MUSICAL TRAINING INFLUENCES AUDITORY TEMPORAL PROCESSING

Saravanan Elangovan, Nicole Payne, Jacek Smurzynski, Marc Fagelson

Department of Audiology and Speech-Language Pathology, East Tennessee State University, Johnson City, TN, USA

Corresponding author: Saravanan Elangovan, Department of Audiology and Speech-Language Pathology, East Tennessee State University, Johnson City, TN, USA, e-mail: elangova@etsu.edu

Abstract

Background: A link between musical expertise and auditory temporal processing abilities was examined.

Material and methods: Trained musicians (*n*=13) and non-musicians (*n*=12) were tested on speech tasks (phonetic identification, speech recognition in noise) and non-speech tasks (temporal gap detection).

Results: Results indicated musicians had shorter between-channel gap detection thresholds and sharper phonetic identification functions, suggesting that perceptual reorganization following musical training assists basic temporal auditory processes.

Conclusions: In general, our results provide a conceptual advance in understanding how musical training influences speech processing, an ability which, when impaired, can affect speech and reading competency.

Key words: music • auditory cortex • phonetics • speech perception • auditory temporal processing • gap detection • categorical perception • speech recognition in noise

---

EDUCACIÓN MÚSICA DE CAPACIDAD AUDITIVA COMO REFUERZO A LOS PROCESOS DE PROCESAMIENTO TEMPORAL

Resumen

Introducción: El presente estudio investiga la relación entre la educación música y los procesos de procesamiento temporal.

Material y métodos: Se han realizado las pruebas verbales (referentes a la identificación de los fonemas y el reconocimiento del habla en el ruido), a las que han sido sometidos músicos cualificados (en el número *n*=13) y a las personas sin estudios en música (*n*=12). Los participantes de la prueba han realizado tareas no verbales (consistentes en identificación de pausas).

Resultados: Los resultados de las susodichas pruebas demuestran que el tiempo de identificación de pausas es más corto en los músicos. Además, en este grupo se observa una identificación más rápida y precisa de los fonemas. Estos datos sugieren que la reorganización perceptual, que es una consecuencia de la formación música de la capacidad auditiva, refuerza los procesos básicos del procesamiento temporal.

Conclusiones: Las relaciones demostradas durante la realización de la prueba han permitido ampliar el conocimiento sobre el tema de efectos de la formación musical en el procesamiento del habla, cuyos trastornos pueden influir en la capacidad de verbalización y la competencia de lectura.

Palabras clave: música • corteza auditiva • fonética • identificación de pausas • percepción categórica • reconocimiento del habla en el ruido

---

ВОСПИТАНИЕ МУЗЫКАЛЬНОГО СЛУХА ПОДДЕРЖИВАЕТ ПРОЦЕССЫ ВРЕМЕННОЙ ПЕРЕРАБОТКИ

Изложение

Введение: В настоящей работе исследовалась связь между музыкальной деятельностью и процессами временной переработки.

DOI: 10.17430/901913

Материал и методы: Квалифицированные музыканты (числом *n*=13) и люди без музыкального образования (*n*=12) взяли участие в вербальных тестах (касающихся идентификации фонем и распознавания речи в шуме). Исследованные люди выполняли также невербальные задания (включающие обнаружение перерывов).

Результаты: Результаты вышеуказанных тестов показывают, что время обнаружения перерывов – короче у музыкантов. Кроме того, в этой группе замечена более быстрая и более точная идентификация фонем. Согласно этим данным, реорганизация восприятия, являющаяся последствием развития музыкального слуха, поддерживает основные процессы временной переработки.

Выводы: Взаимозависимости, показанные в ходе проведенного исследования, позволили расширить знания о влиянии музыкального обучения на переработку речи, нарушения которой могут влиять на способность вербализации и умение чтения.

Ключевые слова: музыка • слуховая кора • фонетика • обнаружение перерывов • категориальное восприятие • распознавание речи в шуме

KSZTAŁCENIE MUZYCZNE SŁUCHU WSPOMAGA PROCESY PRZETWARZANIA CZASOWEGO

Streszczenie

Wprowadzenie: W niniejszej pracy zbadano związek między przygotowaniem muzycznym i procesami przetwarzania czasowego.

Materiał i metoda: Testom werbalnym (dotyczącym identyfikacji fonemów oraz rozpoznawania mowy w szumie) poddano wykwalifikowanych muzyków (w liczbie *n*=13) oraz osoby bez wykształcenia muzycznego (*n*=12). Badani wykonywali też zadania niewerbalne (obejmujące wykrywanie przerw).

Wyniki: Wyniki w/w testów wskazują, że czas wykrywania przerw jest krótszy u muzyków. Ponadto, w grupie tej zauważa się także szybszą i bardziej precyzyjną identyfikację fonemów. Dane te sugerują, iż reorganizacja percepcyjna, będąca konsekwencją kształcenia muzycznego słuchu, wspomaga podstawowe procesy przetwarzania czasowego.

Wnioski: Zależności wykazane w toku przeprowadzonego badania pozwoliły na poszerzenie wiedzy na temat wpływu kształcenia muzycznego na przetwarzanie mowy, którego zaburzenia mogą wpływać na zdolność werbalizacji oraz umiejętność czytania.

Słowa kluczowe: muzyka • kora słuchowa • fonetyka • wykrywanie przerw • percepcja kategorialna • rozpoznawanie mowy w szumie

Background

The capacity to resolve fine auditory temporal cues is critical for comprehension of complex auditory stimuli, the separation of rapidly presented speech stimuli, and speech recognition, especially within competing noise [1]. Deficits in temporal processing ability have been related to phonological impairments, deficits in the perception of speech rhythm [2] and difficulties with listening in noise, both in children with learning impairments (e.g., dyslexia, auditory processing disorders, and specific language impairment) and in elderly adults.

