Musicians and non-musicians are equally adept at perceiving masked speech

Dana Boebinger, Samuel Evans, Stuart Rosen, César F. Lima, Tom Manly, and Sophie K. Scott

Citation: The Journal of the Acoustical Society of America **137**, 378 (2015); doi: 10.1121/1.4904537
View online: https://doi.org/10.1121/1.4904537
View Table of Contents: http://asa.scitation.org/toc/jas/137/1
Published by the Acoustical Society of America

Musicians and non-musicians are equally adept at perceiving masked speech

Dana Boebinger and Samuel Evans
Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, United Kingdom

Stuart Rosen
Speech, Hearing & Phonetic Sciences, University College London, 2 Wakefield Street, London WC1N 2PF, United Kingdom

César F. Lima
Centre for Psychology at University of Porto, Rua Alfredo Allen, 4200-135 Porto, Portugal

Tom Manly
Medical Research Council Cognition and Brain Sciences Unit, Cambridge, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

Sophie K. Scott[a)]
Division of Psychology and Language Sciences, University College London, Gower Street, London WC1E 6BT, United Kingdom

(Received 9 July 2014; revised 24 October 2014; accepted 24 November 2014)

There is much interest in the idea that musicians perform better than non-musicians in understanding speech in background noise. Research in this area has often used energetic maskers, which have their effects primarily at the auditory periphery. However, masking interference can also occur at more central auditory levels, known as informational masking. This experiment extends existing research by using multiple maskers that vary in their informational content and similarity to speech, in order to examine differences in perception of masked speech between trained musicians ($n = 25$) and non-musicians ($n = 25$). Although musicians outperformed non-musicians on a measure of frequency discrimination, they showed no advantage in perceiving masked speech. Further analysis revealed that non-verbal IQ, rather than musicianship, significantly predicted speech reception thresholds in noise. The results strongly suggest that the contribution of general cognitive abilities needs to be taken into account in any investigations of individual variability for perceiving speech in noise. © *2015 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4904537]

[DB] Pages: 378–387

I. INTRODUCTION

Musical expertise offers potential insights into experience-dependent plasticity of auditory function (Herholz and Zatorre, 2012). While it can be difficult to disambiguate factors that may predispose individuals to take up music from those resulting from training, musicians have been reported to demonstrate enhanced concurrent sound segregation (Soderquist, 1970; Zendel and Alain, 2008), pitch discrimination (Micheyl *et al.*, 2006; Spiegel and Watson, 1984; Tervaniemi *et al.*, 2005), duration discrimination (Jeon and Fricke, 1997; Rammsayer and Altenmüller, 2006), and encoding of melodic contour and interval structure (Fujioka *et al.*, 2004). Of relevance to the current study are musicians' relative abilities in understanding speech in background noise.

Extracting a speaker's voice from background masking noise is challenging, even for young adults with normal hearing (Assmann and Summerfield, 2004). Certain populations seem to be especially vulnerable to degraded speech, such as older adults (Helfer and Freyman, 2008; Pichora-Fuller *et al.*, 1995) and individuals with learning disorders like dyslexia and specific language impairment (Ziegler *et al.*, 2005; Ziegler *et al.*, 2009; but see also Hazan *et al.*, 2013; Messaoud-Galusi *et al.*, 2011). In contrast, better performance has been reported in highly trained musicians on two measures of speech perception in noise, the Hearing-in-Noise Test (HINT) and the QuickSIN (Parbery-Clark *et al.*, 2009). However, not all studies have found a robust musicianship advantage (Fuller *et al.*, 2014; Ruggles *et al.*, 2014). Additionally, some studies have only found such an effect in elderly adults, with no difference between musicians and non-musicians under the age of 40 on measures of speech perception in noise (Zendel and Alain, 2012; Fig. 4). Understanding why some populations, like musicians, might show enhanced speech perception in noise could lead to therapeutic interventions to help populations that show deficits in this area.

Psychoacoustical studies have identified at least two ways in which background sounds can interfere with perception of sound: Energetic and informational masking (Shinn-Cunningham, 2008; Brungart *et al.*, 2001). Broadly, energetic masking effects (inclusive of modulation masking; Stone *et al.*, 2011; Stone *et al.*, 2012) are a consequence of

[a)]Author to whom correspondence should be addressed. Electronic mail: sophie.scott@ucl.ac.uk

 0001-4966/2015/137(1)/378/10/$30.00

the masker energy obscuring the properties of the target signal in the same time/frequency region. This is presumed to arise in the auditory periphery and is most often associated with steady-state noise masking. Informational masking, by contrast, refers to masking effects that cannot be explained solely by energetic masking. It is often associated with higher, more central, levels of auditory and linguistic processing. If musicians show an enhanced ability to perceive masked speech, it is important to understand the extent to which this benefit arises from energetic and/or informational masking mechanisms.

Based on findings that musicians have better pitch perception than non-musicians (Micheyl *et al.,* 2006; Spiegel and Watson, 1984; Tervaniemi *et al.,* 2005), it could be argued that better ability in noise in this group may be due to improved auditory acuity. This implies that musicians would show improved speech perception in noise for all masker conditions—both energetic and informational. Alternatively, a musician advantage for speech perception in noise could be explained by experience-dependent improvements in segregation and selection of target sounds within sound mixtures, meaning that musicians would show a selective benefit for informational masker conditions. The potential interaction between type of masking and musicianship has been examined previously, although results have been mixed. One study found that musicians are less affected by informational masking when the stimuli consisted of sinusoidal tones (Oxenham *et al.,* 2003), although it was unclear whether this benefit would transfer to speech perception. Two recent studies used stimuli that varied in their similarity to speech, and did not find evidence of a systematic musician advantage for either periodic or aperiodic speech (Ruggles *et al.,* 2014) or for sinewave-vocoded speech (Fuller *et al.,* 2014).

It is also possible that musicians' improved speech perception in noise could be due to general cognitive abilities. There is some controversy concerning the relationship between musical training and improved cognitive abilities (Schellenberg and Peretz, 2008). However, while there is disagreement over the existence of a causal relationship between the two, cognitive factors often co-vary with musicianship. Additionally, in a review of 20 studies on speech perception in noise, all but one found an effect of cognitive factors on performance (Akeroyd, 2008). Thus, we sought to examine whether any observed group differences were likely to be related to musicianship or whether general cognitive abilities contributed to the effect.

