Musician Enhancement for Speech-In-Noise

Alexandra Parbery-Clark,1,2 Erika Skoe,1,2 Carrie Lam,1,2 and Nina Kraus1,2,3,4

Objective: To investigate the effect of musical training on speech-in-noise (SIN) performance, a complex task requiring the integration of working memory and stream segmentation as well as the detection of time-varying perceptual cues. Previous research has indicated that, in combination with lifelong experience with musical stream segregation, musicians have better auditory perceptual skills and working memory. It was hypothesized that musicians would benefit from these factors and perform better on speech perception in noise than age-matched nonmusician controls.

Design: The performance of 16 musicians and 15 nonmusicians was compared on clinical measures of speech perception in noise—QuickSIN and Hearing-In-Noise Test (HINT). Working memory capacity and frequency discrimination were also assessed. All participants had normal hearing and were between the ages of 19 and 31 yr. To be categorized as a musician, participants needed to have started musical training before the age of 7 yr, have 10 or more years of consistent musical experience, and have practiced more than three times weekly within the 3 yr before study enrollment. Nonmusicians were categorized by the failure to meet the musician criteria, along with not having received musical training within the 7 yr before the study.

Results: Musicians outperformed the nonmusicians on both QuickSIN and HINT, in addition to having more fine-grained frequency discrimination and better working memory. Years of consistent musical practice correlated positively with QuickSIN, working memory, and frequency discrimination but not HINT. The results also indicate that working memory and frequency discrimination are more important for QuickSIN than for HINT.

Conclusions: Musical experience appears to enhance the ability to hear speech in challenging listening environments. Large group differences were found for QuickSIN, and the results also suggest that this enhancement is derived in part from musicians’ enhanced working memory and frequency discrimination. For HINT, in which performance was not linked to frequency discrimination ability and was only moderately linked to working memory, musicians still performed significantly better than the nonmusicians. The group differences for HINT were evident in the most difficult condition in which the speech and noise were presented from the same location and not spatially segregated. Understanding which cognitive and psychoacoustic factors as well as which lifelong experiences contribute to SIN may lead to more effective remediation programs for clinical populations for whom SIN is therefore an essential aspect of SIN perception. Concurrently presented auditory units may be represented as separate, parallel sensory traces that are not completely independent of each other (Fujikawa et al. 2005, 2008). This not only highlights the auditory system’s ability to represent simultaneously presented auditory units as both separate yet integrated sensory streams, but also support the idea that stream segregation is an active, rather than a passive process.

INTRODUCTION

In our acoustic environment, we are rarely confronted with a single auditory signal; rather, our auditory system must process simultaneously occurring complex acoustic signals to extract relevant information. The canonical example of this is listening to speech-in-noise (SIN), a task requiring a complex set of cognitive and perceptual skills, including stream segregation, auditory working memory, and the detection of time-varying perceptual cues. To extract the target acoustic signal, our auditory system must resolve two issues. First, there must be a process that partitions the acoustic input into separate auditory units. Second, there must be a mechanism for appropriately organizing these acoustic units over time. Auditory scene analysis is the term given to the internal process of segregating and subsequent grouping of an auditory stream (Bregman 1990). Auditory scene analysis is based on the notion that preattentive processes use the Gestalt laws of organization (Koffka 1935)—physical similarity, temporal proximity, and good continuity—to group sounds. In acoustic terms, sounds with similar frequency and spatial location are more likely to be grouped together as auditory units. Indeed, listeners take advantage of both frequency and spatial location cues to assist in the perception of SIN. Perceptual streaming, or the ability to hear two streams, is facilitated when concurrently presented complex tones are separated by as little as one semitone. For example, when asked to identify simultaneously presented vowels, performance improved when the fundamental frequencies were different (Scheffers 1983; Assmann & Summerfield 1990). This phenomenon can help explain why speech perception in noise is more difficult when the target and background speakers are of the same sex, and the fundamental frequencies of the different voices are consequently closer in frequency. Even small frequency differences between speakers’ voices can be used as cues to aid speaker differentiation (Treisman 1964; Broxk & Broxk 1979; Broxk & Nooteboom 1982; Brungart 2001).

