The Impact of Musicianship on the Cortical Mechanisms Related to Separating Speech from Background Noise

Benjamin Rich Zendel^{1,2,3,4}, Charles-David Tremblay^{1,4}, Sylvie Belleville^{2,4}, and Isabelle Peretz^{1,3,4}

Abstract

Musicians have enhanced auditory processing abilities. In some studies, these abilities are paralleled by an improved understanding of speech in noisy environments, partially due to more robust encoding of speech signals in noise at the level of the brainstem. Little is known about the impact of musicianship on attention-dependent cortical activity related to lexical access during a speech-in-noise task. To address this issue, we presented musicians and nonmusicians with single words mixed with three levels of background noise, across two conditions, while monitoring electrical brain activity. In the active condition, listeners repeated the words aloud, and in the passive condition, they ignored the words and watched a silent film. When background noise was most intense, musicians repeated more words correctly compared with nonmusicians. Auditory evoked responses were attenuated and delayed with the addition of background noise. In musicians, P1 amplitude

INTRODUCTION

It is well known that musicians have enhanced auditory processing abilities (e.g., Zendel & Alain, 2009; Rammsaver & Altenmüller, 2006; Beauvois & Meddis, 1997), and these benefits are paralleled by an enhanced ability to understand speech in noisy environments (Zendel & Alain, 2012; Parbery-Clark, Strait, & Kraus, 2011; Parbery-Clark, Skoe, & Kraus, 2009; Parbery-Clark, Skoe, Lam, & Kraus, 2009). There is some debate about whether this benefit is real, as Ruggles, Freyman, and Oxenham (2014) found no differences between musicians and nonmusicians on speechin-noise tasks. Nonetheless, neurophysiological evidence demonstrates that musicians encode speech signals presented in difficult listening situations more robustly than nonmusicians at the level of the brainstem (Bidelman & Krishnan, 2010 [speech with reverberation]; Parbery-Clark, Strait, et al., 2011; Parbery-Clark, Skoe, & Kraus, 2009 [speech in multitalker babble noise]). Beyond the brainwas marginally enhanced during active listening and was related to task performance in the most difficult listening condition. By comparing ERPs from the active and passive conditions, we isolated an N400 related to lexical access. The amplitude of the N400 was not influenced by the level of background noise in musicians, whereas N400 amplitude increased with the level of background noise in nonmusicians. In nonmusicians, the increase in N400 amplitude was related to a reduction in task performance. In musicians only, there was a rightward shift of the sources contributing to the N400 as the level of background noise increased. This pattern of results supports the hypothesis that encoding of speech in noise is more robust in musicians and suggests that this facilitates lexical access. Moreover, the shift in sources suggests that musicians, to a greater extent than nonmusicians, may increasingly rely on acoustic cues to understand speech in noise.

stem, understanding speech-in-noise is a complex cognitive process that also relies on matching incoming acoustic information to stored lexical representations of individual words. The process of matching incoming speech information to stored lexical representations is relatively automatic; however, with the addition of background noise, this process likely requires increased attentional and cognitive effort. The impact of musicianship on the cognitive mechanisms involved in matching incoming acoustic information to stored lexical representations is not well understood. Improved understanding of the impact of musical training on cortical mechanisms related to speech processing is of utmost importance because there is growing evidence that musical training may be useful for improving auditory perception in those with hearing difficulties, such as older adults (Alain, Zendel, Hutka, & Bidelman, 2014; Wan & Schlaug, 2010).

Understanding speech-in-noise is a bidirectional hierarchical process that occurs in multiple subcortical and cortical structures. Specific task demands and stimulus characteristics can impact which hierarchical level is used to parse speech from noise (Nahum, Nelken, & Ahissar, 2008). Critically, evidence suggests that processing acoustic information is enhanced in musicians. After acoustic information is transduced into a neural signal in the

© 2015 Massachusetts Institute of Technology

Journal of Cognitive Neuroscience 27:5, pp. 1044–1059 doi:10.1162/jocn_a_00758

¹International Laboratory for Brain, Music and Sound Research (BRAMS), Montréal, Québec, Canada, ²Institut Universitaire de Gériatrie de Montréal, Québec, Canada, ³Centre for Research on Brain, Language and Music (CRBLM), Montréal, Québec, Canada, ⁴Université de Montréal, Québec, Canada

cochlea, it is sent via the vestibulocochlear nerve to the brainstem. At the level of the brainstem, acoustic features are encoded through phase-locked neuronal responses that can be measured electrophysiologically (Chandrasekaran & Kraus, 2010; Young & Sachs, 1979; Marsh, Worden, & Smith, 1970). This encoding occurs automatically; however, top-down influence of these responses is possible (Musacchia, Sams, Nicol, & Kraus, 2006; Lukas, 1981). In musicians, enhanced encoding of speech signals when presented in background noise at the level of the brainstem underlies their enhanced ability to better understand speech-in-noise (Parbery-Clark, Strait, et al., 2011; Bidelman & Krishnan, 2010; Parbery-Clark, Skoe, & Kraus, 2009). Specifically, higher harmonics (or formants) of a speech signal presented in background noise are more faithfully encoded in musicians, and thus, a more robust representation of the speech signal is passed on for cortical processing (Parbery-Clark, Strait, et al., 2011; Bidelman & Krishnan, 2010; Parbery-Clark, Skoe, & Kraus, 2009). Neuroplastic modulation of brainstem encoding of acoustic information in musicians is likely driven via the efferent cortico-fugal pathway (Parbery-Clark, Strait, et al., 2011; Parbery-Clark, Skoe, & Kraus, 2009); however, the impact of musicianship on cortical responses to speech-in-noise remains poorly understood.

A more robust representation of the speech signal likely facilitates the segregation of speech from noise into an auditory stream, and musicians are better at segregating and tracking auditory streams (Zendel & Alain, 2009; Beauvois & Meddis, 1997). The segregation of concurrent sounds relies on grouping spectral components (i.e., harmonics) into auditory objects and then tracking those objects over time (Alain, 2007; Bregman, 1990). One way to isolate the influence of spectral information on the perceptual segregation of concurrent sounds is to present a harmonic complex (i.e., a sound composed of harmonics that are integer multiples of a fundamental frequency) where one of the harmonics is mistuned from its original value. Using this type of stimulus, Zendel and Alain (2009, 2013) demonstrated that musicians can detect smaller levels of mistuning compared with nonmusicians. This enhanced ability to perceptually segregate a mistuned harmonic was paralleled by enhancements to a task-independent ERP known as the object-related negativity and a task-dependent component known as the P400 (Zendel & Alain, 2009, 2013). Once sounds are segregated, they must be tracked over time, and musicians are able to maintain segregated auditory streams in memory longer than nonmusicians (Beauvois & Meddis, 1997). This pattern of results demonstrates that musicians are better able to segregate and track acoustic information over time; however, speech is a unique type of auditory stimulus that relies on specialized cognitive mechanisms to derive meaning. Indeed, lexical access and semantic processing of speech inputs are specifically related to an ERP known as the N400 (Kutas & Federmeier, 2011). These specialized mechanisms may impact how speech information is separated from background noise at the level of the cortex.