The aim of the current study was to elaborate on whether, and in what manner, experienced musicians differ from non-musicians in their perception of masked speech. To this end, we recruited 25 musicians and 25 non-musicians using criteria consistent with previous studies (Parbery-Clark *et al.,* 2009). The current study sought to extend previous work by using intelligible sentences embedded in multiple maskers that vary in their energetic and informational masking effects, and to investigate whether any observed group differences were likely to be related to musicianship or linked with other factors that may co-vary with musicianship, such as general cognitive abilities.

II. METHOD

A. Participants

Fifty native British English speaking adults (mean age = 27.2 yr, SD = 6.9 yr; 21 females) took part in the study. All participants reported normal hearing and no history of neurological or psychological disorders or language learning impairments. The study was approved by the University College London Research Psychology and Language Science ethics committee, and written informed consent was obtained from all participants.

Participants were categorized as musicians ($n = 25$) if they had started training before the age of seven, had at least 10 yr of musical experience, and reported practicing consistently (at least three times per week) over the previous 3 yr (Parbery-Clark *et al.,* 2009). Non-musicians ($n = 25$) failed to meet these musicianship criteria and reported less than 3 yr of total musical experience, none of which occurred in the year prior to the experiment. The two participant groups did not differ significantly in age [$t(48) = -1.453$, $p = 0.153$], gender [$\chi^2(1, N = 50) = 0.739$, $p = 0.390$], or years of post-secondary education [$t(45.265) = -1.012\, p = 0.317$].

B. Materials

1. Musicianship questionnaire

Before participating in the experiment, all participants completed a questionnaire detailing their language, education, extracurricular, and musical experiences. Details of all participants' musical backgrounds are displayed in Table I.

2. Masking task

Target speech stimuli consisted of sentences from the Bamford–Kowal–Bench (BKB) lists (Bench *et al.,* 1979). These are short sentences that have simple syntax and semantic content that were developed using the speech of children who have hearing impairments, i.e., "SHE CUT with her KNIFE." Each sentence included three key words (capitalized above). The sentences were produced by a female Southern British English speaker and recorded in an anechoic chamber at 44 100 Hz with 16-bit quantization and down-sampled to 22 050 Hz. All stimuli, including the target and masking sounds, were low-pass filtered at 3.8 kHz to equate the spectral range across conditions.

The target sentences were embedded in four different kinds of maskers that varied parametrically in their informational content: clear speech, spectrally rotated speech, speech-amplitude modulated noise, and speech-spectrum steady-state noise. Waveforms and spectrograms of all four maskers can be seen in Fig. 1. The clear speech was taken from a male speaker from the EUROM database of British English speech (Chan *et al.,* 1995). All the other conditions were derived from the clear speech condition. Using a female target speaker and a male masking speaker meant that informational masking effects were reduced, compared to the levels of informational masking seen when two same sex speakers are used (Brungart, 2001). This choice was made to enable the instruction "listen to the female speaker,"

TABLE I. Details of all participants' musical experience.

	Participant ID	Years of training	Age of onset (years)	Instrument
Musicians	1	19	4	Violin, viola
	2	24	7	Saxophone, guitar
	3	14	7	Violin, piano
	4	22	7	Piano, guitar
	5	33	6	Violin, guitar
	6	25	6	Piano, drums
	7	24	7	Guitar
	8	26	4	Piano
	9	17	6	Piano, trombone
	10	14	6	Clarinet, saxophone
	11	36	6	Cello, guitar
	12	35	7	Piano, sax, flute, clarinet
	13	22	6	Piano
	14	25	4	Piano, guitar
	15	19	6	Piano, drums
	16	36	6	Piano, flute, drums, guitar
	17	33	5	Accordion
	18	31	5	Piano, violin, guitar, horn
	19	14	6	Flute
	20	15	6	Flute
	21	24	5	Piano
	22	13	5	Piano
	23	14	7	Piano
	24	21	7	Clarinet, sax, flute, piano, bassoon
	25	12	6	Piano, flute
	Mean (SD)	22.7 (7.8)	5.9 (1.0)	
Non-musicians	1	0	-	
	2	0	-	
	3	2	12	Piano
	4	0	-	
	5	1	10	Piano
	6	0	-	
	7	0	-	
	8	0	-	
	9	0	-	
	10	0	-	
	11	0	-	
	12	0	-	
	13	1	18	Guitar
	14	0	-	
	15	0	-	
	16	0	-	
	16	0	-	
	17	0	-	
	18	0	-	
	19	2	11	Violin, piano
	20	0	-	
	21	0	-	
	22	0	-	
	23	0	-	
	24	0	-	
	25	0	-	
	Mean (SD)	0.2 (0.6)	12.8 (3.6)	

so training participants to identify the target speaker would not be necessary (Scott *et al.*, 2004), and has been shown to result in widespread cortical responses to the masking speech, relative to energetic masking conditions (Scott *et al.*, 2004; Scott *et al.*, 2009).

Rotated speech was generated by inverting the frequency spectrum around 2 kHz, using a digital version of the simple modulation technique described by Blesser (1969, 1972). The speech signal was first equalized with a filter (essentially high-pass) that gave the rotated signal

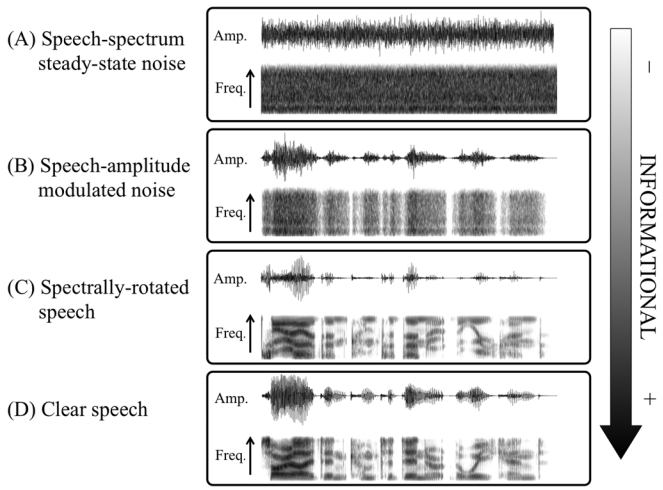

FIG. 1. Waveforms and spectrograms of the four maskers, arranged from least informational masking to most. For the spectrograms, time is represented on the x axis (0−1.7 s) and frequency on the y axis (0−4 kHz). (A) Speech-spectrum steady-state noise has no temporal modulations or spectro-temporal dynamics. (B) Speech-amplitude modulated noise has temporal modulations but no spectro-temporal dynamics. (C) Spectrally rotated speech contains similar temporal modulations and spectro-temporal dynamics as clear speech, but is unintelligible. (D) Clear speech contains both temporal modulations and spectro-temporal dynamics, and is intelligible.