Location is another important cue for speech perception in noise. Sounds emanating from different locations are more likely to be heard as distinct auditory units (Bregman 1990). Furthermore, once the location of an auditory unit has been determined, this can be a powerful tool for tracking it over time. For example, the discrimination of target sentences is greatly improved when competing background messages are presented from different locations (Cherry 1953; Bronkhorst & Plomp 1990, 1992; Arborgast et al. 2002). Additionally, in the case of reduced frequency information, such as when the competing and target sentences are both uttered by female talkers, spatial location becomes the primary cue listeners used to track one voice over another (Freyman et al. 1999).

The ability to properly group, represent, and store auditory units over time is fundamental to forming auditory streams and is therefore an essential aspect of SIN perception. Concurrently presented auditory units may be represented as separate, parallel sensory traces that are not completely independent of each other (Fujikawa et al. 2005, 2008). This not only highlights the auditory system's ability to represent simultaneously presented auditory units as both separate yet integrated sensory streams (Fujikawa et al. 2005, 2008) but also support the idea that stream segregation is an active, rather than a passive process.
Musical experience. Sin tests and that this advantage is mediated by their listening situations such as Sin. Therefore, we hypothesized that musicians would perform better on clinically relevant tasks. Alternatively, because perceptual cues are also important for auditory stream segregation, listeners with better auditory perceptual skills may detect more subtle acoustic cues and thereby improve their ability to separate and group the target signal from the background noise. A clearer representation of the acoustic stream would in turn reduce the attentional demands, leaving more resources available for the rehearsal and recall of the target words.

Musicians spend hours attending to and manipulating complex auditory signals that comprise multiple streams. In addition to processing concurrent auditory units (i.e., simultaneously occurring melodies), musicians must also analyze the vertical relationships between streams (i.e., harmony). In addition to this online auditory scene analysis musicians also hone their abilities to conceive, plan, and perform music in real time. Previous work has documented that musical training improves basic auditory perceptual skills resulting in enhanced behavioral (Jeon & Fricke 1997; Koelsch et al. 1999; Oxenham et al. 2003; Tervaniemi et al. 2005; Micheyel et al. 2006; Rammayer & Altenmüller 2006) and neurophysiological responses (Brattico et al. 2001; Pantev et al. 2001; Schneider et al. 2002; Shahin et al. 2003, 2007; Trainor et al. 2003; Tervaniemi et al. 2005; Kuriki et al. 2006; Kraus et al. 2009). Moreover, it would seem that musicians are able to use these perceptual benefits to facilitate concurrent sound segregation (Zendel & Alain 2009). Musical training not only enhances aspects that are specific to musical perception, but these enhancements also cross over to other domains, particularly language, suggesting shared neural resources for language and music processing (Patel 2003, 2007; Kraus & Banaï 2007; Koelsch et al. 2008; Steinbeis & Koelsch 2008a,b). For example, lifelong musical experience is linked to improved subcortical and cortical representations of acoustic features important for speech encoding and vocal communication (Magne et al. 2003, 2006; Schon et al. 2004; Marques et al. 2007; Musacchia et al. 2007, 2008; Wong et al. 2007; Chandrasekaran et al. 2008; Moreno et al. 2008; Strait et al. 2009).

Likewise, musical experience has been shown to improve verbal ability (Forgerd et al. 2008), verbal working memory, and verbal recall (Chan et al. 1998; Brandler & Rammayer 2003; Ho et al. 2003; Jackobsen et al. 2003). This study brings these ideas together. As a combined consequence of their extensive experience with auditory stream analysis within the context of music, more honed auditory perceptual skills as well as greater working memory capacity, musicians seem well equipped to cope with the demands of adverse listening situations such as Sin. Therefore, we hypothesized that musicians would perform better on clinically relevant Sin tests and that this advantage is mediated by their musical experience.