At the level of the cortex, ERPs to sounds presented in noise are influenced by the type of target sound (tones, speech sounds, words, etc.), the type of noise (white noise, spectrally shaped noise, multitalker babble, etc.), the attentional focus of the listener, and the signal-tonoise ratio (SNR), among other factors. The most basic signal in noise detection task involves processing pure tones in white noise. Billings, Tremblay, Stecker, and Tolin (2009) reported that, when the incoming sounds were not task relevant, the N1 and P2 components decreased in amplitude and increased in latency as SNR decreased. Other studies have replicated this finding for the N1 using speech phonemes embedded in white noise (Kaplan-Neeman, Kishon-Rabin, Henkin, & Muchnik, 2006; Martin, Sigal, Kurtzberg, & Stapells, 1997) and pure tones embedded in speech, music, and intermittent white noise (Hari & Makela, 1988). These results suggest that, as noise level increases, the auditory system has increasing difficulty automatically extracting the signal from the background noise, likely because of a decreased change in the amplitude envelope of the combined target and noise signal when the background noise and target signal have similar amplitudes. In a follow-up study, Billings, Bennett, Molis, and Leek (2011) specifically tested how stimulus and noise type influenced the auditory evoked response. The signals were either pure tones or speech syllables, and the background noise was either multitalker babble or white noise. When white noise was used, a decreasing SNR resulted in a reduced and delayed N1 that was similar for both the pure tones and speech sounds. On the other hand, when multitalker babble was used as noise, the N1 was reduced for the speech sound compared with the pure tone. This finding suggests that informational masking (from competing speech signals) and physical/ energetic masking have separable effects, even when attention is directed away from the auditory scene, and demonstrates the unique nature of separating speech from background noise that is also speech. Importantly, the amplitude of the P1 and N1 components is related to the fidelity of the encoding of the speech sound at the level of the brainstem (Parbery-Clark, Marmel, Bair, & Kraus, 2011; Musacchia, Strait, & Kraus, 2008), suggesting that the P1 or N1 may be enhanced in musicians for processing speech in noise.

Whereas the influence of musicianship on auditory evoked responses to speech in noise is poorly understood, the impact of musical training on the auditory evoked response to sounds presented in isolation is well understood. Previous work has shown enhanced P1, N1, and P2 responses in musicians (Baumann, Meyer, & Jäncke, 2008; Musacchia et al., 2008; Kuriki, Kanda, & Hirata, 2006; Shahin, Bosnyak, Trainor, & Roberts, 2003; Pantev et al., 1998). For the N1–P2 response, these enhancements were especially evident for spectrally rich sounds (Shahin, Roberts, Pantev, Trainor, & Ross, 2005) or sounds that match the musicians instrument of training (Pantev, Roberts, Schulz, Engelien, & Ross, 2001). An enhancement in musicians for P1 was observed for speech sounds (Musacchia et al., 2008); however, a reduction in P1 has been observed in musicians for single tones, chords, and harmonic complexes (Zendel & Alain, 2013; Kuriki et al., 2006). Accordingly, it is likely that P1 is sensitive to musical training. The differential impact of musical training on P1 may be related to the incoming stimulus. Enhanced P1 for speech signals may be related to increased top-down influence from higher cortical structures, whereas the reduced P1 for nonspeech signals may be related to more efficient bottom-up encoding of acoustic information. Enhancements to the N1-P2 seem to be specific to musical specialization as these enhancements were limited to acoustic signals that have similar timbral qualities to instruments the musicians were trained on (Pantev et al., 2001).

Given that the processes of encoding speech sounds in noise (Parbery-Clark, Strait, et al., 2011; Parbery-Clark, Skoe, & Kraus, 2009) or reverberation (Bidelman & Krishnan, 2010) and separating concurrent sounds (Zendel & Alain, 2009, 2013) are enhanced in musicians, it is likely that downstream processing of this information will be facilitated. Matching the neural representation of the incoming acoustic signal to a stored lexical representation is a critical process in understanding speech. It is therefore likely that the process of matching incoming acoustic information to stored lexical representations will be facilitated in musicians. This lexical access is associated with an ERP known as the N400 (Kutas & Federmeier, 2011). The N400 likely represents the comparison between the neural representation of the incoming word and stored lexical representations of words; however, it may also represent semantic integration of the target word with the sentence context in which the word appears (Lau, Phillips, & Poeppel, 2008). Evidence to support the lexical access hypothesis comes from studies that have shown that commonly used words evoke a smaller N400 compared with less common words (Allen, Badecker, & Osterhout, 2003; Van Petten & Kutas, 1990), repeated words evoke a smaller N400 than the first word (Rugg, 1985), and lexically primed words evoke a smaller N400 compared with the same word when not primed (Franklin, Dien, Neely, Huber, & Waterson, 2007). The reduction of the N400 to words that are common or primed supports the lexical access hypothesis of the N400 because it demonstrates that, when the process of matching an acoustic input to a stored representation is facilitated, the N400 is reduced. On the other hand, when the process of matching an acoustic input to a stored representation is impeded, greater neural activity is required to find a lexical match. Moreover, increased difficulty in finding a lexical match will likely lead to more errors in matching the acoustic input to the mental lexicon, and thus, speech understanding will decline. This suggests that the N400 response should increase in amplitude with the addition of background noise. Romei, Wambacq, Besing, Koehnke, and Jerger (2011) investigated this by presenting listeners with word triplets with or without multitalker babble noise. The word triplets were either semantically related or unrelated. The N400 recorded over central and anterior sites increased in amplitude as the level of background noise increased; however, this modulation to the N400 response was only for the second word of the sequence. Moreover, participants were told to attend to all three words; thus, the enhanced N400 response may have been related to priming effects and not the cognitive process of separating the word from background noise. To isolate the specific impact of noise on matching speech to its stored lexical representation, one must manipulate how a listener attends to the auditory environment. Connolly, Phillips, Stewart, and Brake (1992) examined the influence of attentional focus on understanding the terminal word of a sentence when presented in background noise. In two conditions, participants either made an immediate semantic judgment about the sentence or remembered the sentence to answer questions about it later. In both conditions, the addition of background noise delayed the N400, but there was no impact of the attentional manipulation. This is likely because both conditions required the listener to understand the sentence and, therefore, to separate the speech from background noise. These studies suggest that, to isolate the impact of background noise on the N400, a more extreme attentional manipulation is needed.

One way to do this would be to compare N400 responses when the speech is task relevant (i.e., focused attention) with the N400 when the speech is task irrelevant (i.e., ignored). Comparing the N400 when the speech signal is task relevant with when the speech signal is task irrelevant is particularly important because the N400 can be evoked without attention focused on the incoming stimulus (Deacon & Shelly-Tremblay, 2000). Accordingly, the difference between the N400 in task-relevant and taskirrelevant situations would isolate the neural activity related to the increased cognitive effort required to match incoming speech to stored lexical representations. The impact of attention on separating speech from noise is poorly understood, but it is likely that automatic lexical processing of an incoming word will be impacted by the level of background noise. The addition of background noise will likely reduce the quality of the neural encoding of the incoming acoustic stimulus. This "messy" input will likely require greater neural effort to find a lexical match. Accordingly, the attention-dependent component of the N400 response should be larger when background noise is more intense, reflecting greater neural effort, whereas the automatic aspect of the N400 will decrease as noise level increases because of physical masking of the auditory input. At the same time, musicianship should attenuate the background noise-related increase in the N400 response because the encoding of speech sounds in noise is more robust in musicians, and musicians are better able to use the encoded spectral information to separate the speech from background noise. This putative difference between musicians and nonmusicians would provide support for the lexical access model of the N400 because it would demonstrate that decreased neural effort is needed at the lexical access stage when earlier encoding and sound segregation are enhanced. In addition, we expect to observe enhancements to the P1 or N1 response in musicians because they are related to the fidelity of the incoming speech signal. Differences between musicians and nonmusicians in these responses would provide empirical support that the speech sounds were more robustly encoded in the current sample of musicians; however, given that musicians and nonmusicians have similar experiences with speech, an enhanced N1 may not be present. Behaviorally, we only expect to observe an advantage for musicians when background noise is the most intense.

