10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Do Musicians Speak Differently? Preliminary Results of a Production Study

Noam Amir[1], *Sharon Bolle Fridman*[2], *Ortal Shakeman*[1], *Nofar Shuli*[1], *Avi Karni*[2]

[1]Department of Communication Disorders, Sackler Faculty of Medicine, Tel Aviv University, Israel
[2]Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, Israel

noama@tauex.tau.ac.il, sasharonf@gmail.com, ortalc131@gmail.com, nofarshuli@mail.tau.ac.il, avi.karni@yahoo.com

Abstract

Previous research has shown that professional musicians demonstrate superior auditory skills in a range of psychoacoustic and musically related auditory tasks. This heightened acuity has been shown to carry over to speech-oriented auditory tasks. Here, however, we set out to examine whether a strong musical background affects speech production, specifically, in expressing contrastive narrow focus when answering a question. Eight musicians and eight non-musicians were recorded answering 24 questions by 4-word sentences with an elicited narrow focus on one of the words. The productions were evaluated by two panels of listeners (experienced and inexperienced) who were asked to judge whether a specific word was emphasized in each utterance, and to what degree. Results showed a significant difference in judgments between the two groups of listeners, and a significant interaction with word position within the utterance. Regardless of listening group and word position, a consistent trend was observed; emphasis was judged stronger for musicians, although the difference was not statistically significant. We ascribe the lack of significance mainly to the small sample size, and intend to extend the study.

Index Terms: narrow focus, musicians, perception, speech production, prominence

1. Introduction

Professional musicians spend large amounts of time practicing, tuning, performing and listening to music, Therefore, it is not surprising that they develop superior skills on a range of auditory tasks. This has been demonstrated in numerous studies.

Musicians, as compared to non-musicians, were found to have superior perceptual auditory skills in frequency discrimination [1] [2], loudness sensitivity [3], timbre sensitivity [4] [5], sound localization [6] [7], processing in noise [8] and time discrimination [9].

In tasks more directly related to music, musicians were found to outperform non-musicians in rhythm processing [7] [10], detecting rhythm changes [11], detecting harmonic and melodic changes [12] [13], and tone naming [14].

As for musical production, non-singer musicians were found to be more accurate than non-musicians in repeating tones they were presented with [15].

This set of skills has been shown to carry over to speech and language oriented tasks. Musicians have demonstrated advantages in tasks related to speech and language, such as detecting tone changes in speech [16] [17] and prosody and intonation comprehension [18] [19]. Musicians have demonstrated better verbal memory [20], and acquired syntactic rules more efficiently in an invented language [21]. Musical training was found beneficial for accurate prosody imitation [22]. Moreover, musicians were better able to detect, distinguish, and mimic tonal foreign-language phonemes and tones (Mandarin Chinese), even though the features did not exist in their native language [23].

1.1. Rationale and research questions

Despite the large body of literature cited above, it appears that there have been only a few studies examining whether speech produced by musicians, either elicited or spontaneous, differs from that of non-musicians. Of course this is a very broad question, which must be tackled through some kind of systematic approach. The initial approach we chose in the current study was to compare the production of narrow focus by musicians vs. non-musicians.

Narrow focus has been studied extensively, both acoustically and perceptually ([24] [25] among many others), and has been shown to be related to prosodic cues. Elicitation of narrow focus (here we group narrow focus and contrastive focus together) is relatively straightforward to implement, and amenable to acoustic and perceptual analysis. In this study we therefore examined utterances containing narrow focus produced by musicians vs. non-musicians. In the initial analysis presented below we determined whether we could find any *perceptual* differences in the respective productions, before moving on to acoustic analysis. Thus, we addressed three questions:

1. Do musicians produce narrow focus that is perceptually different from productions by non-musicians?
2. Does extensive training in speech and hearing influence perception of narrow focus?
3. Does production/perception of narrow focus differ when focus is present on different word locations in an utterance?

2. Method

This study was carried out within the framework of a larger study examining a range of musical/linguistic abilities in musicians vs. non-musicians, addressing both perception and production. Here, we present only the specific methods we employed in obtaining narrow focus productions and evaluations. Since our study is based on perceptual evaluations,

10.21437/SpeechProsody.2020-121

an additional aspect, we addressed, was whether the *listeners'* background would have an effect on the result.