### Participants and Methods

#### Participants

Participants consisted of 31 right-handed native English-speaking adults (mean age 23 ± 3 yr, 19 women) with normal hearing (≤15 dB HL pure-tone thresholds from 250 to 8000 Hz) and no history of neurological disorders. Participants were also screened for normal intelligence using the Test of Nonverbal Intelligence-3 (Brown et al. 1997). All participants gave their informed consent before participating in this study in accordance with the Northwestern University Institutional Review Board regulations. Participants categorized as musicians (N = 16) were self-identified, began playing an instrument before the age of 7 yr, had 10 or more years of musical experience, and had continued to practice consistently three times a week within the 3 yr before participation in the study. Nonmusicians (N = 15) were categorized by the failure to meet the musician criteria, along with not having received musical training within the 7 yr before the study (Table 1).

<table>
<thead>
<tr>
<th>Musician</th>
<th>Years of training</th>
<th>Age onset, yr</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17</td>
<td>3</td>
<td>Violin</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>3</td>
<td>Violin</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>5</td>
<td>Violin</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
<td>5</td>
<td>Violin</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>7</td>
<td>Violin</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
<td>3</td>
<td>Piano</td>
</tr>
<tr>
<td>7</td>
<td>19</td>
<td>3</td>
<td>Piano</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>3</td>
<td>Piano</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>4</td>
<td>Piano</td>
</tr>
<tr>
<td>10</td>
<td>14</td>
<td>5</td>
<td>Piano</td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>7</td>
<td>Piano</td>
</tr>
<tr>
<td>12</td>
<td>17</td>
<td>5</td>
<td>Piano</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>5</td>
<td>Piano</td>
</tr>
<tr>
<td>14</td>
<td>10</td>
<td>6</td>
<td>Piano</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>5</td>
<td>Piano/bassoon</td>
</tr>
<tr>
<td>16</td>
<td>18</td>
<td>6</td>
<td>Piano/voice</td>
</tr>
<tr>
<td>Mean</td>
<td>16</td>
<td>4.68</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nonmusician</th>
<th>Years of training</th>
<th>Age onset, yr</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>3</td>
<td>9</td>
<td>Piano</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>13</td>
<td>Piano</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>11</td>
<td>Flute</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>9</td>
<td>Recorder</td>
</tr>
<tr>
<td>21</td>
<td>2</td>
<td>12</td>
<td>Trumpet</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>28</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>29</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>31</td>
<td>0</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>Mean</td>
<td>2</td>
<td>10.8</td>
<td></td>
</tr>
</tbody>
</table>

### Notes

- Years of musical training, age at which training began, and major instrument are indicated for all participants with musical experience. Means for years of musical training and age at onset for the nonmusicians were calculated from the five participants who had musical experience.
Materials and Procedures

Speech-in-noise

Hearing-In-Noise Test • The Hearing In Noise Test (HINT; Biologic Systems Corp., Mundelein, IL; Nilsson et al. 1994) is an adaptive test of speech recognition that measures speech-perception ability in speech-shaped noise. During the HINT, the participant is required to repeat short semantically and syntactically simple sentences (e.g., she stood near the window) presented in a speech-shaped background noise that matches the spectra of the test sentences. The speech stimuli consist of Bamford–Kowal–Bench (1979) sentences (12 lists of 20 sentences) spoken by a man and presented in free field via two loud speakers positioned at a 90° angle to each other. The participant sat 1-m equidistant from the two loudspeakers, and the target sentences originate from a source location at 0° (directly in front of the participant) for each condition. There are three SIN conditions that differ in the location of the noise: noise delivered from 0° (HINT-F), 90° to the right (HINT-R), and 90° to the left (HINT-L). The noise presentation level was fixed at 65 dB SPL. The program adjusted the difficulty by increasing or decreasing the intensity level of the target sentences until the threshold signal to noise ratio (SNR)—defined as the difference, in decibels (dB), between the speech and noise presentation levels for which 50% of sentences are correctly repeated—was determined. Three threshold SNRs were calculated, one for each of the three noise conditions. A more negative SNR indicates a greater ability to perceive speech in more adverse listening conditions.

