Chapter 8
Communicating in Challenging Environments: Noise and Reverberation

Gavin M. Bidelman

Abstract In everyday listening situations, speech perception is challenged by interfering noise and other adverse room acoustics (e.g., reverberation). These intrusions hinder verbal communication and prevent audible access to salient cues by masking (noise) and smearing ("reverb") spectrotemporal features of the speech signal. The brainstem frequency-following response (FFR) provides a detailed window into the early neural transcription of complex sounds and how normal and degraded speech signals are coded by the human auditory nervous system. This chapter provides an overview of noise-related and reverb-related changes in brainstem representations for speech as reflected in the scalp-recorded FFR. Although noise and reverberation affect behavior to a similar extent, they have a differential effect on neural speech representations, noise being a larger detriment to the speech code than reverberation. Acoustic interferences also produce distinct effects within the speech signal: the neural encoding of "timbre" is more affected than voice "pitch" cues. Applications for the FFR as a "biomarker" for understanding the neural basis and individual differences in degraded speech perception skills are also discussed.

Keywords Auditory scene analysis · Autocorrelogram · Degraded speech processing · FFR · Figure-ground analysis · Frequency-following response · SIN · Speech-in-noise perception · Voice pitch · Voice timbre

8.1 Introduction

In nearly all real-world listening environments, acoustic interferences hinder the successful extraction of speech information. This chapter concerns the neural basis of human communication in adverse listening conditions and focuses on the effects

---

G.M. Bidelman (✉)
School of Communication Sciences and Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN 38152, USA
e-mail: g.bidelman@memphis.edu

N. Kraus et al. (eds.), *The Frequency-Following Response*,
Springer Handbook of Auditory Research 61,
DOI 10.1007/978-3-319-47944-6_8

of “energetic masking” on neurophysiological speech processing. This problem, classically known as the “cocktail party scenario,” represents a fundamental challenge for the auditory system and a barrier to effective communication. Speech-in-noise (SIN) understanding is often exacerbated in cases of hearing impairment, but the issues typically persist even after restoring audiometric thresholds via hearing aids (for review, see Popelka et al. 2016). Moreover, SIN comprehension is problematic even for individuals without substantial hearing loss (Middelweerd et al. 1990; Song et al. 2011). These findings have led to the increasing notion that speech intelligibility and SIN listening skills are determined by more than simple audibility (i.e., peripheral hearing status) (Humes and Roberts 1990). In particular, recent interest in the physiological basis of speech processing has focused on the role of central auditory brain mechanisms in SIN listening and how robust neural coding supports successful listening skills. In this regard, the human frequency-following response (FFR) has provided considerable insight into human communication and central auditory processing in adverse listening environments.

8.2 Listening at the “Cocktail Party”

8.2.1 Acoustical Consequences of Noise and Reverberation

Listeners face two primary challenges when extracting speech from the auditory scene: *noise* and *reverberation*. Each has a distinct effect on the speech signal, yet both hinder intelligibility (Nabelek and Dagenais 1986; Helfer and Wilber 1990). Noise is caused by the addition of external competing sound(s) to target speech and acts as a simultaneous masker, obscuring less intense portions of the speech signal and reducing its *signal-to-noise ratio* (SNR). In contrast, reverberation (reverb) is an interference caused by the internal room acoustics of an enclosed space (Kinsler et al. 2000). Formally defined, reverberation is the persistence of acoustic energy in the sound field after it is produced. Reflected sound waves are exaggerated in reverberant settings (e.g., a concrete stairwell), resulting in a slow decay of energy and a temporal overlap of incident and reflected wave fronts. The overlap between direct and indirect sounds results in a “smearing” of the signal’s spectrum.

There are qualitative differences in the way in which noise and reverberation obstruct signals of interest. Reverberation is based on reflection and absorption characteristics of materials within an enclosed space (Sabine 1962). Hence, the effectiveness of reverberation to occlude a signal is largely determined by the acoustic properties of the room itself. However, in the case of additive noise, signal occlusion is mainly determined by the similarity between the masker and the signal spectra.

The systematic effects of noise and reverberation on speech perception can be studied by parametrically changing the amount of interference added to a “clean” speech signal (containing no interference). Conveniently, the degree of noise and reverberation superimposed onto a target signal can be quantified by similar metrics. For additive noise, the relative contribution of "noise" and "signal" are quantified via the SNR. Specified in decibels (dB), positive SNRs reflect more favorable noise conditions (i.e., signal > noise), whereas negative SNRs reflect listening conditions where the noise dominates and masks the signal (i.e., noise > signal). Similarly, the proportion of acoustic energy attributable to signal and reverberant energy can be characterized (in dB) by a metric called the *direct-to-reverberant energy ratio* (D/R) (von Békésy 1938; Zahorik 2002). Other metrics can be used to characterize reverberation including *reverberation time* (RT_{60}), a measure describing how long it takes for reverberant sound energy to decay before attenuating by 60 dB. However, D/R is most comparable to SNR and allows the most direct comparison between the two forms of interference. Sometimes D/R is referred to as "wet-to-dry" ratio. Behavioral studies in human listeners have shown that the *just noticeable difference* (JND) for D/R sensitivity is on the order of 5–6 dB (Zahorik 2002; Larsen et al. 2008), slightly higher than the 3 dB JND for noise SNR (McShefferty et al. 2015).

In a reverberant space, D/R decreases with increasing source-to-receiver distance as direct ("dry") energy becomes swamped by sound energy from indirect ("wet") specular (mirror-like) reflections. In signal processing, a system is fully described by its impulse response (i.e., response to a broadband transient). Similarly, the reverberant characteristics of a particular room are described by its acoustic impulse response, which can be recorded with a microphone at different source–receiver distances to the presentation of the impulsive sound (e.g., a balloon pop). By convolving a room's impulse response with a signal (e.g., speech), the resulting output is heard as if the target sound were recorded in the reverberant space. By employing impulse responses measured at different source–receiver distances, D/R can be parametrically manipulated akin to varying the SNR for noise.

Figure 8.1 shows the effects of additive noise and reverberation to the vowel token /i/ (Bidelman and Krishnan 2010) at comparable signal-to-interference levels (i.e., SNR ≈ D/R of ±5 dB). While the relative intensity of the signal to noise/reverberation is identical in this example, it is clear that the two forms of interference have different acoustic effects on the speech signal. With increasing reverberation, the dynamic (i.e., time-varying) change in voice fundamental frequency (F_0) and its harmonics (integer-related frequencies) show a smearing effect; portions of the signal persist, distorting the sequencing of spectral cues in speech as it unfolds in time. This overlap results in a spectrotemporal smearing that distorts ongoing and subsequent speech information. In reverberant settings, target cues essentially act as their own forward maskers (Nabelek et al. 1989; Wang and Brown 2006); yet, the strong harmonic structure of speech and its F_0 contour (cues that convey voice pitch) are largely preserved. Static (i.e., steady-state) signal features would be even less affected by reverberation. Contrastively, increasing levels of broadband noise are seen to "fill in" the peaks and troughs of the spectrum, reducing the spectral contrast of important speech cues necessary for proper identification (e.g., formants). Thus, despite having comparable relative intensity

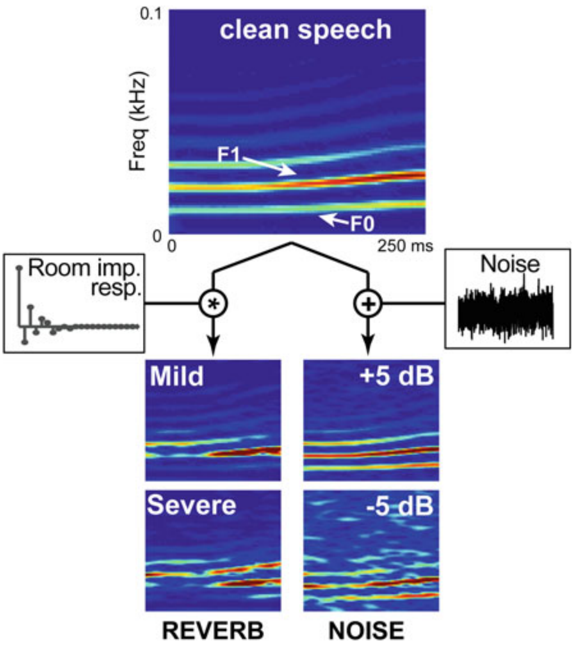


Fig. 8.1 Comparison of noise and reverb-related changes in speech acoustics. Spectrogram of the clean speech vowel token /i/ (*top*) containing a time-varying F_0 (∼100 Hz) and fixed formant frequencies of F1 = 300; F2 = 2500, F3 = 3500 and F4 = 4530 Hz (Bidelman and Krishnan 2010). The parametric effects of reverberation on speech processing can be studied by convolving the clean speech signal with various room impulse responses recorded in reverberant spaces at different source-receiver distances (e.g., Watkins 2005). Larger source-receiver distances decrease the direct-to-reverberant energy ratio (D/R), a metric akin to signal-to-noise ratio (SNR). Similarly, the parametric effect of noise is studied by varying the SNR of additive noise superimposed on clean speech. Acoustically, reverberation has the effect of "smearing" the speech spectrum, distorting the spectrotemporal timing of acoustic landmarks. In contrast, noise "fills in" the signal's spectrum with decreasing SNR, reducing spectral contrast between signal and noise. *FO*, F_0; *imp resp*, impulse response

between the signal and interference, it is clear that noise and reverberation act very differently and even differentially across the various cues important to speech perception (e.g., voice pitch versus timbre; see Sect. 8.3.2).

Lastly, it should be noted that unlike noise, which is nearly always considered to be a negative interference, reverberation can sometimes provide positive benefits to auditory perception. In fact, reverberation is tolerable (and often desirable) in concert music halls (Lifshitz 1925; Backus 1977), where pitch dominates the signal, but not in classrooms (Yang and Bradley 2009), where target acoustics are geared toward speech intelligibility (see Sect. 8.3.2). Beyond aesthetic considerations, the D/R of a reverberant signal can also facilitate the perceptual judgments of source distance when interaural level and intensity cues are ambiguous (Zahorik 2002; Larsen et al. 2008).

