Title: Speech-in-speech perception, non-verbal selective attention, and musical training

Authors: Adam Tierney[a†], Stuart Rosen[b], Fred Dick[a,b]

Affiliations: [a]Birkbeck College, University of London, Malet Street, London, England WC1E 7HX [b]UCL, Gower Street, London, England WC1E 6BT

†Corresponding author

Adam Tierney, Ph.D.

Birkbeck College

a.tierney@bbk.ac.uk

Abstract

Speech is more difficult to understand when it is presented concurrently with a distractor speech stream. One source of this difficulty is that competing speech can act as an attentional lure, requiring listeners to exert attentional control to ensure that attention does not drift away from the target. Stronger attentional control may enable listeners to more successfully ignore distracting speech, and so individual differences in selective attention may be one factor driving the ability to perceive speech in complex environments. However, the lack of a paradigm for measuring non-verbal sustained selective attention to sound has made this hypothesis difficult to test. Here we find that individuals who are better able to attend to a stream of tones and respond to occasional repeated sequences while ignoring a distractor tone stream are also better able to perceive speech masked by a single distractor talker. We also find that participants who have undergone more musical training show better performance on both verbal and non-verbal selective attention tasks, and this musician advantage is greater in older participants. This suggests that one source of a potential musician advantage for speech perception in complex environments may be experience or skill in directing and maintaining attention to a single auditory object.

Keywords: selective attention, speech, cocktail party, musicians, executive function

Word count: 6669

1. Introduction

Perceiving speech requires listeners to rapidly perform phonetic and prosodic categorization while simultaneously adjusting for inter- and intra-speaker variability. This feat is impressive even in the acoustically pristine environment of the laboratory, but even more extraordinary in the noisy, crowded acoustic environment of day-to-day life, where competing sound sources can interfere with speech comprehension. This interference can take several forms. One is so-called 'energetic masking' where there is spectrotemporal overlap between target speech and a competing sound, thereby potentially obscuring crucial acoustic information in the speech stream. Another is ‘modulation masking’, also requiring spectrotemporal overlap, whereby modulations in the masker interfere with the perception of modulations in the target stream (Stone et al., 2012). A third is 'informational masking' (Brungart, 2000; Brungart et al., 2001; Kidd et al., 2008), which can be defined as any interference from the competing signal which is not driven by energetic or modulation masking. One possible source of informational masking in competing talker situations is the attentional demand associated with focusing on the target talker in the presence of multiple salient competing signals (object selection, Shinn-Cunningham, 2008). Research has shown that when speech is presented with competing speech, masking arising from spectrotemporal overlap plays only a secondary role (Brungart et al., 2006), suggesting that informational masking may interfere with speech perception in everyday listening situations.

Informational masking is less severe if sound sources can be perceptually separated along a particular acoustic dimension. For example, listeners perform better when listening to speech presented in the presence of competing speech if the talkers differ substantially in pitch (Darwin et al., 2003, Lee and Humes, 2012), speech rate (Gordon-Salant and Fitzgibbons, 2004), onset time (Kitterick et al., 2010), or spatial location (Gatehouse and Akeroyd, 2008). Individual differences in the ability to separate sound sources along different acoustic dimensions may, therefore, be one source of variability in speech perception in complex environments.

Individual differences in the ability to establish and direct attention to auditory objects could reflect variability in both bottom-up and top-down processes. On the one hand, listeners whose encoding of a particular auditory dimension is imprecise or blurred may struggle to separate perceptual objects which differ along this dimension (Shinn-Cunningham and Best, 2008). The theory that impaired perceptual precision can worsen informational masking is supported by studies showing that listeners with hearing loss are less able to separate streams based on their perceptual characteristics (Grose and Hall, 1996; Mackersie et al., 2001) and show less attentional modulation of cortical responses to sound (Dai et al. 2018). This notion is also supported by links between the ability to perceive speech in competing speech and the robustness of subcortical encoding of sound (Ruggles et al., 2012) as well as between temporal coding fidelity and spatial selective listening (Bharadwaj et al., 2015). On the other hand, the ability to direct and maintain attention to a particular auditory dimension—and, specifically, to a particular range of values along an auditory dimension—may also be a foundational skill for speech perception in complex environments. This theory is supported by findings that reported listening difficulties in complex environments are linked to impaired attention switching (Dhamani et al., 2013; Sharma et al., 2014) and that the ability to understand speech in multi-talker babble or noise correlates with performance on tests of attentional control (Fullgrabe et al., 2015; Heinrich et al., 2015, Neher et al., 2009; Neher et al., 2011; Oberfeld and Klockner-Nowotny, 2016; Yeend et al., 2017). However, other studies have reported a lack of relationship between attentional control and speech-in-speech perception (Gatehouse and Akeroyd, 2008; Heinrich et al., 2016; Schoof and Rosen, 2014).

This inconsistency may be due to the fact that, as pointed out by Oberfeld and Klockner-Nowotny (2016), the attentional skills that have been assessed in these studies are far removed from the demands of listening in complex environments. With the exception of Oberfeld and Klockner-Nowotny (2016), who used a flanker task and a test of intensity discrimination under backward masking as visual and auditory selective attention tests, all of these studies assessed attention using the Test of Everyday Attention (Robertson et al., 1996). This test battery assesses attention in ways that are not obviously relevant to the perception of speech in complex environments, including tests of visual search, tone counting, addition and subtraction on the basis of visual or auditory information, and visual search in the presence of distractors. Only one subtest assesses the ability to process auditory information in the presence of auditory distractors: the “Elevator Counting With Distraction” subtest, in which participants are asked to count low tones while ignoring high tones. However, of the abovementioned studies, only Gatehouse and Akeroyd (2008) and Neher et al. (2009) included this subtest in their test battery, and neither of these studies reported significant correlations between performance on this subtest and speech listening in complex environments. Moreover, due to the simplistic nature of the stimuli in this subtest it is far less demanding of attentional and cognitive resources compared to listening to speech in complex environments.

Our understanding of the role of auditory selective attention in the perception of speech with distractors has, therefore, been held back by the lack of an appropriate paradigm for measuring non-verbal sustained auditory selective attention. In this study we assessed attention using the sustained auditory selective attention (SASA) task (Dick et al., 2017; Holt et al., 2018). In this task, participants are asked to listen to a target stream of short tone sequences and listen for occasional tone sequence repeats, while ignoring repeats in a distractor stream. This paradigm requires the maintenance of attention to a target sound stream, inhibition of attention to a distractor sound stream, and integration of information within the target sound stream (to ensure that even transient lapses in attention are detrimental to performance, as would be the case for conversation at a cocktail party). The unique combination of stream segregation, stream selection, and within-stream integration assessed by this test, which is also characteristic of cocktail party listening, has not been featured in the tasks used in prior studies of selective attention and speech-in-speech perception. In the “Elevator Counting With Distraction” test, for example, low tones and high tones are not presented at a high enough rate that they form two separate streams; rather, participants are asked to categorize and count elements presented one at a time within a single stream. In the intensity discrimination under backward masking task used in Oberfeld & Klockner-Nowotny (2016), although listeners need to direct attention towards a target sound and away from a distractor sound, they are not asked to integrate information from a target stream over time.

In the current study, performance on this task was compared to performance on the coordinate response measure (CRM) paradigm, in which participants listen to a target talker who indicates a colour-number combination, while ignoring similar information spoken by a distractor talker (Bolia et al. 2000). This task and the SASA task are closely matched on cognitive demands, with both presenting stimuli drawn from a small closed set and both requiring integrative listening, but with the tasks differing in the use of verbal versus non-verbal stimuli. The use of these closely matched tasks enabled us to investigate whether there exist stable domain-general individual differences in the ability to direct attention to specific auditory dimensions. For each task we included two conditions, one in which the two sound streams were primarily separable by frequency, and another in which the streams were primarily separable by time. We hypothesized that correlations would be strongest between the verbal and non-verbal frequency-separation tasks and between the two temporal-separation tasks, reflecting the influence of the precision of spectral and temporal auditory processing on the ability to perform auditory object separation. To examine the specificity of links between attention and speech perception in the context of distractors, we also included a test of executive function which does not require attentional selection of auditory objects (an auditory Stroop test, Green and Barber, 1981).

A second goal of this study was to investigate transfer of musical training to the perception of speech in complex environments. There have been many reports of a musician advantage for speech perception in speech maskers (Baskent and Gaudrain, 2016; Clayton et al., 2016; Deroche et al., 2017; Du and Zatorre, 2017; Meha-Bettison et al., 2017; Morse-Fortier et al., 2017; Parbery-Clark et al., 2009; Slater and Kraus, 2016; Swaminathan et al., 2015; Yeend et al., 2017; Zendel and Alain, 2012; for a review see Coffey et al., 2017).