QuickSIN • The Quick Speech-In-Noise test (Etymotic Research, Elk Grove, IL; Killion et al. 2004), a nonadaptive test of speech perception in four-talker babble (three women and one man), was presented binaurally to participants through insert earphones (ER-2, Etymotic Research, Elk Grove Village, IL). Four lists were selected from a corpus of 20 with each list consisting of six sentences with five target words per sentence. Sentences were presented at 70 dB SPL, with the first sentence starting at an SNR of 25 dB and each subsequent sentence being presented with a -5 dB SNR reduction down to 0 dB SNR. The sentences are syntactically correct yet do not contain many semantic or contextual cues (Wilson et al. 2007). Participants were instructed to repeat back each sentence, and their SNR loss was based on the number of target words correctly recalled. Sample sentences, with target words italicized, included “The square peg will settle in the round hole,” and “The sense of smell is better than that of touch.” The total number of key words correctly recalled in the list (30 in total) was subtracted from 25.5 to give the final SNR loss (see Killion et al. 2004 and QuickSIN User’s Manual [Etymotic Research 2001] for further details). The final score is the average SNR loss scores from the four lists. As with the HINT scores, a more negative SNR loss value is indicative of better performance.

Working Memory: Woodcock-Johnson test • The Woodcock-Johnson III Cognitive test (Woodcock et al. 2001) was administered to all participants to measure working memory ability. The overall working memory score was composed of scores from the Numbers Reversed and Auditory Working Memory subtests, both of which required participants to process and reorder aurally presented information stored in their short-term memory.

Numbers Reversed • In the Numbers Reversed test, participants repeated a sequence of numbers in a backward order. The most difficult item contained eight digits, that is, 9, 6, 1, 3, 7, 4, 5, 2 which backward would be 2, 5, 4, 7, 3, 1, 6, 9.

Auditory Working Memory • In the Auditory Working Memory test, participants reordered a dictated series containing digits and words, by first repeating the words in sequential order and then the digits in sequential order. For example, the correct ordering of the following sequence 4, salt, fox, stove, boot and 4, 7, 2, 9.

Auditory acuity: Psychophysical frequency discrimination test • An adaptive staircase method was used to determine the frequency discrimination detection threshold (79%) (Levitt 1971). Pure tones, 250 msec in duration, were presented binaurally at 70 dB SPL through headphones via the Cogent toolbox controlled by Matlab 7.0 (Mathworks, Natick, MA). In each trial, two stimuli were presented, the standard tone (1000 Hz) and a variable tone ranging between 1002 and 1600 Hz; the participant indicated which tone was higher in frequency. The order of standard and variable tones was randomized. Ten practice trials were presented, and then the participant completed the discrimination task three times to determine an average threshold.

Statistical Methods

All statistical analyses were conducted with SPSS (SPSS Inc., Chicago, IL). The Levene test for homogeneity revealed homogeneity of variance for QuickSIN and HINT scores, but not for working memory ability or frequency discrimination thresholds. Therefore, the nonparametric Mann-Whitney U test was used for all between-group comparisons for working memory and frequency discrimination. For all other group comparisons, one-way multivariate analyses of variance were conducted, and the assumptions for normality, linearity, outliers, and multicollinearity were met. To determine the effects of working memory and frequency discrimination on SIN performance, a series of multiple hierarchical regressions was performed, and the adjusted $R^2$ values were reported. Because we had 31 participants, we limited the independent predictor variables to only two (one predictor for every 15 participants; Stevens 1996). We were also interested in investigating the role of musical training on working memory, frequency discrimination, and SIN performance. The relationships between these variables were examined with a set of Pearson $r$ correlations using all participants regardless of group ($N = 31$) and only the musician group ($N = 16$). In interpreting the results, the $\alpha$ level for the correlations was corrected for multiple comparisons.