8.2.2 Behavioral Basis and Individual Variability in Degraded Speech Perception

8.2.2.1 Speech Perception in Noise

The effect of noise on speech perception has enjoyed a long history in the hearing sciences. Early work on speech audiometry recognized two components of hearing impairment: (1) loss of acuity (*audibility*) and (2) loss of clarity (*distortion*) (Carhart 1951; Plomp 1978). Audibility (i.e., signal attenuation) is a linear process and easily predicted from the pure-tone audiogram or articulation index (French and Steinberg 1947; ANSI 1969). In contrast, the distortion is a non-linear component of hearing loss and, more problematically, is poorly predicted from the pure-tone audiogram or word recognition scores (in quiet). Degraded speech perception tests therefore became routine in the late 1960s as a means to quantify the distortion component of hearing (Carhart and Tillman 1970) and to address a common complaint of hearing impaired listeners: poor speech recognition despite restored audibility through hearing aids (for review, see Wilson and McArdle 2005).

SIN perception is now measured using a number of standardized audiological tests, for example: Hearing-in-Noise test (HINT), Nilsson et al. (1994; QuickSIN^{TM}, Killion et al. (2004). SIN tests vary in the semantic and contextual cues they offer the listener (e.g., sentences versus words; high versus low predictability). Nonetheless, at their core the basic premise of these tests is similar: listeners are presented with speech stimuli and are asked to detect certain key words. The SNR is varied at fixed intervals or adaptively to obtain the individual's speech reception threshold in noise.

While the acoustic effects of additive noise on speech perception are somewhat predictable, the behavioral consequences are all but trivial. Speech intelligibility in adverse conditions is influenced by a number of factors including the spectrotemporal characteristics of the noise, its semantic content, whether or not the noise is stationary, modulated or continuous, presented monaurally or binaurally, and the spatial proximity of the noise to the target signal (reviewed by Assmann and Summerfield 2004). Irrespective of lexical-semantic or contextual cues, noise reduces SNR, obscuring the perceptually salient cues of speech by masking contrastive portions of the signal's spectrum. However, in addition to spectral masking effects, noise can have detrimental effects on temporal aspects of the speech signal. For instance, one prominent finding of perceptual studies is that listeners exploit the temporal envelope of speech (i.e., slow amplitude fluctuations) for robust comprehension. This is true even when the speech signal's "fine-structure" (carrier) is noise containing no spectral cues or, in cases of cochlear implant signal processing, where only envelope cues are delivered to the stimulating electrodes (Shannon et al. 1995; Swaminathan and Heinz 2012). This has led to the prevailing view that the reduction in speech intelligibility in noise results in degradations to the speech envelope. In noisy conditions, speech fine structure is often exploited to help aid spoken word recognition (e.g., Lorenzi et al. 2006; Swaminathan and Heinz 2012).

8.2.2.2 Speech Perception in Reverberation

The deleterious effects of reverberation on speech intelligibility can be ascribed to consequences of both "overlap-masking" (i.e., forward) and "self-masking" (Nabelek et al. 1989). As segments of the speech signal reflect in a reverberant space they act as forward maskers, overlapping subsequent syllables and inhibiting their discrimination. In addition, reflections concurrent with the incident (i.e., direct) sound dramatically change the dynamics of speech by blurring the waveform's fine-structure. When acting on a time-varying signal, this "temporal smearing" tends to transfer spectral features of the signal from one time epoch into later ones, inducing a smearing effect in the spectrogram (Wang and Brown 2006). As a consequence, this internal temporal smearing distorts the energy within each phoneme such that a signal can effectively act as its own masker (i.e., self-masking). With such distortions, normal hearing listeners have difficulty identifying and discriminating consonantal features (Gelfand and Silman 1979; Nabelek et al. 1989), vowels (Nabelek and Letowski 1988; Drgas and Blaszak 2009), and time-varying formant cues (Nabelek and Dagenais 1986) in reverberant listening conditions. It should be noted that speech confusions in reverberation are further exacerbated with hearing impairment (Nabelek and Letowski 1985; Nabelek and Dagenais 1986; Nabelek 1988).

Despite its difference from noise SNR (see Sect. 8.2.1), decreased D/R for reverberation has a similar negative effect on speech intelligibility. Figure 8.2A shows closed set vowel identification performance in noise (0 dB SNR) and reverberation (RT_{60} = 1.2 s) for listeners with binaural sensorineural hearing loss (Nabelek and Dagenais 1986). Vowels are typically highly identifiable in noise and reverberation for normal-hearing listeners. As seen in the figure, both noise and reverberation interference reduce speech identification in hearing impaired listeners by a similar magnitude ($\sim$15–20%). However, the pattern of specific vowel confusion errors is typically not the same in noise as it is in reverberation. Nabelek and Dagenais (1986) suggest that in noise, misidentifications are related to the spectral proximity of formant frequencies for confused pairs. In contrast, confusions in reverberation are probably attributable to changes in the relative weighting of the formants as they are smeared in time and frequency by the prolonged reverberant energy.

8.2.2.3 Comparisons Between Noise and Reverberation on Speech Perception

While behavioral studies have mainly focused on the independent consequences of noise and reverberation on speech perception, it should be noted that real-world listening environments typically contain both multiple noise sources and reverberant acoustics. Thus, in most cases, noise and reverberation occur simultaneously

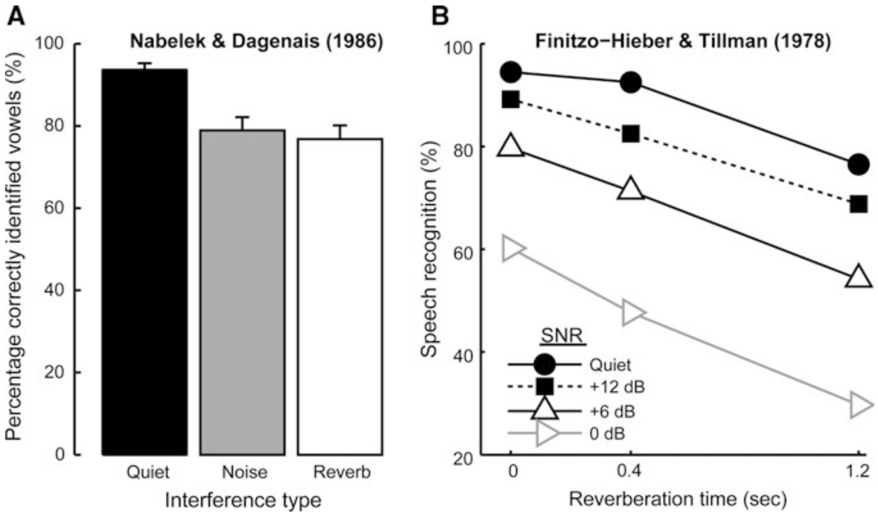


Fig. 8.2 Perceptual speech comprehension in noise and reverberation. (**A**) Closed-set vowel identification performance in noise (0 dB SNR) and comparable degrees of reverberation (RT = 1.2 s) reported by Nabelek and Dagenais (1986). In the absence of contextual speech cues, both noise and reverberation reduce speech intelligibility. (**B**) Interaction of noise and reverberation on speech recognition (Finitzo-Hieber and Tillman 1978). While both noise and reverberation impair speech understanding independently, their combined effect yields poorer speech intelligibility than either factor alone. *SNR*, signal-to-noise ratio

and can act synergistically to further impair understanding (George et al. 2008). Figure 8.2B illustrates the combined effects of noise and reverberation on speech recognition scores reported by Finitzo-Hieber and Tillman (1978) for normal hearing children. Both the SNR of additive noise and reverberation time (*RT*) were manipulated in the experiment. Although RT was the independent variable for reverberation in this experiment, higher RTs correspond with decreased D/R, and hence, less favorable reverberation levels. Both main effects of noise and reverberation were observed on speech recognition scores when each type of interference was considered alone. However, Finitzo-Hieber and Tillman (1978) also noted an interaction between reverberation and noise. That is, the influence of noise or reverberation in isolation was further increased by the introduction of the other interference.

8.2.2.4 Individual Differences and Normal Variability in SIN

Current hearing aids provide little benefit for SIN understanding despite restoring audiometric thresholds (Chmiel and Jerger 1996). Consequently, it is now well accepted that SIN perception cannot be reliably predicted from the audiogram (Killion and Niquette 2000). This might not be entirely surprising in light of the

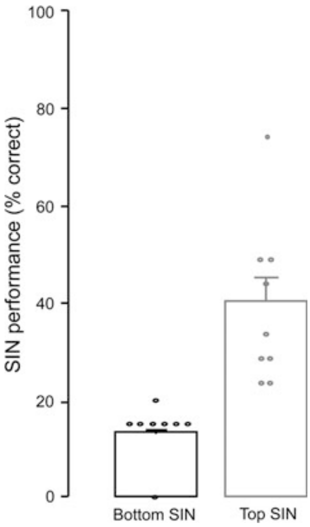


Fig. 8.3 Individual differences in speech-in-noise (SIN) perception. Grand average (*bars*) and individual participant responses (*points*) on the QuickSIN^{TM} for audiometrically normal-hearing young adults (Song et al. 2011). Despite having normal hearing thresholds, there is considerable variability in SIN performance, which can be categorized into top (≥25%, $n = 9$) and bottom (<25%, $n = 8$) SIN perceiving groups. In the (presumed) absence of differences in peripheral hearing function, individual differences in degraded speech perception might be related to differences in the neural encoding of speech (also see Fig. 8.8). (Data from Song et al. 2011)

"distortion" that often accompanies reduced audibility in cases of hearing impairment (see Sect. 8.2.2.1). Furthermore, SIN perception is problematic and performance varies considerably among individuals without substantial hearing impairment (Middelweerd et al. 1990; Frisina and Frisina 1997). Even normal-hearing young adults show individual variability in SIN in the absence of any known audiological or peripheral hearing deficit (Song et al. 2011, 2012). In particular, Song et al. (2011) measured behavioral performance on the QuickSIN^{TM} in normal-hearing young adults with normal pure tone hearing thresholds (≤20 dB HL from 125 to 8000 Hz) (Fig. 8.3). Despite normal hearing, no listener performed the task at ceiling and in fact, there was considerable variability between listeners (i.e., individual differences), ranging from 0 to 75% speech recognition.

These findings challenge conventional and longstanding views that speech intelligibility is determined solely by audibility, i.e., peripheral hearing status (Plomp 1986; Humes and Christopherson 1991). Rather, hearing sensitivity alone seems to be inadequate to account for SIN perception issues (Humes and Christopherson 1991; Parbery-Clark et al. 2011). Consequently, a growing body of evidence suggests that central auditory processing—as early as the brainstem—plays a critical role in mediating robust perceptual SIN abilities.