However, there have also been several which did not demonstrate this putative musician advantage (Boebinger et al., 2014; Fuller et al., 2014; Madsen et al., 2017; Ruggles et al., 2014). Moreover, the mechanism underlying this potential musician advantage is poorly understood. One possibility is that a musician advantage may stem from more precise auditory encoding. For example, musicians show a large advantage for pitch discrimination (Micheyl et al., 2006; Carey et al., 2015) and more precise neural phase-locking to pitch (Wong et al., 2007). Given the usefulness of pitch for separating sound streams (Darwin et al., 2003; Lee and Humes, 2012), the potential musician advantage for speech perception in speech maskers could stem from more precise pitch perception, facilitating auditory object segregation. However, several studies have shown that benefits for F0 separation in speech-on-speech perception are no larger in musicians than in non-musicians (Deroche et al., 2017; Madsen et al., 2017). Similarly, the potential musician advantage may stem from more precise encoding of syllabic timing, given that speech rate can also be used to segregate sound sources (Gordon-Salant and Fitzgibbons, 2004; Kitterick et al., 2010) and musicians display advantages for the perception of sound timing (Repp, 2010), although this possibility has not been directly tested.

Another possibility is that a musician advantage for perceiving speech in the presence of distractors stems from an enhanced ability to direct and maintain attention to a sound source while resisting the pull of distracting information. When playing in large ensembles musicians must learn to direct their attention to a single sound source within a highly complex acoustic environment using melodic and rhythmic information, an experience that may lead to domain-general enhancements in the control of auditory attention (Rodrigues et al., 2013; Tervaniemi et al., 2005; Tervaniemi et al., 2009). Moreover, musicians show less informational masking during tone detection (Oxenham et al., 2003) and the musician advantage for speech-in-speech perception is larger in high-informational-masking conditions (Morse-Fortier et al., 2017; but see Swaminathan et al., 2015, who showed a musician advantage only in a low-informational-masking condition). Carey et al., (2015) however, found no musician advantage in a more naturalistic non-verbal informational masking paradigm using environmental sound scenes.

Musicians' experience with musical beat perception and production may enhance their attentional control as well. According to an influential model of musical rhythm, the perception of musical beats involves a regular waxing and waning of temporal expectation over time (Large and Kolen, 1994). This theory is supported by findings that reaction times are faster to stimuli that occur at anticipated on-beat versus off-beat times (Barnes and Jones, 2000; Bolger et al., 2013; Brochard et al., 2013; McAuley and Jones, 2003; Miller et al., 2012). Complex musical rhythms lacking a single clear pulse may require the listener to actively attend before an internal beat can be established. In particular, perceiving a beat in musical stimuli which consist of multiple sound streams produced by multiple instruments may require initial stream segregation following by stream selection, as listeners assess the rhythmic patterns of each stream and compare them in order to determine the likely location of the beat. Supporting this idea, selective attention to cycling musical rhythms modulates the extent of increase in basal ganglia activity over time (Chapin et al., 2010). Moreover, the ability to perceive and produce beats is linked to individual differences in sustained attention (Birkett and Talcott, 2012; Tierney and Kraus, 2013), and children with ADHD have difficulty synchronizing to beats (Pitcher et al., 2002; Toplak and Tannock, 2005; Ben-Pazi et al., 2006). Thus, the ability to accurately and precisely perceive and produce musical beats may place strong demands upon attentional control; therefore, years of experience with musical beat processing may lead to attentional enhancements. However, no prior study has investigated whether musical beat perception is associated with auditory selective attention ability.

Here we examined relationships between instrumental musical training and verbal and non-verbal selective attention by relating performance on the SASA and CRM task batteries with the degree of musical training in an adult population. If precise pitch encoding facilitating auditory object segregation is the primary factor driving the musician advantage, then only the SASA and CRM conditions in which frequency information is most useful for separating the sound sources should be associated with musical training. On the other hand, if all of the SASA and CRM conditions are associated with the degree of musical training, this would suggest either that musician advantages for auditory encoding which are relevant to auditory object selection extend across both encoding of pitch and of syllabic timing, or that the musician advantage stems from an enhanced ability to direct selective attention to auditory objects. We also examined relationships between CRM scores and performance on the SASA test, as well as a test of musical beat perception. Relationships between these music tests and CRM performance could provide evidence of overlap between components of music training and verbal selective attention, and such overlap is a key ingredient necessary for transfer of musical experience to linguistic skills (Patel, 2011). Specifically, we predicted that beat perception would relate to performance on the CRM and SASA conditions in which temporal information is more useful for separating sound sources.

2. Methods

2.1 Participants

93 participants were recruited from a variety of online sources, including social media and the Sona Experiment Management System. They were run on the online experiment platform Gorilla, and required to use headphones for all tasks. Participants were asked to rate their proficiency between 1 (low) and 7 (high) for English speaking, reading, and writing. To ensure that all participants were reasonably fluent in English, participants who scored less than 6 on average across all three questions (n = 22) were excluded a priori before any other aspect of their data was considered. No participants reported diagnosis of a hearing loss. Furthermore, two participants were excluded based on their performance on the Stroop task, one because their high degree of variability in reaction times in the Stroop task (> 2 SD from mean) suggested a lack of dedication to the task and the other because the participant answered every trial incorrectly during the incongruent condition, indicating a lack of understanding of the task instructions. 69 participants remained (32 female). This sample size was sufficient to supply 82% power to detect a medium effect size, i.e. a correlation of 0.3

(power analysis conducted using G*Power, Faul, Erdfelder, Lang, & Buchner, 2007). Their mean age was 32.5 years (standard deviation 11.2), ranging from 16 to 65 years. On average participants reported 13.2 (14.0) years of instrumental musical training, ranging from 0 years to 54 years of training. Of these participants, 50 were native speakers of English, while the average age of acquisition of English in the remaining 19 participants was 12 years (standard deviation 8 years). The Ethics Committee in the Department of Psychological Sciences at Birkbeck, University of London approved all experimental procedures. Informed consent was obtained from all participants. Participants were not compensated for their participation.

2.2 Auditory Stroop Task

Participants were given an *Auditory Stroop Task* (Green and Barber, 1981; Miyake et al., 2000) which measured the extent to which participants could inhibit their response to the meaning of an auditorily presented word. Two male talkers and two female talkers were recorded saying four words: “man”, “woman”, “game”, and “leave”. During this test, participants heard six tokens of each combination of talker and word, for a total of 96 stimuli. For each word, participants were asked to indicate as rapidly as possible whether the talker was male or female by pressing either the “m” or “w” keys. The Stroop effect was measured as the difference in reaction times between the congruent (when the speaker matched the gender of the word) and incongruent (when the speaker and word gender did not match) conditions.

2.3 Non-verbal Selective Attention Task

We measured the extent to which participants could direct and maintain attention to a non-verbal sound source in the presence of distraction using the *Sustained Auditory Selective Attention* paradigm (SASA, Dick et al., 2017; Holt et al., in press). In this task participants were presented with two streams of tone sequences of three elements, one in a higher band and another in a lower band. For each band the tones which made up the sequences were selected from a pool of three possible tones. Within each band, tone sequences were concatenated, with each sequence consisting of three 125-ms tones (each separated by 125 ms of silence) followed by 250 ms of silence. The total duration of each sequence, therefore, was 1 second. Each block consisted of 30 sequences. Within each block, in each band, there occurred between 3 and 6 repeats, i.e. cases in which a sequence was identical to the previous sequence. Repeating sequences were always separated by at least one non-repeating sequence. Participants were asked to attend to either the low or high band and click the mouse whenever they detected one of these repeats, ignoring the other band. Correct mouse clicks were rewarded by a green check mark. Incorrect mouse clicks and repeats to which the participant did not respond were indicated by a red X.

The SASA task contained two conditions. In the *Simultaneous* condition, sequences in the high and low band began simultaneously. The lowest tone for the low band was set at 370 Hz, and the other two tones were two and four semitones higher, while the lowest tone for the high band was set at 1110 Hz, and the other two tones were two and four semitones higher. In this condition, therefore, although the simultaneous presentation makes sequence separation on the basis of temporal information impossible, the sequences can be easily separated by frequency. In the *Interleaved* condition, sequences in the low band began 125 ms before sequences in the high band. As a result, when a tone was present in one band, the other band was silent. The tones in the low band were identical to those in the Simultaneous condition, but the lowest tone for the high band was 520 Hz, and the other two tones were two and four semitones higher. In this condition, therefore, sequences could be easily separated by time but less easily separated by frequency. In each condition, participants completed eight blocks in which they were asked to attend to the higher band and eight blocks in which they were asked to attend to the lower band, for a total of 32 blocks. Scores were converted to d-prime prior to analysis, with any click occurring in a window other than a target window considered to be a false alarm, with the total number of windows not containing a target then equal to the number of distractors.