2.1. Participants

2.1.1. Speakers

The speakers (right handed, native Hebrew speakers) were composed of two groups: **8 musicians** (5 males, 3 females, aged 21-46y, mean 36y; on average 18 years of formal education) and **8 non-musicians** (3 males, 5 females, 26-38y, mean 30y; on average 16 years of formal education). At the time of participation, all musicians were employed as professionals playing classical music, with an average musical experience of 28y. The non-musicians had no formal education in music for more than 6 months, did not read musical notation and have not participated in any regular playing/singing activities. All participants underwent a hearing screening test demonstrating hearing thresholds better than 15 dB (250-8000Hz).

2.1.2. Listeners

There were two groups of listeners. One group of 10 listeners (10 females) was composed of 4th year undergraduate students enrolled in the Communications Disorders program, at Tel Aviv University. We surmised that students at this level, with formal and practical training in phonetics and phonology would possibly be more sensitive to finer differences in speech production than the general population. We therefore termed this group as **experienced listeners.** The second group, **inexperienced listeners**, was composed of 10 students (2 males, 8 females) enrolled in undergraduate programs unrelated to speech/hearing.

Listener's ages were in the range of 20-32y. All were native Hebrew speakers with no known hearing problems and no musical background.

2.2. Procedure

2.2.1. Recordings

Six four-word sentences were constructed, of the form /subject-verb-object-adjective/. Four contrastive questions were composed for each such sentence, resulting in 24 questions overall. The questions were recorded by a young woman and normalized for intensity.

Speakers were presented with the 24 questions in a randomized order. The questions were presented aurally via headphones. The speakers were directed to answer the questions by using the appropriate sentence, as written, emphasizing the word which resolved the specific question. Answers were recorded with an Audio-Technica AT892 microphone, connected to a Centrance Micport-Pro USB soundcard, using Audacity software on a laptop computer, in a quiet room.

For clarity, we present an example sentence and associated questions, translated to English. The target sentence was: "The lion saw a big cat." (In Hebrew this is a four-word sentence). The questions were: 1) Did **the dog** see a big cat? 2) Did the lion **eat** a big cat? 3) Did the lion see a **small** cat? 4) Did the lion see a big **pig**? The respective answers that were expected: 1) **The lion** saw a big cat; 2) The lion **saw** a big cat; 3) The lion saw a **big** cat; 4) The lion saw a big **cat.** The speakers received a compensation of 100NIS for their efforts. A total 24 sentences X 16 speakers were recorded, giving 384 recordings. Due to technical issues, only 366 of these were used in the perceptual tests.

2.2.2. Perceptual tests

Custom written software in Matlab [26] was used to run the experiment. The software played the 366 recordings in random order, at a comfortable volume level.

After playing each utterance, the software presented a screen in which the listener was asked to click on a radio-button in response to the question: "Which is the emphasized word (out of the four words in the sentence)?" The listeners could choose between words one to four, or indicate that no word was emphasized. In the case that one of the words was marked as emphasized, a follow-up question was presented: "Rate the level of emphasis on a scale of 1-5, from 1=weak emphasis to 5=strong emphasis." Thus, each individual recording received two scores from each of the 10 listeners: (1) which word, out of four, was perceived as prominent, or no word at all; (2) The degree of prominence, if present, on a scale of 1 to 5. A supplementary button allowed the listeners to replay the current sentence. Average session duration was approximately 60 minutes.

2.3. Data handling

Two variables for statistical analysis were derived as follows: first, Correct Identification (CI) was derived by counting the number of times the correct word (expected to be emphasized by the speaker) was identified as the emphasized one. This was counted separately for experienced vs. non-experienced listeners. The maximum possible score was 10, achieved if all the listeners in a group identified the correct word. Second, Intensity (INT) was derived by averaging the intensity scores given when the correct word was judged as emphasized. In cases where the wrong word was chosen, or no word at all was judged to be emphasized, the intensity value was entered into the average as zero.

The above procedure resulted in two sets (one for experienced, one for inexperienced, listeners) of 24 (6 sentences x 4 word positions) CI and 24 INT scores per speaker. The scores per word position were also averaged over the six target sentences, generating four scores of CI and four scores of INT, per speaker, per group of listeners.

