What subcortical–cortical relationships tell us about processing speech in noise

Alexandra Parbery-Clark,¹,² Frederic Marmel,¹ Julia Bai³ and Nina Kraus¹,²,⁴,⁵
¹Auditory Neuroscience Laboratory, Frances Searle Building, 2240 Campus Drive, Evanston, IL 60208-2952, USA
²Department of Communication Sciences, Northwestern University, Evanston, IL, USA
³Department of Neuroscience, Oberlin College, Oberlin, OH, USA
⁴Department of Neurobiology and Physiology, Northwestern University, Evanston, IL, USA
⁵Department of Otolaryngology, Northwestern University, Evanston, IL, USA

Keywords: auditory brainstem response, cortical auditory-evoked potentials, N1, speech in noise

Abstract

To advance our understanding of the biological basis of speech-in-noise perception, we investigated the effects of background noise on both subcortical- and cortical-evoked responses, and the relationships between them, in normal hearing young adults. The addition of background noise modulated subcortical and cortical response morphology. In noise, subcortical responses were later, smaller in amplitude and demonstrated decreased neural precision in encoding the speech sound. Cortical responses were also delayed by noise, yet the amplitudes of the major peaks (N1, P2) were affected differently, with N1 increasing and P2 decreasing. Relationships between neural measures and speech-in-noise ability were identified, with earlier subcortical responses, higher subcortical response fidelity and greater cortical N1 response magnitude all relating to better speech-in-noise perception. Furthermore, it was only with the addition of background noise that relationships between subcortical and cortical encoding of speech and the behavioral measures of speech in noise emerged. Results illustrate that human brainstem responses and N1 cortical response amplitude reflect coordinated processes with regards to the perception of speech in noise, thereby acting as a functional index of speech-in-noise perception.

Introduction

Background noise infiltrates everyday communication, requiring our auditory system to focus on relevant information while suppressing irrelevant sounds. Successful perception of speech in noise is dependent on cognitive factors as well as sound processing at peripheral, subcortical and cortical levels, making it one of the most complex aspects of human communication (Shinn-Cunningham & Best, 2008; Kujala & Brattico, 2009). In the present study, we focus on the relationships between subcortical and cortical auditory processing, as well as perception of speech in noise in normal hearing young adults.

It is generally accepted that sound processing relies on the auditory system’s ability to represent acoustic information in both a fine-grained and integrated manner (Langner, 1992; Winer et al., 2005; Chechik et al., 2006; Nelken & Ahissar, 2006). Subcortical encoding, as reflected by the scalp-recorded auditory brainstem response (hereafter ABR or brainstem response), provides a fine-grained representation of the acoustic information. The ABR arises from transient and phase-locked neural activity within subcortical nuclei (reviewed in Chandrasekaran & Kraus, 2010); the ABR also represents stimulus features with high fidelity in that the neural response physically resembles the acoustic characteristics of the stimulus (Galbraith et al., 1995; Kraus & Nicol, 2005). The addition of background noise, however, degrades the ABR, resulting in delayed response timing, diminished response amplitude and reduced subcortical response fidelity to the stimulus (Don & Eggermont, 1978; Burkard & Hecox, 1983; Cunningham et al., 2001; Oates & Purdy, 2001; Russo et al., 2004; Parbery-Clark et al., 2009a; Anderson & Kraus, 2010; Anderson et al., 2010; Song et al., in press). The degree of degradation caused by the addition of noise on the brainstem response to speech relates to speech perception in noise, with earlier and more robust subcortical responses correlating with better perception of speech in noise (Cunningham et al., 2001; Parbery-Clark et al., 2009a; Anderson et al., 2010).