2.4 Verbal Selective Attention Task

We measured the extent to which participants could direct and maintain attention to a target talker in the presence of a single distractor talker using a modified version of the *Coordinate Response Measure* paradigm (CRM, Bolia et al., 2000). In each trial of this task participants heard two talkers say the sentence “Show the [animal] where the [color] [number] is”, where [animal] could be “dog” or “cat”, [color] could be “black”, “blue”, “green”, “pink”, or “red”, and [number] could be 1, 2, 3, 4, 5, 6, 8, or 9. All sentences were duration normalized using Praat (Boersma et al. 2018) such that each sentence’s duration was set to the mean duration of all sentences in the corpus (1.95 s). After the sentences were played, a grid of 40 numbers appeared on the screen, one for each color/number combination. Participants were asked to click on the color/number combination indicated by the talker who said the word “dog”. This talker was always the same (male) talker, and so participants knew in advance which speaker they needed to attend to. Target and distractor sentences always contained different colours and numbers. The target sentence was presented at an SNR of 0 dB with respect to the distractor sentence. Both the target and distractor sentences were produced by talkers who spoke standard southern British English.

The CRM task contained two conditions. In the *Different Gender* condition, the talker who spoke the sentence containing the word “dog” (i.e. the target talker) was male, while the distractor talker was female, and the sentences began simultaneously. In this condition, therefore, the talkers could be segregated by fundamental frequency (F0) but not by time. In the *Same Gender* condition, both talkers were male, but the talker who spoke the sentence containing the word “dog” began 150 ms before the distractor talker. In this condition, therefore, the talkers could be segregated by time but much less so by talker F0. In each condition, participants completed 40 trials (one for each color-number combination), for a total of 80 trials.

2.5 *Musical Beat Perception Task*

We measured the extent to which participants could track the beat of a piece of instrumental music using a modified version of the Beat Alignment Test (Iversen and Patel, 2008). Participants were presented with brief audio clips of instrumental music. Superimposed on each of the clips was a sequence of tones. These tones were either aligned with the beat of the music or shifted away from the beat by 30% of the interval between beats. Participants were asked to indicate whether or not the tones were aligned with the beat of the music by pressing one of two buttons with the mouse. Scores were converted to d-prime prior to analysis, with the loglinear approach used to avoid infinite scores (Hautus, 1995).

2.6 Analysis

There was a wide range of ages in the dataset (SD = 11.8), raising the possibility of a confounding relationship with age. To examine this possibility, we ran pairwise non-parametric correlations between age and performance across all of the outcome measures. The only significant correlation was with performance on the Stroop task, $r = 0.28$, $p = 0.020$, such that older participants showed greater effects of gender-word congruency on reaction times; all other correlations with age had r values of less than 0.2. As a result, it is unlikely that any relationships between performance on SASA and CRM tasks, as well as relationships between years of musical training and cognitive performance, reflect a confounding relationship with age. Nevertheless, we covaried for age when conducting correlational analyses to ensure that age is not driving any relationships reported.

First, we used partial Spearman correlations, covarying for age, to examine the relationship between non-verbal selective attention, verbal selective attention, musical beat perception, the size of the Stroop effect, and years of musical training. (Data were non-normally distributed, so non-parametric methods were used). False Discovery Rate was used to correct for multiple comparisons, with p-values across all correlations entered into the correction (Benjamini-Hochberg procedure, Benjamini and Hochberg, 1995). All p-values from correlations reported in the remainder of the paper, both in text and in tables, are corrected p-values. Statistical significance of differences between correlations was tested using Fisher's r-to-z transformation.

Second, we conducted a logistic regression to examine the extent to which several different predictors explained variability in speech-in-speech perception. A number of transformations were applied to the data prior to regression analysis. Years of musical training were converted to a dichotomous variable, with 0 indicating no musical training and 1 indicating at least one year of training. To minimize the problem of multicollinearity, we used a generalized least squares factor analysis with varimax rotation to uncover latent variables reflecting shared variance across the four behavioral predictors, SASA Interleaved, SASA Simultaneous, Beat Perception, and Stroop Effect. We extracted two factors, the first of which accounted for 54.6% of variance, and the second of which accounted for an additional 25.4% of variance. Table 3 lists factor loadings for these two factors. The first factor shows high loadings for SASA Simultaneous, SASA Interleaved, and Beat Perception; we interpret this factor as reflecting Selective Attention ability. The second factor shows high loadings for Stroop Effect scores only. Finally, the regression analysis was conducted by first including the Selective Attention factor, the Stroop factor, age, musical training, and the interaction between musical training and the three continuous variables as predictors and summed CRM performance across both different-gender and same-gender conditions as the outcome variable. Model selection with was performed by sequentially excising from the model terms that were not significant, using a quasibinomial approach in order to deal with overdispersion.

Given our use of an online experiment design, we could not control the testing environment to the extent which in-lab testing makes possible. This could be a source of cross-subject variability which is consistent across tasks, potentially leading to spurious correlations. To examine this possibility, as a measure of distractibility we calculated reaction time variability across all conditions of the auditory Stroop task, and correlated this metric with performance on the other measures. All correlations failed to reach significance, $p > 0.05$. These results suggest that variability in testing environment is not the primary factor driving our results. Nevertheless, it remains possible that the distractibility of the environment may have changed between tests, and so we cannot rule out the possibility that environmental variability may have added a certain amount of noise to our data. However, research directly comparing online testing to in-lab testing has revealed that for cognitive and perceptual measurement online testing is no less reliable (Germine et al., 2012). Furthermore, online testing provides several benefits over in-lab testing, including the ability to recruit larger numbers of participants, as well as participant populations which are more diverse and more representative of the general population (Gosling, Vazire, Srivastava, & John, 2004; Buhrmester, Kwang, & Gosling, 2011).

3. Results

3.1 Correlations between cognitive tasks

Summary measures for all behavioral measures are listed in Table 1. Reliability, measured using split-half correlation, was high for the CRM tests, with the Different Gender condition showing a reliability of 0.78, while the Same Gender showed a reliability of 0.93. Nevertheless, visual inspection of score distributions revealed a ceiling effect for some participants on the CRM task in the Different Gender condition, but a floor effect for some participants in the Same Gender condition (see Figure 1), which could potentially limit the possible size of inter-task correlations. Reliability was similarly high for the SASA Interleaved (0.87) and SASA Simultaneous (0.91) tests. However, reliability was lower for the Beat Perception (0.53) and Stroop Effect (0.40) tests. As a result, any null effects for these tests should be interpreted with caution.

Participants who performed better on the non-verbal selective attention tasks tended to perform better on the verbal selective attention tests. Total performance across both SASA conditions, for example, correlated with total performance across both CRM conditions (rho = 0.31; see Figure 2). Contrary to our hypothesis, however, correlations between verbal and non-verbal attention were no stronger for conditions in which the relative usefulness of spectral versus temporal information in segregating the sound streams was matched. For example, the correlation between CRM Same Gender (where talkers were temporally offset) and SASA Simultaneous (rho = 0.17) was not significantly different (> 0.05) from the correlation between

CRM Different Gender (simultaneous talker onset) and SASA Simultaneous (rho = 0.25). The correlation between CRM Same Gender and SASA Interleaved (rho = 0.31) also did not significantly differ from the correlation between CRM Different Gender and SASA Interleaved (rho = 0.36). The size of the Stroop effect did not significantly correlate with performance in any of the selective attention tests (all $p > 0.1$). Correlations between scores on all cognitive tasks can be found in Table 2.

Participants who performed better on the test of musical beat perception also performed better on both verbal and non-verbal selective attention tests. For example, beat perception correlated with total performance across both SASA conditions (rho = 0.52) and total performance across both CRM conditions (rho = 0.34). However, and again contra our hypotheses, correlations between beat perception and selective attention were not stronger for the conditions of the attention tests that relied on the ability to use temporal information to segregate the sound streams. For example, the correlation between beat perception and SASA Interleaved (rho = 0.48) was not significantly greater than the correlation between beat perception and SASA Simultaneous (rho = 0.48). Similarly, the correlation between beat perception and CRM Same Gender (rho = 0.3) was not significantly greater than the correlation between beat perception and CRM Different Gender (rho = 0.39).