3. Results

Means and Standard deviations for CI (% correct) are shown in Figures 1 and 2, for both groups of speakers (musicians vs. non-musicians); judged by experienced listeners in Figure 1 and inexperienced listeners in Figure 2. Figures 3 and 4 present the corresponding data for mean ratings of INT; experienced listeners in Figure 3 and inexperienced listeners in Figure 4. In all 4 figures, blue bars present musicians' speech, and red bars present non-musicians' speech.

Several trends are consistently present in both figures: (1) experienced listeners tended to give higher scores than inexperienced listeners, for both variables (CI, INT); (2) musicians tended to receive higher scores than non-musicians, for both variables; (3) scores tended to become progressively lower for the third and fourth word positions.

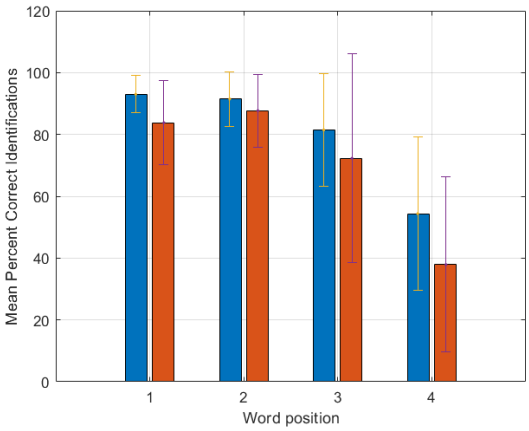

Figure 1: *Mean percent correct of emphasis detection, by experienced listeners. Blue: musicians' speech; Red: non-musicians' speech.*

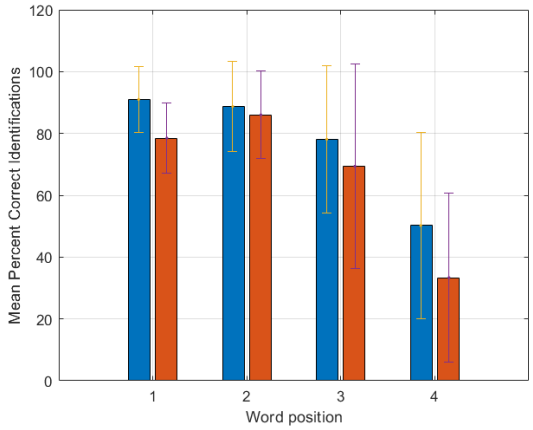

Figure 2: *Mean percent correct of emphasis detection, by inexperienced listeners. Blue: musicians' speech; Red: non-musicians' speech.*

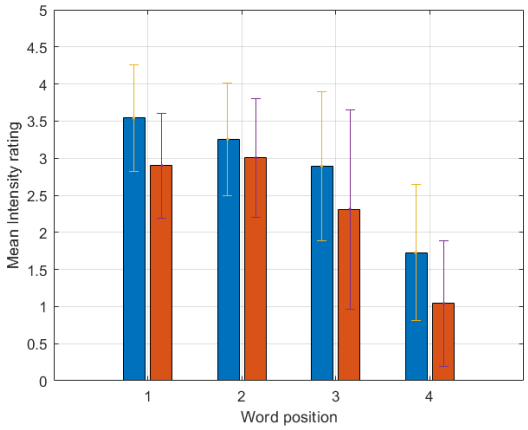

Figure 3: *Mean intensity rating, by experienced listeners. Blue: musicians' speech; Red: non-musicians' speech.*

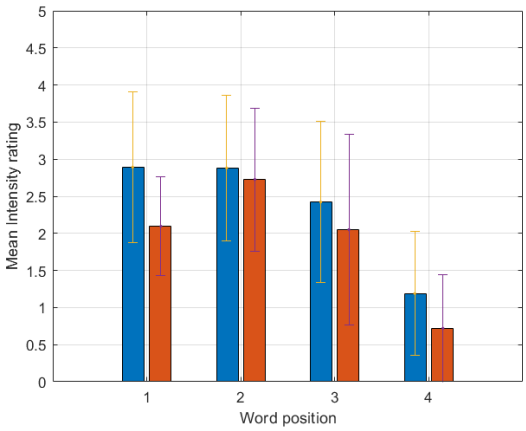

Figure 4: *Mean intensity rating, by inexperienced listeners. Blue: musicians' speech; Red: non-musicians' speech.*

3.1. Statistical Analysis

Two analyses were carried out to determine whether the trends observed above were significant. ANOVA with repeated measures was performed on each of the two dependent variables (CI, INT). Within-subject factors were Listener Group and Word Position, and the between-subject factor was Speaker Group.