Compared with ABRs, cortical-evoked responses represent stimulus properties in a more abstract fashion (Las et al., 2005; Hickok & Poeppel, 2007). Background noise prolongs cortical response timing but its effect on cortical response magnitude is less clear, with studies reporting N1–P2 amplitude decrease (Whiting et al., 1998; Androulidakis & Jones, 2006; Billings et al., 2009), increase (Alain et al., 2009) or both (Kaplan-Neeman et al., 2006). The N1 response is primarily considered to be an onset response generated by onset, pause and offset neurons within the primary auditory cortex. The N1 response has both exogenous and endogenous properties, reflecting
the presence of an audible stimulus (Naätänen & Picton, 1987) and becoming more pronounced with attention (Hillyard et al., 1973; Hall, 2007). To better understand the neural bases underlying the perception of speech in noise, we recorded subcortical and cortical responses to speech with and without competing background noise. We sought to delineate the relationships between subcortical and cortical encoding of speech, and to determine how background noise modulates these relationships in normal hearing young adults. We were particularly interested in the relationships between neural and behavioral measures of speech in noise. Delineation of these relationships in the normal system may serve as a framework within which to view patterns of disruption in clinical populations who have excessive difficulties with speech perception in noise.

Materials and methods
Participants
Twenty-two adults (age range 19–34 years; 14 females) participated in this study. All participants were right-handed, had binaural hearing thresholds (air and bone thresholds) ≤ 20 dB HL at octave frequencies from 125 to 8000 Hz and had normal ABRs to a click and speech syllable as measured by BioMARK (Biological Marker of Auditory Processing; Natus Medical, Mundelein, IL, USA). Participants reported no cognitive or neurological deficits. The research protocol was approved by the Institutional Review Board of Northwestern University. Informed consent was obtained for all evaluation procedures, and participants were compensated for their participation in the study.

Stimuli
The speech syllable /da/ was a 170-ms five-formant speech sound synthesized using a Klatt synthesizer (Klatt, 1980) at a 20-kHz sampling rate. Except for the initial 5-ms stop burst, this syllable is voiced throughout with a steady fundamental frequency \( f_0 = 100 \) Hz. This consonant–vowel syllable is characterized by a 50-ms formant transition period (transition between the stop burst and the vowel) followed by a 120-ms steady-state (unchanging formants) period corresponding to /a/ (Fig. 1, top). During the formant transition transition...
period, the first formant rises linearly from 400 to 720 Hz, while the second and third formants fall linearly from 1700 to 1240 Hz and 2580 to 2500 Hz, respectively. The fourth, fifth and sixth formants remain constant at 3330, 3750 and 4900 Hz for the entire syllable. We chose to investigate the neural representation of /da/ as previous research indicates that the perception of consonants, particularly stop consonants, is compromised in background noise (Brandt & Rosen, 1980).

The background noise consisted of multi-talker babble created by the superimposition of nonsensical sentences spoken by six different speakers (two males and four females). These sentences were recorded for a previous experiment and the specific recording parameters can be found in Smiljanic & Bradlow (2005). The noise file was 45 s in duration. Multi-talker babble was chosen as the competing background noise because of its known deleterious effect on speech perception (Sperry et al., 1997) and because multi-talker babble results in different neural activation patterns relative to white noise maskers (Scott et al., 2004; Kozou et al., 2005).

Procedure

Subcortical responses

The speech syllable /da/ was presented binaurally through insert earphones (ER-3; Etymotic Research, Elk Grove Village, IL, USA) via the stimulus presentation software NEUROSCAN STIM2 (Compumedics, Charlotte, NC, USA) at 80 dB SPL with an inter-stimulus interval (ISI) of 83 ms. The speech syllable was presented in alternating polarities, a technique commonly used in brainstem recordings to minimize the contribution of stimulus artifact and cochlear microphonic. Because the stimulus artifact and cochlear microphonic follow the phase of the stimulus, when the responses to alternating polarities are added together the artifacts are reduced, leaving the phase-invariant component of the response intact (see Skoe & Kraus, 2010 for more details). To change a stimulus from one polarity to another, the stimulus waveform was inverted by 180°. In the noise condition, the /da/ was presented at a +10 dB signal-to-noise ratio (SNR) relative to the multi-talker babble, which was looped for the duration of the condition. This SNR was chosen because in noisy environments a person typically projects 10 dB above the background noise (Pearsons et al., 1977). Furthermore, pilot testing indicated that lower SNRs (more difficult) resulted in greatly reduced onset and transition peaks, thus complicating the quantification of background noise on the neural encoding of the onset and consonant–vowel transition.