3.2 Musical training and cognitive performance

Participants who reported more years of musical training performed better on selective attention tasks. For verbal selective attention, greater musical training was linked to better performance in both the Same Gender (rho = 0.27, $p < 0.05$) and Different Gender (rho = 0.37, $p < 0.01$) conditions, as well as total performance (rho = 0.31, $p < 0.05$; see Figure 3, top). For non-verbal selective attention, greater musical training was linked to better performance in the Interleaved condition (rho = 0.45, $p < 0.001$) and the Simultaneous condition (rho = 0.44, $p < 0.001$), as well as to total performance (rho = 0.47, $p < 0.001$; see Figure 3, bottom). There was no significant relationship between amount of musical training and the size of the Stroop effect (rho = -0.10, $p > 0.1$). Finally, participants with a greater degree of musical experience were better able to detect the beat of music (rho = 0.50, $p < 0.001$). Correlations between years of musical training and performance across all cognitive tasks can be found in Table 2.

3.3 Correlations in native English-speaking participants

To investigate whether the above-reported relationships might partially reflect a confounding relationship with age of acquisition of English, these correlations were re-run on a smaller dataset consisting of only those participants who were native speakers of English (n = 50). The only notable change in the pattern of results was that the relationship between verbal selective attention performance in the Same Gender condition and non-verbal attention total performance was no longer significant after FDR correction (rho = 0.30, $p = 0.053$). Otherwise, all significant relationships as reported in sections 3.2 and 3.1 remained significant, and all relationships which did not previously reach significance remained below threshold.

3.4 Multiple Regression

A logistic regression was conducted with CRM performance as the outcome variable with predictors of age, musical training, Selective Attention and Stroop (the two extracted factors described above), and the interaction between musical training and the three continuous variables. Terms that were not significant were excised sequentially. The resulting model included age, musicianship, and the interaction between age and musicianship as predictors (see Table 4), with the interaction term highly significant (F test comparing models with and without the interaction, $F = 8.24$, $p < 0.006$). To interpret this interaction, we ran two follow-up logistic regressions with age as predictor and CRM performance as outcome variable in musicians and non-musicians. This analysis revealed that CRM performance does not change with age in the musicians ($F = 0.29$, $p > 0.5$), whereas it worsens for the nonmusicians ($F = 8.53$, $p < 0.01$). See Figure 4 for a scatterplot illustrating the age by musicianship interaction. No other variables emerged as significant predictors.

4. Discussion

We found that participants who were better able to selectively attend to a target tone stream while ignoring a distractor tone stream were also better able to perceive spoken sentences in the presence of a speech distractor. This finding, along with previous research showing that intensity discrimination under backward masking is associated with speech-in-speech perception (Oberfeld and Klockner-Nowotny 2016), suggests that the ability to direct attention towards auditory objects (and away from distractor objects) is one factor determining successful speech comprehension in the distracting environments which are commonly encountered in everyday life.

Regression analysis, however, revealed that when attention, Stroop performance, age, musicianship, and the interaction between musicianship and the other three predictors were all entered into a regression model, the only predictors which emerged as significant were age, musicianship, and their interaction. There are several conclusions which can be drawn from these findings. First, the interaction between age and musicianship suggests that musical training may have a particularly beneficial effect for speech-in-speech perception later in life. This is in line with previous reports that age-related declines in speech-in-noise perception are smaller in musicians (Zendel & Alain, 2012), and that older adults assigned music training showed gains in speech-in-noise perception relative to a control group who were assigned to learn a visuo-spatial video game (Zendel, West, Belleville, & Peretz, 2019). With respect to the putative link between attention and speech-in-speech perception mentioned just above, there are two possible interpretations of the results of the regression analysis. The first is that the link between attention and speech-in-speech perception is spurious, resulting from the influence of a third variable: musical training. According to this interpretation, musical training enhances executive function, and also enhances speech-in-speech perception, but for reasons unrelated to attention (such as improved bottom-up segregation of sound sources due to more robust subcortical representation of sound, Bidelman & Alain, 2015). An alternate interpretation of these results is that musical training enhances speech-in-speech perception by enhancing attention. Both of these interpretations are consistent with previous reports that musical training is linked to enhanced executive function skills, including attention, in older adulthood (Amer, Kalender, Hasher, Trehub, & Wong, 2013; Zendel & Alain, 2013; Moussard, Bermudez, Alain, Tays, & Moreno, 2016). Future research could adjudicate between these two interpretations by examining the effects of non-verbal attention training versus control training in older adults.

More musical training was linked to enhanced performance across both frequency-selective and temporally-selective conditions of the CRM tests. There are several possible interpretations of this result. One is that musicians benefit from more precise perception of both temporal and spectral information, boosting the ability to segregate sound sources which are separable in either dimension. Alternatively, domain-general enhancements in the control of auditory selective attention may enable musicians to better direct attention towards target sound streams and away from distractor sound streams, regardless of whether the stimuli are verbal or non-verbal and regardless of the acoustic dimensions which are most useful in separating them. These two possibilities are not mutually exclusive, and in practice the musician advantage (if one exists) may stem from a mix of bottom-up and top-down processes. That correlations between the two frequency-selective conditions and between the two temporally-selective conditions were not stronger than correlations across dimensional selectivity conditions, however, suggests that individual differences in SASA and CRM performance may have been primarily driven by domain-general differences in attentional control, which supports a stronger role for top-down versus bottom-up processing in any putative musician advantage for auditory selective attention. These results do not support accounts of the musician advantage for speech perception in the presence of distractors which posit that that these enhancements stem from increased precision in the perception of a single acoustic dimension. For example, if precise pitch perception were the main factor driving the musician advantage for speech-in-speech perception then the musician advantage should have been attenuated or non-existent in conditions in which the sound sources were less separable by frequency (such as the same-gender condition of the CRM test).

An alternate explanation—which we cannot rule out—is that there are pre-existing differences in auditory selective attention which help determine whether individuals choose to engage in musical training in the first place. Although relationships between degree of musical training and perceptual and cognitive skills have been used to argue for a causal relationship between musical training and cognitive enhancement (Strait and Kraus, 2011), it remains possible instead that perceptual and cognitive factors predict whether an individual will continue to pursue training rather than losing interest (Corrigall et al., 2013; Corrigall and Schellenberg, 2015). For example, links between musical training and speech-in-speech perception may indicate that individuals with heightened auditory attention abilities may be more likely to succeed in music, encouraging them to stay involved in music performance. Alternately, the link between musical training and speech-in-speech perception may reflect a more general cognitive trait which predicts both commitment to musical training and speech-in-speech perception, such as IQ. These factors could be disentangled in future longitudinal music training studies examining selective nonverbal auditory attention, general cognitive factors such as IQ and working memory, and speech-in-speech perception.

The ability to perceive the beat of music was linked to verbal and non-verbal selective attention performance, and correlations were similar for the frequency-selective and temporally-selective conditions. This result suggests that musical beat perception makes demands on attentional control, in accordance with prior findings that musical beat perception relies upon selective attention (Chapin et al., 2010). Thus, listening to speech in competing speech correlates with two different aspects of musical training, attending to one melodic line while ignoring another (i.e. SASA) and perceiving musical beats. Future work could test the hypothesis that these two different aspects of musical training are key in boosting verbal selective attention by examining selective attention and speech-in-speech perception in musicians whose training has focused on different aspects of music perception. For example, drummers have been reported to show a particular advantage for perceiving speech in complex environments, even when compared to vocalists (Slater and Kraus, 2016). An additional possibility which could be tested by future research is that musicians who perform primarily in groups versus solo and who transcribe melodic lines by ear versus primarily learning written material may also display a larger advantage for verbal selective attention.

We find that the musician advantage is specific to measures of auditory selective attention and does not extend to executive function more generally, as measured using an auditory Stroop test (Green and Barber, 1981), suggesting that musical experience specifically enhances attentional control. This finding contradicts previous reports that musicians display an advantage in inhibitory control, as measured using a variety of measures, importantly including Stroop tasks (Amer et al., 2013; Bialystok and DePape, 2009; Holochwost et al., 2017; Schroeder et al., 2016; Travis et al., 2011). However, there have also been a number of reports that inhibitory control does not differ between musicians and nonmusicians (D’Souza et al., 2018; Slevc et al., 2016; Vasuki et al., 2016; Zuk et al., 2014). There are several contributing factors which might explain this inconsistently reported relationship between musical training and inhibitory control. Slater et al. (2017) reported that percussionists but not vocalists displayed an advantage in inhibitory control relative to non-musicians, suggesting that an inhibitory control advantage may be linked to experience with only certain instruments or certain genres. The inconsistency of findings regarding musical training and Stroop performance may also be due to the fact that this is a derived measure, leading to low reliability. In the present study, for example, the Stroop test featured the lowest reliability of all of the predictors, which could help explain the lack of a relationship with musical training. Whether musical training enhances inhibitory control may also depend upon the amount of training and the age at which training was begun, which differs across the musician groups defined in these studies. These inconsistent findings may also relate to cross-study differences in statistical power, although across this body of work there is no obvious relationship between the number of participants tested and whether or not an effect of musicianship was found.