approximately the same long-term spectrum as the original. The equalized signal was then amplitude modulated by a sinusoid at 4 kHz, followed by low-pass filtering at 3.8 kHz. Rotated speech preserves much of the spectro-temporal complexity of speech, e.g., it contains amplitude modulations, and a formant and quasi-harmonic structure, but is unintelligible. Speech-amplitude modulated noise was created by modulating a speech-spectrum noise with envelopes extracted from the original speech by full-wave rectification and low-pass filtering at 30 Hz. Speech-amplitude modulated noise has the amplitude modulations of speech but does not contain spectro-temporal dynamics such as formant or harmonic structure. Speech-spectrum steady-state noise was generated by synthesizing a white noise and filtering it to have the same long-term average spectrum as the clear speech. Although having the same long-term spectrum as speech, this steady-state noise does not have amplitude modulation or any structured spectro-temporal dynamics.

The target sentences were assigned to the four masking conditions so that each sentence was presented only once during the course of the experiment, and the order of conditions was randomized between participants using a Latin square. During the task, participants were required to listen to the target sentences embedded in the masker and repeat them back to the experimenter. Responses were scored for the number of correct key words (0 to 3).

A one-up one-down adaptive staircase procedure was used to determine participants' speech reception thresholds (SRTs) (Levitt, 1971). The stimuli were presented over Sennheiser HD25 headphones. The maskers were presented at 71 dB sound pressure level (SPL) as measured by a Bruel & Kjaer 4153 artificial ear. The level of the target was varied initially in a 10-dB step size, which reduced to 6.5-dB after the first reversal, and to 3-dB steps after the second reversal. For each trial, the target and masker were mixed at the appropriate signal-to-noise ratio (SNR), which was constructed by changing the intensity of the target while maintaining a constant level of the masker. The termination criterion was 18 reversals, and if this number of reversals was not achieved, the track ended after 28 trials. A 50% SRT was calculated for each condition by omitting the first two reversals and averaging all remaining reversals. To prevent biasing the thresholds, in the case that there were an odd number of reversals remaining after omitting the first two, the third reversal was also omitted so that the number of averaged reversals was always even. Each participant completed two adaptive procedure tasks for each masking condition, and the two SRTs for each condition were averaged to increase the reliability of the threshold estimate. Levene's test of equality of variance showed there to be no evidence of a difference in the measure of variability in the number of reversals between groups, $F(1, 48) = 0.046$, $p = 0.831$. There was also no group difference in the number of reversals between musicians ($M = 13.2$, SD = 0.8) and non-musicians ($M = 12.9$, SD = 0.8), $F(1, 48) = 1.148$, $p = 0.289$, partial $\eta^2 = 0.023$, or group by masking condition interaction, $F(3, 144) = 0.856$, $p = 0.466$, partial $\eta^2 = 0.18$.

3. Psychoacoustic measures

Participants' frequency discrimination and duration discrimination thresholds were determined using an adaptive staircase procedure (Grassi and Soranzo, 2009). For each trial in the frequency discrimination task, participants were presented with three 250-ms pure tones and asked to indicate which of the three tones was highest. Two of the tones were presented at the same frequency (1-kHz) and one at a higher frequency by a specified Δf, ranging from 2 to 256 Hz. In each trial, the three tones were presented in a random order, and the participant was asked to indicate which of the three tones was highest. Similarly, in the duration discrimination task, participants were presented with three 1-kHz pure tones and had to determine which of the three tones was the longest. Two of the tones were the same length (250 ms), and one of the tones was longer by Δt, which was varied adaptively from 8 to 256 ms.

All tone onsets and offsets were gated on and off with two 10-ms raised cosine ramps. Both adaptive procedures were transformed "two-down one-up," staircases tracking the 70.7% point on the psychometric function. The Δf or Δt was adaptively changed throughout the experiment, such that the Δf or Δt was changed by a factor of 2 for the first four reversals, and by a factor of $\sqrt{2}$ for the next eight reversals, with the track ending after 12 reversals. The threshold was calculated by averaging the values of the last eight reversals.

4. Auditory working memory

The forward and backward Digit Span subtest of the Wechsler Adult Intelligence Scale (WAIS) (Wechsler, 1997) was used to measure participants' auditory short-term and

working memory, respectively. Participants' scaled scores were used in all further analyses.

5. Non-verbal IQ

The Matrix Reasoning subtest of the Wechsler Abbreviated Scale of Intelligence (WASI) was used (Wechsler, 1999) to estimate participants' non-verbal IQ. Participants' scaled scores were used in all further analyses.

6. Selective attention

The Stroop task is a commonly used assessment of selective attention and inhibition control (Stroop, 1935). In a control task, participants were shown a list of groups of X's (i.e., XXXX XXXX XXXX XXXX) printed in various colors and had to name the ink color. In the experimental task, participants were shown a series of color words printed in a mismatching ink color (i.e., "blue" printed in pink ink). They were instructed to go down the list and name the ink color of the words as quickly and as accurately as possible, while ignoring the word meaning. The amount of interference was calculated as the ratio between the experimental task and the control task performance times (Jensen, 1965; Macleod, 1991).

7. Mental flexibility

The trail making test has been extensively used in neuropsychological testing to evaluate participants' motor speed (Gaudino *et al.*, 1995), fluid intelligence (Salthouse, 2011), and task-switching (Wecker *et al.*, 2005). In Part A, the participant draws a line to connect numbers in sequential order as quickly as possible. In Part B, the participant draws a line to connect circled numbers and letters in sequential and alphabetical order, alternating between numbers and letters as quickly as possible (i.e., 1-A-2-B-3-C, etc.). The ratio between times for Part B and Part A was used as the dependent measure, because this greatly reduces the effects of motor speed (Salthouse, 2011).