Recently, studies have documented the positive benefits of musical training on cognitive and perceptual processes (see [3] and [4] for reviews). Speech and music have many similarities in both fine temporal and spectral structure, and they both require efficient neural processing in order to resolve dynamic timing cues [5]. Despite complex abilities required in both domains, linguistic and musical competencies develop spontaneously in normally developing children, without conscious effort or even formal instruction.

Experience-dependent plasticity associated with musical training results in changes throughout the auditory system [6] allowing musicians to perform better on listening tasks beyond just music, in later years. Consistent musical training from a young age has been thought to provide functional benefits in speech and language processing at both cortical [7,8] and subcortical [9–11] levels. Specifically, musicians outperform non-musicians on tasks such as auditory working memory and attention [12–14], speech perception in noise [1,15], phonetic categorization of consonant-vowel syllables [16], phonemic discrimination [17,18], and phonological awareness [5,19]. These findings support the use of musical training as a rehabilitation strategy to overcome deficits in auditory processing and speech perception, as seen in individuals with dyslexia or age-related decline [6] or with listening and language problems. Trained musicians have also been shown to possess superior temporal processing skills than non-musicians, as demonstrated with backward masking tasks [3] and temporal order judgments [20]. Recent studies have also demonstrated that musicians have finer temporal acuity skills and do better in detecting short temporal gaps in acoustic stimuli [6,21], a task known as auditory gap detection.

DOI: 10.17430/901913

Psychophysical [22,23] and clinical experiments [24] suggest that the spectrotemporal attributes of a sound in which a gap is embedded will engage specific gap-discrimination operations and therefore affect gap detection thresholds (GDTs). Two forms of gap detection have been distinguished, each mediated by different neural mechanisms [25], namely within-channel (WC) and between-channel (BC) gap detection paradigms. In a WC paradigm, the noise markers demarcating the gap have the same spectral and temporal properties, while in a BC paradigm, the noise markers are not identical and vary in either spectral and/or temporal properties. A number of investigations have shown that gap performance assessed through BC paradigms, but not WC, is highly correlated with voice onset time (VOT) phonetic boundaries [22,23,26], and that VOT relates to phonological reading skills in children [27] and reflects a decline in auditory processing with advanced age [28] or cerebral injury [29].

To date there have been only a few studies [6,21] that have examined gap detection performance in musicians. Further, neither of these studies examined the relation between gap detection performance and functional speech perceptual skills in musicians, particularly at a sub-segmental level. In the present study, we investigate whether musical training alters primary temporal perceptual processing abilities that have been shown to be common for speech processing (e.g., speech recognition in noise and phonetic identification) and for non-speech (i.e., temporal) processing.

Material and methods

Participants

Thirteen musicians and 12 non-musicians participated in this study. The sample size was determined based on a power analysis (G*Power 3.1 analysis software) with a power set at 0.80, alpha at 0.05, and an anticipated mild *a priori* effect size predicted by group differences in between-channel gap detection recorded in a pilot study. All participants were native English speakers with ages ranging from 18 to 28 years (mean ages: 20.5 years for musicians and 22.8 years for non-musicians). Participants categorized as musicians started musical training before the age of 12 years, had 8 or more years of musical experience, and consistently practiced for at least 1 hour per day and at least three times per week over the last 10 years. Although a few participants in the non-musicians group (n=5) reported some experience of playing music during school years, none reported having more than 3 years of formal musical training and all failed to meet the musician-group inclusion criteria. All participants had normal pure tone hearing thresholds (modified Hughson-Westlake method) from 250 to 8000 Hz (<15 dB HL) and no history of middle ear pathology/dysfunction, cognitive or neurological deficits, or any previously diagnosed learning, listening, or auditory processing disorder.

All participants were screened for amusia through the online Montreal Battery of the Evaluation of Amusia (MBEA; 30) test. Amusia, also known as tone deafness, refers to a musical disorder that presents as defects in processing pitch, musical memory, and musical recognition, and is believed to affect approximately 4% of the general population (35). The MBEA assesses six music-processing components: scale, contour, interval, rhythm, metric, and musical memory. The MBEA has been shown to be sensitive to deficits in perception of musical tones in both normal [30] and neurological disordered patients [31]. The results of the MBEA test did not identify any participants from either group as amusic. Further, a t-test comparison of their MBEA performances revealed no significant difference between the musician (mean score=25.5; SD=2.9) and the non-musician (mean score=22.3; SD=3.8) groups (F=0.39; p=0.28).

Stimuli

The gap detection and phonetic perception stimuli and methods employed in the present study are comparable to those reported in Elangovan and Stuart (2008; [22]).

Gap detection

Gap markers for the between-channel (BC) and within-channel (WC) gap detection tasks were synthesized from noise stimuli generated by SigGen 32 (v3.1) and a Tucker-Davis Technologies DD1 32-bit resolution digital-to-analog converter with 20 µs sampling period (TDT, Alachua, FL, USA). To prevent aliasing, the synthesized stimuli were low-pass filtered (TDT FT6-2) and then attenuated (TDT PA4) before being digitally filtered (TDT PF1). These stimulus tokens were power amplified (TDT HB6) to give a calibrated 80 dB peak SPL when presented through ER-3A insert earphones (Etymotic Research, Elk Grove, IL). Each trial comprised three stimulus sequences: two control sequences and one test sequence which were randomly presented, and the subject's task was to press a button to indicate which sequence sounded different (see Figure 1 for a schematic of the test and control sequence). Each control sequence consisted of a leading marker and a trailing marker separated by a brief (1.0 ms) inaudible gap to ensure that gating transients were similar for both the control and test stimuli. This was done to reduce the possibility of creating an extraneous cue for the test sequence. The leading and trailing markers of the test sequence were separated by a gap that was varied adaptively by a tracking procedure. For the BC gap detection task, the leading marker was a short (10 ms), wideband (10–20,000 Hz) noise stimulus with 1 ms rise/fall times. The trailing marker was a relatively long (300 ms), narrowband (half-octave; filter roll-off of 48 dB/octave) noise stimulus; the center frequency of the trailing marker for the between-channel condition was either 2 or 8 kHz. For the WC gap detection task, both the leading and trailing noise markers were identical. They were half-octave narrow-band (half-octave; filter roll-off of 48 dB/octave) noise bursts with a center frequency of 1000 Hz. The duration of these stimuli were 200 ms with rise/fall times of 10 ms.