METHODS

Participants

Twenty-six participants were recruited for the study and provided formal informed consent in accordance with the research ethics board of the Quebec Neuroimaging Group. Participant demographics are presented in Table 1. All participants were right-handed, native French speakers, and bilingual (all had English competency). All participants had normal audiometric thresholds (i.e., below 25-dB hearing level for frequencies between 250 and 8000 Hz), and a 2 (musician, nonmusician) \times 2 (left ear, right ear) \times 6 (frequency: 250, 500, 1000, 2000, 4000, and 8000 Hz) ANOVA revealed that pure-tone thresholds were similar for both musicians and nonmusicians (F(1, 24) = 0.04, p =.95). Participants were recruited from the university community through advertisements posted around the campus. All musicians had formal training, began training by the age of 15 years, had at least 10 years of musical experience, and practiced a minimum of 10 hr per week in the year the testing took place. The principal instruments varied but included both instrumentalists and vocalists. Nonmusicians had less than 1 year of formal lessons and did not regularly play a musical instrument.

Stimuli

Stimuli were 150 French words spoken by a male native speaker of Quebec French. These words were taken from a test used by audiologists in Quebec to measure audio-

metric speech thresholds (Benfante et al., 1966). The words were rated as being familiar by people from six distinct regions of Quebec, were from all parts of speech, were monosyllabic consonant-vowel-consonant structure, and were phonetically representative of Quebec French (plus, cirque, forte, grande, trouve, etc.; see Picard, 1984). Words were presented binaurally at 75-dB sound pressure level, through insert earphones (Etymotic ER-2, Elk Grove Village, IL), as determined by a sound level meter (Quest Technologies, Medley, FL) that measured the amplitude of the stimuli presented from the left insert earphone. In two of the conditions, multitalker babble noise was presented with the words at 60 (\pm 3)- and 75 (\pm 3)-dB sound pressure level, yielding an SNR of 15 and 0 dB, respectively. The multitalker babble was created by individually recording four native speakers of Quebec French (two women, two men), each reading a rehearsed monologue in a sound-attenuated room for 10 min. The recordings were made at a sampling rate of 44.1 kHz at 16 bits, using an Audio-Technica 4040 condenser microphone. The individual recordings of each monologue were amplitude normalized and combined into a single monaural sound file using Adobe Audition (Version 10; San Jose, CA). The 10-min multitalker babble noise was looped repeatedly during listening conditions where the multitalker babble was present.

Procedure

All 150 words were presented in a random order, in each of the three levels of multitalker babble noise. In the "none" condition, words were presented without multitalker babble noise. In the "SNR-15" condition, words were presented with multitalker babble noise that was 15 dB below the level of word (i.e., 15-dB SNR), whereas in the more difficult "SNR-0" condition, words were presented with multitalker babble noise that was at the same level as the word (i.e., 0-dB SNR). In addition, all three noise levels were presented in two listening conditions, active and passive. In the passive condition, participants were told to ignore the words and watched a self-selected silent subtitled movie; words were presented with an SOA that was randomized between 2500 and 3500 msec. The use of muted subtitled movies has been shown to effectively capture attention without interfering with

Table 1. Participant Demographics

Group	Age (Years)	Gender	Education (Years)	Music Training Onset (Age)	Music Experience (Years)	Music Practice (Hours per Week)	
Musicians	18-35 (M = 23.4, SD = 4.3)	Five women, eight men	14-21 (M = 16.7, SD = 1.9)	2-15 (M = 7.8, SD = 3.6)	10-28 (M = 15.5, SD = 5.1)	11-85 (M = 23.4, SD = 4.3)	
Nonmusicians	19-27 (M = 21.9, SD = 2.6)	Nine women, four men	14–18 ($M = 15.2$, SD = 1.1)	_	_	-	

auditory processing (Pettigrew et al., 2004). In the active condition, participants were told to repeat the word aloud and did not watch a movie. To avoid muscle artifacts in the ERPs, participants were told to delay their response until they saw a small LED light flash 2000 msec after the presentation of the word. Word correctness was judged online by a native French speaker. The subsequent word was presented 2000 msec after the response judgment was made. We chose to use word repetitions because it required an accurate lexical match of the incoming word to correctly repeat it back, and pilot testing confirmed that the delayed oral response did not contaminate the ERPs with muscle artifacts. An alternative would have been to use a forced-choice procedure; however, this would likely create a biased estimate of word understanding because the presentation of choices limits what a participant can report and may bias their performance if they were able to hear part of the word. The active SNR-0 condition was always presented first, followed by the active SNR-15 and active none conditions; these were then followed by the passive conditions, which were presented in a random order. This order was chosen to ensure that performance accuracy in the SNR-0 condition (the most difficult condition) was not biased by prior exposure to the words. Performance in the SNR-15 and none conditions were both expected to be near ceiling; thus, any prior exposure to the words would be identical for each participant, making performance in these two conditions comparable across participants.

Recording and Averaging of Electrical Brain Activity

Neuroelectric brain activity was digitized continuously from 71 active electrodes at a sampling rate of 1024 Hz, with a high-pass filter set at 0.1 Hz, using a Biosemi ActiveTwo system (Biosemi, Inc., Amsterdam, Netherlands). Six electrodes were placed bilaterally at mastoid, inferior ocular, and lateral ocular sites (M1, M2, IO1, IO2, LO1, and LO2). All averages were computed using BESA (Version 5.2). The analysis epoch included 200 msec of prestimulus activity and 1500 msec of poststimulus activity. Trials containing excessive noise (>120 µV) at electrodes not adjacent to the eyes (i.e., IO1, IO2, LO1, LO2, FP1, FP2, FPz, FP9, and FP10) were rejected before averaging. Continuous EEG was then averaged separately for each condition, into six ERPs-active: none, SNR-15, and SNR-0; passive: none, SNR-15, and SNR-0-at each electrode site. Prototypical eye blinks and eye movements were recorded before the start of the study. A PCA of these averaged recordings provided a set of components that best explained the eye movements. These components were then decomposed into a linear combination along with topographical components that reflect brain activity. This linear combination allowed the scalp projections of the artifact components to be subtracted from the experimental ERPs to minimize ocular contamination

such as blinks, vertical, and lateral eye movements for each individual average with minimal effects on brain activity (Berg & Scherg, 1994). After this correction, trials with greater than 100 μ V of activity were considered artifacts and excluded from further analysis. In addition, during active listening, trials where the participant did not correctly repeat the word were excluded from the analysis. Averaged ERPs were then band-pass filtered to attenuate frequencies below 0.1 Hz, and above 15 Hz, and referenced to the linked mastoid.