III. RESULTS

A. Cognitive tasks

Two-tailed independent sample *t*-tests showed that the musician and non-musician groups did not differ on any of the cognitive measures, including digit span [$t(40.903) = -0.928$, $p = 0.359$], Matrix Reasoning [$t(48) = -1.342$, $p = 0.186$], the Stroop task [$t(48) = 0.432$, $p = 0.668$], and the trail making task [$t(48) = -0.089$, $p = 0.929$]. Group means for each task are presented in Table II.

B. Frequency and duration acuity

The Shapiro–Wilk test of normality indicated that the distributions of the pitch and duration discrimination tasks were positively skewed, p's < 0.001. The log transform often used in psychoacoustic research was not sufficient to normalize the data, so according to Tukey's ladder of powers (Tukey, 1977), the nonlinear monotonic transformation $-1/(\sqrt{x})$ was used. After transformation, the Shapiro–Wilk test was no longer significant for either the pitch or duration measure, p's > 0.3.

Independent samples *t*-tests showed that musicians ($M = 10.9$ Hz, SD = 19.2 Hz) performed significantly better than non-musicians ($M = 41.1$ Hz, SD = 73.0 Hz) on the pitch discrimination task, $t(48) = 3.757$, $p < 0.001$, but there was no difference between the groups on the duration discrimination task, $t(48) = 1.394$, $p = 0.170$ (musicians: $M = 28.4$ ms, SD = 11.3 ms, non-musicians: $M = 35.0$ ms, SD = 17.8 ms; Fig. 2). Note that for both measures, lower scores indicate lower thresholds and thus better performance.

C. Masking task

A mixed effects analysis of variance (ANOVA), with a within-subjects factor of masker condition (clear speech, spectrally rotated speech, speech-amplitude modulated noise, and speech-spectrum steady-state noise) and a between-subjects factor of musicianship (musician, non-musician) revealed a significant main effect of masker condition, $F(3, 144) = 320.194$, $p < 0.001$, partial $\eta^2 = 0.870$, with all pairwise comparisons reaching statistical significance (all p's < 0.001, Bonferroni corrected). Notably, there was no significant main effect of musicianship, $F(1, 48) = 1.413$, $p = 0.240$, partial $\eta^2 = 0.029$, or a masker by musicianship interaction, $F(3, 144) = 0.187$, $p = 0.905$, partial $\eta^2 = 0.004$. Group means are presented in Table III, and results can be seen in Fig. 3.

In order to better understand the observed lack of evidence of an effect of musicianship on the masked speech task, we generated confidence intervals for the effect size of musicianship (Hentschke and Stüttgen, 2011). The confidence interval was shown to include zero (Hedges's $g = 0.330$, 95% CI = [−0.221, 0.878]). Further, when we generated confidence intervals for the mean difference between the groups in dB, we could have a 95% confidence that the mean difference between the groups was between a 0.34 dB advantage for non-musicians and a 1.3 dB advantage for musicians (95% CI = [−0.34 dB, 1.3 dB]). If the observed effect size accurately describes the true effect size in the population, a sample size of over 230 participants

TABLE II. Descriptive statistics and *t*-tests for cognitive tasks, separated by group.

	Musicians ($N = 25$)	Non-musicians ($N = 25$)	Total ($N = 50$)	*t*-Test results
Cognitive task	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)	*t*-value (*p*-value)
Digit span	12.0 (2.2)	11.2 (3.4)	11.6 (2.9)	−.928 (0.359)
Matrix reasoning	58.7 (5.5)	56.4 (6.8)	57.5 (6.3)	−1.342 (0.186)
Stroop task	1.5 (0.2)	1.5 (0.2)	1.5 (0.2)	0.432 (0.668)
Trail making task	2.0 (0.6)	1.9 (0.7)	2.0 (0.6)	−0.089 (0.929)

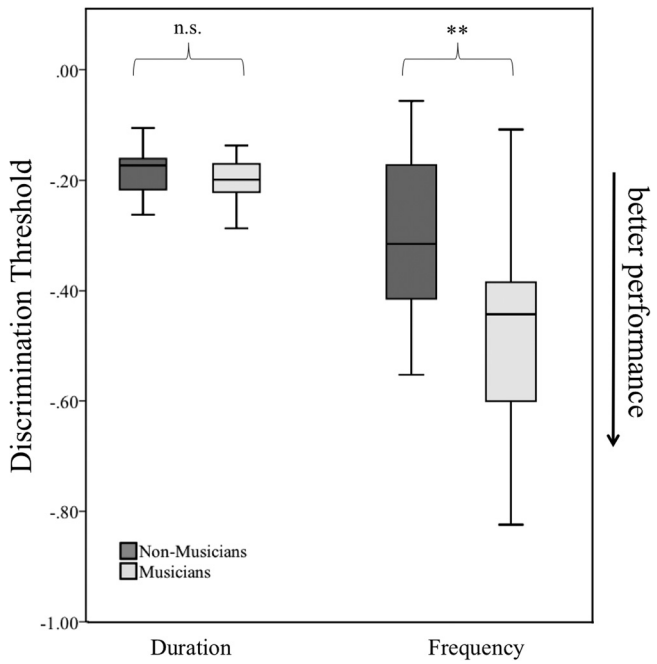

FIG. 2. Boxplots showing the transformed data for the duration (left) and frequency discrimination (right) tasks separated by group, ** = $p < 0.001$. Lower thresholds indicate better performance.