Phonetic identification

A series of consonant-vowel (CV) syllables was synthesized such that the end points were perceived categorically as a voiced bilabial stop /*ba*/ and a voiceless /*pa*/. This was achieved by parameterizing a single acoustic temporal dimension – viz., voice onset time (VOT), across

DOI: 10.17430/901913

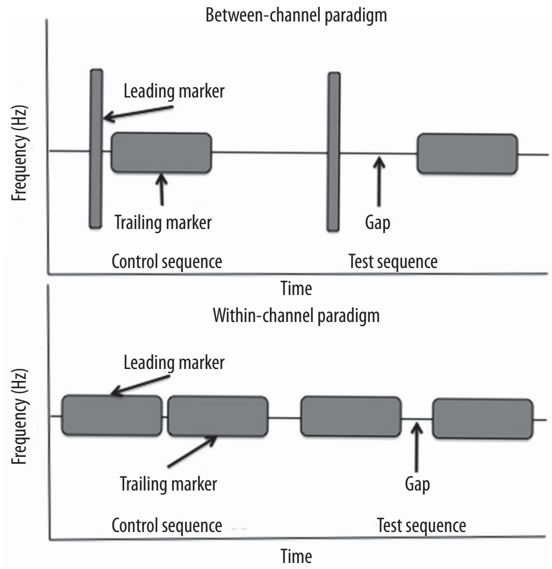

Figure 1. Schematic representation of the control stimuli sequences and test stimuli sequences of the within-channel and between-channel gap detection paradigms

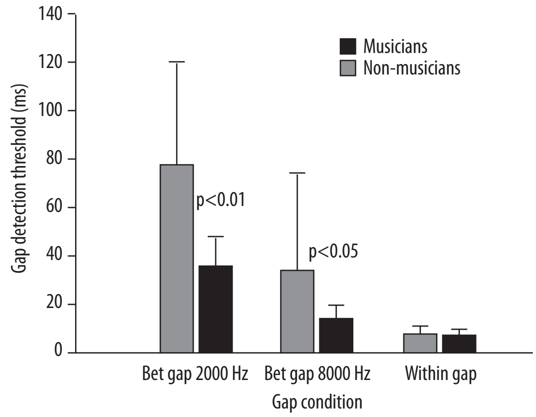

Figure 2. Mean gap detection thresholds for the four between-channel (BC) gap detection threshold tasks (at 2000 and 8000 Hz) and the within-channel (WC) gap detection task. Error bars indicate ± one standard deviation of the mean

the synthesized stimuli. These stimuli were created using an HLSyn (version 2.2) Klatt synthesizer (Sensimetrics Corporation) running on a Dell Optiplex computer with a 16-bit sound card. The stimuli were spectrally identical and differed only in VOT values that separated the onset of the stop burst and that of subsequent voicing. The VOT values varied in 5 ms steps and ranged from 0 to 60 ms, thereby creating a 13-stimuli continuum. The HLsyn is a quasi-articulatory synthesizer that generates speech units by altering time-varying parameters. A noise source 10 ms long and 60 dB SPL in amplitude simulated the initial stop burst. A 40 ms vowel /*a*/ formant (F) transition followed this stop burst. The onset frequencies (and bandwidths, BWs) of the F transitions were F1=438 Hz (200 Hz), F2=1025 Hz (70 Hz), F3=2425 Hz (130 Hz), F4=3250 Hz (350 Hz), and F5=4500 Hz (500 Hz). The fundamental frequency of each token varied over the steady-state portion of the vowel, beginning at 120 Hz and ending at 100 Hz. The steady-state portion of the /*a*/ varied in duration relative to the VOT, while maintaining a constant overall token duration of 320 ms. The F (and BW) values for the vowel were as follows: F1 = 700 Hz (200 Hz); F2=1200 Hz (70 Hz); F3=2600 Hz (160 Hz); F4=3300 Hz (350 Hz); and F5=4500 Hz (500 Hz). Illustrative spectrograms, prepared with Signalyze software (Linguist Plus, Inc; version 3.12), of seven tokens of these synthesized stimuli are shown in Figure 2 of Elangovan and Stuart [22].

Speech recognition in noise

The stimuli and procedure used in the word recognition in noise task were similar to those used by Elangovan and Stuart [22]. The test stimuli consisted of a custom 2-channel compact disc recording of 50 monosyllabic words from lists 1–4 of the Northwestern University Auditory Test No. 6 (NU-6) and competing continuous or interrupted broadband noise. The word lists were edited to remove the carrier phrase and to reduce the inter-stimulus intervals from 4.2 to 3.0 seconds. The continuous noise consisted of a 10-second segment of noise with a flat spectrum (within ±1 dB) from 100 to 8000 Hz. The competing interrupted broadband noise consisted of noise bursts, with silent periods between them, both of which had durations varying randomly from 5 to 95 ms. All speech and noise stimuli were normalized to have equal power.