Brainstem responses were collected using NEUROSCAN ACQUIRE 4.3 (Compumedics). A vertical electrode montage, consisting of four Ag–AgCl electrodes placed at Cz (active), forehead (ground) and both earlobes (linked-references), was used. Contact impedance for all
electrodes was < 5 kΩ. The continuous file was offline bandpass filtered from 70 to 2000 Hz (12 dB/octave, zero phase-shift filters). The continuous file was epoched (40 ms of pre-stimulus activity and 213 ms of post-stimulus activity), and an artifact rejection criterion of ±35 μV was applied. Six thousand artifact-free sweeps were recorded for each condition, half of each polarity, which were then added together. Each condition lasted between 23 and 25 min. We employed a traditional passive-listening paradigm in which participants watched a movie of their choice to facilitate a restful but wakeful state. Because the stimulus was presented binaurally, the soundtrack was muted and subjects followed the movie by way of closed-captions.

Cortical responses
The same hardware, software and earphones were used for the collection of cortical responses. Stimulus presentation was identical, except that the ISI was 863 ms. Cortical responses were collected with a 31-channel tin electrode cap (Electrocap International, Eaton, OH, USA). Additional electrodes were placed on the earlobes and superior and outer canthi of the left eye, thereby acting as reference and eye-blink monitors, respectively. Contact impedance for all electrodes was < 5 kΩ. To isolate the cortical contributions, the continuous file was offline bandpass filtered from 1 to 20 Hz (12 dB/octave, zero phase-shift filters). The removal of eye-blink artifacts was conducted using the NEUROSCAN EDIT 4.3 spatial filtering algorithm. The continuous file was epoched (100 ms of pre-stimulus activity and 500 ms of post-stimulus activity), and an artifact rejection criterion of ±50 μV was applied. Lastly, the responses were subjected to a noise reduction algorithm described in detail in Abrams et al. (2008). Briefly, the algorithm computes the degree of similarity between each epoch and the average of all epochs using Pearson’s correlations. The sweeps are ranked according to their Pearson r-values, and the bottom (lowest) 30% of r-scores are discarded. The remaining 70% epochs are averaged and re-referenced to a common average reference computed across all electrodes excluding the eye-blink channels. Six-hundred and fifty artifact-free sweeps were collected for each condition, with each condition lasting between 10 and 12 min. After processing, an individual’s averaged response was formed from approximately 450 (±30) sweeps.

Behavioral component – speech in noise
The Hearing in Noise Test (HINT; Bio-logic Systems; Nilsson et al., 1994) is an adaptive test that measures speech perception ability in speech-shaped noise. For the HINT, participants repeated short semantically and syntactically simple sentences spoken by a male (e.g., ‘She stood near the window’). Participants sat 1 m away from the speaker from which both the target sentences and the background noise were delivered. The noise presentation level was fixed at 65 dB SPL, with the target sentence intensity level adaptively increasing or decreasing depending on the individual’s performance. A final threshold SNR, defined as the difference in dB between the speech and noise presentation levels at which 50% of sentences are correctly repeated, was calculated. Thus, lower SNRs indicate better performance on the task.