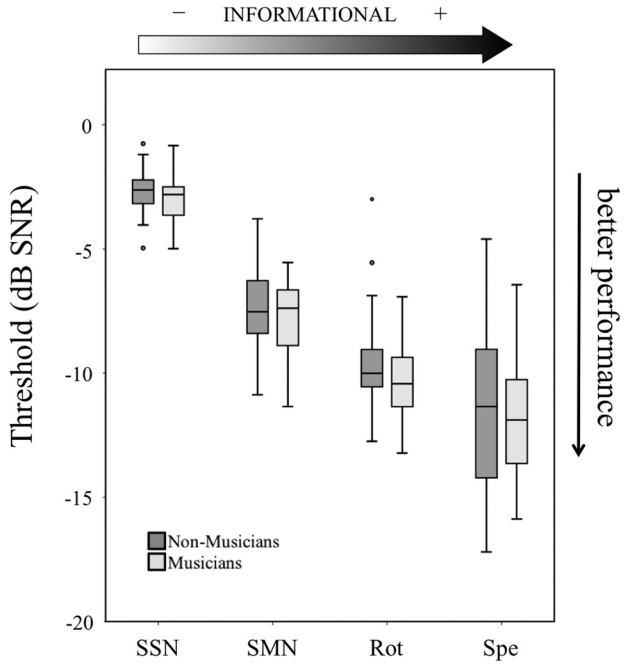

FIG. 3. Boxplots of SRTs for each masking condition. Note that lower thresholds indicate better performance. SSN = speech-spectrum steady-state noise, SMN = speech-amplitude modulated noise, Rot = rotated speech, Spe = clear speech.

[$n = 115$ per group, one-tailed test; (Faul *et al.*, 2007)] would be needed to obtain a significant result with statistical power at the recommended 0.80 level.

Last, the data from musicians was considered separately and the age of onset of musical training was added as a covariate. No significant main effect of age of onset of musical training was found, $F(1, 23) = 1.651$, $p = 0.212$, partial $\eta^2 = 0.067$. There was also no significant interaction between age of onset of musical training and masking condition, $F(3, 69) = 0.704$, $p = 0.553$, partial $\eta^2 = 0.030$. Additionally, within musicians the correlation between number of years of musical training and average SRT was not significant, $R = -0.348$, $p = 0.089$. Also, frequency discrimination thresholds did not correlate with SRTs in any of the four masker conditions, both within musicians (all p's > 0.250) and across all participants (all p's > 0.152).

D. Predictors of masked speech thresholds

As there was no evidence of a difference between musicians' and non-musicians' speech perception in noise abilities, group membership, pitch discrimination, duration discrimination, Matrix Reasoning score, Trail Making Task score, digit span score, and Stroop task score were used in a stepwise multiple regression analysis to predict participants' SRTs. To increase the stability of the threshold estimates we averaged the SRTs across conditions and included this as the dependent measure in the regression. We used a factor analysis to validate this approach, which indicated a single factor (with an eigenvalue greater than one) with high and similar loadings on each condition (loadings range = 0.690−0.852). The resulting factor scores correlated exceptionally highly with the average SRT ($R^2 = 0.98$, $p < 0.001$). Using the average SRT as the dependent measure in the stepwise regression, the resulting model contained only the Matrix Reasoning task as a predictor, $F(1, 48) = 8.209$, $p = 0.006$, $R^2 = 0.146$. No other variable entered into the model (all p's > 0.108), indicating that no other variables accounted for additional variance. The relationship between participants' Matrix Reasoning score and average SRT can be seen in Fig. 4. Note that equivalent results were attained in the stepwise regression using the factor scores rather than the average SRT.

We also used the more fine-grained measurement of participants' years of musical experience as a predictor,

TABLE III. Participants' mean SRTs for all four masking conditions, separated by group.

	Musicians ($N = 25$)	Non-musicians ($N = 25$)	Total ($N = 50$)
Masker condition	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)
Speech	−11.9 (2.4)	−11.5 (3.2)	−11.7 (2.8)
Rotated speech	−10.3 (1.7)	−9.5 (2.2)	−9.9 (2.0)
Speech-amplitude modulated noise	−7.8 (1.5)	−7.4 (1.7)	−7.6 (1.6)
Speech-spectrum steady-state noise	−3.0 (1.0)	−2.7 (0.9)	−2.9 (1.0)
Overall mean (*SE*)	−8.3 (0.2)	−7.8 (0.3)	

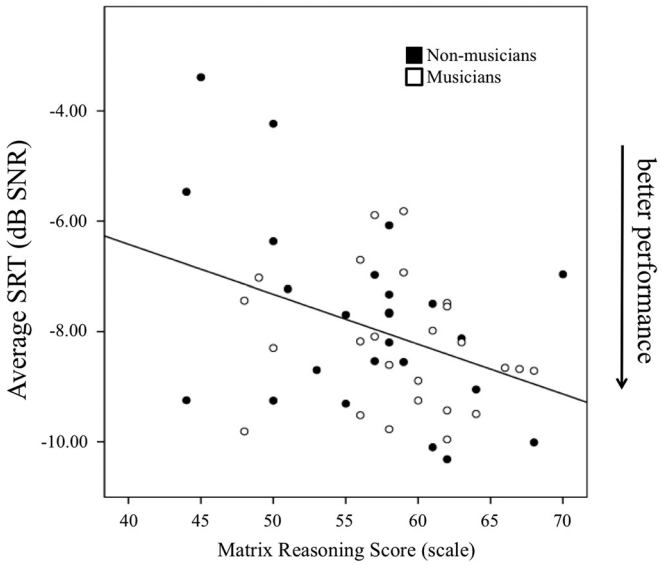

FIG. 4. Participants' average SRTs as predicted by their WASI Matrix Reasoning score. Note that lower thresholds indicate better performance.

rather than their membership in the musician/non-musician groups. However, this did not affect the significance of the findings. Collinearity diagnostic tests produced variance inflation factor values all less than 1.252, indicating that there is little redundancy among the predictor variables and multicollinearity was not a problem. Further analyses found that a 1 dB improvement of participants' SRTs corresponded to an 11.1-point increase on the Matrix Reasoning scale score. Overall, speech perception thresholds were thus primarily predicted by non-verbal IQ, rather than musical experience or psychoacoustic abilities.

To ascertain whether the slopes of the regressions for the relationship between Matrix Reasoning scores and average SRT were different between the groups, we constructed a linear model predicting SRT scores with a categorical group factor (musician/non-musician) and the continuous Matrix Reasoning predictor. There was no evidence of a main effect of musicianship, $F(3, 46) = 1.488$, $p = 0.229$, partial $\eta^2 = 0.031$, or interaction between musicianship and Matrix Reasoning score, $F(3, 46) = 1.321$, $p = 0.256$, partial $\eta^2 = 0.028$, but there was a significant main effect of Matrix Reasoning Score, $F(3, 46) = 5.629$, $p = 0.022$, partial $\eta^2 = 0.109$. This indicates that there was no evidence of a difference in the relationship between average SRT and Matrix Reasoning Scores between the groups. Furthermore, Levene's test of equality of variance showed that there was no evidence of a difference between the groups in the measure of variability in average SRT, $F(1,48) = 1.968$, $p = 0.167$, and Matrix Reasoning scores, $F(1,48) = 1.031$, $p = 0.315$.