Procedure

Gap detection

To estimate the gap detection thresholds (GDTs), a two-down, one-up, three-interval forced choice adaptive procedure was set to yield a 70.7% performance level [32]. The experiments were controlled by Psychosig version 3.11 (TDT system II) and interfaced with a GSI-16 (Grason-Stadler) audiometer via a TDT RP2 interface. Gap stimuli were presented binaurally at 80 dB peak SPL with ER-3A insert earphones. Following presentation of each trial (each including three stimuli sequences with two control and one test sequence presented in a random order), the subjects pressed a button to indicate which stimuli sequence had the longer gap. Prior to the test trails, all participants completed practice sessions to acquaint themselves with the stimuli and task. For these sessions, the WC and BC gap stimuli had longer gaps and visual feedback was provided after each trial. In general, the participants required more practice to become familiar with the BC tasks in comparison to the WC tasks, which is consistent with what is reported in the literature [25]. On average, these practice sessions lasted approximately 15–20 minutes for both groups of participants. Test trials began after participants demonstrated familiarity with the task, achieving relatively stable performance and smaller variations at short gap values.

After a participant's response was recorded, a new trial began 300 ms later. Gap duration was adaptively changed and each run was terminated after 14 reversals or after

DOI: 10.17430/901913

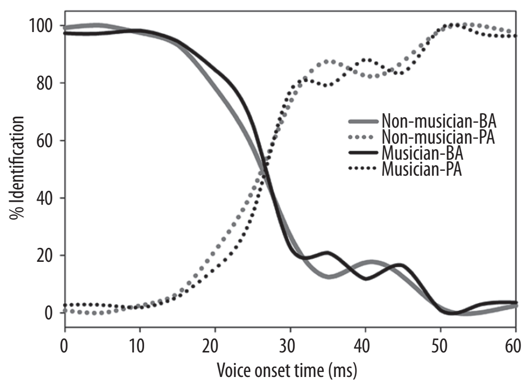

Figure 3. Identification functions showing the group mean percent correct */ba/* responses and */pa/* responses as a function of VOT value (0 to 60 ms) for the musician group (black) and the non-musician group (grey). Error bars indicate ± one standard deviation of the mean

a maximum of 60 trials. The GDT was estimated as the arithmetic mean of the last six reversals and was defined as the silent interval, exclusive of rise/fall time [25]. For the BC gap detection tasks, GDTs were determined separately for each trailing marker center frequency (2 and 8 kHz).

Phonetic identification

The VOT series of 13 stimuli were routed from the computer to the GSI-61 audiometer and presented at 80 dB peak SPL binaurally with ER-3A insert earphones. A set of 10 stimulus blocks (each with the 13-stimuli continuum in random order) was used to generate identification functions using the single-interval forced-choice paradigm. In a short practice session, all participants listened to identified endpoint /*ba*/ and /*pa*/ stimuli to familiarize themselves with the stimuli. Participants pressed a corresponding button on a response pad to indicate whether they heard /*ba*/ or /*pa*/.

Speech recognition in noise

The speech stimuli for the quiet condition were presented binaurally through insert earphones (ER-3A) at 40 dB SL re: SRT. For the noise conditions, the speech stimuli were presented binaurally at a constant level while the noise was varied with signal-to-noise (S/N) ratios of −5, −10, −15, −20, −25, and −30 dB. Presentation order of the word lists and noise conditions was counterbalanced and the S/N ratios were randomized across participants. The participants listened and repeated the target words in quiet and in competing continuous or interrupted broadband noise. The responses for the different conditions were scored as whole word percent correct.

Results

Gap detection

The group mean GDTs and standard deviations (SDs) for the different conditions for both groups are shown in Figure 3. For the musician group, the thresholds were 7.1 ms (SD=2.9) for the WC, 35.9 ms (SD=4.2) for the 2000 Hz BC, and 14.0 ms (SD=3.5) for the 8000 Hz BC. For the non-musician group, the corresponding mean GDTs (and SDs) were 7.9 ms (SD=2.9), 77.3 ms (SD=5.3), and 34.3 ms (SD=3.3). The following trends can be observed. First, for both groups the BC condition generally produced higher GDTs and inter-subject variability than the WC condition. Second, the difference between the center frequency of the leading and trailing markers affected the BC gap thresholds, with better discrimination for the 8000 Hz condition than the 2000 Hz condition. Most significantly, however, the musicians outperformed the non-musicians for the BC gap detection tasks. A one-way repeated measures analysis of variance (ANOVA) indicated that the musicians had significantly shorter BC GDTs than the non-musicians [$F(1,20)=11.10$, $p=0.003$ for the 2000 Hz condition; $F(1,20)=8.23$, $p=0.045$ for the 8000 Hz condition)]. No significant differences were observed between the groups for the WC GDTs [$F(1,20)=0.53$, $p=0.48$].

Phonetic identification

The VOT categorical boundary was determined as the VOT value that resulted in a 50% probability of a /*ba*/ response, and was determined for each participant using the Spearman-Kärber equation. The mean categorical boundary was 30.2 ms (SD=2.6 ms) for the musician group and 29.6 ms (SD = 4.8 ms) for the non-musician group. These categorical values are comparable to those reported in other studies [33–35] for young, normal hearing, native English listeners for the same (*viz.*, voicing) phonetic contrast. As can be observed in Figure 3, phonetic identification functions for the musicians were steeper (i.e., more categorical) than non-musicians; statistically however, there was no significant group difference [$F(1,24)=0.119$, $p=0.073$, $\eta^2=0.005$, $\theta=0.063$] between the slope of the regression functions when tested with a logistic regression model.