Analysis
Subcortical response – timing and amplitude
The neural response to the onset of the stimulus (onset peak) and the formant transition (transition peak) are represented by large positive peaks occurring between 9–11 ms and 43–45 ms post-stimulus onset (0 ms), respectively (Fig. 1, peaks 1 and 2). The onset and transition peaks were defined as the largest peak before the trough within these time ranges. These peaks were identified independently by the first author and a second rater. To facilitate peak picking, each judge used grand averages along with a comparison of an individual’s response in the quiet and noise conditions. In the case of disagreement over peak identification, the advice of a third rater was sought. All participants had a distinct onset peak in the quiet condition, yet in the noise condition this peak was not identifiable in three participants. Statistical analyses for onset peak latency and amplitude were limited to the participants who had identifiable peaks in both quiet and noise (n = 19). The transition peak reported here has been previously shown to relate to speech-in-noise perception (Parbery-Clark et al., 2009a; Anderson et al., 2010), and was clearly identifiable in all participants for both conditions (n = 22). For examples of individual subcortical response waveforms, see Supporting Information Fig. S1.

Subcortical response – stimulus-to-response correlations
To gage the effect of noise on the neural response, the stimulus and response waveforms were compared via cross-correlation. The degree of similarity was calculated by shifting the stimulus waveform in time by 8–12 ms relative to the response until a maximum correlation was found between the vowel portion of the stimulus (50–170 ms) and the region of the response corresponding to the vowel (60–180 ms). This time lag (8–12 ms) was chosen because it encompassed the stimulus transmission delay (from the ER-3 transducer and ear insert approximately 1.1 ms) and the neural lag between the cochlea and the rostral brainstem. This calculation resulted in Pearson’s r-values for both the quiet and noise conditions, which were subsequently transformed to z-scores for statistical analyses.

Cortical response – timing and amplitude
Latencies and amplitudes of P1, N1 and P2 were visually identified by the two first authors independently. The response from the Cz electrode was analysed to facilitate the comparison of our results to published literature and for its relevance in clinical settings. N1 was defined as the largest negative peak between 90 and 122 ms in quiet and between 130 and 188 ms in noise. P2 was defined as the largest positive peak between 158 and 226 ms in quiet and between 190 and 308 ms in noise. P1, which is known to be small in adults (Martin et al., 2008), was reliably identified in quiet but not in noise, and thus P1 results are not reported. Grand average responses, global field power traces and a comparison of an individual’s response in the quiet and noise condition were used to aid peak picking. In the case of disagreement over peak identification, the advice of a third rater was sought. All subjects had distinct peaks for N1 and P2 in both quiet and noise conditions (n = 22). Grand average responses (the average of all the individual subjects’ responses) are plotted in Fig. 2.

Statistical analysis
Response amplitudes and latencies were analysed with repeated-measures one-way ANOVAs, with Background (Quiet/Noise) as the within-participant factor. Two separate MANOVAs were conducted – one for the subcortical response measures and one for the cortical response measures. A series of Pearson r correlations were employed to investigate the relationship among the cortical, subcortical and perceptual measures. All statistical analyses were conducted with SPSS (SPSS, Chicago, IL, USA). Only statistically significant
analyses are reported in the main text below, additional analyses are reported in the Supporting Information Tables S1 to S3.

Results

Summary

The addition of background babble noise profoundly altered subcortical and cortical responses as well as the relationships between them. Brainstem encoding of speech was delayed and less precise in the presence of background noise. Similarly, cortical responses demonstrated a significant increase in latency for both N1 and P2. The amplitudes of the two cortical peaks were affected differently by noise, with N1 tending to be larger and P2 being attenuated. In quiet, the degree of stimulus fidelity in the subcortical response was related with measures of cortical response timing (N1 and P2 latency) as well as P2 amplitude. With the addition of noise, however, the degree of subcortical fidelity and brainstem response timing related only to N1 amplitude. Furthermore, it was only with the addition of noise that relationships between subcortical and cortical encoding of speech and the behavioral measures of speech in noise emerged.