In conclusion, our findings suggest that when perceiving speech in the presence of distracting information, listeners may rely on their ability to maintain control over their attentional focus. Musical training may help sharpen this attentional control, leading to advantages in listening in complex environments that extend to both musical and linguistic contexts.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We would like to thank Monica Whiting, Peter Stavrou, Paula Ramon, Genoveva Berecki, and Francesca Seri for assistance with data collection, as well as Giada Guerra and Aeron Laffere for assistance with construction of the Auditory Stroop task.

References

Amer, T., Kalender, B., Hasher, L., Trehub, S., Wong, Y., 2013. Do older professional musicians have cognitive advantages? PLoS ONE 8, e71630.

Barnes, R., Jones, M. R, 2000. Expectancy, Attention, and Time. Cognitive Psychology 41, 254-311.

Baskent, D., Gaudrain, E., 2016. Musician advantage for speech-on-speech perception. JASA 139, EL51-EL56.

Ben-Pazi, H., Shalev, R., Gross-Tsur, V., Bergman, H., 2006. Age and medication effects on rhythmic responses in ADHD: Possible oscillatory mechanisms? Neuropsychologia 44, 412-416.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society 57, 289-300.

Bharadwaj, H., Masud, S., Mehraei, G., Verhulst, S., Shinn-Cunningham, B., 2015. Individual differences reveal correlates of hidden hearing deficits. Journal of Neuroscience 35, 2161-2172.

Bialystok, E., DePape, A., 2009. Musical experience, bilingualism, and executive functioning. JEP:HPP 35, 565-572.

Bidelman, G., Alain, C., 2015. Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. Journal of Neuroscience 35, 1240-1249.

Birkett, E., Talcott, J., 2012. Interval timing in children: effects of auditory and visual pacing stimuli and relationships with reading and attention variables. PLoS ONE 7, e42820.

Boebinger, D., Evans, S., Rosen, S., Lima, C., Manly, T., Scott, S., 2014. Musicians and non-musicians are equally adept at perceiving masked speech. JASA 137, 378-387.

Bolia, R. S., Nelson, W. T., Ericson, M. A., Simpson, B. D., 2000. A speech corpus for multitalker communications research. Journal of the Acoustical Society of America 107, 1065-1066.

Bolger, D., Trost, W., Schön, D., 2013. Rhythm implicitly affects temporal orienting of attention across modalities. Acta Psychologica 142, 238-244.

Brochard, R., Tassin, M., Zagar, D., 2013. Got rhythm… for better and for worse. Cross-modal effects of auditory rhythm on visual word recognition. Cognition 127, 214-219.

Brungart, D., 2000. Informational and energetic masking effects in the perception of two simultaneous talkers. JASA 109, 1101-1109.

Brungart, D., Simpson, B., Ericson, M., Scott, K., 2001. Informational and energetic masking effects in the perception of multiple simultaneous talkers. JASA 110, 2527-2538.

Brungart, D., Chang, P., Simpson, B., Wang, D., 2006. Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. JASA 120, 4007-4018.

Buhrmester, M., Kwang, T., Gosling, S., 2011. Amazon’s Mechanical Turk: a new source of inexpensive, yet high-quality, data? Perspectives on Psychological Science 6, 3-5.

Carey, D., Rosen, S., Krishnan, S., Pearce, M., Shepherd, A., Aydelott, J., Dick, F., 2015. Generality and specificity in the effects of musical expertise on perception and cognition. Cognition 137, 81-105.

Chapin, H., Zanto, T., Jantzen, J., Kelso, S., Steinberg, F., Large, E., 2010. Neural responses to complex auditory rhythms: the role of attending. Frontiers in Psychology 1, 224.

Clayton, K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A., Kidd, G., 2016. Executive function, visual attention and the cocktail party problem in musicians and non-musicians. PLoS ONE 11, e0157638.

Coffey, E., Mogilever, N., Zatorre, R., 2017. Speech-in-noise perception in musicians: a review. Hearing Research 352, 49-69.

Corrigall, K., Schellenberg, E., Misura, N., 2013. Music training, cognition, and personality. Frontiers in Psychology 4, 222.

Corrigall, K., Schellenberg, E., 2015. Predicting who takes music lessons: parent and child characteristics. Frontiers in Psychology 6, 282.

D’Souza, A., Moradzadeh, L., Wiseheart, M., 2018. Musical training, bilingualism, and executive function: working memory and inhibitory control. Cognitive Research: Principles and Implications 3, 11.

Dai, L., Best, V., Shinn-Cunningham, B., 2018. Sensorineural hearing loss degrades behavioral and physiological measures of human spatial selective auditory attention. PNAS 115, E3286-E3295.

Darwin, C., Brungart, D., Simpson, B., 2003. Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. JASA 114, 2913-2922.

Deroche, M., Limb, C., Chatterjee, M., Gracco, V., 2017. Similar abilities of musicians and non-musicians to segregate voices by fundamental frequency. JASA 142, 1739-1755.

Dhamani, I., Leung, J., Carlile, S., Sharma, M., 2013. Switch attention to listen. Scientific Reports 3, 1297.

Dick, F., Lehet, M., Callaghan, M., Keller, T., Sereno, M., Holt, L., 2017. Extensive tonotopic mapping across auditory cortex is recapitulated by spectrally directed attention and systematically related to cortical myeloarchitecture. Journal of Neuroscience 37, 12187-12201.

Du, Y., Zatorre, R., 2017. Musical training sharpens and bonds ears and tongue to hear speech better. PNAS 114, 13579-13584.

Faul, F., Erdfelder, E., Lang, A., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods 39, 175-191.

Fuller, C., Galvin, J., Maat, B., Free, R., Baskent, D., 2014. The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? Frontiers in Neuroscience 8, 179.

Fullgrabe, C., Moore, B., Stone, M., 2015. Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. Frontiers in Aging Neuroscience 6, 347.

Gatehouse, S., Akeroyd, M., 2008. The effects of cueing temporal and spatial attention on word recognition in a complex listening task in hearing-impaired listeners. Trends in Amplification 12, 145-161.

Germine, L., Nakayama, K., Duchaine, B., Chabris, C., Chatterjee, G., Wilmer, J., 2012. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. Psychon Bull Rev 19, 847-857.

Gosling, S., Vazire, S., Srivastava, S., John, O., 2004. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. American Psychologist 59, 93-104.

Gordon-Salant, S., Fitzgibbons, P., 2004. Effects of stimulus and noise rate variability on speech perception by younger and older adults. JASA 4, 1808-1817.

Green, E., Barber, P., 1981. An auditory Stroop effect with judgments of speaker gender. Perception and Psychophysics, 30, 459-466.

Grose, J., Hall, J., 1996. Perceptual organization of sequential stimuli in listeners with cochlear hearing loss. JSLHR, 39, 1149-1158.

Hautus, M., 1995. Corrections for extreme proportions and their biasing effects on estimated values of d'. Behavior Research Methods, 27, 46-51.

Heinrich, A., Henshaw, H., Ferguson, M., 2015. The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests. Frontiers in Psychology 6, 782.

Heinrich, A., Henshaw, H., Ferguson, M., 2016. Only behavioral but not self-report measures of speech perception correlate with cognitive abilities. Frontiers in Psychology 7, 576.

Holochwost, S., Propper, C., Wolf, D., Willoughby, M., Fisher, K., Kolacz, J., Volpe, V., Jaffee, S., 2017. Music education, academic achievement, and executive functions. Psychology of Aesthetics, Creativity, and the Arts 11, 147-166.

Holt, L., Tierney, A., Guerra, G., Laffere, A., Dick., F., 2018. The dynamic interplay of attention and context underlying speech processing. Hearing Research 366, 50-64.

Iversen, J., Patel, A., 2008. The Beat Alignment Test (BAT): Surveying beat processing abilities in the general population. In K. Miyazaki, M. Adachi, Y Hiraga, Y Nakajima, & M. Tsuzaki (Eds.), Proceedings of the 10th International Conference on Music Perception & Cognition (ICMPC10) (pp. 465–468). Adelaide, Australia: Causal Productions.