8.3 Effects of Acoustic Interference on FFR Representations of Speech

The brainstem FFR has provided critical insight toward understanding the neurobiological encoding of clean and degraded speech from a subcortical perspective (Parbery-Clark et al. 2009; Bidelman and Krishnan 2010; Song et al. 2011). Although there are multiple sources of FFRs throughout the hearing pathway (e.g., cochlear microphonic, Sohmer and Pratt 1977; auditory nerve, Bidelman 2015b; and brainstem, Sohmer et al. 1977; Bidelman 2015b), the inferior colliculus of the midbrain is considered the primary generator of the scalp-recorded FFR (Sohmer et al. 1977; Bidelman 2015b). The brainstem FFR is also distinct in its response characteristics from the more conventional click-evoked auditory brainstem response (ABR) familiar to audiologists, differing in rate susceptibility (Krizman et al. 2010), frequency specificity (Picton et al. 1977), spectral content (Bidelman 2015b), susceptibility to noise masking (Cunningham et al. 2002; Russo et al. 2004), and latency-intensity changes (Akhoun et al. 2008). These response properties make the FFR a unique window into auditory neurophysiological function that is distinct from the traditional brainstem ABR.

Also different from the transient ABR, FFRs code dynamic, spectrotemporal features of periodic acoustic stimuli. This unique feature makes FFRs a quasi "neural fingerprint" of the acoustic signal within the human EEG. Indeed, the remarkable fidelity of FFRs is evident in listening experiments in which the neural responses are replayed to human listeners as audio signals and can be reliably identified as intelligible speech (Galbraith et al. 1995; Weiss and Bidelman 2015). Moreover, although debated, brainstem responses are largely unaffected by attention (Woods and Hillyard 1978; Galbraith et al. 2003; also see Shinn-Cunningham, Varghese, Wang, and Bharadwaj, Chap. 7). Thus, unlike their cortical event-related potential (ERP) counterparts that are highly malleable to subject state, habituation, and overlap with endogenous neural activity, FFRs provide a stable window into the neural transcription of speech signals that can be obtained under passive listening paradigms. Consequently, FFRs have provided important insight into individual differences in SIN listening skills and the neural encoding of speech in normal and clinical populations who cannot participate in traditional, behavioral (i.e., subjective) auditory assessments (e.g., Cunningham et al. 2001).

8.3.1 Noise-Related Changes in Brainstem Speech Processing

Noise-related changes in the FFR elicited by complex sounds are evident in both the time and frequency domain. Figure 8.4A shows brainstem FFRs recorded in response to a complex tone containing the 12th–17th harmonics of a 90 Hz F_0 (unresolved harmonics) (Smalt et al. 2012). In the Smalt et al. (2012) study,

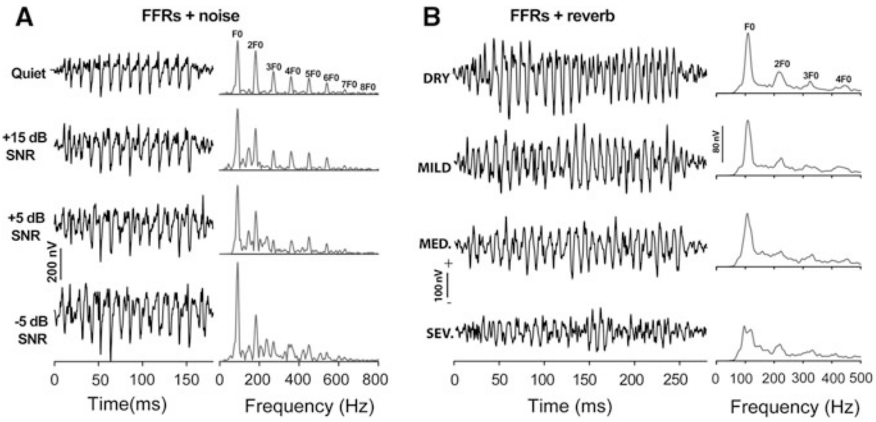


Fig. 8.4 Brainstem FFRs as a function of signal-to-noise ratio (SNR) (Smalt et al. 2012) and reverberation level (Bidelman and Krishnan 2010). (**A**) With increasing noise, phase-locking to the stimulus diminished. Most notable is the reduction in periodicity at higher harmonics of the F_0 frequency (i.e., $2F_0$–$8F_0$), components useful for the calculation of the common pitch. In contrast, the F_0 component remains largely intact or is enhanced in noise. (**B**) Similarly, speech FFRs evoked by the token /i/ (see Fig. 8.1) show a reduction in overall response energy with notable declines in neural synchronization (reduced periodicity) with increasing reverberation, particularly at higher frequencies. Neural coding at F_0 is largely maintained until the most severe levels of reverberation. *FO*, F_0; *MED*, medium; *SEV*, severe. (Reprinted from Smalt et al. 2012; Bidelman and Krishnan 2010, with permissions from Elsevier)

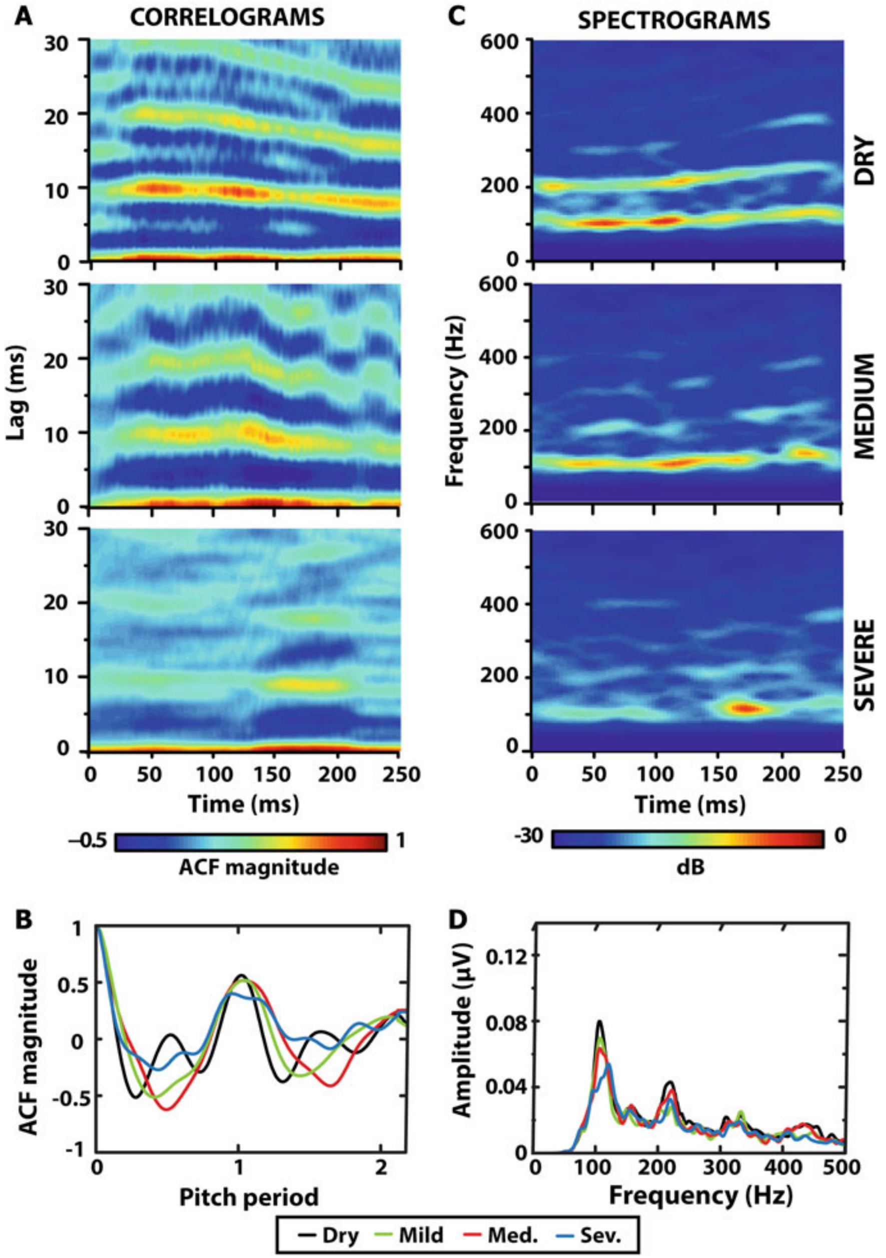
lowpass filtered noise was used to mask audible distortion products. The first observation apparent from these FFRs is the prominent neural energy at the F_0 and its lower-order harmonics (F_0, $3F_0$, etc.), despite the fact that these components did not occur in the stimulus. The presence of response energy at the F_0 indicates that the FFR phase locks at the common periodicity of the stimulus, providing a neural correlate of the missing fundamental (Greenberg et al. 1987). Secondly, with the addition of noise, it is apparent that the neural encoding of the sustained F_0 periodicity (i.e., stimulus envelope) is well-maintained at decreasing (poorer) SNRs; little degradation in FFR F_0 is observed even at higher noise levels. In contrast, broadband white noise often delays, attenuates, or even eradicates the onset components of the transient ABR (Burkard and Hecox 1983; Russo et al. 2004). The resilience of the brainstem FFR at F_0 (but not its higher harmonics or onset) in the presence of noise has been noted by a number of investigators (e.g., Russo et al. 2004; Li and Jeng 2011; Prevost et al. 2013) and suggests that neural synchronization at the fundamental F_0 periodicity is relatively robust to acoustic interference. In speech perception, F_0 provides a correlate of voice pitch and a robust cue for stream segregation and identification of the number of sources in complex auditory scenes (Assmann and Summerfield 1990). Thus, the low susceptibility of the F_0 steady-state portion of the FFR is consistent with the notion that pitch remains a robust cue for segregating target speech from a sound mixture (Assmann and Summerfield 1990). In contrast to the F_0 of speech, higher spectral components captured by the FFR (e.g., formant-related harmonics) are systematically degraded with noise, paralleling their rapid deterioration behaviorally (Liu and Kewley-Port 2004).