Speech recognition in noise

There was no significant difference between groups for word recognition in quiet (musician mean: 93.67±3.05%; non-musician mean: 94.83±3.66%; $F=0.72$, $p=0.406$). Both groups demonstrated a significant release from masking in the interrupted noise condition as observed in Figure 4. However, no significant group differences were observed in the magnitude of the masking release. These results are in contrast to those reported by other research investigations utilizing different speech-in-noise paradigms that have consistently shown that musicians have enhanced word recognition in the presence of competing noise [1,6]. However, an interesting and important result was found when, for the musician and non-musician group, independently, we performed a series of Pearson product-moment correlations between the different speech-in-noise conditions (13 levels: speech in quiet, 6 continuous-noise conditions, and 6 interrupted noise conditions) and gap detection performance (2 levels: BC and WC thresholds). These results revealed, for the musician group, a significant negative correlation between the interrupted noise at −30 dB SNR condition (Int −30 SNR) and the BC 8 kHz gap threshold ($r[12]=-0.59$, $p=0.04$). In other words, musicians who performed better (higher word recognition

DOI: 10.17430/901913

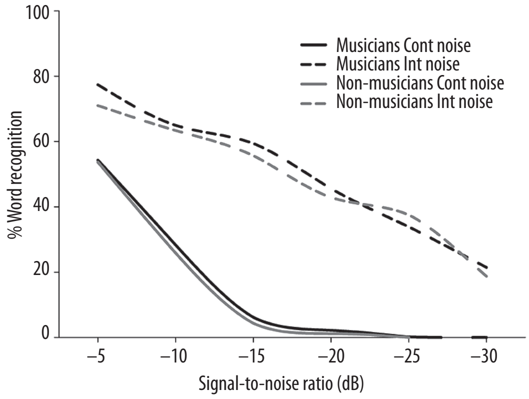

Figure 4. Mean word recognition scores as a function of signal to noise ratio for continuous noise (full lines) and interrupted noise (dashed lines) for both the musician group (black) and non-musician group (grey)

scores) under the most stringent SNR condition with competing interrupted noise also had better (shorter) BC gap detection thresholds, at least for the 8 kHz condition. All the other comparisons were non-significant for both the musician and non-musician groups.

Discussion

The hypothesis that musical training enhances the ability to resolve temporal processing cues in both speech and non-speech stimuli was tested by comparing performance of musicians and non-musicians for three behavioral tasks: gap detection, phonetic identification, and word recognition in noise. We postulated that if musical training shapes the auditory system at the level of basic processing, then this might have an effect on the perception of distinct phonemes and the ability to resolve speech fragments that fall in the brief silent gaps between interrupted noise. For these reasons, we expected to see shorter gap detection thresholds (GDTs), steeper (i.e., more categorical) phonetic identification functions, and enhanced word recognition in the presence of interrupted noise.

While our methodology does not allow causal inferences to be made, our study findings reinforce the proposition that musical training enhances some temporal processing skills of non-speech stimuli; it also potentially affects the ability to categorize phonetic contrasts, at least those varying in a single temporal dimension (i.e., voice onset time). Our results do not reveal any superiority of the musicians over the non-musician group with regard to release of masking and speech perception in noise. However, we do acknowledge that our participant sample size was relatively small for both groups. Considering the greater inter-group variability with word recognition and phonetic characterization, it is possible that this may have affected our results. Future investigations in this area will need to include larger sample sizes.

Gap detection

The task of gap detection as a perceptual operation takes on special importance from a speech perception standpoint, especially because such parses or gaps are ubiquitous in natural speech at both the segmental and suprasegmental level. In fact, some of the most interesting "gaps" in the stream of speech, such as VOT, are phonetically relevant cues [23] in most languages. Elangovan and Stuart [22] contend that the auditory perceptual task underlying a between-channel (BC) exercise must be processed at a more central level, since the peripheral auditory system contains no neural machinery capable of mediating such cross-channel comparisons. Thus the argument can be made that the two types of gap detection paradigms (i.e., BC and WC) potentially activate fundamentally independent timing perceptual operations. The timing mechanism required for performing a WC task appears to be "simple discontinuity detection in the auditory channels" [22] activated by the stimuli, a processing strategy likely to be achievable at the periphery. In contrast, the higher GDTs in the BC paradigm is likely to reflect complex central representation of the markers bounding the gap, and these would require relatively longer computation times, with inherently higher inter-subject variability, than the peripheral processing of WC gaps. Our present results accord with earlier researchers who reported elevated GDTs and greater inter-subject variability for BC paradigms compared to WC gap thresholds [22,26], which we observed for both the musician and non-musician groups.

Recent investigations have revealed that neuroplasticity induced by musical learning should be reflected as enhanced cognitive auditory endogenous responses in both electrophysiological [36–38] and magnetoencephalographic measures [39]; the same measures correlate with structural changes in cortical auditory and motor areas [8]. One important finding of the present study was that the musicians had significantly lower thresholds for the two BC gap detection tasks (at 2000 and 8000 Hz) than the non-musicians. However, no such group differences were observed for the WC gap detection task. These findings are similar to those reported by other investigators with musicians from diverse genres, and supports the notion that a BC gap detection test paradigm is more sensitive to relatively subtle neuroplastic changes secondary to musical training.