RESULTS

There were no significant group differences (musicians versus nonmusicians) in age, pure-tone averages, or nonverbal I.Q. (one-way multivariate analyses of variance, all $p > 0.3$). Previous research has indicated that women perform better in SIN (Dubno et al. 1997) and verbal tasks (Mann et al. 1990); however, our participants did not show sex differences for the SIN test, working memory ability, or frequency discrimination.
Behavioral Differences Between Musicians and Nonmusicians

The musician group performed better with both QuickSIN (F[1,29] = 9.776, p = 0.004) and HINT-F (F[1,29] = 8.170, p = 0.008), meaning that musicians were able to repeat sentences presented at a lower, more challenging SNR than nonmusicians. On the other hand, both groups performed equivalently when the speech and the noise were presented from separate spatial locations in HINT-L and HINT-R (F[1,29] = 0.235, p = 0.632 and F[1,29] = 2.015, p = 0.166, respectively). Musicians also demonstrated better perceptual skills as evidenced by smaller frequency discrimination thresholds (U = 36, p = 0.001) and greater working memory capacity (U = 47.5, p = 0.004), see Figure 1.

Relationships Among Behavioral Measures, SIN, and Musical Experience

Surprisingly, HINT-F and QuickSIN scores were not correlated (r = 0.298, p = 0.104), suggesting that they are not measuring the same skills. Working memory scores (Fig. 2) correlated strongly with QuickSIN performance (r = −0.578, p < 0.001) and moderately with HINT-F (r = −0.369, p = 0.041). Moreover, frequency discrimination was highly correlated with QuickSIN (r = 0.511, p = 0.003) but not with HINT-F (r = −0.155, p = 0.404; Fig. 4).

Across all individuals, years of consistent practice (Fig. 3) correlated with QuickSIN (r = −0.580, p = 0.001) and working memory (r = 0.614, p < 0.0005) scores (Fig. 4). However, duration of training did not predict HINT-L or HINT-R scores (both p > 0.1). When considering only the musician group (N = 16), years of consistent musical practice correlated with QuickSIN (r = −0.579, p = 0.019) and working memory ability (r = 0.494, p = 0.052). In the case of HINT-F, although there was an overall group correlation between years practiced and HINT-F scores (r = 0.494, p = 0.005), this result was driven by the five nonmusicians with limited musical training; the correlation was not significant when restricted to the musician group then (r = −0.259, p = 0.333). Again, no relationship was found for years practiced and HINT-L or HINT-R scores within the musician group (both p > 0.1). These results suggest that while improved QuickSIN and working memory scores are related to the number of years practiced, the musician advantage seen for HINT-F is reliant on other factors.

A series of hierarchical multiple regressions were performed to determine the effects of the behavioral measures and musical experience on SIN performance. Across all partici-
both of which showed a musician advantage. A greater extent by working memory and frequency discrimination ability, working memory correlated highly with QuickSIN and moderately with frequency discrimination correlated with QuickSIN but not with HINT-F. These results suggest that performance on QuickSIN is mediated by working memory, frequency discrimination, and the speech-in-noise tests, not reported in this figure. When considering the relationship among working memory, frequency discrimination, and the SIN measures. The correlation coefficients are based on all participants (N = 31). The results imply that the more years a person practices a musical instrument, the better his or her working memory ability (r = 0.614, p < 0.001). When considering just the musician group, years of consistent practice correlated with QuickSIN (r = −0.579, p = 0.019) and working memory (r = 0.494, p = 0.052). The total variance explained by the five nonmusicians with limited musical training and, hence, not reported in this figure. When considering the relationship among frequency discrimination, working memory, and the speech-in-noise tests, frequency discrimination correlated with QuickSIN but not with HINT-F. Working memory correlated highly with QuickSIN and moderately with HINT-F. These results suggest that performance on QuickSIN is mediated to a greater extent by working memory and frequency discrimination ability, both of which showed a musician advantage. **, p < 0.01; *, p < 0.05.