IV. DISCUSSION

This study was conducted in an attempt to replicate the finding that musical training confers a beneficial effect on speech perception in noise, and to interrogate whether this difference was largely due to energetic or informational masking effects. Notably, however, we found no advantage for musicians' masked speech perception over that of non-musicians, and this did not vary by masker type.

As would be expected based on previous findings (Micheyl *et al.*, 2006; Spiegel and Watson, 1984; Tervaniemi *et al.*, 2005), musicians were found to outperform non-musicians in tests of frequency discrimination. Importantly, though, within musicians and across all participants, frequency discrimination did not correlate with participants' SRTs in any of the four masker conditions. Thus, it seems that low-level auditory abilities like frequency discrimination do not confer an advantage when it comes to higher-level speech perception. This confirms findings from previous studies showing speech perception to be unrelated to performance on low-level psychoacoustic tasks (e.g., Surprenant and Watson, 2001), and suggests that there is no simple shared mechanism between frequency discrimination and speech perception in noise abilities. These results also support previous studies that have found a large musician advantage for tasks that rely heavily on frequency perception, but only a small and inconsistent (Fuller *et al.*, 2014), or non-existent (Ruggles *et al.*, 2014), musician advantage in speech perception tasks.

The variability of empirical demonstrations of musicians' advantage in masking tasks may be due to the types of stimuli used. Oxenham *et al.* (2003) used sinusoidal tones embedded in both continuous noise and two types of multitone maskers with different predictabilities, and found musicians to be less affected by informational masking than non-musicians. However, this task relies on being able to selectively attend to a narrow frequency band, which could be aided by musicians' well-documented improved frequency discrimination and might not extend to perception of more complex speech stimuli. Further, a recent study by Fuller *et al.* (2014) suggests that although musicians outperform non-musicians on tasks that rely on frequency discrimination, these benefits do not carry over to the perception of degraded speech. In the speech perception task used by Fuller *et al.* (2014), musicians and non-musicians listened to single words in speech-spectrum steady state noise (at multiple SNRs), as well as sentences in speech-spectrum steady state noise, speech-spectrum fluctuating noise, or six-talker babble. Additionally, all stimuli were presented both unprocessed and processed by an eight-channel sinewave vocoder, which degrades the spectro-temporal information in normal speech. No musician advantage was observed in any condition for the perception of sentences, and the only condition in which musicians showed improved performance was the most challenging condition: Vocoded single word identification when the background noise was presented at +5 dB SNR. Indeed, Parbery-Clark *et al.* (2009) only found differences between musicians and non-musicians when masking and target sounds were co-located, rather than in conditions where the sounds were separated in space, suggesting that differences between groups may only arise when masking tasks require greatest effort. However, Ruggles *et al.* (2014) presented musicians and non-musicians with voiced and whispered sentences, in either continuous or fluctuating noise, in order to investigate whether musicians' improved performance in previous studies was due to improved encoding of periodicity in normal speech (Parbery-Clark *et al.*,

2011). They found no significant effect of musicianship with co-located sounds even when repeating the exact clinical assessments used by Parbery-Clark *et al.* (2009).

In our masked speech task, we found no musicianship advantage or any condition-specific differences between the groups, despite using similar target sentences and a co-located speech-spectrum steady-state noise masker with adaptive tracking, as had been used by Parbery-Clark *et al.* (2009). Parbery Clark *et al.* (2009) demonstrated a difference between groups with both a speech-spectrum noise (HINT) and a four-talker babble (Quick SIN). Both of these sounds are mostly energetic in nature, and compared to a single competing talker, are low in informational masking. To extend these findings, we sought to test for differences between musicians and non-musicians using maskers that differed more widely in their energetic and informational masking properties, in order to gain a better mechanistic understanding of the musicianship advantage. By including a steady-state and modulated noises, as well as maskers with and without a clear pitch, we could identify whether putative differences between the groups were associated with exploiting "glimpses" of the target (which may have reflected a temporal processing advantage) and/or pitch processing. However, we did not find a condition by musicianship interaction which would have supported these proposed mechanisms.

The lack of observed group differences in the current study was despite a larger sample size and nearly identical criteria for recruiting musicians as those used previously in studies reporting significant effects of musicianship (Parbery-Clark *et al.,* 2009). To further understand the discrepancy between our results and those of past studies, we calculated a 95% confidence interval around our observed effect size for the main effect of musical experience on SRTs. The size of the confidence interval means that results could differ considerably across samples from "no effect" to a "large effect," leading to occasions where group differences in speech perception are not observed (Fuller *et al.,* 2014; Ruggles *et al.,* 2014) as well as occasions when they are (Parbery-Clark *et al.,* 2009). It should be noted that in previous studies, the absolute difference between musicians' and non-musicians' scores on clinical speech-in-noise measures was small (<1 dB), even when these differences were statistically significant (Parbery-Clark *et al.,* 2009). This small advantage could be an additional reason for the lack of replication in our experiment. Furthermore, if the difference between musicians' and non-musicians' SRTs is less than 1 dB, it is unlikely to have much significance in a non-experimental setting.

We did, however, find a significant relationship between non-verbal IQ (as measured by the WASI Matrix Reasoning task) and overall masked speech thresholds, such that participants with higher non-verbal IQ scores had lower average SRTs (indicative of better performance). In general, performance on any range of tests sensitive to individual differences in the general population will positively correlate with each other, and the first principle component accounting for that shared variance will be general cognitive abilities (Spearman, 1927). Thus, to claim a specific relationship (e.g., between musicianship and speech perception in noise) it must be demonstrated that this does not simply stem from common loading on cognitive abilities. One way to overcome this is to match musician and non-musician participants on tests known to be highly correlated with general cognitive abilities and which have minimal overlap with the demands of understanding speech in noise. The task used in this experiment, the Matrix Reasoning subtest of the WASI (Wechsler, 1999), is a good example because it has little ostensive requirement for auditory processing or verbal working memory. In addition, we also gathered data about the number of years participants had spent in post-secondary education (linked both to IQ and opportunity), and measured a range of other cognitive functions including auditory working memory, selective attention, and mental flexibility. Thus, it is not entirely surprising that we find a significant effect of general cognitive abilities on our measure of masked speech.