Phonetic identification

Categorical perception of consonants exemplifies our unique speech perceptual skill of effortlessly mapping smooth, continuous acoustical features into discrete, phonetic units. This fundamental ability enhances speech comprehension by promoting perceptual constancy in the face of talker variability within different acoustical dimensions such as pitch and VOT. Although categorical perception is an innate ability [40], linguistic investigations reveal that categorical phonetic boundaries are modified early in life by native language [32] and are malleable to later life language experiences, as in the case of bilinguals. From a clinical perspective, the importance of categorical perception is supported by a number of studies of phonological reading impairments in developmental dyslexia [41,42]. These have shown an abnormally reduced sensitivity to acoustic differences across phonetic boundaries (i.e., deficits in phonetic precision) during reading acquisition, a deficit which is a prime factor in dyslexia. In addition, an effect of age on phoneme category boundaries has already been

DOI: 10.17430/901913

demonstrated, with older, normal-hearing listeners requiring different (longer) target temporal cues [43], or having different identification slope functions [34], than younger adults when forming judgments about the phonetic category of a stimulus. Such investigations validate the use of a phonetic identification task – one which varies in terms of a single temporal cue – as a good test for distinguishing perceptual processing abilities between listener groups.

Our results did not reveal a statistically significant difference between the categorical boundary and/or identification function slopes of the musician group compared to that of the non-musicians. However, as can be observed in Figure 3, the identification functions were steeper (i.e., more "categorical") for the musicians (black curve) than the non-musicians (gray curve). Although this finding was not statistically significant and may be speculative, it fits with recent electrophysiological and imaging investigations which have suggested that trained musicians can make discernible phonetic distinctions in acoustic variations within a phonetic category while maintaining robust mental representations of phonological contrasts, thus reflecting an exceptional ability to encode and analyze spectrotemporal features of auditory stimuli [10,16,18,44]. Processing temporally brief speech elements is not only important for understanding speech in demanding listening environments, it is also imperative for language learning and phonological awareness (i.e., reading ability). Further, there is consistent evidence that categorical perception is related to phoneme awareness and word reading performance. Thus, the results of the present study support the recommendation of using musical training as one of the remediation strategies to improve phoneme awareness in children with developmental dyslexia.

Speech recognition in noise

The mean word recognition score in quiet was similar for both the musician and non-musician groups. Both groups demonstrated a significant release from masking in the interrupted noise condition, as indicated in Figure 4. However, no significant group differences were observed in the magnitude of the masking release. The overall superior performance in the interrupted noise condition was consistent with previous findings suggesting that enhanced understanding in noise depends on temporal resolution, so that favorable signal-to-noise (S/N) ratio signals are recoverable from the brief "silent" gaps between noise bursts [23,45].

However, our word recognition in noise results did not reveal a robust difference between the musician and non-musician groups. This is in contrast to a number of other studies that conclude that musicians show functional advantages when listening in noise, as indicated by their performance on traditional speech-in-noise tasks (e.g., QuickSIN, HINT, WIN, and Sentences in Portuguese Lists Test [1,38,46]. These findings are in accord with those reported in recent reports by Ruggles et al. [47] and Boebinger et al. [48] who both found that musicians outperformed non-musicians on measures of frequency discrimination but showed no advantage in perceiving masked speech.

There are some potential differences in our methodology and characteristics of our participants which may have contributed making our outcomes somewhat different from other reports in the literature. The differences include more stringent "musician" inclusion criteria related to age at which training started, duration of musical training intensity (hours/week of practice), instructions given, variables related to the target stimulus and/or competing stimuli, and other task-related variables, such as stationary or spatially separated competing messages. For example, Parbery-Clark et al. [1], who documented a significantly better performance in their musician group for the hearing in noise test (HINT) and QuickSIN, specified an average of 16 years of formal musical training and a self-reported practice regimen of 5 hours a week. In comparison, in our study the musician group had an average of 10.2 years of musical training and typically reported at least 3 hours/week of practice.

Another critical variance in methodology that could explain our results was the nature of the speech stimulus employed to assess speech in noise (SIN) performance. Most studies that have demonstrated a "musician advantage" have used either commercial versions (HINT or QuickSIN; e.g., 1,47) or custom-developed [49] versions of the speech-in-noise tests that used *sentences* as the target material. In contrast, the target stimuli in our study were open-set words (NU-6 words) that provided minimal contextual cues. It is known that SIN performances with a longer target stimuli would probably be more affected by variances in cognitive skills, such as auditory working memory [50,51] and attention [52], differences that would not necessarily be accessed when employing an open-set word SIN task [53]. This contention is actually supported by recent findings by Strait et al. [54], who also investigated SIN performances of musically trained children, compared to non-musician children, using both sentence (HINT) and word (Words in Noise test; [55]) target material. Their results revealed a musical advantage for only the HINT task, while no group differences were observed for the WIN task. It is plausible that the results obtained with the word task used in this study would be more reflective of peripheral aspects of hearing function in noise [50,53] and as such, are less likely to be representative of the more "central" plasticity believed to be influenced by musical training.

An isolated but albeit interesting finding that emerged in our study was the finding for the musician group of a relationship between performances in the between-channel gap detection task and word recognition scores with interrupted competing noise presented at –30 dB SNR. The negative correlation between these two measures seems to suggest that those with lower gap detection scores also had a larger release from masking, particularly in the most challenging condition. We did not observe any significant correlations for the rest of the comparisons in either group. To the best of our knowledge, this is the first investigation that has shown a direct relationship between gap detection performance and speech recognition in noise for any given listener group, clinical or non-clinical. However, since this finding is isolated (i.e., no particular trend existed for other competing interrupted or continuous noise conditions and gap detection in both groups), further research is warranted before any strong conclusions can be drawn.