Some studies have even reported a facilitation of F_0 with additive noise (Smalt et al. 2012; Prevost et al. 2013). Two possible mechanisms have been proposed to account for the enhancement of F_0 in noise. It has been suggested that the maintenance and/or strengthening of the response at the F_0 in noise occurs as the result of *stochastic resonance* in the auditory system (Henry 1999; Cunningham et al. 2002). As stated by Prevost et al. (2013), this phenomenon is described as "improved detection and physiological representation of a weak periodic signal by the addition of noise." In other words, degraded F_0 representations can actually be counteracted (enhanced) by further neural entrainment to the signal in the presence of a noise masker. The specific neuronal mechanisms responsible for this phenomenon are unknown. However, stochastic resonance has been reported in the envelope following responses of auditory nerve fibers in the Mongolian gerbil (*Meriones unguiculatis*) (Henry 1999), so it is conceivable that resonance occurs as a result of cochlear nonlinearities (Jaramillo and Wiesenfeld 1998). Alternatively, the higher stimulus presentation levels used in typical brainstem experiments (>70–80 dB SPL) mean that FFRs reported in most studies reflect contributions from a wide range of the cochlear partition due to upward spread of excitation with increasing level (Dau 2003). Thus, the robustness of F_0 within the FFR amid noise may not reflect stochastic resonance per se, but instead reflects the additional engagement of low-frequency "tails" of basal, high-frequency neurons as they begin to phase lock to the common, high intensity F_0 across cochlear channels (Kiang and Moxon 1973). Multiple points of phase locking at F_0 across the cochlear array would tend to reinforce one another and, consequently, offer some resilience or redundancy in a pitch cue in the presence of noise interference.

8.3.2 Reverberation-Related Changes in Brainstem Speech Processing

To date, only a single study has investigated changes in the FFR under reverberation. Bidelman and Krishnan (2010) measured speech FFRs in response to the speech token /i/ (250 ms; time-varying F_0) presented in dry, mild, medium, and severe levels of reverberation (see Fig. 8.1). FFRs showed a systematic degradation in neural periodicity with increasing levels of reverberation (Fig. 8.4B). Thus, as with additive noise, reverberation degrades the normal phase-locking capacity of the FFR and reduces the response's ability to "tag" acoustic features of the speech signal.

Spectral analyses are typically more informative than time-domain analysis for sustained evoked potentials and can reveal specific stimulus-related changes in the FFR not apparent in raw waveforms. In the time domain, response autocorrelograms (ACGs) can be computed to index variation in neural periodicities over the



duration of the response (Fig. 8.5A). ACGs represent the short term (i.e., running) autocorrelation function of windowed frames of a compound signal:

◀ **Fig. 8.5** Temporal and spectral changes in speech-FFR with increasing reverberation. (**A**) Autocorrelograms, (**B**) time-averaged autocorrelation functions, (**C**) spectrograms, and (**D**) time-averaged FFTs derived from FFR waveforms in response to the vowel /i/ in various amounts of reverberation. As indexed by the invariance of the autocorrelation function (*ACF*) magnitude at the fundamental period (i.e., F_0 pitch period = 1), increasing levels of reverberation have little effect on the neural encoding of pitch-relevant information (**A–B**). The effect of reverberation on FFR encoding of the formant-related harmonics is much more pronounced (**C–D**). As with temporal measures, the representation of F_0 (100–130 Hz) remains more intact across conditions than higher harmonics (>200 Hz), which are smeared and intermittently lost in more severe amounts of reverberation (e.g., compare strength of F_0 to strength of harmonics across conditions). *Med.*, medium; *Sev.*, severe. (Reprinted from Bidelman and Krishnan 2010, with permission from Elsevier)

$$ACG(\tau, t) = X(t) \times X(t - \tau)$$

for each time t and time-lag τ. It is a three-dimensional plot quantifying the variations in periodicity and “neural pitch strength” (i.e., degree of phase locking) as a function of time. The horizontal axis represents the time at which single autocorrelation function (ACF) “slices” are computed while the vertical axis represents their corresponding time-lags (i.e., periods). The intensity of each point in the image represents the instantaneous ACF magnitude computed at a given time within the FFR response. Mathematically, the running autocorrelogram is the time-domain analog to the frequency-domain spectrogram (Fig. 8.5C). In terms of neurophysiology, it represents the running distribution of all-order interspike intervals present in the population neural activity (Cariani and Delgutte 1996; Sayles and Winter 2008), which may underlie the farfield FFR. From the time-varying ACG, global neural periodicity strength can be obtained by pooling the running ACG. This results in a summary ACF (Fig. 8.5B), which is a time-domain analog to a spectrum in the frequency domain (Fig. 8.5D). The magnitude of the ACF at the lag of the fundamental pitch period (i.e., $\tau = 1/T_0$) has been used as a unitary measure of FFR pitch strength (Krishnan et al. 2005; Bidelman and Krishnan 2010).

Temporal and spectral analyses of FFR recordings to reverberant speech (Bidelman and Krishnan 2010) are shown in Fig. 8.5. Apparent from the FFR’s ACF is the strong phase-locked response at the fundamental pitch period ($T_0 \approx 10$ ms; $F_0 = 100$ Hz) and its subharmonic periodicities both with and without reverberation. Time-averaged ACFs more clearly show that the FFR representation of F_0 remains largely intact with the addition of reverberant energy. That is, the magnitude of summary ACF energy at fundamental period (i.e., voice pitch) is invariant to increasing reverberation (Fig. 8.5B). Only in the most severe reverberation tested (D/R = −12 dB; RT_{60} = 900 ms) was there a noticeable change in response magnitude at F_0. In other words, pitch cues remain largely intact in the presence of reverberation consistent with single unit responses in the cochlear nucleus (Sayles and Winter 2008). In contrast, spectral analysis of the FFR (Fig. 8.5C–D) reveals a smearing of the response spectrum with increasing reverberation, particularly in higher harmonics of the signal’s spectrum. The weaker, more diffuse encoding of higher spectral components in the brainstem FFR is particularly evident near harmonics proximal to the first formant ($\sim$300 Hz).

These examples demonstrate that the brainstem's ability to encode speech cues is not an all or nothing phenomenon. Rather, pitch (F_0) and timbre (F1) cues that are differentially affected by reverberation at the acoustic level are similarly differentially encoded in subcortical FFRs. As observed for noise-degraded speech (Cunningham et al. 2002; Russo et al. 2004), FFRs largely preserve F_0 cues with increasing reverberation, whereas formant cues are rapidly degraded. Interestingly, these neural effects appear to parallel listeners' behavioral responses. Perceptual discrimination thresholds (difference limens) for speech F_0 are largely invariant to increasing reverberation, whereas F1 discrimination thresholds worsen in even minimal amounts of reverberation (Bidelman and Krishnan 2010). A larger impact on lower compared to higher spectral components of speech may at least partly be due to the effects of harmonic resolvability. Lower-order harmonics that are said to be "resolved" by the peripheral auditory filters dominate the FFR and are also more resilient to noise than higher, "unresolved" harmonics (Laroche et al. 2013). Conceivably, the differential perception of voice pitch and timbre cues in the presence of reverberation (and resilience of the former) might be driven by the differential encoding of pitch and timbre cues and/or resolved versus unresolved harmonics at subcortical, sensory levels of auditory processing.

The dissociable effects of reverberation (and noise) on the neural encoding of F_0 versus higher speech harmonics suggest that acoustic interferences have a differential effect on the 'source-filter' components of speech (Fant 1960). The *source-filter theory* of speech postulates that speech acoustics result from the glottal source being filtered by the vocal tract's articulators. The fundamental frequency at which the vocal folds vibrate determines the pitch of a talker independently from the configuration of the vocal tract and oral cavity, which determine formant structure (i.e., voice quality) (Fant 1960). Together, voice pitch and timbre cues provide adequate information for identifying who is speaking (e.g., male versus female talker) and what is being said (e.g., distinguishing vowel sounds) (Assmann and Summerfield 1989, 1990). Cast in terms of the source-filter model, it appears that source-related response components (i.e., F_0) coded in the FFR are relatively immune to additive background interferences, whereas filter-related components (i.e., formant structure/upper harmonics) are easily degraded (Russo et al. 2004; Bidelman and Krishnan 2010). This differential effect may be one reason, for instance, why reverberation is desirable in concert music halls (Lifshitz 1925; Backus 1977) but not in classrooms designed for speech intelligibility (Yang and Bradley 2009).

8.3.3 Direct Comparisons Between Noise-Related and Reverberation-Related Changes in FFR

To date, there has not been a direct comparison in a single study between speech-evoked FFRs recorded in noise and reverberation. Nevertheless, one can compare across studies to appreciate potential differences in how these two forms of interference have both similar and unique effects on the subcortical encoding of speech. A direct comparison between noise-related and reverb-related changes in FFR would also be useful in light of the differential acoustic and perceptual effects that these two forms of interference have on the speech signal (see Sect. 8.2).

Figure 8.6 illustrates speech-evoked FFRs recorded in clean, reverberant, and noise-degraded listening conditions. The noise and reverberation levels reflect an SNR and D/R of roughly +5dB and thus represent a comparable degree of masking between the two classes of interference. Despite a similar relative level between target speech and the masker, it is clear that noise and reverberation produce unique changes in the FFR. As observed acoustically, noise largely “fills in” the spectral peaks and troughs of the speech signal, masking the spectral contrast between perceptually salient cues (e.g., formant peaks) and the noise floor. In contrast, comparable levels of reverberation blur the speech spectrogram, although harmonics are still clearly visible in the response. These differences are more apparent in time-averaged spectra, which show (at least qualitatively) clearer neural representation of the speech spectrum in reverberation compared to speech presented in noise.