DISCUSSION

In line with our hypothesis, musicians performed better on standardized measures of speech perception in noise. Our results are consistent with previous behavioral and neurophysiological studies showing that skills honed through musical training transfer to speech (Anvari et al. 2002; Magne et al. 2003, 2006; Schon et al. 2004; Besson et al. 2007; Marques et al. 2007; Musacchia et al. 2007, 2008; Wong et al. 2007; Chandrasekaran et al. 2008). Thus, our findings are also in accordance with the growing body of literature supporting the existence of shared resources for music and language processing (Patel 2003, 2007; Koelsch et al. 2005, 2008; Kraus & Banai 2007; Steinbeis & Koelsch 2008a,b). In addition, we demonstrate that musicians have superior working memory skills and that this is a significant driving force behind the group’s SIN performance.

Working Memory and SIN

Although musicians performed better on both QuickSIN and HINT-F tasks, QuickSIN scores were more highly correlated to working memory ability than HINT-F. This difference may be because of the different demands each test places on auditory working memory. Both QuickSIN and HINT require the repetition of sentences, which is more cognitively demanding in terms of working memory than remembering single words or digits (McArdle et al. 2005). In a comparative study of SIN tests, HINT was found to be easier than QuickSIN and provided less differentiation between normal-hearing and hearing-impaired groups (Wilson et al. 2007). The HINT has shorter sentences (mean 5.3 words), and the vocabulary used is appropriate of a first-grade level (e.g., It’s time to go to bed). In contrast, QuickSIN uses longer, less semantically predictable sentences (mean 8.6 words) with more advanced vocabulary (e.g., A cruise in warm waters in a sleek yacht is fun). By increasing the semantic load and sentence length, the listener is forced to rely more on acoustic cues and working memory. In line with the differing sentence complexity between the tests, we found a stronger relationship between working memory and QuickSIN than working memory and HINT-F; this supports the idea of working memory having a significant contribution to enhanced SIN abilities in musicians especially when the SIN test uses longer more complex sentences.

Previous research highlights the importance of working memory for language comprehension even in quiet settings (Daneman & Merikle 1996; Walters & Caplan 2005; Wingfield & Tun 2007) and suggests that a person’s working memory span is reduced in challenging listening environments (Rabbit 1968; Pichora-Fuller & Souza 2003). In the case of SIN, the presence of background noise may also increase the attentional load, resulting in fewer resources being available for the rehearsal and recollection of target words (Heinrich et al. 2008). The Reverse Hierarchy Theory, developed originally to explain visual processing, has recently been applied to SIN perception (Nahum et al. 2008). This model suggests that as the SIN task becomes more difficult perception becomes more reliant on low-level acoustic information. However, this lower level acoustic information is only accessible to higher-order
cognitive processes by way of a backward (top-down) search, which impedes the concurrent perception of the ongoing auditory stream. Interpreting our results within the framework of the Reverse Hierarchy Theory model, it is plausible that having better working memory skills might offset the disruptive backward search, resulting in heightened SIN performance. Alternatively, if musicians have more distinct acoustic representations, the need for backward searches is reduced, allowing musicians to focus on the higher-level representations that are related to comprehension.

Perceptual anchoring (Ahissar 2007) is another relevant theory that can be invoked to explain musicians’ better SIN perception. Perceptual anchors are stable internal references that are formed in response to a repeated stimulus. For example, during frequency discrimination tasks, if the reference tone is invariant, better thresholds are obtained than when the reference tone is constantly roving (Harris 1948). One argument put forth for the perceptual benefit is that an invariant reference tone gives rise to the formation of a strong perceptual anchor against which the test tone can be compared, thus resulting in improved perceptual discrimination (Harris 1948; Ahissar et al. 2006; Ahissar 2007). The ability to create a perceptual anchor of the speaker’s voice is considered a key to improve signal perception (Best et al. 2008). Dyslexics have been found to have difficulty in creating perceptual anchors, and this may underlie their known working memory (Briscoe & Rankin 2008; Ramus & Szenkovits 2008) and SIN performance deficits (Ahissar et al. 2006; Ahissar 2007; Hornickel et al. 2009). Therefore, it may be the case that musicians are better at creating perceptual anchors which places them at the other end of the auditory processing spectrum from dyslexics.