Kidd, G., Mason, C., Richards, V., Gallun, F., Durlach, N., 2008. Informational Masking. In: Yost, W., Popper, A., Fay, R. (Eds.), Auditory Perception of Sound Sources. Springer Science + Business Media, New York, pp. 143-190.

Kitterick, P., Bailey, P., Summerfield, A., 2010. Benefits of knowing who, where, and when in multi-talker listening. JASA 127, 2498-2508.

Large, E., Kolen, J., 1994. Resonance and the perception of musical meter. Connection Science 6, 177-208.

Lee, J., Humes, L., 2012. Effect of fundamental-frequency and sentence-onset differences on speech-identification performance of young and older adults in a competing-talker background. JASA 132, 1700-1717.

Mackersie, C., Prida, T., Stiles, D., 2001. The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss. JSLHR 44, 19-28.

Madsen, S., Whiteford, K., Oxenham, A., 2017. Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. Scientific Reports 7, 12624.

McAuley, J., Jones, M. R., 2003. Modeling effects of rhythmic context on perceived duration: a comparison of interval and entrainment approaches to short-interval timing. Journal of Experimental Psychology: Human Perception and Performance 29, 1102-1125.

Meha-Bettison, K., Sharma, M., Ibrahim, R., Vasuki, P., 2017. Enhanced speech perception in noise and cortical auditory evoked potentials in professional musicians. International Journal of Audiology 57, 40-52.

Micheyl, C., Delhommeau, K., Perrot, X., Oxenham, A., 2006. Influence of musical and psychoacoustical training on pitch discrimination. Hearing Research 219, 36-47.

Miller, J., Carlson, L., McAuley, J., 2012. When what you hear influences when you see: listening to an auditory rhythm influences the temporal allocation of visual attention. Psychological Science 24, 11-18.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., Wager, T. D., 2000. The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. Cognitive psychology 41, 49-100.

Morse-Fortier, C., Parrish, M., Baran, J., Freyman, R., 2017. The effects of musical training on speech detection in the presence of informational and energetic masking. Trends in Hearing 21, 1-12.

Moussard, A., Bermudez, P., Alain, C., Tays, W., Moreno, S., 2016. Life-long music practice and executive control in older adults: an event-related potential study. Brain Research 1642, 146-153.

Neher, T., Behrens, T., Carlile, S., Jin, C., Kragelund, L., Petersen, A., Schaik, A., 2009. Benefit from spatial segregation of multiple talkers in bilateral hearing-aid users: effects of hearing loss, age, and cognition. International Journal of Audiology 48, 758-774.

Neher, T., Laugesen, S., Jensen, N., Kragelund, L., 2011. Can basic auditory and cognitive measures predict hearing-impaired listeners’ localization and spatial speech recognition abilities? JASA 130, 1542-1558.

Oberfeld, D., Klockner-Nowotny, F., 2016. Individual differences in selective attention predict speech identification at a cocktail party. eLife 5, e16747.

Oxenham, A., Fligor, B., Mason, C., Kidd, G., 2003. Informational masking and musical training. JASA 114, 1543-1549.

Parbery-Clark, A., Skoe, E., Lam, C., Kraus, N., 2009. Musician enhancement for speech in noise. Ear and Hearing 30(6): 653-661.

Patel., A., 2011. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Frontiers in Psychology 2, 142.

Payton, M., Greenstone, M., Schenker, N., 2003. Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? Journal of Insect Science 3, 34.

Pitcher, T., Piek, J., Barrett, N., 2002. Timing and force control in boys with attention deficit hyperactivity disorder: subtype differences and the effect of comorbid developmental coordination disorder. Human Movement Science 21, 919-945.

Repp B (2010). Sensorimotor synchronization and perception of timing: effects of music training and task experience. Human Movement Science 29, 200-213.

Robertson, I., Ward, T., Ridgeway, V., Nimmo-Smith, I., 2009. The structure of normal human attention: The Test of Everyday Attention. Journal of the International Neuropsychological Society 2, 525-534.

Rodrigues, A., Loureiro, M., Caramelli, P., 2013. Long-term musical training may improve different forms of visual attention ability. Brain and Cognition 82, 229-235.

Ruggles, D., Bharadwaj, H., Shinn-Cunningham, B., 2012. Why middle-aged listeners have trouble hearing in everyday settings. Current Biology 22, 1417-1422.

Ruggles, D., Freyman, R., Oxenham, A., 2014. Influence of musical training on understanding voiced and whispered speech in noise. PLoS ONE 9, e86980.

Schoof, T., Rosen, S., 2014. The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. Frontiers in Aging Neuroscience 6, 307.

Schroeder, S., Marian, V., Shook, A., Bartolotti, J., 2016. Bilingualism and musicianship enhance cognitive control. Neural Plasticity 2016, 4058620.

Sharma, M., Dhamani, I., Leung, J., Carlile, S., 2014. Attention, memory, and auditory processing in 10- to 15-year-old children with listening difficulties. JSLHR 57, 2308-2321.

Shinn-Cunningham, B., 2008. Object-based auditory and visual attention. Trends in Cognitive Sciences 12, 182-186.

Shinn-Cunningham, B., Best, V., 2008. Selective attention in normal and impaired hearing. Trends in Amplification 12, 283-299.

Slater, J., Kraus, N., 2016. The role of rhythm in perceiving speech in noise: a comparison of percussionists, vocalists and non-musicians. Cognitive Processes 17, 79-87.

Slater, J., Azem, A., Nicol, T., Swedenborg, B., Kraus, N., 2017. Variations on the theme of musical expertise: cognitive and sensory processing in percussionists, vocalists and non-musicians. European Journal of Neuroscience 45, 952-963.

Slevc, L., Davey, N., Buschkeuhl, M., Jaeggi, S., 2016. Tuning the mind: exploring the connections between musical ability and executive functions. Cognition 152, 199-211.

Stone, M., Fullgrabe, C., Moore, B., 2012. Notionally steady background noise acts primarily as a modulation masker of speech. JASA 132, 317-326.

Swaminathan, J., Mason, C., Streeter, T., Best, V., Kidd, G., Patel, A., 2015. Musical training, individual differences and the cocktail party problem. Scientific Reports 5, 11628.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., Schroger, E., 2005. Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. Exp Brain Res 161, 1-10.

Tervaniemi, M., Kruck, S., De Baene, W., Schroger, E., Alter, K., Friederici, A., 2009. Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. European Journal of Neuroscience 30, 1636-1642.

Tierney, A., Kraus, N., 2013. The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. Brain and Language 124, 225-231.

Toplak, M., Tannock, R., 2005. Tapping and anticipation performance in attention deficit hyperactivity disorder. Perceptual and Motor Skills 100, 659-675.

Travis, F., Harung, H., Lagrosen, Y., 2011. Moral development, executive functioning, peak experiences and brain patterns in professional and amateur classical musicians: interpreted in light of a Unified Theory of Performance. Consciousness and Cognition 20, 1256-1264.

Vasuki, P., Sharma, M., Demuth, K., Arciuli, J., 2016. Musicians' edge: a comparison of auditory processing, cognitive abilities and statistical learning. Hearing Research 342, 112-123.

Wong, P., Skoe, E., Russo, N., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nature Neuroscience 10, 420-422.

Yeend, I., Beach, E., Sharma, M., Dillon, H., 2017. The effects of noise exposure and musical training on suprathreshold auditory processing and speech perception in noise. Hearing Research 353, 224-236.

Zendel, B., Alain, C., 2012. Musicians experience less age-related decline in central auditory processing. Psychology and Aging 27, 410-417.

Zendel, B., Alain, C., 2013. The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation. Journal of Cognitive Neuroscience 25, 503-516.

Zendel, B., Alain, C., 2014. Enhanced attention-dependent activity in the auditory cortex of older musicians. Neurobiology of Aging 35, 55-63.

Zendel, B., West, G., Belleville, S., Peretz, I., 2019. Music training improves the ability to understand speech-in-noise in older adults. Neurobiology of Aging. doi: 10.1016/j.neurobiolaging.2019.05.015

Zuk, J., Benjamin, C., Kenyon, A., Gaab, N., 2014. Behavioral and neural correlates of executive functioning in musicians and non-musicians. PLoS ONE 9, e99868.