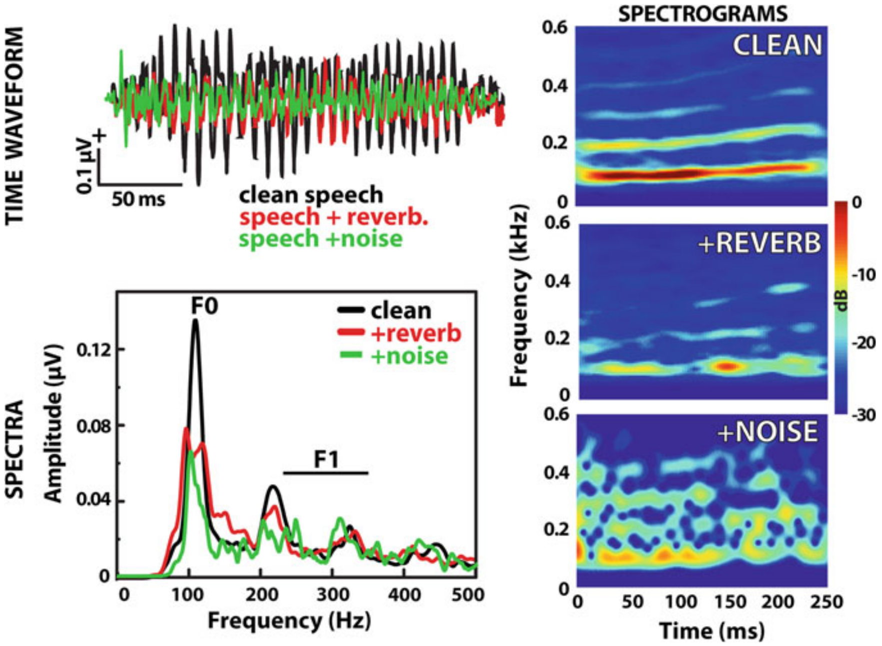


Fig. 8.6 Direct comparison between noise-related and reverb-related changes in speech-FFRs. Response waveforms (*upper left*) elicited by clean, noisy, and reverberant speech (both SNRs ≈ +5 dB). Response spectra (*bottom left*) show reduced encoding of voice pitch (F_0 ≈ 100 Hz) and higher spectral information in noise/reverb. FFR spectrograms illustrate that voice “pitch” (F_0) and “timbre” (F1) are hindered more by noise than reverberation. [Based on data from Bidelman and Krishnan (2010) and unpublished data]

Though a more comprehensive characterization is needed, these findings point to a qualitative difference (and even uniqueness) in how noise and reverberation interference affect the subcortical transcription of speech. A differential neural coding in noise compared to reverberation may help account for the challenges observed by hearing impaired listeners in certain acoustic environments (but not others) as well as the unique types of perceptual confusions listeners experience in noise compared to reverberation (Nabelek and Dagenais 1986). It is possible that signal processing performed by hearing aids and other listener devices may need to be specifically tailored to restore the neurophysiological representation for speech, dependent on the type of interference in a given listening situation (i.e., noisy versus reverberant setting). Future studies are needed (particularly examining reverberation) to better characterize the potentially nuanced effects of different acoustic interferences on speech coding. Given the limited number of FFR studies examining speech in reverberation (Bidelman and Krishnan 2010), the remainder of this chapter will focus mainly on the psychophysiological processing of speech amidst additive noise.

8.3.4 Brain-Behavior Connections Underlying Perceptual SIN

To satisfy a meaningful neural correlate of SIN perceptual skills, changes in the subcortical encoding of speech, as reflected in the FFR, should parallel listener's perceptual performance in noise-degraded listening tasks. Indeed, a growing number of studies have investigated noise-related changes in the speech FFR concurrent with individuals' behavioral SIN measures and have reported robust predictive relationships. Figure 8.7 illustrates brain-behavior correlations between FFR F_0 amplitude (a proxy measure for pitch encoding) and SIN performance on the *QuickSIN^{TM}* test in young, audiometrically normal listeners (Song et al. 2011). Song et al. (2011) showed that FFR F_0 magnitudes positively predicted SIN performance: "Top SIN" performers (see Fig. 8.3) had more robust subcortical responses than "Bottom SIN" performers, who had both weaker neural representation of the speech F_0 and poorer perceptual scores. Complementary findings were reported by Anderson et al. (2010) who showed that poorer (i.e., lower median) SIN listeners experienced greater ($\sim$0.5–1 ms) noise-related shifts in the timing of their speech FFR from quiet to noise than top performing listeners. Collectively, these results suggest a strong relationship between both the magnitude and timing of the brainstem FFR and perceptual SIN skills, whereby faster and more robust subcortical speech encoding is associated with better behavioral outcomes. Bidelman and Krishnan (2010) similarly showed a relationship between FFR encoding of speech in reverberation and behavioral discrimination for pitch and timbre cues whereby more robust neural responses predicted higher perceptual sensitivity.

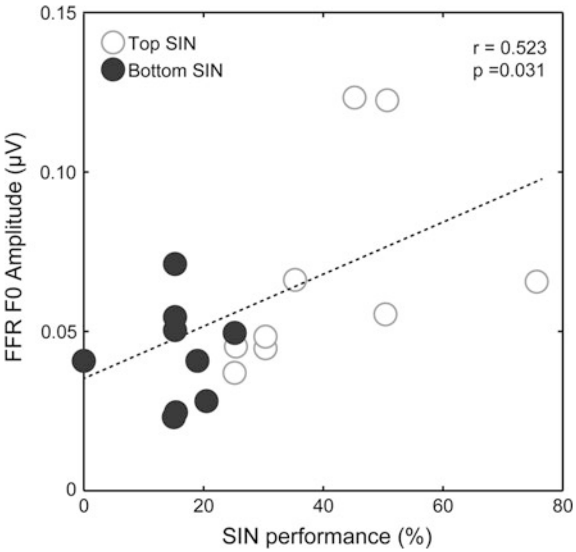


Fig. 8.7 Correlations between speech-in-noise (SIN) listening skills and FFR response properties. FFR F_0 amplitude reflects the brainstem response amplitude during the 40-ms formant transition of the 170-ms syllable/da/, presented in six-talker babble. SIN performance on the QuickSIN^{TM} test (see Fig. 8.3) is associated with larger, more robust subcortical responses to the speech F_0. Top performers on SIN perception behaviorally also show larger brainstem responses to noise-degraded speech. This suggests that individual differences in SIN processing are at least partially accounted for based on how well the auditory nervous system transcribes speech at pre-attentive levels of brain processing. (Data based on Song et al. 2011)

These findings provide evidence for at least a relationship between subcortical auditory processing and SIN perception. However, most studies to date examining the FFR and SIN perception have been correlational in nature. Consequently, it is currently unclear if improved neural encoding of speech (as reflected in the brainstem FFR) causes improved listening skills in noise (but see Sects. 8.4.3, 8.4.4). Moreover, investigators typically manipulate the amount of acoustic information in the stimulus (e.g., SNR) and observe parallel changes in neural responses. In such experimental designs, modulations in the evoked response and human behavior both covary with the acoustic properties of the signal. This confounding of variables further obscures if changes in FFR reflect a true neural correlate of the auditory percept or merely reflect properties of the stimulus itself. This is an important distinction as recent studies employing stimuli that dissociate acoustics from the actual auditory percept suggest that the FFR may not reflect a true neural correlate of the auditory percept but rather reflects more exogenous stimulus properties (Gockel et al. 2011; Bidelman et al. 2013). Regardless of whether FFRs reflect a perceptual correlate of speech phenomena, it is clear that the FFR can be a useful tool in examining individual differences in stimulus coding and auditory function prior to the cerebral cortex.

Although beyond the scope of this chapter, it should also be noted that there are other factors involved in SIN behaviors, including cognitive (post-perceptual) mechanisms. In addition to the encoding (pre-perceptual) factors reviewed here, SIN listening skills are also dependent on certain cognitive abilities, including working memory (Anderson et al. 2010) and selective auditory attention (Ruggles et al. 2011). Nevertheless, the clear connection between FFR response properties and behavioral SIN abilities in young, normal hearing adults implies that at least some of the individual variation in auditory scene analysis is partly determined by how well acoustic information is transcribed at early (and pre-attentive) levels of sensory processing—before the engagement of these cognitive mechanisms (for a review of cognitive and sensory factors, see Anderson et al. 2013a).

8.3.5 Brainstem Versus Cortical Encoding of Degraded Speech

Noise-induced changes in the magnitude and timing of the auditory cortical ERPs have been reported by comparing responses to clean speech sounds relative to noise-degraded speech sounds. The cortical encoding of auditory stimuli amidst noise reflects a complex interaction between the types of signal/noise, as well as the evoking stimulus paradigm (e.g., sequential versus oddball paradigm) (Billings et al. 2010). Component waves of the ERPs (e.g., P1-N1-P2) can be suppressed (i.e., delayed and reduced in amplitude) (Billings et al. 2009, 2010) or facilitated (i.e., enhanced in amplitude) (Alain et al. 2009; Bidelman and Dexter 2015; Bidelman and Howell 2016) depending on the type and effectiveness of a concurrent noise in masking the target signal (e.g., white noise, multi-talker babble). Moreover, like the correlation observed between the FFR and behavior, several studies have shown a relationship between the cortical N1 component latency/amplitude and better SIN perception (Parbery-Clark et al. 2011; Billings et al. 2013; Bidelman and Howell 2016), in the form of earlier/larger responses that are associated with improved behavior. It is conceivable that this type of degraded signal analysis in early auditory cortex is at least partially inherited or influenced by structures much lower in the auditory pathway, including the brainstem (Bidelman and Krishnan 2010; Song et al. 2011) and/or cochlea (Bidelman and Bhagat 2015). This raises the intriguing question of how subcortical FFRs relate to cortical ERPs in the processing of normal and noise-degraded speech.

A handful of studies have directly investigated the correspondence and interactions between brainstem and cortical-evoked activity for clean speech signals (Bidelman et al. 2013, 2014), but few have examined concurrent brainstem-cortical responses under noisy conditions (Parbery-Clark et al. 2011). Parbery-Clark et al. (2011) recorded FFRs and ERPs in normal-hearing listeners in response to the speech syllable /da/presented in quiet and +10 dB SNR of noise (multi-talker babble). Direct comparisons between brainstem and cortical measures showed strong correlations between FFR fidelity (measured via stimulus-to-response correlation) and the cortical N1 magnitude, both of which also related to behavioral SIN perception (i.e., HINT scores). Parbery-Clark et al. (2011) posited that background noise produces a "system-wide" degradation in the neural representation of speech observable across the auditory pathway. Moreover, the strong link between FFR and ERP activity suggested that the neural representations for speech at a cortical level are partially determined (i.e., inherited) from encoding at the level of the brainstem.