**QuickSIN versus HINT**

Our study shows that working memory is not the only contributor to SIN performance. QuickSIN and HINT-F, although having different presentation parameters (multitalker babble versus speech-shaped noise and earphones versus speakers, respectively), do share the common characteristic of the speech target and background noise originating from the same source. For the conditions in which the noise and the target were spatially separated (HINT-L and HINT-R), the two groups performed equally, suggesting that both benefited from the spatial cues. However, when the target speech and the background noise came from the same location, such that spatial cues were no longer available, there was a separation in group performance. The superior performance on both QuickSIN and HINT-F suggests that the musicians were able to use acoustic cues other than localization for stream segregation.

Frequency discrimination ability was found to be an important predictor of performance on QuickSIN. Frequency discrimination thresholds can be considered indicative of more honed auditory perception, and it is well documented that musicians are better at detecting frequency, timbre, and timing differences (Jeon & Fricke 1997; Koelsch et al. 1999; Tervaniemi et al. 2005; Michely et al. 2006; Rammsayer & Altenmüller 2006). As would be suggested by previous work (Koelsch et al. 1999; Zendel & Alain 2009), musicians may pay more attention to different acoustic cues or may be better at detecting subtle acoustic differences, which could improve their faculty for segregating and grouping concurrent sounds. As a result of their improved ability for auditory stream analysis, musicians may not have to allocate as many resources to attending to “hearing out” words, resulting in more resources being available for rehearsal (working memory).

In this study, frequency discrimination was found to correlate with QuickSIN but not with HINT-F. A possible explanation is that the two tests use different types of background noise. HINT-F uses a speech-shaped noise, which matches the spectra of the target sentences and acts as an energetic masker. This type of noise has a flat envelope that maintains a fairly consistent SNR across a given sentence and so for HINT, frequency cues do not help to parse out the noise. In contrast, multitalker babble, which is used in the QuickSIN test, is composed of different speakers with unique continuous frequency properties. Furthermore, the target and the multitalker babble have energy and spectral composition that are comparable only in the long-term average, whereas on the short term both the energy and the spectra may fluctuate significantly. Therefore, in the case of QuickSIN, both the continuous frequency properties of the speakers (target and competing voices) as well as the energy and spectrum fluctuations of the multitalker babble could be used by the listener to parse and track the target voice effectively.

Another important consideration is that of the temporal envelope characteristics of the different background noises used by HINT and QuickSIN. The temporal envelope of the different speakers within the multitalker babble may provide cues for segregating speakers and thus aid the tracking of the target voice. Furthermore, the natural dips in the temporal envelopes may provide a listener with the opportunity to “listen in the valleys” (Buus 1985). Indeed, when confronted with one or two competing talkers, normal-hearing listeners are able to use gaps in the masking noise to distinguish the target speaker, a phenomenon known as glimpsing, although this benefit is reduced when more talkers are added to the babble (Miller 1947; Carhart et al. 1969). The QuickSIN babble noise comprised four talkers that would provide more masking and less gaps in the temporal envelope than one or two competing talker babble, but listeners may still be able to glean some benefit from the time-varying cues of the background babble speech envelope and the target talkers. On the other hand, HINT uses a flat-envelope spectrum noise in which the opportunity for glimpsing is considerably reduced, thus resulting in a greater energetic masking effect on the target speech signal at the level of the periphery. This may also be a contributing factor to why the HINT scores did not demonstrate a strong relationship with the higher-level cognitive measure of working memory. However, as discussed earlier, QuickSIN and HINT also place different demands on auditory working memory because of the differing complexity of the sentences. Therefore, to elucidate the relationship among musical training, working memory, and the type of background maskers, future research comparing different types of maskers should use target sentences equated for length and semantic predictability.

