1,61
female musicians	22,90	27,59	25,03	1,24
male non-musicians	22,70	27,54	24,65	1,36
female non-musicians	23,43	27,97	25,30	1,32

Table 6 shows the correlation coefficients with all participants and broken down to subgroups by gender and musicianship. Neither the whole listener group, nor any of the subgroup demonstrated statistically significant correlation, although the *p* value at the male musician group approached statistical significance.
The data suggest that participants were unable to differentiate speakers by BMI, but male musicians were non-significantly better. Again, BMI estimations between the

subgroups strongly and significantly correlated (in every case $r > .700$, $p < .001$), which reflects a strong agreement.

Table 6. Pearson's correlation coefficients between actual and estimated BMI

listener group	*r*	*p*
all listeners	.321	.126
male musicians	.398	.054
female musicians	.295	.161
male non-musicians	.302	.151
female non-musicians	.243	.252

Again, no significant ANOVA model can be expected since actual and estimated BMI did not correlate in any of the groups. This was confirmed, neither the main effects of the between-subjects factors (gender: $F(1,81) = 1.896$, $p = .172$, musicianship: $F(1, 81) = .233$, $p = .631$), nor gender×musicianship interaction ($F(1, 81) = .11$, $p = .917$) was found to explain the variance of BMI estimation accuracy.

DISCUSSION

The purpose of this study was to examine if musical training is an advantage in forming voice-based impressions about the speaker's body size. Listeners were in general able to recognise which speakers were shorter or taller (non-musician females non-significantly, but they also showed this trend), while only male musicians could correctly differentiate them by weight. BMI, used here as a possible measure of impression about body size, was not a reliable parameter in differentiating speakers by body size, but with BMI, male musicians still performed better than the other participants. Musical training has not yet been examined in this context, but these results are at least in part in line with previous studies, in which male participants were significantly (Charlton et al., 2013), or non-significantly (Rendall et al., 2008) better at voice-based body size judgments than females, while they contradict the findings of Pisanski et al. (2016) who did not find significant differences between sexes. Nevertheless, the strong correlation between the estimations across the groups indicate that listeners of both sexes share common, but usually inaccurate stereotypes about what body size an unseen speaker may have. Regarding speaker size estimation accuracy, no significant effects were found. As a general impression about the results, male and to some degree, female musicians seemed to demonstrate a trend to be better body size estimators, while non-musician females appeared to perform poorer in these tasks.

Major limitations of this study include the fact that only male speakers were used, and the role of acoustic parameters has not been examined. In another study, Charlton et al., (2013) found that lower pitch improves the ability of listeners to perform acoustic size judgments. Research into the way musicians and non-musicians use acoustic parameters for their judgments would also contribute to a better understanding of acoustic size estimations. Although Lima and Castro (2011) found that musicians and non-musicians use the same acoustic properties of speech to the same degree in emotion recognition, it is not known if the same applies to other areas of voice-based social perception, including size judgments. Finally, one also needs to note that this is a cross-sectional study, only presenting differences between listener groups. Causal inferences can only be drawn from longitudinal research.

CONCLUSION

The differences between musicians and musically untrained listeners, both the signifincant ones and also the non-significant trends indicate that this area deserves attention for future research. These findings may contribute to a better understanding of the transfer effects of music learning and possibly the development of social competences and attitudes towards humans in musicians.

REFERENCES

1. Banai, K., Fisher, S., & Ganot, R. (2012). The effects of context and musical training on auditory temporal-interval discrimination. *Hearing Research*, *284*(1–2), 59–66. https://doi.org/10.1016/j.heares.2011.12.002
2. Boersma, P., & Weenik, D. (2019). *Praat: Doing phonetics by computer* (6.0.52). computer software. www.praat.org
3. Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, *82*, 55–59.
4. Charlton, B. D., Taylor, A. M., & Reby, D. (2013). Are men better than women at acoustic size judgements? *Biology Letters*, *9*(20130270). http://dx.doi.org/10.1098/rsbl.2013.0270
5. Deguchi, C., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., & Colombo, L. (2012). Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: Behavioral, electrophysiological and psychoacoustic study. *Brain Research*, *1455*, 75–89. https://doi.org/10.1016/j.brainres.2012.03.034
6. Gocsál, Á. (2018). Speaker age estimation by musicians and non-musicians. In M. Gósy & T. E. Gráczi (Ed.), *Challenges in Analysis and Processing of Spontaneous Speech* (pp. 185–205). MTA Nyelvtudományi Intézet. http://real.mtak.hu/81368/1/CAPSS-Gocsal.pdf
7. Gósy, M., Gyarmathy, D., Horváth, V., Gráczi, T. E., Beke, A., Neuberger, T., & Nikléczy, P. (2012). BEA: Beszélt Nyelvi Adatbázis. In M. Gósy (Ed.), *Beszéd, adatbázis, kutatások* (pp. 9–24). Akadémiai Kiadó.

8. Hallam, S. (2010). The power of music: Its impact on the intellectual, social and personal development of children and young people. *International Journal of Music Education*, *28*(3), 269–289. https://doi.org/10.1177/0255761410370658
9. Hallam, S., & Himonides, E. (2022). *The Power of Music: An Exploration of the Evidence*. Open Book Publishers. https://doi.org/10.11647/OBP.0292
10. Krauss, R. M., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology*, *38*, 618–625.
11. Laver, J. (2003). Three semiotic layers of spoken communication. *Journal of Phonetics*, *31*, 413–415.
12. Liang, C., Earl, B., Thompson, I., Whitaker, K., Cahn, S., Xiang, J., Fu, Q.-J., & Zhang, F. (2016). Musicians are better than non-musicians in frequency change detection: behavioral and electrophysiological evidence. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00464
13. Lima, C., & Castro, S. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, *11*(5), 1021–1031.
14. Marques, C., Moreno, S., Luís Castro, S., & Besson, M. (2007). Musicians Detect Pitch Violation in a Foreign Language Better Than Nonmusicians: Behavioral and Electrophysiological Evidence. *Journal of Cognitive Neuroscience*, *19*(9), 1453–1463. https://doi.org/10.1162/jocn.2007.19.9.1453
15. Neves, L., Correia, A. I., Castro, S. L., Martins, D., & Lima, C. F. (2022). Does music training enhance auditory and linguistic processing? A systematic review and meta-analysis of behavioral and brain evidence. *Neuroscience and Biobehavioral Reviews*, *140*(104777), 1–21. https://doi.org/10.1016/j.neubiorev.2022.104777
16. Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J. M., & Feinberg, D. R. (2014). Return to Oz: Voice pitch facilitates assessments of men's body size. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1316–1331.
17. Pisanski, K., Oleszkiewicz, A., & Sorokowska, A. (2016). Can blind persons accurately assess body size from the voice? *Biology Letters*, *12*(4), 20160063. https://doi.org/10.1098/rsbl.2016.0063
18. Rendall, D., Vokey, J. R., & Nemeth, C. (2007). Lifting the curtain on the wizard of Oz: Biased voice-based impressions of speaker size. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1208–1219.
19. Skoog Waller, S., Eriksson, M., & Sörqvist, P. (2015). Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. *Frontiers in Psychology*, *6*(978). https://doi.org/10.3389/fpsyg.2015.00978
20. Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, Jr, G., & Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Scientific Reports*, *5*(1), 11628. https://doi.org/10.1038/srep11628

21. Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, *4*(1), 46–64. https://doi.org/10.1037/1528-3542.4.1.46

Contact

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