Yet, interpreting the relationship between brainstem and cortical speech processing is all but straightforward. While the FFR is generated primarily by deep, brainstem nuclei (Smith et al. 1975; Bidelman 2015b), the cortical ERPs are more diffuse, reflecting overlapping activity generated from multiple sources including the auditory cortices and contributions from the frontal lobes (Knight et al. 1999; Picton et al. 1999). This blurring of sources in the scalp potentials precludes firm interpretations between auditory ERPs and SIN perception, as behaviors cannot be definitively attributed to changes in a single auditory (or nonauditory) brain structure. Indeed, analysis of source-level (as opposed to sensory-level) neural responses reveals that both primary and inferior frontal regions (e.g., proximal to Broca's area) are engaged during noise-degraded speech processing (Bidelman and Dexter 2015; Bidelman and Howell 2016). Moreover, it is often neural activity in these frontal (rather than lemniscal) auditory brain regions that best predicts SIN behaviors (Wong et al. 2008; Bidelman and Dexter 2015). How brainstem activity exactly relates to auditory and perhaps nonauditory cerebral processing remains to be seen.

8.4 Experiential and Clinical Influences on SIN Abilities

8.4.1 Hearing Impairment and Auditory Disorders

Perceiving speech in noise is particularly problematic for older adults and listeners with hearing loss who often experience poorer speech understanding than would be predicted by their hearing thresholds (Glasberg and Moore 1989; Gordon-Salant and Fitzgibbons 1993; also see Anderson, Chap. 11). Disentangling the contributions of aging and hearing loss to auditory function is challenging as these factors typically covary over the lifespan and are often comorbid with other age-related deficits (e.g., cognitive decline). Nevertheless, several studies suggest that age and hearing loss have independent effects on the FFR encoding of complex sounds (Clinard et al. 2010; Marmel et al. 2013; Bidelman et al. 2014). Aging alone tends to weaken and delay the neural encoding of voice pitch and timbre cues coded by the FFR (Anderson et al. 2012; Bidelman et al. 2014), resulting in a less precise neural transcription of the speech waveform. These age-related changes in FFR phase locking (often most prominent at F_0) reflect a reduction in neural synchronization. Age-related changes in the precision of synchronization may be related to declines in neural inhibition (Caspary et al. 2008) and increased deafferentation (Kujawa and Liberman 2006; Makary et al. 2011) that occur along the aging mammalian auditory nervous system.

Hearing impaired listeners' difficulty understanding speech in noise is thought to reflect an inability to use temporal fine structure of speech (Lorenzi et al. 2006), a prominent cue for pitch perception and signal extraction in noise. Given that the FFR reflects phase-locked activity to the spectral fine structure of complex sounds, it provides a useful objective assay to track changes in auditory function that accompany hearing loss. Indeed, FFR studies reveal that even for clean speech, listeners with sensorineural hearing loss show poorer representations for the onset (Anderson et al. 2010; Bidelman et al. 2014), formant transitions (Plyler and Ananthanarayan 2001; Anderson et al. 2010), and upper harmonics of speech sounds (Anderson et al. 2013c; Bidelman et al. 2014). In addition, hearing loss has a differential impact on how the auditory system extracts envelope versus fine-structure cues. Paralleling single-unit responses in animal models (Kale and Heinz 2010), the population FFR response shows an over-exaggeration in the stimulus envelope and imbalance in the normal encoding of envelope and fine-structure cues in hearing impaired humans (Anderson et al. 2013c). Distortions between speech envelope and fine structure in terms of both their neural encoding and perception may partly underlie deficits in SIN perception in hearing-impaired listeners that are not accounted for by purely reduced audibility.

Deficiencies in brainstem encoding for noise-degraded speech have also been observed in individuals with auditory processing disorders (Billiet and Bellis 2011; also see Schochat, Rocha-Muniz, and Filippini, Chap. 9), language-learning impairments (Cunningham et al. 2001; Banai et al. 2009), and autism (Russo et al. 2008). These deficits are not always apparent in neural responses to clean speech. From a clinical standpoint, this suggests that some central auditory processing disorders might be best detected when the auditory system is taxed by challenging listening conditions (e.g., noise or reverberation) or is required to perform difficult figure-ground analysis (see Reetzke, Xie, and Chandrasekaran, Chap. 10). In this regard, the FFR may provide a useful "biomarker" for identifying certain central auditory processing deficits using an objective electrophysiological technique.

8.4.2 Language Experience

Seminal FFR studies have demonstrated that extensive auditory experiences introduce functional reorganization in the human midbrain, as reflected in the FFR. First examined in the context of language, studies have shown that long-term experience with a tonal language—in which changes in pitch alter word meaning—enhances the subcortical representation of pitch-relevant information as indicated by the smoother, more robust voice fundamental-frequency tracking in the FFRs of native Chinese relative to English-speaking listeners (Krishnan et al. 2005; Krishnan and Gandour 2009). Krishnan et al. (2005) provided the first evidence that (1) long-term auditory experience can enhance subcortical auditory processing in an experience-dependent manner, and (2) that neuroplasticity extends below cerebral cortex. The effects of life-long language experience on subcortical FFRs are treated more thoroughly elsewhere in this volume (see Krishnan and Gandour, Chap. 3). However, an interesting question that emerges from those studies is whether these cross-linguistic enhancements in brainstem auditory processing confer any advantage to signal-in-noise extraction and SIN perception.

To address this question, Krishnan et al. (2010) measured FFRs in native Chinese and English-speaking listeners in response to a dynamic pitch pattern modeled after the rising lexical tone in the Mandarin language (i.e., “T2”). The tone was synthesized using *iterated rippled noise* (IRN), which allowed for the parametric control of the tone’s degree of periodicity (“noisiness”) by varying the number of iterations in the IRN generation circuit (Yost 1996). Both clean and degraded T2 responses were recorded and “pitch tracking accuracy” was used to quantify the degree to which neural FFRs followed the stimulus pitch contour. Results showed that Chinese FFRs show more faithful response tracking of the T2 contour in clean and noise-degraded conditions compared to English listeners (Fig. 8.8). This illustrates that long-term language experience enhances not only the fidelity of brainstem pitch processing for clean signals but also confers advantages to extracting degraded speech information. The fact that noise-related enhancements were not observed in non-native listeners

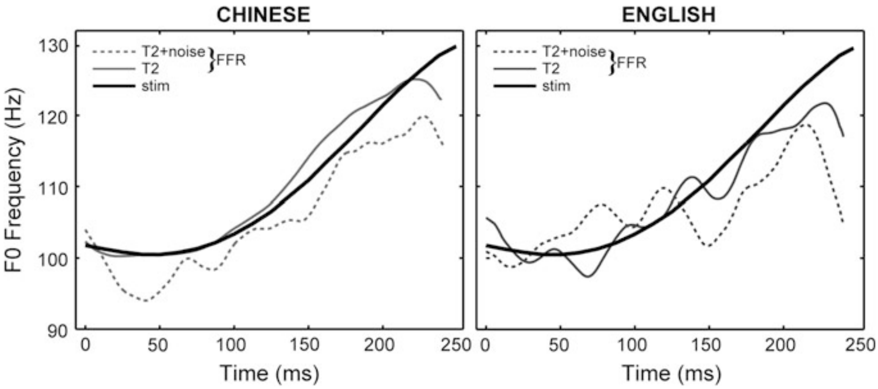


Fig. 8.8 Tone-language experience improves the fidelity of linguistic pitch processing and its extraction from noise. FFRs were measured in response to rising linguistic pitch pattern (T2) found in Mandarin Chinese. Linguistic pitches were created using dynamic iterated rippled noise (IRN) stimuli and parametrically varied in their “noisiness” by varying the number of iterations in the IRN generation. Fewer iterations disrupt the periodicity of the IRN pitch and result in a noisy pitch percept; more iterations result in clearer pitch percepts. Traces represent the pitch track (i.e., F_0 contour) extracted from the stimulus (*thick line*) and corresponding FFRs (*thin and dashed lines*) recorded from native Chinese and nonnative English speaking listeners elicited by clean and noise-degraded versions of the Mandarin tone. Chinese FFRs show more faithful response tracking of the T2 contour in clean and noise-degraded conditions compared to English listeners. This illustrates that long-term language experience can enhance the fidelity with which the brainstem encodes important acoustic features of listeners’ native language. *FO*, F_0; *stim*, stimulus. (Data based on Krishnan et al. 2010)

further suggests that brainstem enhancements are restricted to specific features of speech that are most congruent with a listeners' native language (Krishnan et al. 2009; Bidelman et al. 2011).

Somewhat at odds with the aforementioned brainstem findings are behavioral and cortical evoked-potential studies, which often show the opposite effect: a clear bilingual disadvantage for SIN processing. Behavioral studies, for instance, have consistently shown that bilinguals show much poorer (i.e., upwards of ~10 dB SNR difference) SIN perception for their second (non-native) language relative to monolinguals (Rogers et al. 2006; Bidelman and Dexter 2015). Moreover, in response to noise-degraded speech contrasts, cortical mismatch negativity responses—reflecting the brain's automatic differentiation of speech—are both prolonged and weaker in bilingual compared to monolingual speakers (Bidelman and Dexter 2015). Thus, bilinguals' more faithful encoding of noise-degraded speech observed at the level of the brainstem (Krishnan et al. 2010; Krizman et al. 2012) does not seem to be paralleled in higher level responses at a cortical level or perception in non-native listeners.

The dissociable effects of SIN processing between brainstem and cortex indicate that bilinguals' sensory gains in auditory processing are, by themselves, likely insufficient to improve perceptual SIN. Differences in brainstem-cortical speech processing could reflect the fact that complex listening skills like SIN require additional engagement of higher-order cortical brain areas more specialized for language processing that override any sensory gains from the brainstem. Alternatively, if FFR representations reflect primarily stimulus properties rather than the behavioral percept (Gockel et al. 2011; see Sect. 8.3.5), dissociations between brainstem and cortical speech processing might be expected. More research is needed using integrative brainstem-cortical recording paradigms (e.g., Bidelman 2015a) to clarify the role of subcortical and cortical activity in SIN processing as well as potential transformations between these two stages of auditory signal processing (Bidelman et al. 2013).