	Mean	*Min*	*Max*	*Skewness*	*Reliability*
CRM Same Gender	45.6%	0%	100%	-0.00363	0.93
CRM Different Gender	81.7%	2.5%	100%	-1.91	0.78
SASA Interleaved	2.12	0.61	4.78	0.446	0.87
SASA Simultaneous	1.83	0.16	4.3	0.733	0.91
Beat Perception	1.96	-0.42	3.46	-0.672	0.53
Auditory Stroop	128.9	-14.7	398.5	1.29	0.4

Table 1. Summary statistics for all behavioural measures. Reliability was calculated using split-half reliability (odd versus even trials).

	CRM Same Gender	*CRM Different Gender*	*CRM Total*	*SASA Simul*	*SASA Inter*	*SASA Total*	*Beat Perception*	*Stroop Effect*
CRM Different Gender	**0.69** **<0.001**							
CRM Total	**0.95** **<0.001**	**0.85** **<0.001**						
SASA Simul	0.17 0.21	0.25 0.053	0.22 0.096					
SASA Inter	**0.31** **0.017**	**0.36** **0.0054**	**0.36** **0.0061**	**0.85** **<0.001**				
SASA Total	**0.26** **0.042**	**0.32** **0.013**	**0.31** **0.017**	**0.95** **<0.001**	**0.96** **<0.001**			
Beat Perception	**0.30** **0.020**	**0.39** **0.0028**	**0.34** **0.0095**	**0.48** **<0.001**	**0.48** **<0.001**	**0.52** **<0.001**		
Stroop Effect	0.15 0.26	0.07 0.59	0.11 0.42	0.07 0.60	0.11 0.42	0.09 0.48	0.06 0.61	
Musical Training	**0.27** **0.034**	**0.37** **0.0046**	**0.31** **0.017**	**0.44** **<0.001**	**0.45** **<0.001**	**0.47** **<0.001**	**0.50** **<0.001**	-0.1 0.46

Table 2. Spearman's rho and p-values for partial correlations between all study variables, controlling for age. P-values have been FDR corrected using the Benjamini and Hochberg algorithm. Significant correlations are indicated via boldface.

	Attention Factor	*Stroop Factor*
SASA Interleaved	**0.94**	0
SASA Simultaneous	**0.91**	-0.07
Beat Perception	**0.68**	-0.19
Stroop Effect	-0.10	**0.99**

Table 3. Factor loadings resulting from exploratory factor analysis on behavioral predictors. Boldface indicates loadings of greater than 0.5.

	Coef. Estimate	*Standard Error*	*t value*	*p value*
(Intercept)	2.38	0.78	3.05	0.003
Age	-0.06	0.02	-2.97	0.004
Musicianship	-1.85	0.92	-2.02	0.048
Age:Musicianship	0.07	0.03	2.75	0.008

Table 4. Parameter estimates from a logistic regression predicting CRM performance summed across both Different Gender and Same Gender conditions from Age and Musicianship. Note that statistics quoted in the text were drawn from direct comparisons of models with and without the predictor whose significance is being addressed.

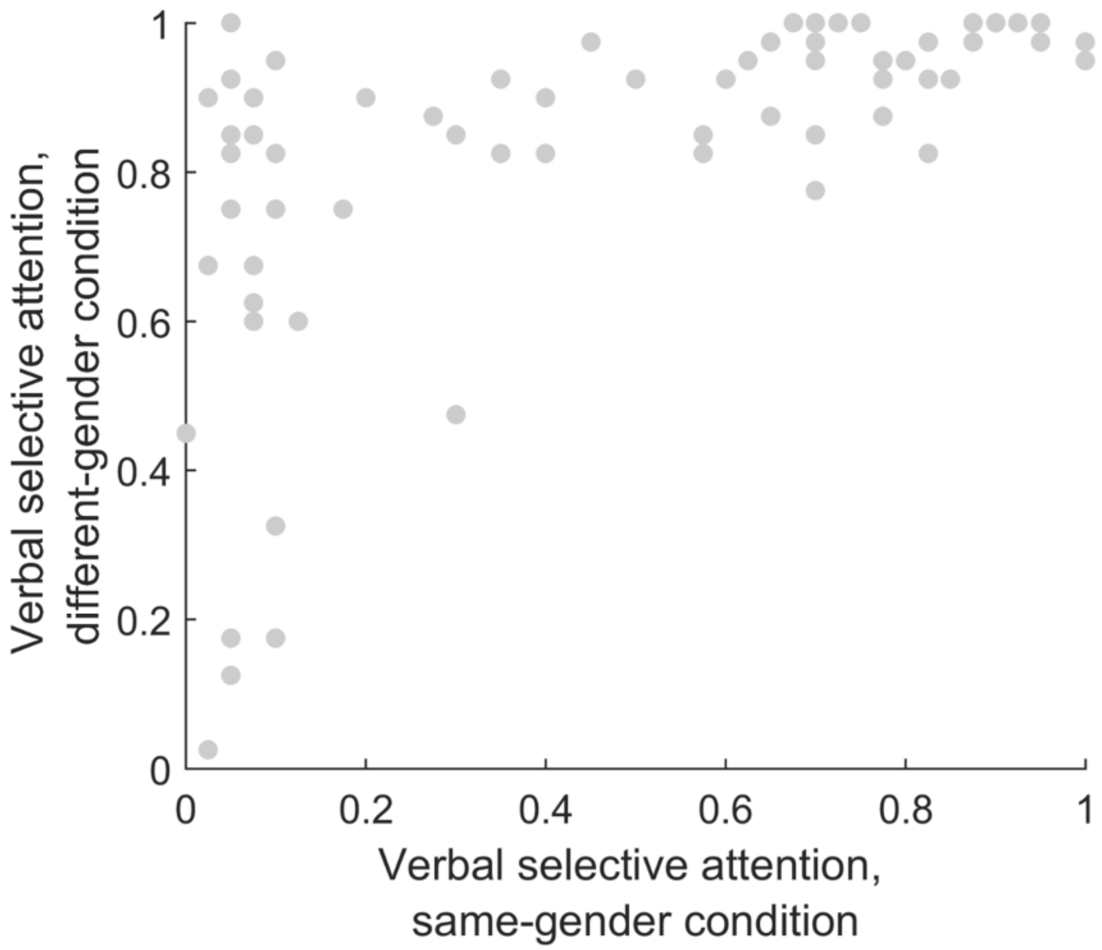

Figure 1. Scatterplot displaying performance on verbal selective attention (CRM) in same-gender and different-gender conditions.

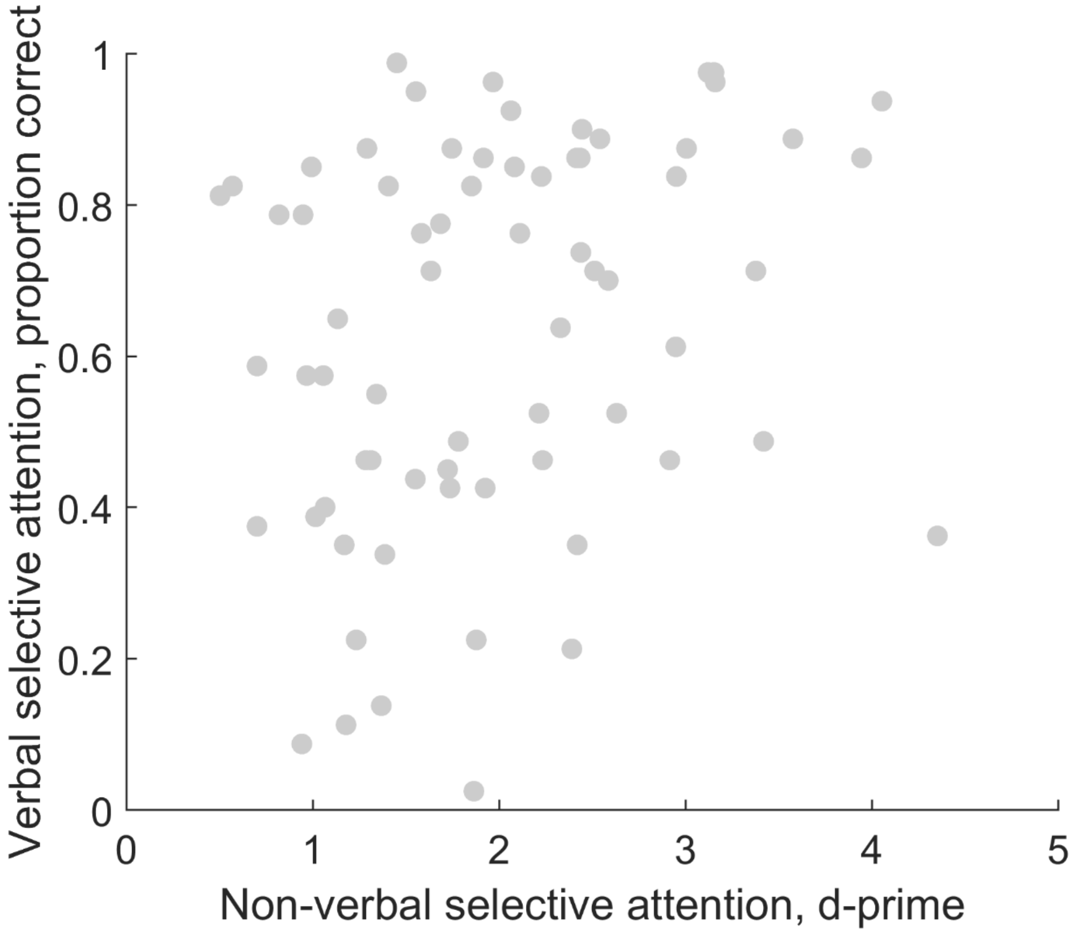

Figure 2. Scatterplot displaying performance on tests of verbal selective attention (CRM, both conditions) versus non-verbal selective attention (SASA, both conditions).