**Clinical Implications and Future Directions**

There is already a vast literature, and a general consensus, that some populations have more difficulty comprehending SIN than others. For example, although speech perception in noise can be difficult for normal-hearing adults, the deleterious effects of noise on speech comprehension are amplified in
young children (Hetu et al. 1990), children with learning disorders (Bradlow et al. 2003; Ziegler et al. 2005; Cunningham et al. 2001; Hornickel et al. 2009), people with hearing loss (Gordon-Salant & Fitzgibbons 1995), non-native speakers (Mayo et al. 1997; Bradlow & Alexander 2007), and older adults (Pichora-Fuller et al. 1995; Pichora-Fuller 2006; Shinn-Cunningham & Best 2008). Research to date has focused either on populations who experience difficulty with SIN or investigated this ability across the life span by comparing young and old participants. Indeed, the results of these studies are often contradictory, with some researchers finding that hearing thresholds and SIN performance are correlated, whereas others find the opposite (Dubno et al. 1984; Harris & Sweson 1990; Killion & Niquette 2000; Humes 2007; Souza et al. 2007). This mismatch between pure-tone audiometry and SIN test results may occur because SIN performance depends on other factors such as age, cognitive ability, and perceptual acuity in addition to hearing thresholds (Pichora-Fuller 2006; Oxenham 2008), a viewpoint also supported by our data. Thus, by looking at a young, nonclinical population who demonstrate enhanced perception of SIN, we are able to examine this skill from a different angle, which may ultimately lead to greater insight into speech perception difficulties in clinical populations. Likewise, understanding the elements that contribute positively to SIN may lead to better, more effective remediation programs for impaired populations.

SIN tests are designed to measure how well a person copes with competing background noise. Words- and digits-in-noise tests may provide a benchmark of SIN performance without context, but they are not as ecologically valid as the sentence material used here. By using sentences, tests such as QuickSIN and HINT draw on linguistic knowledge and experience (native, non-native, and bilingual) as well as cognitive abilities. Our results imply that QuickSIN measures both SIN and working memory abilities. Most notably, we found that musically trained participants had better QuickSIN scores, which could be largely attributable to heightened working memory. By extension, our results underscore the important roles that external factors such as musical experience play when interpreting SIN test outcomes and suggest the need for normative ranges for different populations.

The results of this study do not speak to which aspects of musical training enhance the working memory and SIN results. Our participants were highly trained musicians with many years of experience. Future work is needed to determine which aspects of musical training are important for improving SIN performance, the time course of the improvements, and the transfer effects. With this information, an auditory training program using music-related tasks could be envisioned to enhance SIN performance. An additional question concerns the effect of instrument played. For example is the effect of musical training enhancing SIN equivalent with monophonic (e.g., flute) or polyphonic (e.g., piano) instruments? Likewise, does instrumental ensemble work also contribute? One of the limitations of this study is that enrollment of participants was restricted to those musicians who had started their instrument before the age of 7 yr, resulting in the exclusion of musicians who played instruments that are not typically started until later in development such as the tuba, trombone, and clarinet. Further studies are needed to investigate whether age started is also a contributing factor or whether benefits can be seen with years of consistent practice later in life.

CONCLUSION

The results from this study suggest that musical experience enhances the ability to hear speech in challenging listening environments. SIN performance is a complex task requiring perceptual cue detection, stream segmentation, and working memory. Musicians performed better than nonmusicians in conditions where the target and the background noise were presented from the same source, meaning parsing was more reliant on the acoustic cues present in the stream. We also found a strong relationship between working memory ability and QuickSIN, which were both correlated with years of musical practice, unlike HINT-F. Although we cannot determine causality, the working memory enhancement of musicians seems to mediate their better performance in QuickSIN. These results provide further evidence for musical training transferring to nonmusical domains and highlight the importance of taking musical training into consideration when evaluating SIN ability.

ACKNOWLEDGMENTS

The authors thank Prof. Steven Zecker for his advice on appropriate statistical treatment of these data, as well as Trent Nicol and Dr. Susan Erler for their insightful comments on an earlier version of this manuscript. They also thank the people who participated in this study.

This work was supported by NSF SGER 0842376.

Address for correspondence: Alexandra Parbery-Clark, Northwestern University, Auditory Neuroscience Laboratory (www.brainvolts.northwestern.edu), 2240 Campus Drive, Evanston, IL 60208-2952. E-mail: apc@u.northwestern.edu.

Received January 2, 2009; accepted June 19, 2009.

REFERENCES