Previous studies on musicians' speech perception in noise have not consistently examined the effect of cognitive factors in depth, often only examining cognitive variables by testing the difference between group means using a *t*-test (Parbery-Clark *et al.,* 2011; Strait *et al.,* 2012). The assumption behind this is that because the groups do not differ on cognitive measures, any observed difference between them on the masked speech task is due to their musical experience. However, without doing a multiple regression and including the cognitive measure(s) as a continuous regressor, the roles of cognitive measures versus musical experience cannot be established. Strikingly, in a study which included working memory as a factor in a hierarchical regression analysis, the addition of years of musical experience as a predictor explained only an additional 6% of the variance in clinical speech-in-noise scores (Parbery-Clark *et al.,* 2009). Also, in this particular study, the number of years of musical experience only correlated with one (QuickSIN) of the two clinical speech-in-noise measures used (QuickSIN and HINT). Parbery-Clark *et al.* (2009) acknowledged that musicians' advantage on the QuickSIN test could be largely attributable to better working memory.

The results of this experiment suggest that there is no significant difference between the cognitive abilities of musicians and non-musicians, but that across all participants, non-verbal IQ is a significant predictor of performance on masked speech tasks. These results are in contrast to the recent study by Ruggles *et al.* (2014), which found a significant difference in the full scale IQs of musicians and non-musicians, but no significant relationship between IQ and SRTs. This inconsistency is further evidence that the relationship between musical training and general cognitive abilities is complex and remains controversial (Schellenberg and Peretz, 2008). Cross-sectional designs make drawing inferences about musicianship problematic, as children with better cognitive abilities may be more predisposed to begin music lessons, and children who opt to take up music lessons may exhibit pre-existing personality differences compared to those who do not (Corrigall *et al.,* 2013). Longitudinal studies examining the effect of musical training may help to address these issues in the future (Kraus *et al.,* 2014). Despite the ambiguity of the causal relationship between the

two, cognitive abilities often co-vary with musical training, making it an important factor to consider in any study with musicians. The results of the current study support a relationship between non-verbal IQ and speech perception in noise, which is consistent with previous observations (Akeroyd, 2008). Consequently, future studies must account for general cognitive abilities when investigating any potential differences between musicians and non-musicians, and more detailed experiments are necessary to identify specific aspects of cognitive and personality differences that may underlie any observed individual differences in speech perception in noise.

There clearly are differences across individuals in how well they cope with masked speech. The results from this experiment demonstrate the necessity of accounting for general cognitive abilities before attributing observed group differences to musical training. An advantage for perceiving masked speech has potential implications for therapeutic interventions for individuals who exhibit difficulties with comprehending speech in noisy environments. Efforts to devise treatments using music have already begun, and show promise (Kraus, 2012; Song *et al.*, 2008; Song *et al.*, 2011); however, more effective therapies would result from a better understanding of the specific aspects of musical training, such as working memory or selective attention, that lead to superior performance—if indeed musical training leads to better performance at all.

ACKNOWLEDGMENTS

This work is supported by grants from the Wellcome Trust (WT090961MA) and the US–UK Fulbright Commission.

Akeroyd, M. A. (**2008**). "Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults," Int. J. Audiol. **47**, S53–S71.

Assmann, P., and Summerfield, Q. (**2004**). "The perception of speech under adverse conditions," in *Springer Handbook of Auditory Research*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, New York), pp. 231–308.

Bench, J., Kowal, A., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Br. J. Audiol. **13**, 108–112.

Blesser, B. A. (**1969**). "Perception of spectrally rotated speech," Doctoral dissertation, Massachusetts Institute of Technology, pp. 1–234.

Blesser, B. A. (**1972**). "Speech Perception under conditions of spectral transformation: I phonetic characteristics," J. Speech, Lang. Hear. Res. **15**, 5–41.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Chan, D., Fourcin, A., Gibbon, D., Granstrom, B., Huckvale, M., Kokkinakis, G., Kvale, K., Lamel, L., Lindberg, B., Moreno, A., Mouropoulos, J., Senia, F., Trancoso, I., Veld, C., and Zeiliger, J. (**1995**). "EUROM—A spoken language resource for the EU," *Proc. 4th Eur. Conf. Speech Commun. Speech Technol.*, pp. 867–870.

Corrigall, K. A., Schellenberg, E. G., and Misura, N. M. (**2013**). "Music training, cognition, and personality," Front. Psychol. **4**, 222.

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (**2007**). "G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," Behav. Res. Methods **39**, 175–191.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., and Pantev, C. (**2004**). "Musical training enhances automatic encoding of melodic contour and interval structure," J. Cogn. Neurosci. **16**, 1010–1021.

Fuller, C. D., Galvin, J. J., Maat, B., Free, R. H., and Başkent, D. (**2014**). "The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations?," Front. Neurosci. **8**, 179.

Gaudino, E. A., Geisler, M. W., and Squires, N. K. (**1995**). "Construct validity in the Trail Making Test: What makes Part B harder?," J. Clin. Exp. Neuropsychol. **17**, 529–535.

Grassi, M., and Soranzo, A. (**2009**). "MLP: A MATLAB toolbox for rapid and reliable auditory threshold estimation," Behav. Res. Methods **41**, 20–28.

Hazan, V., Messaoud-Galusi, S., and Rosen, S. (**2013**). "The effect of talker and intonation variability on speech perception in noise in children with dyslexia," J. Speech, Lang. Hear. Res. **56**, 44–62.

Helfer, K. S., and Freyman, R. L. (**2008**). "Aging and speech-on-speech masking," Ear Hear. **29**, 87–98.

Hentschke, H., and Stüttgen, M. C. (**2011**). "Computation of measures of effect size for neuroscience data sets," Eur. J. Neurosci. **34**, 1887–1894.

Herholz, S. C., and Zatorre, R. J. (**2012**). "Musical training as a framework for brain plasticity: Behavior, function, and structure," Neuron **76**, 486–502.