Effect of noise on subcortical and cortical responses

The addition of background noise resulted in prolonged subcortical onset ($F_{1,21} = 84.271, P < 0.005$) and transition ($F_{1,21} = 16.604, P = 0.001$) response peaks (Fig. 1). Noise diminished the amplitude of the subcortical onset response ($F_{1,21} = 106.797, P < 0.005$), but not the transition response peak (43 ms) amplitude ($F_{1,21} = 1.011, P = 0.328$). Noise also reduced the precision of the neural representation of the speech sound as indicated by lower stimulus-to-response correlations relative to quiet ($F_{1,21} = 6.154, P = 0.023$). For cortical responses, background noise resulted in later response timing for N1 ($F_{1,21} = 293.663, P < 0.001$) and P2 ($F_{1,21} = 85.916, P < 0.001$) peaks. The effect of noise on the magnitude of these peaks was not consistent, with N1 amplitude tending to increase ($F_{1,21} = 4.020, P = 0.058$) and P2 amplitude being significantly reduced ($F_{1,21} = 56.470, P < 0.001$; Fig. 2).

Subcortical–cortical relationships

Subcortical-cortical responses in noise

In the noise condition, N1 amplitude significantly correlated with the precision with which the brainstem represented stimulus characteristics in noise ($r = -0.608, P = 0.013$; Fig. 3, top) as calculated by a stimulus-to-response correlation. N1 amplitude also correlated with brainstem response timing to speech in noise, with greater N1 amplitude associated with earlier subcortical onset ($r = 0.603, P = 0.003$; Fig. 3, middle) and transition ($r = 0.583, P = 0.002$; Fig. 3, bottom) peaks. Together, these results suggest that earlier response timing and more precise subcortical encoding of the speech signal in noise relate to greater N1 response amplitude. On the other hand, N1 and P2 latency measures as well as P2 amplitude in noise did not correlate with any of the brainstem measures in noise (Table 1).

Subcortical-cortical responses in quiet

In the quiet condition, N1 and P2 latency as well as P2 amplitude correlated with the degree of subcortical fidelity. Higher subcortical fidelity was associated with earlier N1 latency (Pearson’s $r = -0.438, P = 0.041$), earlier P2 latency ($r = -0.540, P = 0.009$) and greater P2 amplitude ($r = 0.492, P = 0.020$), suggesting that more precise brainstem encoding of speech in quiet is related to earlier cortical responses and greater P2 amplitude. P2 amplitude was also related to transition latency in that earlier peak latencies correlated with greater P2 amplitudes ($r = -0.572, P = 0.005$). N1 amplitude in quiet demonstrated no relationships with any of the brainstem measures in quiet (Supporting Information Table S1).

![Neural correlates of speech in noise](553)
Neural encoding and behavioral relationships

Subcortical–behavioral relationship

Speech-in-noise perception was correlated with all three subcortical response measures in noise. Better speech-in-noise performance correlated with greater subcortical fidelity ($r = -0.560$, $P = 0.007$; Fig. 4, top), earlier onset ($r = 0.530$, $P = 0.020$) and earlier transition ($r = 0.535$, $P = 0.013$) peak response timing (Table 2). No relationships were observed between any measure of brainstem encoding of speech in quiet and speech-in-noise perception (Supporting Information Table S2).

Cortical–behavioral relationship

In the noise condition, N1 amplitude positively correlated with speech perception in noise ($r = 0.520$, $P = 0.013$; Fig. 4, bottom), such that better performance on HINT was associated with greater N1 amplitude. Other cortical measures in noise, N1 and P2 latency and P2 amplitude, were not correlated with speech-in-noise perception (Table 2). Likewise, no relationships were found between cortical measures in quiet and HINT (Supporting Information Table S2).

Discussion

Numerous studies have investigated the effects of noise on auditory-evoked activity, but little research has focused on the functional coupling between subcortical and cortical activity in normal hearing adults. In the present study, we aimed to delineate the relationships between the neural encoding of speech in noise at subcortical and cortical levels in normal hearing young adults. We found that, with the exception of N1 amplitude, background noise reduced subcortical and cortical neural response amplitude, and prolonged neural response timing. Furthermore, in the noise condition, N1 amplitude was the only cortical component that correlated with subcortical encoding and behavioral perception of speech in noise. Together, these results suggest that in young adults N1 amplitude may provide a functional index of perception and subcortical neural encoding of speech in noise under the conditions reported here.