Analysis of CI showed a significant main effect for Listener Group ($F(1,14)=7.300$, $p=0.017$), Word Position ($F(3,12)=25.536$, $p<0.001$) but not for Speaker Group ($F(1,14)=1.493$, $p=0.242$). Contrast analysis over Word Position revealed a significant quadratic trend ($p<0.001$).

Analysis of INT gave similar results. A significant main effect was found for Listener Group ($F(1,14)=160.691$, $p<0.001$), Word Position ($F(3,12)=36.515$, $p<0.001$) but not for Speaker Group ($F(1,14)=1.542$, $p=0.235$). Contrast analysis over Word Position revealed a significant quadratic trend ($p<0.001$).

4. Discussion

Starting from the third research question, a result that stands out is that judgments of emphasis decreased towards the end of an utterance. Out of the four words in each sentence, judgments of CI and INT were similar for the first two words, then lower for the third word and lower still for the fourth. It cannot be inferred directly from this experiment whether *speakers* have more difficulty in *producing* emphasis towards the end of an utterance, or whether *listeners* have more difficulty in *perceiving* it. Acoustic analysis might shed some further light on this issue, though it is not clear whether similar acoustic cues will produce similar judgments of emphasis at the beginning and end of an utterance. One possible way to disentangle production from perception in this respect would be to synthesize a series of utterances with gradually varying prosodic properties and examine their effect on judgments of emphasis.

Concerning the second research question, we found differences in judgments between the two groups of listeners. Neither group was accustomed to making specific judgments of

the kind performed here. However, the group of experienced listeners had approximately three years of studies in a communication disorders program, including theoretical studies and practicum experience in speech and hearing clinics. One can assume that over this period, members of this group would develop a heightened awareness of phonology and pragmatics. This was borne out in the results, where the experienced listeners consistently gave higher ratings of CI and INT, regardless of speaker type or word position. This issue should definitely need further validation. One possibility is to test even more experienced listeners, such as speech therapists with several years of professional experience. Another, is to obtain judgments from a group of musicians. As stated in the introduction, musicians have been shown to have heightened perceptual capabilities that might indeed carry over to judgments related to prosodic characteristics.

Lastly, the results related to the first research question were inconclusive. While the mean scores for musicians' speech (productions) were on average higher than those for non-musicians in all the 16 comparisons in graphs 1-4 (i.e. in each pair of blue vs. red bars, the blue bars were higher in every case), this trend was not found to be statistically significant. Our cautious suggestion is that this is possibly due to the relatively small sample size (8 participants) in each group of speakers. However, we do find the preliminary results presented here to be promising and unique in nature. We are currently in the process of extending the groups to 16 musicians and 16 non-musicians.

5. References

[1] L. Kishon-Rabin, O. Amir, Y. Vexler and Y, Zaltz, "Pitch discrimination: are professional musicians better than non-musicians?" *Journal of Basic and Clinical Physiology and Pharmacology*, vol. 12, no. 2, 2001.

[2] C. Liang, B. Earl, I. Thompson, K. Whitaker, S. Cahn, J. Xiang, QJ. Fu and F. Zhang, "Musicians are better than non-musicians in frequency change detection: behavioral and electrophysiological evidence," *Frontiers in Neuroscience*, vol. 25, 2016

[3] P. Bailey and S. Dobinson, "Pitch and loudness memory in musicians and non-musicians," *Journal of Acoustical Society of America*, vol. 112, no. 5, 2002.

[4] C. Micheyl, O. Carbonnel and L Collet, "Medial olivocochlear system and loudness adaptation: difference between musicians and non-musicians," *Brain and Cognition*, vol. 29, 1995.

[5] C. Pantev, L.E. Roberts, M. Schulz, A. Engelien and B. Ross, "Timbre specific enhancement of auditory cortical representations," *Neuroreport* vol. 112 , no. 1, 2001.