DOI: 10.17430/901913

Conclusion

To the best of our knowledge, this is the first study that has investigated non-speech temporal processing in trained musicians using both WC and BC gap detection paradigms and has related these findings to speech perceptual measures, particularly sub-segmental (VOT) processing. Our results indicate that musicians have enhanced temporal processing ability as demonstrated by shorter GDTs and steeper phonetic identification functions. Taken together, these findings suggest that musical training results in a better perceptual reorganization at the level of fundamental temporal processes, an improvement benefitting music, speech, and language. Our results therefore complement a growing literature which suggests that musical training can produce functional benefits in listening and language skills, as well as influence auditory processing at cortical and sub-cortical levels. While most of these earlier investigations have demonstrated beneficial effects of musical training in the processing of suprasegmental aspects of speech, such as pitch and prosody, our results reveal that neuroplasticity extends to subsegmental (e.g., VOT) processing.

Our findings do not indicate the causal direction of this relationship. That is, it is still not clear whether inherently good auditory processing skills lead to enhanced musical aptitude or, conversely, that musical training leads to improved processing ability. Although there is as yet no clear evidence for what factor is responsible for providing a musician with their distinct skill set, several interpretations imply that the processing of temporal structure in both music and speech may rely on common mechanisms [19] that share the same pool of neural resources [56,57]. This speculation is supported by recent functional magnetic resonance imaging results showing that Brodmann Area 47 of the left inferior frontal gyrus and the temporal cortices of both hemispheres are involved in processing temporal structure in both music and speech [58]. Further, a few longitudinal investigations support the notion that the neural and perceptual differences between musicians and nonmusicians are due to experience-dependent plastic brain changes rather than self-selection due to preexisting genetic differences [8,39]. Understanding how speech signals are translated from a time-partitioned external acoustical event into internal neural objects, and how psychoacoustic factors and/or individual experience affects this process, is essential for the design of more effective rehabilitation programs aimed at improving, or at least maintaining, speech listening abilities in impaired subjects.

Acknowledgments

This research was funded by an East Tennessee State University research development committee award to SE. NP presented portions of this research as a recipient of the Student Research Forum award at the American Academy of Audiology Annual Convention, Boston, MA, March 2012. A portion of this work was also awarded the best presentation award to NP at the 2012 Appalachian Student Research Forum, Johnson City, TN, USA.

References:

1. Parbery-Clark A, Skoe E, Lam C, Kraus N. Musician enhancement for speech-in-noise. Ear Hear, 2009; 30: 653–61.
2. Bishop-Liebler P, Welch G, Huss M, Thomson JM, Goswami U. Auditory temporal processing skills in musicians with dyslexia. Dyslexia, 2014; 20: 261–79.
3. Strait DL, Kraus N, Parbery-Clark A, Ashley R. Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. Hear Res, 2010; 261: 22–29.
4. Beck DL, Chasin M (eds). Music and hearing loss (special edition). Hear Rev, 2014; 21(8): 4–40.
5. Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N. Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. Proc Natl Acad Sci USA, 2009; 106: 13022–27.
6. Zendel BR, Alain C. Musicians experience less age-related decline in central auditory processing. Psychol Aging, 2012; 27: 410–17.
7. Gaser C, Schlaug G. Brain structures differ between musicians and non-musicians. J Neurosci, 2003; 23: 9240–45.
8. Hyde KL1, Lerch J, Norton A, Forgeard M, Winner E, Evans AC et al. Musical training shapes structural brain development. J Neurosci, 2009; 29: 3019–25.
9. Chandrasekaran B, Krishnan A, Gandour JT. Relative influence of musical and linguistic experience on early cortical processing of pitch contours. Brain Lang, 2009; 108: 1–9.
10. Musacchia G, Strait D, Kraus N. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hear Res, 2008; 241: 34–42.
11. Wong PCM, Skoe E, Russo NM, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat Neurosci, 2007; 10: 420–22.
12. George EM, Coch D. Music training and working memory: An ERP study. Neuropsychologia, 2011; 49: 1083–94.
13. Pallesen KJ, Brattico E, Bailey CJ, Korvenoja A, Koivisto J, Gjedde A et al. Cognitive control in auditory working memory is enhanced in musicians. PLoS One, 2010; 5: e11120.
14. Tierney AT, Bergeson-Dana TR, Pisoni DB. Effects of early musical experience on auditory sequence memory. Empir Musicol Rev, 2008; 3: 178–86.
15. Soncini F, Costa MJ. Efeito da prática musical no reconhecimento da fala no silêncio e no ruído. Pró-Fono Rev Atualização Científica, 2006; 18: 161–70 [in Italian].
16. Elmer S, Meyer M, Jäncke L. Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. Cereb Cortex, 2012; 22: 650–58.
17. Anderson S, Kraus N. Sensory-cognitive interaction in the neural encoding of speech in noise: A review. J Am Acad Audiol, 2010; 21: 575–85.
18. Ott CG, Langer N, Oechslin MS, Meyer M, Jäncke L et al: Processing of voiced and unvoiced acoustic stimuli in musicians. Front Psychol, 2011; 2: 195.
19. Besson M, Chobert J, Marie C. Transfer of training between music and speech: common processing, attention, and memory. Front Psychol, 2011; 2: 94.