Data Analysis (Electrophysiological)

To account for the effect of noise on the latency of the P1, N1, and P2 responses (see Figure 1A), peak amplitude and latency were extracted from different analysis epochs for each noise condition. Peak amplitude (i.e., the maximum amplitude of the ERP during the analysis epoch) was chosen as an appropriate measure because of the significant effects of noise on the latency of the evoked responses. The influence of noise level on the latency of the P1-N1-P2 response was also why different windows were chosen to search for each peak in each noise level. Windows were chosen based on a visual inspection of the averaged data. For the none condition, the window was 40-140 msec for P1, 75-175 msec for N1, and 140-260 msec for P2. For the SNR-15 condition, the window was 40-140 msec for P1, 115-215 msec for N1, and 190-310 msec for P2. For the SNR-0 condition, the window was 40-140 msec for P1, 130-280 msec for N1, and 230-330 msec for P2. The P1-N1-P2 complex were maximal at electrode Cz; therefore, to ensure a stable and reliable estimate of the P1-N1-P2, a montage of nine central electrodes was used (FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, and CP2). Separate ANOVAs were carried out on the amplitude and latency data using Noise level (none, SNR-15, and SNR-0), Listening condition (active, passive), and Electrode as within-subject factors and Musicianship (musician, nonmusician) as a between-subject factor.

To isolate the influence of attention on hearing a word in noise, difference waves were calculated by subtracting ERPs recorded during the passive listening task from the ERPs recorded during the active listening task. In the none condition, lexical matching can occur automatically, regardless of attention (Deacon & Shelly-Tremblay, 2000). As noise level increased, greater cognitive resources were required to understand each word. Thus, by examining difference waves, we can isolate the additional attention-dependent cognitive mechanisms required to understand a word when presented in noise. This analysis focused on the peak amplitude and latency of the difference waves between 325 and 900 msec. Like the P1-N1-P2 data, peak amplitude was chosen as the best measure because of the impact of background noise on the latency of the response. Peak amplitude and latency were analyzed in separate mixed-design ANOVAs with Noise level and Musicianship as factors. As a follow-up to



Figure 1. Percentage of words repeated correctly for musicians and nonmusicians.

this analysis, we conducted the same analysis on the evoked responses (i.e., not the difference waves; see Figure 4) during both active and passive listening. For this analysis, we will refer to the N400 as the N400e ([e]voked) to maintain the distinction between the N400 derived from the difference wave. Peak amplitude for the N400e was analyzed separately using a mixed-design ANOVA with Noise level, Listening condition, and Musicianship as factors. Alpha for all statistical tests was set at 0.05, and p values between .05 and .1 were considered significant at a trend level.

Post Hoc Analyses

Given the interaction between musicianship and noise level on the N400 response and its topography (see Results), we wanted to examine if there was a shift in the underlying sources of the N400 in musicians. To determine the distribution of sources that contribute to the peak of the difference wave (i.e., N400), we calculated a local autoregressive average (LAURA) source analysis using BESA (Version 5.2; Gräfelfing, Germany) for each noise level and for each group (Menendez, Andino, Lantz, Michel, & Landis, 2001). The LAURA technique is a weighted minimum norm method that uses a spatial weighting function in the form of a local autoregressive function to compute a distributed inverse solution to scalp-recorded electrical brain activity. The analysis assumed a four-shell ellipsoidal head model with relative conductivities of 0.33, 0.33, 0.0042, and 1 for the head, scalp, bone, and cerebrospinal fluid, respectively, and sizes of 85 mm (radius), 6 mm (thickness), 7 mm (thickness), and 1 mm (thickness). This analvsis locates broad areas of activation, with local maxima presented in the order of their strength, such that the first local maximum is the strongest point of activation in the brain for that group/condition. These sources were calculated on the group average and were therefore not quantified statistically. Individual variability in distributed source analyses when using EEG data makes it difficult to make comparisons between participants. Specifically, because of individual variability, the exact sources will be unique for each participant. Moreover, the main purpose of this analysis was to estimate putative brain regions involved in understanding speech in background noise and to observe if these sources are differently impacted by background noise in musicians compared with nonmusicians. Working with the averaged data provides the clearest and most robust estimate of these regions.

A second group of post hoc analyses focused on various correlations within the data. First, we looked at the group of musicians and determined if both behavioral performance and neurophysiological measures were related to the age at which musical training began, the years of musical training, and the hours of weekly practice. The second group of correlations examined the relationship between task performance and neurophysiological responses that were modulated by both musicianship and noise level, to demonstrate a relationship between brain activity and behavior. We calculated two types of correlations for this analysis, between-subject and within-subject correlations. Between-subject correlations determine the relationship between behavioral performance and brain activity in a given condition. These correlations tell us if the overall magnitude of a neurophysiological measurement predicts performance. Within-subject correlations determine if the impact of noise level in a neurophysiological measure is correlated to the impact of noise level on performance. To do this, a Pearson r is calculated between task performance and a neurophysiological measure from the same condition, for each participant. Then, at the group level, these values are compared against the null hypotheses (i.e., a Pearson r of 0) using a one-sample t test. A final group of within-subject correlations was calculated between the latency of the evoked responses and the latency of the N400 to determine if the noiserelated change in latency for the P1-N1-P2 response predicted the noise-related change in the N400.

RESULTS

Behavioral Data

The percentage of words repeated correctly is shown in Figure 1. Noise level had a significant impact on the number of incorrect words (F(2, 48) = 237.186, p < .001). Follow-up pairwise comparisons revealed significantly more errors in the SNR-0 condition compared with SNR-15 condition (p < .001) and more errors in the SNR-15 condition compared with the None condition (p < .001). Musicians repeated more words correctly overall (F(1, 24) = 5.13, p = .03); however, the influence of Musicianship interacted with Noise level (F(2, 48) = 6.72, p = .003). Follow-up comparisons revealed that musicians repeated more words correctly in the SNR-0 condition (p = .02) but not in the SNR-15 or none conditions (ps = .76 and .26, respectively).

Difference Waves

The main purpose of this experiment was the interaction between musicianship and noise level on attentiondependent cognitive activity related to understanding speech in noise. Accordingly, the first analysis focused on the comparison between active and passive listening after the P1-N1-P2 waves. A difference wave was calculated between active and passive listening for each participant and each noise level. This difference wave is presented in Figure 2, and the overall scalp topography for the 10-msec window surrounding the N400 peak is presented in Figure 3. The analysis revealed a peak that occurred around 450 msec for both musicians and nonmusicians in the none condition. For the nonmusicians, the amplitude of this peak increased as noise level increased. For both groups, the latency increased with noise level. This effect may be because of a shift in the underlying sources for this activity in the musicians, as the scalp topography appears to shift as noise level increased in musicians, whereas the topography remained relatively stable in nonmusicians (see Figure 3).

For N400 amplitude, the main effect of Noise level was significant at a trend level (F(2, 48) = 2.82, p = .07); however, there was a significant Noise level × Musician-

ship interaction (F(2, 48) = 3.99, p = .03). A follow-up simple main effects analysis found an effect of Noise level in nonmusicians (F(2, 24) = 11.73, p < .001) but no effect of Noise level in the musicians (F(2, 24) = 1.15, p = .33). In nonmusicians, polynomial decompositions revealed a linear effect of increasing noise level on N400 amplitude (F(1, 12) = 15.19, p = .002). The linear trend was not significant in musicians (F(1, 12) = 1.06, p = .32).