[6] W. Nager, C. Kohlmetz, E. Altenmuller, A. Rodriguez-Fornells and T.F. Munte, "The fate of sounds in conductors' brains: an ERP study," *Cognitive Brain Research*, vol.17, no. 1, 2003.

[7] T.F. Munte, W. Nager, T. Beiss, C. Schroeder and E. Altenmuller, "Specialization of the Specialized: Electrophysiological investigations in professional musicians," *Annals of the New York Academy of Sciences, vol.* 999, 2003.

[8] F. Zhang, C. Roland, D. Rasul, S. Cahn, C. Liang and G. Calencia, "Comparing musicians and non-musicians in signal-in-noise perception," *International Journal of Audiology* vol. 58, no. 11, 2019.

[9] Y. Zaltz, E. Globerson and N. Amir, "Auditory perceptual abilities are associated with specific auditory experience," *Frontiers in Psychology*, vol. 8, 2017.

[10] L. Parsons and M. Thaut, "Functional neuroanatomy of the perception of musical rhythm in musicians and non-musicians," *Neuroimage*, vol. 13 no. 6, part 2, 2001.

[11] A. Habibi, V. Wiranatana and A. Starr, "Cortical activity during perception of musical rhythm; comparing musicians and non-musicians," *Psychomusicology*, vol. 1:24, no. 2, 2014.

[12] S. Koelsch, E., Schroger and M. Tervaniemi, "Superior pre-attentive auditory processing in musicians," *Neuroreport*, vol. 10, 1999.

[13] M. Tervaniemi, M. Rytkonen, E. Schroger, R.J. Ilmoniemi and R. Naatanen, "Superior formation of cortical memory traces for melodic patterns in musicians," *Learning and Memory*, vol.8, 2001.

[14] S. Baharloo, P.A. Johnson, S.K. Service, J. Getschier and N.B. Freimer, "Absolute pitch: an approach for identification of genetic and nongenetic components," *American Journal of Genetics*, vol. 62, 1998.

[15] O. Amir, N. Amir, L. Kishon-Rabin, "The effect of superior auditory skills on vocal accuracy," The Journal of the Acoustical Society of America, Vol. 113, 2003.

[16] D. Schon, C. Magne and M. Besson, " The music of speech: music training facilitates pitch processing in both music and language," *Psychophysiology*, vol. 41, 2004.

[17] I. Zioga, C. Di Bernardi Luft and J. Bhattacharya, "Musical training shapes neural responses to melodic and prosodic expectation," *Brain Research*, vol.1650, 2016.

[18] W.F. Thompson, E.G. Schellenberg and G. Husain, "Decoding speech prosody: Do music lessons help?" *Emotion*, vol. 4, no. 1, 2004.

[19] W.F. Thompson, E.G. Schellenberg and G. Husain, "Perceiving prosody in speech - effects of music lessons", *Annals of the New York Academy of Sciences*, vol. 999, 2003.

[20] A.S. Chan, Y.C. Ho and M.C. Cheung, "Music training Improves verbal memory," *Nature*, vol. 396, 1998.

[21] G. Brod and B. Opitz, "Does it really matter? separating the effects of musical rraining on syntax acquisition," *Frontiers in Psychology*, vol. 13, no. 3, 2012.

[22] N. Copeland, "The Relationship Between Melody and Prosody: Perception and Production Capabilities of Musicians and Non-Musicians," Ph.D., Dissertation, University of Illinois at Urbana-Champaign, 2010.

[23] T. Gottfried and G. Ouyang, "Production of Mandarin tone contrasts by musicians and non-musicians," *Journal of The Acoustical Society of America*, vol. 118, no. 3, pp. 2025-2025, 2005.

[24] S. J. Eady, W. E. Cooper, G. V. Klouda and D. W. Lotts, "Acoustical characterization of sentential focus: Narrow vs. broad and single vs. dual focus environments," *Language and Speech*, Vol. 29, 1987.

[25] Y. Xu and C. X. Xu, "Phonetic realization of focus in English declarative intonation," *Journal of Phonetics*, Vol. 33, No. 2, 2005.

[26] Matlab 2018a, The MathWorks Inc., Natick Massachusetts.