DOI: 10.17430/901913

20. Gaab N, Tallal P, Kim H, Lakshminarayanan K, Archie JJ, Glover GH et al. Neural correlates of rapid spectrotemporal processing in musicians and nonmusicians. Ann NY Acad Sci, 2005; 1060: 82–88.
21. Mishra SK, Panda MR, Herbert C. Enhanced auditory temporal gap detection in listeners with musical training. J Acoust Soc Am, 2014; 136: EL173–78.
22. Elangovan S, Stuart A. Natural boundaries in gap detection are related to categorical perception of stop consonants. Ear Hear, 2008; 29: 761–74.
23. Phillips DP, Taylor TL, Hall SE, Carr MM, Mossop JE. Detection of silent intervals between noises activating different perceptual channels: Some properties of "central" auditory gap detection. J Acoust Soc Am, 1997; 101: 3694–705.
24. Phillips DP, Comeau M, Andrus JN. Auditory temporal gap detection in children with and without auditory processing disorder. J Am Acad Audiol, 2010; 21: 404–8.
25. Phillips DP, Hall SE. Independence of frequency channels in auditory temporal gap detection. J Acoust Soc Am, 2000; 108: 2957–63.
26. Phillips DP, Smith JC. Correlations among within-channel and between-channel auditory gap-detection thresholds in normal listeners. Perception, 2004; 33: 371–78.
27. Walker KMM, Hall SE, Klein RM, Phillips DP. Development of perceptual correlates of reading performance. Brain Res, 2006; 1124: 126–41.
28. Lister J, Besing J, Koehnke J. Effects of age and frequency disparity on gap discrimination. J Acoust Soc Am, 2002; 111: 2793–800.
29. Stefanatos GA, Braitman LE, Madigan S. Fine grain temporal analysis in aphasia: Evidence from auditory gap detection. Neuropsychologia, 2007; 45: 1127–33.
30. Ayotte J, Peretz I, Rousseau I, Bard C, Bojanowski M. Patterns of music agnosia associated with middle cerebral artery infarcts. Brain J Neurol, 2000; 123 (Pt 9): 1926–38.
31. Ayotte J, Peretz I, Hyde K. Congenital amusia: A group study of adults afflicted with a music-specific disorder. Brain J Neurol, 2002; 125: 238–51.
32. Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. Science, 1992; 255: 606–8.
33. Sharma A, Marsh CM, Dorman MF. Relationship between N1 evoked potential morphology and the perception of voicing. J Acoust Soc Am, 2000; 108: 3030–35.
34. Strouse A, Ashmead DH, Ohde RN, Grantham DW. Temporal processing in the aging auditory system. J Acoust Soc Am, 1998; 104: 2385–99.
35. Tyler RS, Summerfield Q, Wood EJ, Fernandes MA. Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners. J Acoust Soc Am, 1982; 72: 740–52.
36. Sanju HK, Nikhil J, Kumar, P. Effect of carnatic vocal music training and experience on cortical auditory evoked potentials, J Hear Sci, 2016; 6: 40–47.
37. Nikjeh DA, Lister JJ, Frisch SA. Hearing of note: An electrophysiologic and psychoacoustic comparison of pitch discrimination between vocal and instrumental musicians. Psychophysiol, 2008; 45: 994–1007.
38. Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J Neurosci, 2003; 23: 5545–52.
39. Fujioka T, Ross B, Kakigi R, Pantev C, Trainor LJ. One year of musical training affects development of auditory cortical-evoked fields in young children. Brain J Neurol, 2006; 129: 2593–608.
40. Eimas PD, Siqueland ER, Jusczyk P, Vigorito J. Speech perception in infants. Science, 1971; 171: 303–6.
41. Werker JF, Tees RC. Speech perception in severely disabled and average reading children. Can J Psychol, 1987; 41: 48–61.
42. Werker JF, Tees RC. Speech perception as a window for understanding plasticity and commitment in language systems of the brain. Dev Psychobiol, 2005; 46: 233–51.
43. Gordon-Salant S, Yeni-Komshian GH, Fitzgibbons PJ, Barrett J. Age-related differences in identification and discrimination of temporal cues in speech segments. J Acoust Soc Am, 2006; 119: 2455–66.
44. Musacchia G, Sams M, Skoe E, Kraus N. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc Natl Acad Sci USA, 2007; 104: 15894–98.
45. Stuart A, Phillips DP. Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusic listeners. Ear Hear, 1996; 17: 478–89.
46. Shahin AJ. Neurophysiological influence of musical training on speech perception. Front Psychol, 2011; 2: 126.
47. Ruggles DR, Freyman RL, Oxenham AJ. Influence of musical training on understanding voiced and whispered speech in noise. PloS One, 2014; 9: e86980.
48. Boebinger D, Evans S, Rosen S, Lima CF, Manly T, Scott SK. Musicians and non-musicians are equally adept at perceiving masked speech. J Acoust Soc Am, 2015; 137: 378–87.
49. Swaminathan J, Mason CR, Streeter TM, Best V, Kidd J Gerald, Patel AD. Musical training, individual differences and the cocktail party problem. Sci Rep, 2015; 5: 11628.
50. Cervera TC, Soler MJ, Dasi C, Ruiz JC. Speech recognition and working memory capacity in young-elderly listeners: Effects of hearing sensitivity. Can J Exp Psychol Rev Can Psychol Exp, 2009; 63: 216–26.
51. Parbery-Clark A, Strait DL, Anderson S, Hittner E, Kraus N. Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. PloS One, 2011;6: e18082.
52. Strait D, Kraus N. Playing music for a smarter ear: Cognitive, perceptual and neurobiological evidence. Music Percept, 2011; 29: 133–46.
53. Wilson RH, McArdle RA, Smith SL. An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. J Speech Lang Hear Res, 2007; 50: 844–56
54. Strait DL, Parbery-Clark A, Hittner E, Kraus N. Musical training during early childhood enhances the neural encoding of speech in noise. Brain Lang, 2012; 123: 191–201.
55. Wilson RH. Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. J Am Acad Audiol, 2003; 14: 453–70.
56. Kraus N, Chandrasekaran B. Music training for the development of auditory skills. Nat Rev Neurosci, 2010; 11: 599–605.
57. Patel AD. Language, music, syntax and the brain. Nat Neurosci, 2003; 6: 674–81.
58. Abrams DA, Bhatara A, Ryali S, Balaban E, Levitin DJ, Menon V. Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. Cereb Cortex, 2011; 21: 1507–18.

DOI: 10.17430/901913