The N400 latency increased as noise level increased (F(2, 24) = 4.47, p = .02), and follow-up polynomial decompositions revealed that the increase was linear (F(1, 24) = 6.88, p = .02). The effect of Musicianship and its interaction with Noise level were not significant for N400 latency (ps = .58 and .49, respectively). Accordingly, a linear increase in N400 latency was observed for both musicians (566, 624, and 631 msec for none, SNR-15, and SNR-0, respectively) and nonmusicians (526, 589, and 665 msec for none, SNR-15, and SNR-0, respectively).

Evoked Responses

To further probe the N400 response, we analyzed the peak amplitude of the N400e(voked) response in both



Figure 2. (A) Averaged difference waves (active listening – passive listening) plotted at electrode CP2 separately for musicians and nonmusicians. Each noise level is plotted separately. (B) Peak amplitude of the N400 wave. (C) Latency of the N400 wave.

Figure 3. Scalp topography for the N400. Musicians are plotted on the left; nonmusicians, on the right. Topographical epoch is a 10-msec window around the peak of the N400 for each group/condition.



active and passive listening (Figure 4). Overall, N400e amplitude was larger during active listening compared with passive listening (F(1, 24) = 32.14, p < .001) and was larger in the none condition compared with the SNR-15 and SNR-0 conditions (F(2, 48) = 7.58, p =.001). These two main effects are qualified because of a significant Noise level × Listening condition interaction (F(2, 48) = 4.78, p = .013). Simple main effects revealed a significant decrease in N400e amplitude as noise level increased during passive listening $(F(2, 50) = 18.92, p < 10^{-5})$.001; linear trend: F(1, 25) = 27.35, p < .001). There was no impact of noise level on N400e amplitude during active listening (p = .13). Although there were no significant effects of musicianship, there were some trend level effects. The main effect of Musicianship was significant at a trend level, as the N400e was smaller in musicians compared with nonmusicians (F(1, 24) = 3.01,p = .093). Importantly, this difference may only have been significant during active listening as the Musicianship \times Listening condition interaction was also significant at a trend level (F(1, 24) = 2.99, p = .096).

The next step in the analysis focused on the P1–N1–P2 waves of the auditory ERP. The ERPs are plotted at electrode Cz in Figure 4, and the peak amplitudes and latencies are presented in Figure 5. In the None condition, clear P1–N1–P2 waves can be seen in both active and passive listening. As the noise level increased, the P1–N1–P2 were delayed and reduced in amplitude.

The amplitude of the P1 response decreased as noise level increased (F(2, 48) = 7.32, p = .002) but was not influenced by Listening condition (p = .26). Importantly, the P1 was larger in musicians compared with nonmusicians, although this only reached significance at a trend level (F(1, 24) = 3.04, p = .09). The effect of Musicianship did not interact with Noise level, Listening condition, or the Noise level × Listening condition interaction (ps =.98, .16, and .635, respectively). Although none of the interactions were significant, there appeared to be a trend that P1 was larger in musicians only during active listening (see Figure 5). Accordingly, we calculated simple main effects of Musicianship for P1 amplitude in both active and passive listening. In active listening, the P1 was larger in musicians compared with nonmusicians, but this difference was only significant at a trend level (F(1, 24) = 3.91, p = .06). During passive listening, P1 amplitude was similar in both musicians and nonmusicians (p = .48).

The latency of the P1 response increased as noise level increased (F(2, 48) = 11.49, p < .001) but was not influenced by listening condition (p = .25). Importantly, P1 latency was longer in musicians compared with non-musicians (F(1, 24) = 5.02, p = .035). The effect of Musicianship did not interact with Noise level, Listening condition, or the Noise level × Listening condition interaction (ps = .78, .17, and .23, respectively).

The amplitude of the N1 response decreased as noise level increased (F(2, 48) = 38.36, p < .001) and was larger overall during active listening (F(1, 24) = 18.72, p < .01). The effect of noise level was reduced during active listening (i.e., less noise level related reduction in N1 amplitude during active listening); however, the Noise level × Listening condition interaction only reached a trend level of significance (F(2, 48) = 2.57, p = .09). The main effect of Musicianship was not significant (p = .71) nor was its interaction with Noise level, Listening condition, or the Noise level × Listening condition interaction (ps = .86, .63, and .88, respectively).

The latency of the N1 response increased as noise level increased (F(2, 48) = 224.75, p < .001) but was not influenced by either listening condition or musicianship (ps = .49 and .66, respectively). More importantly, Musicianship did not interact with Noise level, Listening condition, and the Noise level × Listening condition interaction (ps = .56, .73, and .65, respectively).

The amplitude of the P2 response decreased as noise level increased (F(2, 48) = 88.1, p < .001) and was not influenced by Listening condition or Musicianship (ps = .9)

and .96, respectively). In addition, Musicianship did not interact with Noise level, Listening condition, and the Noise level \times Listening condition interaction (ps = .88, .95, and .15, respectively).

The latency of the P2 response increased as noise level increased (F(2, 48) = 182.84, p < .001) and was longer during active listening (F(1, 24) = 21.69, p < .001); however, these factors did not interact (p = .46). The main effect of Musicianship was not significant (p = .1) nor was its interaction with Noise level, Listening condition, and the Noise level × Listening condition interaction (ps = .91, .59, and .17, respectively).

Source Analysis (LAURA)

In musicians, 10 local maxima were identified for each noise level. For nonmusicians, 10 local maxima were identified in the none condition, 8 in the SNR-15 condition, and 10 in the SNR-0 condition. To ease comparisons between the groups, the six most prominent sources for each group in each condition are presented in Table 2. In this table, the approximate brain region is listed in each column, and the group and noise level are listed on each row. In each cell, the image presents the local maxima with a small black square, and the number represents the relative size of that maxima compared with the others in the same group and condition (i.e., along the same row; 1 is the largest). Empty cells represent areas that were not one of the six most prominent sources in that group/condition. Although the pattern is complex, the overall trend suggests that the addition of background noise did not have an impact on the distribution of sources in the nonmusicians. In the musicians, the distribution of sources had a slight rightward shift as noise level increased. Specifically, in the SNR-0 condition, the strongest activation was in the right anterior superior temporal gyrus (STG; Brodmann's area [BA] 38) for the musicians. The strongest peak in the SNR-15 and none conditions in musicians and for all conditions in nonmusicians was in the left anterior STG (BA 38). In addition, the medial portion of the right STG (BA 21/22, near primary auditory cortex) was active for musicians in both



Figure 4. Averaged ERPs for musicians (top) and nonmusicians (bottom) plotted at electrode Cz. The passive listening condition is plotted on the left, and the active listening condition is plotted on the right. Each plot illustrates the three noise levels separately. P1, N1, and P2 are labeled on the top left plot.



Figure 5. Amplitudes and latencies of the P1, N1, and P2 responses. Amplitude is on the left, and latency is on the right. P1 is on the top row, N1 is in the middle row, and P2 is on the bottom. Note that N1 amplitude is plotted in negative microvolts to ease comparison between the components.

the SNR-15 and SNR-0 conditions, and this region was not active in the nonmusicians during the SNR-15 and SNR-0 conditions.