Jensen, A. R. (**1965**). "Scoring the Stroop test," Acta Psychol. **24**, 398–408.

Jeon, J. Y., and Fricke, F. R. (**1997**). "Duration of perceived and performed sounds," Psychol. Music **25**, 70–83.

Kraus, N. (**2012**). "Biological impact of music and software-based auditory training," J. Commun. Disord. **45**, 403–410.

Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T., and White-Schwoch, T. (**2014**). "Music enrichment programs improve the neural encoding of speech in at-risk children," J. Neurosci. **34**, 11913–11918.

Levitt, H. (**1971**). "Transformed up–down methods in psychoacoustics," J. Acoust. Soc. Am. **49**(2), 467–477.

Macleod, C. M. (**1991**). "Half a century of research on the Stroop effect: An integrative review," Psychol. Bull. **109**, 163–203.

Messaoud-Galusi, S., Hazan, V., and Rosen, S. (**2011**). "Investigating speech perception in children with dyslexia: Is there evidence of a consistent deficit in individuals?," J. Speech, Lang. Hear. Res. **54**, 1682–1701.

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (**2006**). "Influence of musical and psychoacoustical training on pitch discrimination," Hear. Res. **219**, 36–47.

Oxenham, A. J., Fligor, B. J., Mason, C. R., and Kidd, G. (**2003**). "Informational masking and musical training," J. Acoust. Soc. Am. **114**, 1543–1549.

Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (**2009**). "Musician enhancement for speech-in-noise," Ear Hear. **30**, 653–661.

Parbery-Clark, A., Strait, D. L., and Kraus, N. (**2011**). "Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians," Neuropsychologia **49**, 3338–3345.

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). "How young and old adults listen to and remember speech in noise," J. Acoust. Soc. Am. **97**, 593–608.

Rammsayer, T., and Altenmüller, E. (**2006**). "Temporal information processing in musicians and nonmusicians," Music Percept. **24**, 37–48.

Ruggles, D. R., Freyman, R. L., and Oxenham, A. J. (**2014**). "Influence of musical training on understanding voiced and whispered speech in noise," PLoS One **9**, e86980.

Salthouse, T. A. (**2011**). "What cognitive abilities are involved in trail-making performance?," Intelligence **39**, 222–232.

Schellenberg, E. G., and Peretz, I. (**2008**). "Music, language and cognition: Unresolved issues," Trends Cogn. Sci. **12**, 45–46.

Scott, S. K., Rosen, S., Beaman, C. P., Davis, J. P., and Wise, R. J. S. (**2009**). "The neural processing of masked speech: Evidence for different mechanisms in the left and right temporal lobes," J. Acoust Soc. Am. **125**, 1737–1743.

Scott, S. K., Rosen, S., Wickham, L., and Wise, R. J. S. (**2004**). "A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception," J. Acoust. Soc. Am. **115**, 813–821.

Shinn-Cunningham, B. G. (**2008**). "Object-based auditory and visual attention," Trends Cogn. Sci. **12**, 182–186.

Soderquist, D. R. (**1970**). "Frequency analysis and the critical band," Psychon. Sci. **21**, 117–119.

Song, J. H., Skoe, E., Banai, K., and Kraus, N. (**2011**). "Training to improve hearing speech in noise: Biological mechanisms," Cereb. Cortex **22**, 1180–1190.

Song, J. H., Skoe, E., Wong, P. C. M., and Kraus, N. (**2008**). "Plasticity in the adult human auditory brainstem following short-term linguistic training," J. Cogn. Neurosci. **20**, 1892–1902.

Spearman, C. (**1927**). *The Abilities of Man, Their Nature and Measurement* (Macmillan, New York), pp. 1–470.

Spiegel, M. F., and Watson, C. S. (**1984**). "Performance on frequency-discrimination tasks by musicians and nonmusicians," J. Acoust. Soc. Am. **76**, 1690–1695.
Stone, M. A., Füllgrabe, C., Mackinnon, R. C., and Moore, B. C. J. (**2011**). "The importance for speech intelligibility of random fluctuations in 'steady' background noise," J. Acoust. Soc. Am. **130**, 2874–2881.
Stone, M. A., Füllgrabe, C., and Moore, B. C. J. (**2012**). "Notionally steady background noise acts primarily as a modulation masker of speech," J. Acoust. Soc. Am. **132**, 317–326.
Strait, D. L., Parbery-Clark, A., Hittner, E., and Kraus, N. (**2012**). "Musical training during early childhood enhances the neural encoding of speech in noise," Brain Lang. **123**, 191–201.
Stroop, J. R. (**1935**). "Studies of interference in serial verbal reactions," J. Exp. Psychol. **18**, 643–662.
Surprenant, A. M., and Watson, C. S. (**2001**). "Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners," J. Acoust. Soc. Am. **110**, 2085–2095.
Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., and Schröger, E. (**2005**). "Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study," Exp. Brain Res. **161**, 1–10.
Tukey, J. W. (**1977**). *Exploratory Data Analysis* (Addison-Wesley, Reading, MA), pp. 1–688.
Wechsler, D. (**1997**). *Wechsler Adult Intelligence Scale*, 3rd Ed. (The Psychological Corporation, San Antonio, TX), pp. 1–237.
Wechsler, D. (**1999**). *Wechsler Abbreviated Scale of Intelligence* (The Psychological Corporation, San Antonio, TX), pp. 1–237.
Wecker, N. S., Kramer, J. H., Hallam, B. J., and Delis, D. C. (**2005**). "Mental flexibility: Age effects on switching," Neuropsychology **19**, 345–352.
Zendel, B. R., and Alain, C. (**2008**). "Concurrent sound segregation is enhanced in musicians," J. Cogn. Neurosci. **21**, 1488–1498.
Zendel, B. R., and Alain, C. (**2012**). "Musicians experience less age-related decline in central auditory processing," Psychol. Aging **27**, 410–417.
Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F.-X., and Lorenzi, C. (**2005**). "Deficits in speech perception predict language learning impairment," Proc. Natl. Acad. Sci. U.S.A. **102**, 14110–14115.
Ziegler, J. C., Pech-Georgel, C., George, F., and Lorenzi, C. (**2009**). "Speech-perception-in-noise deficits in dyslexia," Dev. Sci. **12**, 732–745.