Subcortical- and cortical-evoked responses originate from different regions in the auditory pathway, representing acoustic information via distinct neural codes. Subcortical auditory responses represent the acoustics of the evoking stimulus with high fidelity, while cortical-evoked responses provide a more abstract representation of sound. Consequently, each of these responses provides a unique neural framework within which to objectively assess the biological processes underlying speech-in-noise perception. Our results indicate that the degree of subcortical resilience to the disruption caused by background noise is indexed by N1 amplitude. Specifically in noise, increased brainstem precision (i.e. more accurate representation of the speech sound) related to more robust cortical processing, as indicated by greater N1 amplitude. Alternatively stated, individuals who demonstrated greater disruption of brainstem activity in noise had smaller N1 response amplitudes. Thus, the addition of background noise modulated auditory processing at both subcortical and cortical levels, indicating a pervasive, system-wide impact of noise on neural function. Furthermore, for both subcortical and cortical measures, the degree of resilience to noise related to measures of speech-in-noise perception.

The strong relationship between measures of subcortical encoding, N1 amplitude and speech-in-noise perception suggests that these factors operate in an integrated manner. The N1 response is thought to

---

**Table 2. Correlations between behavioral measures (HINT) and neural encoding of speech in noise**

<table>
<thead>
<tr>
<th>Noise</th>
<th>Subcortical measures</th>
<th>Cortical measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Onset latency</td>
<td>Onset amplitude</td>
</tr>
<tr>
<td>HINT (dB SNR)</td>
<td>0.530*</td>
<td>0.103</td>
</tr>
</tbody>
</table>

HINT, Hearing in Noise Test. *$P < 0.05$; **$P < 0.01$. Latency (ms); Amplitude ($\mu$V).
reflect the time-varying aspects of a stimulus (Näätänen & Picton, 1987), such as the change from a consonant to a vowel (Tremblay et al., 2006), amplitude envelope and spectral changes (Martin & Boothroyd, 2000), as well as voice onset time ( Sharma & Dorman, 1999; Sharma et al., 2000; Hoonhorst et al., 2009). Accurate subcortical and cortical representations of temporal acoustic information are important for speech perception ( Kraus et al., 1995; Eggermont, 2001; Eggermont & Ponton, 2002; Tzounopoulos & Kraus, 2009), and are especially useful for understanding speech in background noise ( Hornickel et al., 2009; Parbery-Clark et al., 2009a; Anderson et al., 2010). Therefore, our findings of greater N1 amplitude being associated with better speech-in-noise performance and more robust brainstem representation of speech in noise may reflect a better neural representation of the important temporal acoustic cues such as the onset and transition. While noise significantly prolonged cortical peak latencies (N1 and P2) and reduced cortical response amplitudes (P2), no relationships were found between these measures and subcortical or behavioral measures of speech in noise. As such, our understanding of how these other cortical measures relate to subcortical encoding and speech-in-noise perception is limited.