Correlations

First, in the group of musicians, we found that the age at which music training started predicted performance in both the SNR-15 and SNR-0 conditions (r(13) = -.72, p = .006 and r = -.56, p = .049) but not in the none condition (p = .33). For the SNR-15 and SNR-0, the younger a musician was when they started playing an instrument, the more words they were able to correctly repeat in both levels of background noise. The overall years of music training and the average hours per week of practice did not predict task performance (p = .13-.87).

The brain-behavior correlations focused on the P1 during active listening and the N400 derived from the difference wave, as these responses were differentially impacted by group and noise level. In the none and SNR-15 conditions, neither P1 amplitude nor latency predicted task performance. In the SNR-0 condition, in-

creased P1 amplitude during active listening predicted improved task performance (r(26) = .50, p = .009). Most critical, when we ran the analysis separately in each group, the correlation was only significant for musicians (r(13) =.67, p = .01) and not for the nonmusicians (r(13) = .15, p = .62). A scatterplot of these data can be seen in Figure 6A. Between-subject correlations revealed no significant relationship between N400 amplitude or latency and task performance.

Within-subject correlations revealed that the change in N400 amplitude was related to the change in task performance. Overall, lower N400 amplitude and N400 latency predicted improved task performance (t(25) = 2.97, p = .006 and t(26) = -3.14, p = .004, respectively). When the same analysis was performed on each group separately, we found that the relationship between N400 amplitude and latency remained significant for nonmusicians (t(12) = 4.45, p = .001; t(12) = -3.53, p = .004) but not for musicians (ps = .42 and .20, respectively). Figure 6B illustrates the relationship between task performance and N400 amplitude and latency, separately for musicians and nonmusicians. For the nonmusicians, a relationship between



STG = superior temporal gyrus; SFG = superior frontal gyrus; AnG = angular gyrus; MTG = middle temporal gyrus; Cereb. = cerebellum; ITG = inferior temporal gyrus; IFG = inferior frontal gyrus; Mus = musician; Non = nonmusician. The number denotes the relative size of the activation compared with other areas for the same group and condition (1 is largest); the black square on each brain indicates the local maxima for that source.

decreasing accuracy and increased N400 amplitude and latency can be seen. This relationship is not present in the musicians. Within-subject correlations were not significant for P1.

Within-subject correlations revealed that the noise-related change in latency for the N1 and P2 predicted the noise-related change in the N400 (t(25) = 2.80, p = .01; t(25) = 2.91, p = .008). P1 latency was not related to N400 latency (p = .5). Interestingly, when we performed the same analysis separately in each group, the relation-ship between N1 and P2 amplitude only remained significant for nonmusicians (t(12) = 3.53, p = .004 and t(12) = 3.1, p = .009, respectively). The relationship between N1/P2 latency with N400 latency was not significant in musicians (ps = .33 and .23, respectively).

DISCUSSION

The main finding from this study was that the background noise-related increase in N400 amplitude was mitigated in musicians compared with nonmusicians. This was paralleled by a noise-related shift in both the topography and the underlying sources of the N400 in musicians but not nonmusicians. In addition, the P1-N1-P2 were delayed and attenuated with the addition of background noise in both groups; however, the P1 was marginally enhanced in musicians across active listening conditions. Behaviorally, musicians were able to repeat more words correctly in the most difficult listening situation, providing support for the idea that younger musicians have enhanced abilities for understanding speech in noisy environments (see Parbery-Clark, Strait, et al., 2011; Parbery-Clark, Skoe, & Kraus, 2009, but see Ruggles et al., 2014, for a null result). Finally, to the best of our knowledge, this is the first study to demonstrate an advantage in French-speaking musicians during a speech-in-noise task.

The P1–N1–P2 is known to index basic processing of a transient acoustic input and is sensitive to a rapid change in acoustic energy (Näätänen & Picton, 1987). Like previous studies that presented stimuli in noise (Billings et al., 2009, 2011; Parbery-Clark, Marmel, et al., 2011; Romei et al.,



Figure 6. (A) Scatterplot illustrating the relationship between accuracy (i.e., percent of words repeated correctly) and P1 amplitude in the SNR-0 condition for musicians and nonmusicians. (B) Within-subject correlations illustrating the relationship between accuracy and N400 amplitude and latency, separately for musicians and nonmusicians.

2011; Connolly et al., 1992), we observed a decrease in the amplitude and increase in the latency of the P1-N1-P2 as noise level increased. Importantly, we observed that the P1 response was delayed and a trend toward it being enhanced during active listening in musicians compared with nonmusicians, which is consistent with previous findings of P1 enhancement in musicians for speech sounds (Musacchia et al., 2008). The P1 enhancement during active listening is suggestive of a top-down influence on early cortical processing of acoustic information, and the delay suggests that this enhanced processing requires extra time. Supporting this proposal is evidence that suggests that P1 is related to subcortical encoding of speech sounds (Musacchia et al., 2008), and subcortical encoding of speech sounds in noise is enhanced in musicians (Parbery-Clark, Strait, et al., 2011; Parbery-Clark, Skoe, & Kraus, 2009). In the current study, the marginally enhanced P1 in musicians may be because of enhanced cognitive processing of basic auditory features that can be used to separate speech from background noise, as the P1 amplitude in the most difficult listening condition predicted task performance. Alternatively, it may be related to an attentiondependent feedback mechanism that can modify the function of the brainstem. This could be one of the mechanisms that drives long-term plasticity related to musicianship via the cortico-fugal pathway to the auditory brainstem (Parbery-Clark, Strait, et al., 2011; Parbery-Clark, Skoe, & Kraus, 2009).

Although a trend for an enhanced P1 was observed in musicians compared with nonmusicians, the N1 and P2 waves were similar in both groups, even as noise level increased. Previous work has shown enhanced N1 and P2 responses in musicians (Shahin et al., 2003; Pantev et al., 1998), and these enhancements were especially evident for spectrally rich sounds (Shahin et al., 2005) or sounds that match the musicians instrument of training (Pantev et al., 2001). Accordingly, some aspects of the N1-P2 waves seem to be input specific, and therefore, the underlying neural processing that occurs during the N1-P2 is not completely shared for processing music and language. Indeed, one of the current theories as to why musical training impacts the processing of speech is that speech processing requires many of the same neural resources as processing music, and many of these overlapping neural processes are strengthened by musical training (Patel, 2011, 2012). Accordingly, when we observe enhancements in musicians for processing speech information, it is likely that the same neural resources are used for processing music. This lack of overlap may explain the lack of effect of musicianship on the N1-P2.