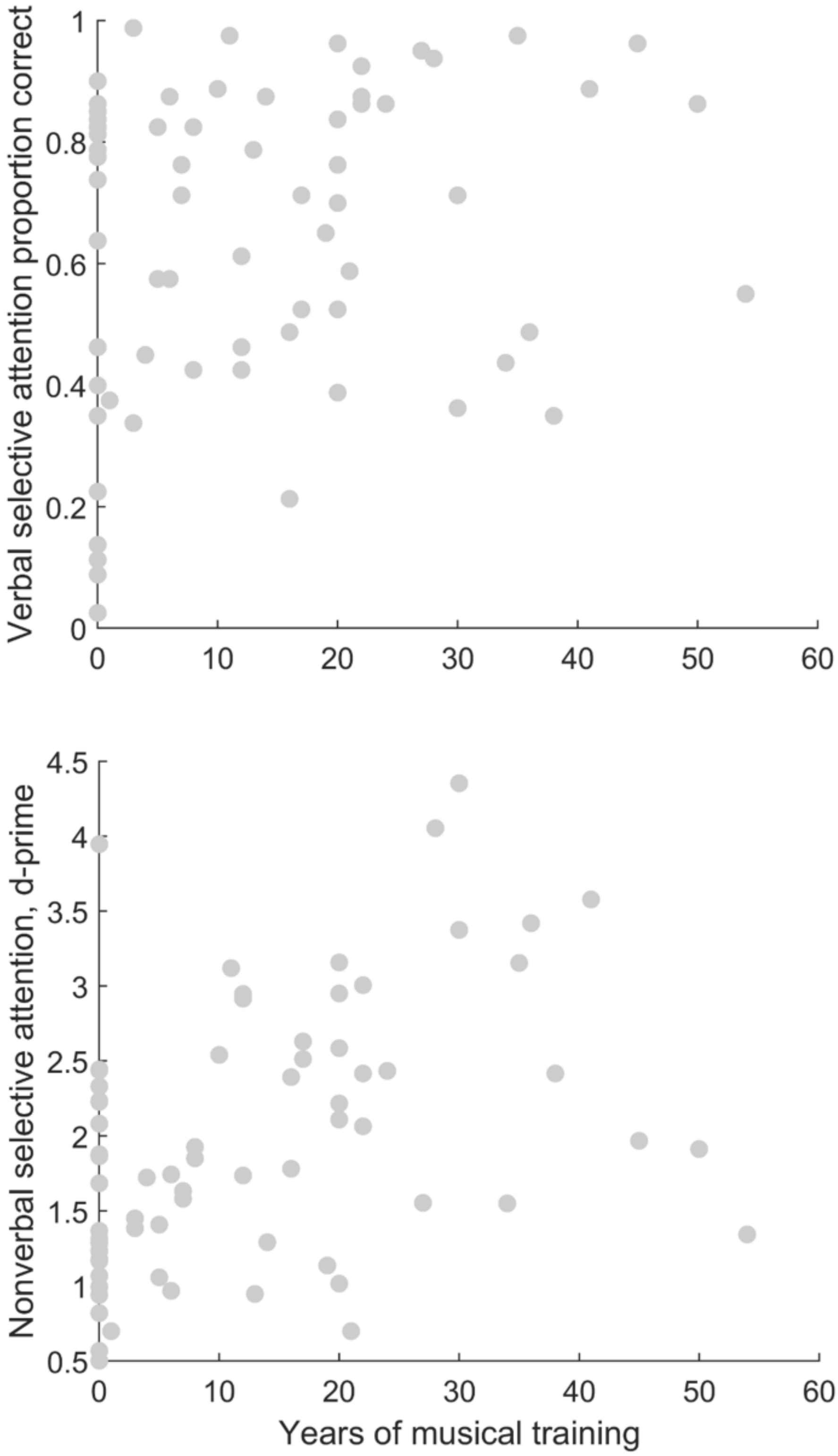

Figure 3. Scatterplots displaying years of musical training versus verbal selective attention performance (CRM, left) and non-verbal selective attention performance (SASA, right).

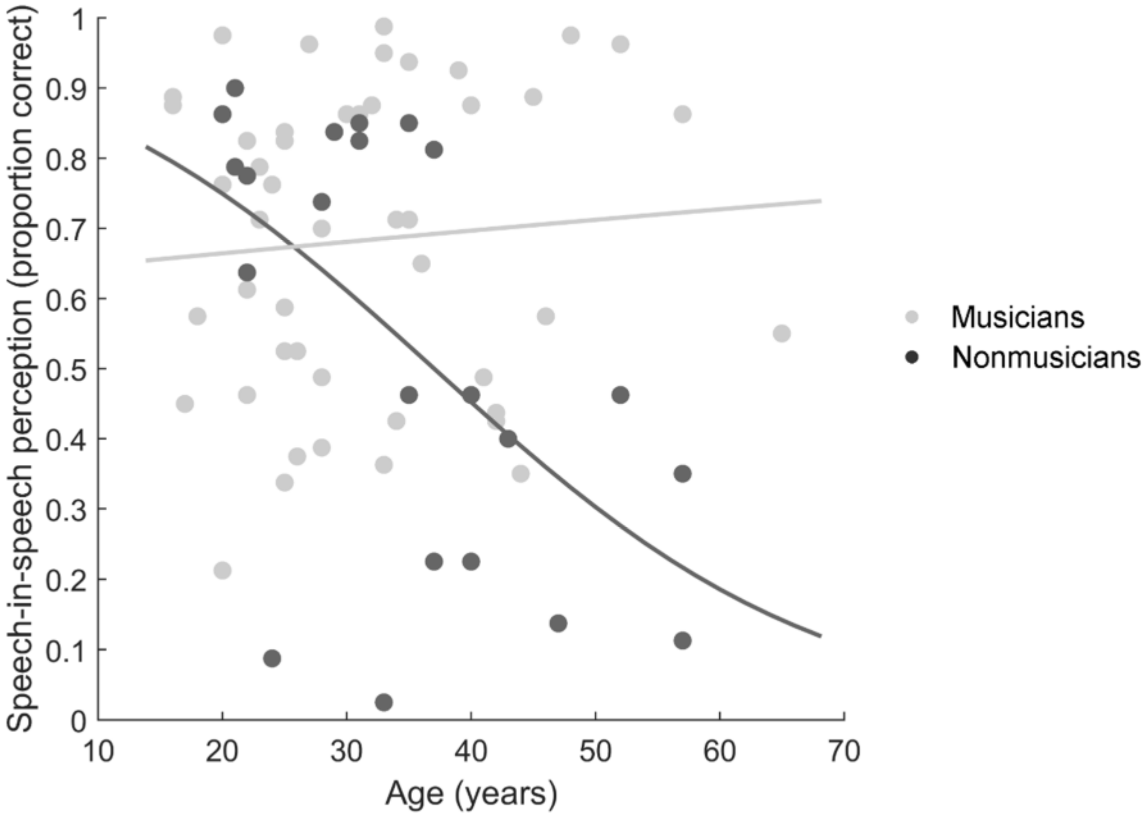

Figure 4. Scatterplot displaying age versus speech-in-speech perception in musicians (light grey) and nonmusicians (dark grey). Fit lines have been generated from the logistic regression model predicting CRM performance from the continuous predictor of age and the categorical predictor of being a musician or non-musician, including the significant interaction term which indicates that the changes of performance with age are different in the two groups.