Low-intensity background noise can enhance cortical responses ( Stufflebeam et al., 2000; Alain et al., 2009). It is thought that the noise-induced increase in N1 amplitude provides a neural metric reflecting an enhanced ability to separate the target from the noise ( Bertoli et al., 2005). Further support for this idea comes from Alain et al. (2009), who suggest that such noise-induced increases in response amplitude may enhance concurrent stream segregation through the activation of the efferent system ( Alain et al., 2009). The efferent system is thought to reduce the negative impact of background noise on neural encoding by increasing the SNR of the target stimulus in adverse listening environments. This is accomplished by a series of feedback loops which originate from cortical centers and terminate at subcortical nuclei and the cochlea. Cortical neurons, which can rapidly adapt their response field properties to improve the encoding of signals ( Fritz et al., 2003, 2005, 2007), also contribute to the sharpening of subcortical signal encoding ( Suga, 2008) by increasing the gain for the relevant auditory information and suppressing the irrelevant noise ( de Boer & Thornton, 2008; Luo et al., 2008; Nahum et al., 2008). Consistent with this idea, animal models have demonstrated enhanced cochlear ( Kawase & Liberman, 1993), auditory nerve ( Kawase et al., 1993) and midbrain ( Selu- kuman et al., 2008) coding of signals in noise with the activation of the efferent system. In humans, a stronger efferent system has been shown to correlate with better speech-in-noise perception ( Giraud et al., 1997; Kim et al., 2006) in studies that measured the extent to which otoacoustic emissions were inhibited during the activation of the medial olivocochlear efferent system using contralateral presentation of noise. Additionally, auditory training aimed at improving speech-in-noise perception saw a concomitant improvement in efferent strength and behavioral perception, again underscoring the relationship between the two ( de Boer & Thornton, 2008). Consequently, it would seem that a stronger cortical influence on subcortical structures would improve subcortical encoding of the signal by minimizing the negative impact of background noise. Therefore, the recent evidence pointing to N1 amplitude increase underscoring an improvement in concurrent stream segregation – a process known to contribute to speech-in-noise perception ( Alain et al., 2009) – is corroborated and extended by our findings which demonstrate a neural-behavioral relationship between N1 amplitude and speech-in-noise perception.

The functional relationship between brainstem and cortical centers mediating speech perception can be elucidated by subcortical and cortical events. Previous research has shown that normal-learning children demonstrate a coordinated relationship between brainstem response timing and cortical response robustness ( Cunningham et al., 2001; Wible et al., 2005), unlike children with language-based learning disorders ( LD) where this relationship is disrupted ( Wible et al., 2005). Children who have LD with brainstem response abnormalities also demonstrate reduced cortical sensitivity to acoustic change ( Banai et al., 2005), and fail to demonstrate the classical pattern of cortical asymmetry for speech processing ( Abrams et al., 2006). Additionally, children with LD can be more susceptible to the deleterious effects of background noise on speech perception ( Bradlow et al., 2003; Ziegler et al., 2005), unlike musicians who appear to have an advantage for speech perception in noise ( Parbery-Clark et al., 2009a). Musicians also demonstrate enhanced neural representation of the acoustic features of speech both subcortically ( Musacchia et al., 2007; Wong et al., 2007; Parbery-Clark et al., 2009a, reviewed in Kraus & Chandrasekaran, 2010) and cortically ( Nikjeh et al., 2009; Tervaniemi et al., 2009). Furthermore, musical training strengthens subcortical-cortical relationships, demonstrating that experience modulates auditory processing in a coordinated manner ( Musacchia et al., 2008). Our findings indicate that N1 amplitude relates to subcortical encoding and behavioral perception of speech in noise, highlighting the functional coupling of subcortical measures and N1 cortical response amplitude within an individual. The extent to which these relationships are upheld in populations with excessive difficulties in background noise ( e.g. older adults, hearing impaired) or populations who demonstrate enhancements ( e.g. musicians) should inform our understanding of the biological mechanisms subserving the impairment or strengthening of speech-in-noise perception.

Conclusion

We report a coordinated relationship between specific aspects of neural encoding of sound at cortical and subcortical levels that directly relate to speech-in-noise perceptual ability. Measures of subcortical timing and cortical response magnitude relate strongly to speech-in-noise performance. These electrophysiological events offer a system-wide view of the key neural facets underlying the perception of speech in noise. This work provides a baseline for viewing impaired processing of speech in noise and for understanding mechanisms of training-related plasticity.

Supporting Information

Additional supporting information may be found in the online version of this article:

Fig. S1. Three individual subcortical responses in noise are plotted. The subjects plotted in the top and middle plots clearly have identifiable onset and transition responses (circled), whereas the subject in the bottom plot demonstrates only a transition response.

Table S1. Correlations between subcortical and cortical responses to speech in quiet.
References