After the basic features of the speech signal are processed, a lexical comparison is possible. For the participants to perform the active task properly, they had to repeat the incoming spoken word. To accomplish this neurophysiologically, a match must be made between the incoming acoustic information and stored representations of the word. Once a match was found, the word could be repeated. It is likely that this lexical matching process was reflected in the N400 component of the auditory evoked response (Kutas & Federmeier, 2011; Lau et al., 2008). In the current study, we observed an increased negativity during active listening that peaked around 450 msec during the "no noise" condition in both musicians and nonmusicians. The overall topography of the N400 response showed a peak at fronto-central locations, and the underlying sources included the superior temporal gyrus, frontal cortex, and angular gyrus. Both the topography and source distribution are consistent with an N400 (Lau et al., 2008; Hickok & Poeppel, 2007). One issue with the source analysis was that we observed sources in the left middle and superior frontal gyri, and not in the left inferior frontal gyrus, a structure that is known to be critical for processing language. This may have been simply a technical limitation with doing source analysis on ERP data, as the large activation we found near the temporal pole suggests that these frontal sources may actually be because of activity in the inferior frontal gyrus. In addition, the largest peak in both musicians and nonmusicians was observed in the left superior temporal gyrus, at the anterior end (BA 38), which is proximal to the inferior frontal gyrus. Given this proximity, it is not surprising that this area is also associated with language processing. Specifically, BA 38 is believed to be related to lexical access, as it is active when processing words compared with acoustically matched noise (Mummery & Ashburner, 1999); is highly atrophied in persons with semantic dementia (Mummery et al., 2000); and is related to semantic processing of written words (Pugh et al., 1996). Given the latency, topography, and source analysis of the response in the current study, we consider the negativity observed after 400 msec in the difference wave to be an N400, and it is therefore likely related to making a lexical match between the incoming acoustic waveform and a stored representation of the word.

In both groups, increasing levels of background noise delayed the N400 response. The noise-related increase in latency for the N400 response in both groups suggests that semantic processing was delayed by the addition of background noise. Although the N400 latency is usually stable (Kutas & Federmeier, 2011), previous work has found that background noise can delay the N400 (Connolly et al., 1992). Here, we demonstrate that this delay was related to the increased latency of the N1–P2 in increasing levels of background noise. Accordingly, this delay in the formation of the word as an auditory object (during the N1–P2 epoch) likely has a cascading effect on subsequent lexical processing of the word.

In the nonmusicians, the amplitude of the N400 response increased with the level of background noise, and this increase in amplitude was related to a decrease in task performance. This is consistent with previous reports of increasing N400 amplitude when word identification was more difficult (Franklin et al., 2007; Allen et al., 2003; Van Petten & Kutas, 1990; Rugg, 1985). The current findings suggest that noise increases the difficulty of segregating a word as a perceptual object and that this difficulty makes the lexical matching process more difficult. That is, the noise cannot be completely perceptually filtered, and thus, the neural representation of the word is distorted. Accordingly, matching the incoming word to a stored lexical representation takes more time and requires greater neural resources. This is reflected in the increased latency and amplitude of the N400 component measured in nonmusicians.

More important was the influence of musical training on the N400. In nonmusicians, the amplitude of the N400 increased as the level of background noise increased, whereas in musicians, the level of background noise had little impact on the N400 amplitude, and the N400 amplitude was not related to task performance. This suggests that the lexical matching process when words were presented in noise was facilitated in musicians, likely because of enhanced processing during the P1, which was related to task performance in musicians during the most difficult listening situation. This finding is consistent with previous studies demonstrating that the encoding of speech sounds in noise (Parbery-Clark, Strait, et al., 2011; Bidelman & Krishnan, 2010; Parbery-Clark, Skoe, & Kraus, 2009) and that the ability to separate concurrent sounds (Zendel & Alain, 2009, 2013) are both enhanced in musicians.

It is also possible that musicians use different brain areas to process speech in noise. In the no-noise condition, the overall distribution of sources was similar to that of nonmusicians; however, as noise level increased, sources in the right hemisphere increasingly contributed to the N400. Specifically, activity in the anterior portion of the right superior temporal gyrus and activity near the right auditory cortex made an increasingly important contribution to the N400 wave in musicians. This rightward shift in sources may be related to differences in how musicians and nonmusicians process speech in noise. It is well established that auditory processing is somewhat specialized in each hemisphere, with the right auditory cortex being specialized for processing spectral information and the left being specialized for processing temporal information (Warrier et al., 2009; Johnsrude, Penhune, & Zatorre, 2000; Liégeois-Chauvel, de Graaf, Laguitton, & Chauvel, 1999). Psychoacoustic studies have demonstrated that spectral information can be used to separate concurrent sounds (Moore, Glasberg, & Peters, 1986), and identification of concurrent vowel sounds has been associated with right-lateralized ERPs (Alain, Reinke, He, Wang, & Labaugh, 2005). Moreover, specialized language processing structures, such as Broca's area, are left lateralized, and current models of speech processing models provide further support that lexical matching is also left lateralized (Lau et al., 2008; Hickok & Poeppel, 2007). On the basis of these models, the current source analysis may have revealed a shift in "neural strategy" for musicians in increasing levels of background noise. The rightward shift suggests that musicians may increasingly rely on spectral acoustic cues to separate speech from noise. Spectral information is critical for segregating simultaneously occurring sounds (Alain, 2007), and musicians are better at separating simultaneous sounds using spectral information (Zendel & Alain, 2009, 2013). In addition, attention-dependent positive electrical brain activity related to the detection of concurrent sounds is enhanced in musicians during the 400-msec poststimulus epoch (Zendel & Alain, 2009, 2013, 2014). Critically, if musicians are treating speech in noise as an acoustic task, the positive polarity of the P400 related to separating concurrent sounds may have counteracted the increase in N400 amplitude during the SNR-15 and SNR-0 conditions. On the other hand, for nonmusicians, the noise-related increase in the N400, coupled with a stable topography and source distribution, suggests that nonmusicians rely more heavily on lexical information to help separate speech from noise; thus, increased activity is observed in left-lateralized regions associated with semantic processing and lexical access.

One important issue to point out is the seemingly opposite effect of noise when looking at the evoked responses, instead of the difference waves. In both groups, the N400e was largest in the none condition and decreased as noise level increased during passive listening. In the case of passive listening, these waveforms represent bottom-up processing of acoustic information; for active listening, they represent both bottom-up and top-down processing. In the none condition, matching a word to a stored lexical representation is likely automatic (Deacon & Shelly-Tremblay, 2000), and a clear N400 peak can be seen in the no-noise condition in both active and passive listening in both groups on Figure 4. Accordingly, even while ignoring the incoming words, it is likely that some lexical processing of the word occurred. As the level of background noise increased, this process requires more cognitive effort; thus, as noise level increased, the N400e decreased during passive listening only. The lack of change in the N400e during active listening suggests increasing cognitive demands that are best reflected in the difference between active and passive listening. Accordingly, focusing on the difference waves highlighted this increase in attention-dependent lexical access.

Conclusion

Enhanced early encoding of speech information and enhanced segregation of simultaneous sounds in musicians likely facilitate downstream lexical processing of speech signals presented in background noise. In support of this proposal, we observed a marginally increased P1 amplitude in musicians compared with nonmusicians, which was related to task performance in the most difficult condition in musicians only. In nonmusicians, the topography and underlying sources contributing to the N400 response remained relatively stable as background noise level increased, and the noise-related change in N400 amplitude was related to task performance. In musicians, the N400 amplitude was little impacted by noise level, there was no relationship between the N400 and task performance, and the topography and underlying sources of the N400 shifted. This pattern of results suggests that the encoding of speech information in noise may be more robust in musicians. Moreover, when there are high levels of background noise, musicians may increasingly rely on acoustic information to understand speech in noise, whereas nonmusicians rely more heavily on lexical information.