8.4.3 Musical Experience

A detailed treatment of the effects of musical experience on brainstem FFRs, speech processing, and listening skills is reviewed elsewhere in this volume (see White-Schwoch and Kraus, Chap. 6). In relation to degraded speech perception, several studies have noted that the neural enhancements in quiet/clean speech encoding observed for musically trained individuals also extend to the processing of acoustically degraded speech sounds (Parbery-Clark et al. 2009; Bidelman and Krishnan 2010; Zendel and Alain 2012). Parbery-Clark et al. (2009) first reported a "musician advantage" for SIN listening. They found correlations between the speech-FFR onset latency, transition latency (i.e., timing during formant transition period), and fidelity (i.e., stimulus-to-response correlations) and behavioral measures of SIN. Similar neural and behavioral advantages have been observed for musicians

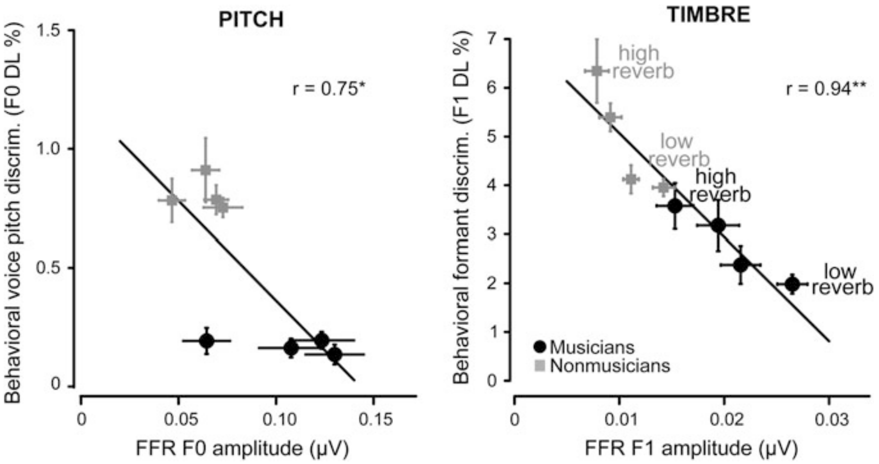


Fig. 8.9 Effects of musical training on behavioral and neural processing of degraded speech. Frequency-following responses (FFRs) were recorded in listeners with (>10 years) and without (<3 years) formal musical training in response to speech sounds varying in the severity of concurrent reverberation (see Fig. 8.4). Additionally, behavioral difference limens (DLs) were obtained for voice fundamental frequency ($FO = F_0$) and first formant (F1) discrimination to index behavior acuity for voice "pitch" and "timbre" elements of speech, respectively. The magnitude of F_0 and F1 encoding in reverberation, as indexed by the FFR, strongly predicts perceptual discrimination for both speech cues; larger, more robust subcortical activity is associated with better (i.e., smaller) DLs. Group comparisons demonstrate that musicians have both stronger neurological coding of these speech cues and improved behavioral discrimination in reverberation. *Points* show the group mean responses across dry, mild, medium, and severe reverberation conditions. *Error bars* reflect ±1 s.e.m. in either the neural or behavioral dimension. (Based on unpublished data from Bidelman and Krishnan 2010)

in reverberation (Bidelman and Krishnan 2010) (Fig. 8.9). Both pitch (F_0) and timbre (F1 formant) magnitudes are stronger in the FFRs of listeners with musical expertise (>10 years formal training) compared to their nonmusician peers (<3 years training). Importantly, neural enhancements in the FFR are well-correlated with behavioral discrimination of these cues for both groups, but they are more highly coupled to behavior in musicians.

Collectively, these studies indicate that musicians might parse and segregate competing signals in complex auditory scenes more effectively (Munte et al. 2001; Zendel and Alain 2009). Musicians are also less influenced by information masking (Oxenham et al. 2003) than nonmusicians. Given the importance of these factors in auditory scene analysis (e.g., "cocktail party" scenarios), it is possible that musical expertise improves important aspects of real-world listening required for robust communication (Alain et al. 2014). While most studies to date have only assessed musicians' SIN advantages via cross-sectional comparisons, newer training studies with randomized group assignments are promising in that they show improvement in FFR and SIN listening skills with 1–2 years of musical training (Kraus et al. 2014; Slater et al. 2015).

8.4.4 Short-Term Auditory Training and Perceptual Learning

The putative benefits of long-term auditory experiences (e.g., bilingualism, musicianship) on the FFR and SIN provoke the question of whether or not short-term training regimens might similarly boost SIN processing skills. In this regard, there have been few investigations of how auditory training impacts SIN perception (Burk and Humes 2007; Yund and Woods 2010). Of the reports showing successful training effects, gains are often modest (improvements of a few dB), fail to generalize to untrained material (lack generalizability outside the laboratory), or quickly fade (no long-term retention) (also see Plack and Carcagno, Chap. 4).

Nevertheless, a handful of recent studies have demonstrated that short-term auditory training can improve FFR encoding of degraded speech as well as SIN perception. In particular, Song et al. (2012) assessed speech FFRs to clean and noise-degraded speech before and after training on the *Listening and Communication Enhancement* (LACE®) program (Sweetow and Sabes 2006). LACE is a commercially available auditory training suite that tests users' open-set speech recognition in various listening conditions and is often provided to hearing aid patients for acclimatization after the first fitting. FFRs and SIN perception were measured before and after LACE training (twenty, 30-min sessions over 4 weeks) as well as 6 months for a long-term follow-up. A group of noncontact, control listeners was measured at similar intervals but did not participant in the training schedule. Results showed improvements in both the neural encoding of voice pitch (F_0) and upper harmonics of speech in the trained group (but not the control group) coupled with improvements in SIN perception (*QuickSIN™* and HINT scores). Interestingly, training-related enhancements were correlated with changes in the FFR and persisted at six months after the cessation of the program. The findings of Song et al. (2012) provide provocative evidence that even short-term auditory training can improve SIN listening skills in young adults and might be mediated, at least partially, by experience-dependent mechanisms of the auditory brainstem.

Whether similar training-related benefits are observed for older adults—for whom SIN issues are more problematic—remains somewhat equivocal. Anderson et al. (2013b) trained older adults on an adaptive, computer-based auditory training program for eight weeks. Results showed that the normal timing deficits (i.e., prolonged responses) observed in older adults' speech-FFR with noise (Anderson et al. 2010) is partially ameliorated with training. Trained older adults showed faster FFR timing, less noise-related change, and less trial-to-trial jitter in brainstem activity to speech sounds than age-matched controls. Unfortunately, subsequent follow-up testing at six months suggested that while training-related improvements in FFR responses were maintained, perceptual SIN gains did not persist (Anderson et al. 2014). Thus, while SIN training might be successful for younger adults, it is currently unclear what dosage of training is required to achieve the same longevity (i.e., "sticking power") of SIN benefits in older listeners. It is possible that the aged auditory nervous system is less plastic than earlier in life (Stiles 2000); hence, auditory training may not yield the same degree of neurological and listening benefits in older individuals compared to younger individuals.

8.5 Summary

This chapter has described several perceptual and neurophysiological consequences of real-world human communication in noise and reverberant environments. Both forms of acoustic interference have similar detriments to behavioral recognition of speech but each challenges the neural coding of speech sounds in different manners. In this regard, the scalp-recorded FFR has provided considerable insight into the neural representations of speech at early, pre-attentive levels of the auditory pathway and how such representations are altered in challenging listening conditions. Noise largely masks the spectral details of speech, reducing the contrast between perceptually salient frequency characteristics (e.g., formants) and the surrounding noise floor. In contrast, the effect of reverberation is to smear spectrotemporal details of a signal, producing a temporal overlap (i.e., self-masking) of time-frequency information in the running speech signal. These acoustic consequences are closely paralleled in human FFRs, which similarly show changes in "neural SNR" and spectral smearing with additive noise and reverberation, respectively. Yet, even within the speech signal, acoustic interferences do not produce uniform impairments. To a certain extent, voice pitch (F_0) cues of speech are largely immune to noise/reverberation interferences, whereas timbral cues (i.e., higher harmonics and spectral envelope) degrade quickly with increasing noise/reverberation.

The subcortical encoding of speech and SIN listening skills are not static. Rather they naturally decline across the lifespan and are impaired with certain auditory disorders. Promisingly, the neural encoding and perception of SIN can be enhanced by aspects of the human experience (e.g., language and musical experience) and short-term auditory training. Neuroplastic changes within training suggest that the brainstem plays an active (but underappreciated) role in molding functional properties of the auditory system. Moreover, perceptual and neurobiological changes in FFR imply that impoverished auditory sensory coding, resulting from aging, hearing impairment, or central auditory processing disorders, might be partially counteracted with certain forms of auditory learning and experience.

Several empirical and theoretical questions of the FFR were raised but remain unanswered. First, while a number of studies have focused on the brainstem encoding of speech in noise (e.g., Song et al. 2011; Prevost et al. 2013), at present there is an unfortunate dearth of studies examining how reverberation changes the FFR coding of speech (Bidelman and Krishnan 2010). Nevertheless, studies reviewed herein reveal a possible differential neural coding of speech in noise compared to reverberation. Conceivably, the unique impact of noise versus reverberation on speech processing may help account for the challenges observed by hearing impaired listeners in certain acoustic environments but not others. Indeed, while both noise and reverberation have a negative impact of similar magnitude on speech perception, their effects are reflected in different types of perceptual errors (e.g., confusions) (Nabelek and Dagenais 1986) and associated differences in neural coding (Bidelman and Krishnan 2010). Future studies are warranted that directly compare noise and reverberation. Such contrasts will surely reveal a more nuanced understanding of the effects of acoustic interference(s) on auditory brainstem processing and might provide new insight into both specific deficits and biomarkers of degraded speech perception abilities.

Secondly, while brainstem response properties often covary with perceptual abilities in SIN tasks, it is often unclear how (and even if) neural enhancements seen in the FFR (e.g., faster timing, more robust magnitudes) are causally related to improvements in behavior. While FFR studies do suggest experience-dependent effects in brainstem, auditory-processing training studies will help clarify cross-sectional findings that dominate the literature. These studies are beginning to reveal causal links between biological changes in degraded speech coding with behavior (e.g., Song et al. 2012; Slater et al. 2015), suggesting that improvements in SIN skills are directly related to enhancements in the brainstem's transcription of speech.

Thirdly, it is possible that perhaps subcortical FFRs do not carry a perceptual code, but rather they reflect mainly stimulus properties (Gockel et al. 2011; Bidelman et al. 2013). Experimental designs that pit stimulus acoustics at least partially orthogonal to the resulting percept (e.g., as in categorical perception: Bidelman et al. 2013; Bidelman and Alain 2015) might be used to clarify exogenous and endogenous contributions to the FFR that are not yet fully clarified.