Acknowledgments

We would like to thank Mihaela Felezeu for her assistance with data collection. This research was supported by grants from the La Fondation Caroline Durand, the Grammy Foundation, and the Natural Sciences and Engineering Research Council of Canada.

Reprint requests should be sent to Benjamin Rich Zendel, BRAMS -Suite 0-120, Pavillon 1420 boul. Mont Royal, Université de Montréal, C.P. 6128 - Station Centre ville, Montreal, Quebec, Canada, H3C 3J7, or via e-mail: benjamin.rich.zendel@umontreal.ca.

REFERENCES

- Alain, C. (2007). Breaking the wave: Effects of attention and learning on concurrent sound perception. *Hearing Research*, 229, 225–236.
- Alain, C., Reinke, K., He, Y., Wang, C., & Labaugh, N. (2005). Hearing two things at once: Neurophysiological indices of speech segregation and identification. *Journal of Cognitive Neuroscience*, 17, 811–818.
- Alain, C., Zendel, B. R., Hutka, S., & Bidelman, G. M. (2014). Turning down the noise: The benefit of musical training on the aging auditory brain. *Hearing Research*, 308, 162–173.
- Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: An ERP study. *Language and Cognitive Processes*, 18, 405–430.
- Baumann, S., Meyer, M., & Jäncke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *Journal of Cognitive Neuroscience*, 20, 2238–2249.
- Beauvois, M. W., & Meddis, R. (1997). Time decay of auditory stream biasing. *Perception & Psychophysics*, 59, 81–86.
- Benfante, H., Charbonneau, R., Areseneault, A., Zinger, A., Marti, A., & Champoux, N. (1966). *Audiométrie vocal*. Montréal, Canada: Hôpital Maisonneuve.
- Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, 90, 229–241.
- Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research*, 1355, 112–125.
- Billings, C. J., Bennett, K. O., Molis, M. R., & Leek, M. R. (2011). Cortical encoding of signals in noise: Effects of stimulus type and recording paradigm. *Ear and Hearing*, *32*, 53–60.
- Billings, C. J., Tremblay, K. L., Stecker, G. C., & Tolin, W. M. (2009). Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hearing Research*, 254, 15–24.

- Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound. Cambridge, MA: The MIT Press.
- Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology*, 47, 236–246.
- Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain* and Language, 43, 1–18.
- Deacon, D., & Shelly-Tremblay, J. (2000). How automatically is meaning accessed: A review of the effects of attention on semantic processing. *Frontiers in Bioscience*, *5*, e82–e94.
- Franklin, M. S., Dien, J., Neely, J. H., Huber, E., & Waterson, L. D. (2007). Semantic priming modulates the N400, N300, and N400RP. *Clinical Neurophysiology*, 118, 1053–1068.
- Hari, R., & Makela, J. P. (1988). Modification of neuromagnetic responses of the human auditory cortex by masking sounds. *Experimental Brain Research*, 71, 87–92.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402.
- Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, 123, 155–163.
- Kaplan-Neeman, R., Kishon-Rabin, L., Henkin, Y., & Muchnik, C. (2006). Identification of syllables in noise: Electrophysiological and behavioral correlates. *The Journal of the Acoustical Society of America*, 120, 926.
- Kuriki, S., Kanda, S., & Hirata, Y. (2006). Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *The Journal of Neuroscience*, 26, 4046–4053.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, 9, 920–933.
- Liégeois-Chauvel, C., de Graaf, J. B., Laguitton, V., & Chauvel, P. (1999). Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cerebral Cortex*, 9, 484–496.
- Lukas, J. H. (1981). The role of efferent inhibition in human auditory attention: An examination of the auditory brainstem potentials. *The International Journal of Neuroscience, 12,* 137–145.
- Marsh, J. T., Worden, F. G., & Smith, J. C. (1970). Auditory frequency-following response: Neural or artifact? *Science*, 169, 1222–1223.
- Martin, B. A., Sigal, A., Kurtzberg, D., & Stapells, D. R. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *The Journal of the Acoustical Society of America*, 101, 1585–1599.
- Menendez, R. G. P., Andino, S. G., Lantz, G., Michel, C. M., & Landis, T. (2001). Noninvasive localization of electromagnetic epileptic activity. I. Method descriptions and simulations. *Brain Topography*, 14, 131–137.
- Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *The Journal of the Acoustical Society* of America, 80, 479–483.
- Mummery, C. J., & Ashburner, J. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *The Journal of the Acoustical Society of America*, 106, 449–457.
- Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S., & Hodges, J. R. (2000). A voxel-based morphometry study of semantic dementia: Relationship

between temporal lobe atrophy and semantic memory. *Annals of Neurology*, *47*, 36–45.

Musacchia, G., Sams, M., Nicol, T., & Kraus, N. (2006). Seeing speech affects acoustic information processing in the human brainstem. *Experimental Brain Research*, *168*, 1–10.

Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research*, 241, 34–42.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*, 375–425.

Nahum, M., Nelken, I., & Ahissar, M. (2008). Low-level information and high-level perception: The case of speech in noise. *PLoS Biology, 6,* e126.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, 392, 811–814.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport*, *12*, 169–174.

Parbery-Clark, A., Marmel, F., Bair, J., & Kraus, N. (2011). What subcortical–cortical relationships tell us about processing speech in noise. *The European Journal of Neuroscience*, 33, 549–557.

Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience, 29*, 14100–14107.

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

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

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology, 2,* 1–14.

Patel, A. D. (2012). The OPERA hypothesis: Assumptions and clarifications. Annals of the New York Academy of Sciences, 1252, 124–128.

Pettigrew, C. M., Murdoch, B. E., Ponton, C. W., Kei, J., Chenery, H. J., & Alku, P. (2004). Subtitled videos and mismatch negativity (MMN) investigations of spoken word processing. *Journal of the American Academy of Audiology*, 15, 469–485.

Picard, M. (1984). L'audiométrie vocale au Québec français. *Audiology, 23,* 337–365.

Pugh, K. R., Shaywitz, B. A., Shaywitz, S. E., Constable, R. T., Skudlarski, P., Fulbright, R. K., et al. (1996). Cerebral organization of component processes in reading. *Brain, 119,* 1221–1238.

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

Romei, L., Wambacq, I. J. A., Besing, J., Koehnke, J., & Jerger, J. (2011). Neural indices of spoken word processing in background multi-talker babble. *International Journal of Audiology*, *50*, 321–333.

Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology*, 22, 642–647.

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

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *The Journal of Neuroscience*, 23, 5545–5552.

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *NeuroReport*, 16, 1781–1785.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, 18, 380–393.

Wan, C. Y., & Schlaug, G. (2010). Music making as a tool for promoting brain plasticity across the life span. *The Neuroscientist*, *16*, 566–577.

Warrier, C., Wong, P., Penhune, V., Zatorre, R., Parrish, T., Abrams, D., et al. (2009). Relating structure to function: Heschl's gyrus and acoustic processing. *The Journal of Neuroscience*, 29, 61–69.

Young, E. D., & Sachs, M. B. (1979). Representation of steadystate vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *The Journal of the Acoustical Society of America*, 66, 1381–1403.

Zendel, B. R., & Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience, 21*, 1488–1498.

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

Zendel, B. R., & Alain, C. (2013). The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation. *Journal of Cognitive Neuroscience*, 25, 503–516.

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