Finally, while the FFR has provided a rich window into how the human auditory system encodes speech acoustics, it is but a single snapshot of the auditory brain at work. A holistic account of human communication in challenging environments would presumably need to account for both sensory and cognitive mechanisms subserving SIN processing (e.g., Anderson et al. 2013a). Moreover, how spectrotemporal features coded by the FFR are transformed as they arrive in auditory cortex and how brainstem following responses relate to the more diffuse waves of the cortical ERPs remain to be elucidated (Bidelman et al. 2013). Integrative electrophysiological paradigms that include simultaneous recordings of brainstem and cortical-evoked activity are needed (Bidelman 2015a) to help clarify how the auditory brainstem works in concert with auditory (and nonauditory) brain regions to analyze the auditory scene.

Acknowledgments Portions of this work were supported by grants from the American Hearing Research Foundation (AHRF) and American Academy of Audiology Foundation (AAF) awarded to G.M.B. The author thanks Amy Fehrenbach, Jill Lowther, and Sarah Lawson for comments on an earlier version of this manuscript.

Compliance with Ethics Requirements G. Bidelman received grants from the American Hearing Research Foundation (AHRF) and American Academy of Audiology Foundation (AAF) to support portions of this work.

References

Akhoun, I., Gallego, S., Moulin, A., Menard, M., et al. (2008). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clinical Neurophysiology, 119*(4), 922–933.

Alain, C., Quan, J., McDonald, K., & Van Roon, P. (2009). Noise-induced increase in human auditory evoked neuromagnetic fields. *European Journal of Neuroscience, 30*(1), 132–142.

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.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., Drehobl, S., & Kraus, N. (2013a). Effects of hearing loss on the subcortical representation of speech cues. *Journal of the Acoustical Society of America, 133*(5), 3030–3038.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2012). Aging affects neural precision of speech encoding. *The Journal of Neuroscience, 32*(41), 14156–14164.

Anderson, S., Skoe, E., Chandrasekaran, B., & Kraus, N. (2010). Neural timing is linked to speech perception in noise. *The Journal of Neuroscience, 30*(14), 4922–4926.

Anderson, S., White-Schwoch, T., Choi, H. J., & Kraus, N. (2014). Partial maintenance of auditory-based cognitive training benefits in older adults. *Neuropsychologia, 62*, 286–296.

Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013b). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hearing Research, 300*, 18–32.

Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013c). Reversal of age-related neural timing delays with training. *Proceedings of the National Academy of Sciences of the USA, 110*(11), 4357–4362.

ANSI. (1969). ANSI American national standards methods for the calculation of the articulation index. ANSI S3.5-69-1969. New York.

Assmann, P. F., & Summerfield, Q. (1989). Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency. *The Journal of the Acoustical Society of America, 85*(1), 327–338.

Assmann, P. F., & Summerfield, Q. (1990). Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *The Journal of the Acoustical Society of America, 88*(2), 680–697.

Assmann, P. F., & Summerfield, A. Q. (Eds.). (2004). *The perception of speech under adverse conditions*. New York: Springer Science+Business Media.

Backus, J. (1977). *The acoustical foundations of music* (2nd ed.). New York: Norton.

Banai, K., Hornickel, J., Skoe, E., Nicol, T., et al. (2009). Reading and subcortical auditory function. *Cerebral Cortex, 19*(11), 2699–2707.

Bidelman, G. M. (2015a). Towards an optimal paradigm for simultaneously recording cortical and brainstem auditory evoked potentials. *The Journal of Neuroscience Methods, 241*, 94–100.

Bidelman, G. M. (2015b). Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR. *Hearing Research, 323*, 68–80.

Bidelman, G. M., & Alain, C. (2015). Hierarchical neurocomputations underlying concurrent sound segregation: Connecting periphery to percept. *Neuropsychologia, 68*, 38–50.

Bidelman, G. M., & Bhagat, S. P. (2015). Right ear advantage drives the link between olivocochlear efferent "antimasking" and speech-in-noise listening benefits. *NeuroReport, 26*, 483–487.

Bidelman, G. M., & Dexter, L. (2015). Bilinguals at the "cocktail party": Dissociable neural activity in auditory-linguistic brain regions reveals neurobiological basis for nonnative listeners' speech-in-noise recognition deficits. *Brain and Language, 143*, 32–41.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience, 23*(2), 425–434.

Bidelman, G. M., & Howell, M. (2016). Functional changes in inter- and intra-hemispheric auditory cortical processing underlying degraded speech perception. *NeuroImage, 124*, 581–590.

Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research, 1355*, 112–125.

Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage, 79*(1), 201–212.

Bidelman, G. M., Villafuerte, J. W., Moreno, S., & Alain, C. (2014). Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiology of Aging, 35* (11), 2526–2540.

Billiet, C. R., & Bellis, T. J. (2011). The relationship between brainstem temporal processing and performance on tests of central auditory function in children with reading disorders. *Journal of Speech, Language, and Hearing Research, 54*, 228–242.

Billings, C. J., Bennett, K. O., Molis, M. R., & Leek, M. R. (2010). Cortical encoding of signals in noise: Effects of stimulus type and recording paradigm. *Ear and Hearing, 32*(1), 53–60.

Billings, C. J., McMillan, G. P., Penman, T. M., & Gille, S. M. (2013). Predicting perception in noise using cortical auditory evoked potentials. *Journal of the Association for Research in Otolaryngology, 14*(6), 891–903.

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*(1–2), 15–24.

Burk, M. H., & Humes, L. E. (2007). Effects of training on speech recognition performance in noise using lexically hard words. *Journal of Speech, Language, and Hearing Research, 50*, 25–40.

Burkard, R., & Hecox, K. (1983). The effect of broadband noise on the human brainstem auditory evoked response. I. Rate and intensity effects. *The Journal of the Acoustical Society of America, 74*(4), 1204–1213.

Carhart, R. (1951). Basic principles of speech audiometry. *Acta Oto-Laryngologica, 40*, 62–71.

Carhart, R., & Tillman, T. W. (1970). Interaction of competing speech signals with hearing losses. *Archives of Otolaryngology, 91*, 274–279.

Cariani, P. A., & Delgutte, B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology, 76*(3), 1698–1716.

Caspary, D. M., Ling, L., Turner, J. G., & Hughes, L. F. (2008). Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *Journal of Experimental Biology and Medicine, 211*, 1781–1791.

Chmiel, R., & Jerger, J. (1996). Hearing aid use, central auditory disorder, and hearing handicap in elderly persons. *Journal of the American Academy of Audiology, 7*(3), 190–202.

Clinard, C. G., Tremblay, K. L., & Krishnan, A. R. (2010). Aging alters the perception and physiological representation of frequency: Evidence from human frequency-following response recordings. *Hearing Research, 264*(1–2), 48–55.

Cunningham, J., Nicol, T., King, C., Zecker, S. G., & Kraus, N. (2002). Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. *Hearing Research, 169*, 97–111.

Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., & Kraus, N. (2001). Neurobiologic responses to speech in noise in children with learning problems: Deficits and strategies for improvement. *Clinical Neurophysiology, 112*(5), 758–767.

Dau, T. (2003). The importance of cochlear processing for the formation of auditory brainstem and frequency following responses. *The Journal of the Acoustical Society of America, 113*(2), 936–950.

Drgas, S., & Blaszak, M. A. (2009). Perceptual consequences of changes in vocoded speech parameters in various reverberation conditions. *Journal of Speech, Language, and Hearing Research, 52*(4), 945–955.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Finitzo-Hieber, T., & Tillman, T. (1978). Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children. *Journal of Speech and Hearing Research, 21*, 440–458.

French, N., & Steinberg, J. (1947). Factors governing the intelligibility of speech sounds. *The Journal of the Acoustical Society of America, 19*, 90–119.

Frisina, D. R., & Frisina, R. D. (1997). Speech recognition in noise and presbycusis: Relations to possible neural mechanisms. *Hearing Research, 106*(1–2), 95–104.

Galbraith, G., Arbagey, P. W., Branski, R., Comerci, N., & Rector, P. M. (1995). Intelligible speech encoded in the human brain stem frequency-following response. *NeuroReport, 6*(17), 2363–2367.

Galbraith, G., Olfman, D. M., & Huffman, T. M. (2003). Selective attention affects human brain stem frequency-following response. *NeuroReport, 14*(5), 735–738.

Gelfand, S. A., & Silman, S. (1979). Effects of small room reverberation upon the recognition of some consonant features. *The Journal of the Acoustical Society of America, 66*(1), 22–29.

George, E. L., Festen, J. M., & Houtgast, T. (2008). The combined effects of reverberation and nonstationary noise on sentence intelligibility. *The Journal of the Acoustical Society of America, 124*(2), 1269–1277.

Glasberg, B. R., & Moore, B. C. J. (1989). Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. *Scandinavian Audiology, 32*, 1–25.

Gockel, H. E., Carlyon, R. P., Mehta, A., & Plack, C. J. (2011). The frequency following response (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. *Journal of the Association for Research in Otolaryngology, 12*(6), 767–782.

Gordon-Salant, S., & Fitzgibbons, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech and Hearing Research, 36*, 1276–1285.

Greenberg, S., Marsh, J. T., Brown, W. S., & Smith, J. C. (1987). Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hearing Research, 25*(2–3), 91–114.

Helfer, K., & Wilber, L. (1990). Hearing loss, aging, and speech perception in reverberation and in noise. *Journal of Speech and Hearing Research, 33*, 149–155.

Henry, K. R. (1999). Noise improves transfer of near-threshold, phase-locked activity of the cochlear nerve: Evidence for stochastic resonance? *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 184*(6), 577–584.

Humes, L. E., & Christopherson, L. (1991). Speech identification difficulties of hearing-impaired elderly persons: The contributions of auditory processing deficits. *Journal of Speech and Hearing Research, 34*(3), 686–693.

Humes, L. E., & Roberts, L. (1990). Speech-recognition difficulties of the hearing-impaired elderly: The contributions of audibility. *Journal of Speech and Hearing Research, 33*(4), 726–735.

Jaramillo, F., & Wiesenfeld, K. (1998). Mechanoelectrical transduction assisted by Brownian motion: A role for noise in the auditory system. *Nature Neuroscience, 1*, 384–388.

Kale, S., & Heinz, M. G. (2010). Envelope coding in auditory nerve fibers following noise-induced hearing loss. *Journal of the Association for Research in Otolaryngology, 11* (4), 657–673.

Kiang, N. Y., & Moxon, E. C. (1973). Tails of tuning curves of auditory-nerve fibers. *The Journal of the Acoustical Society of America, 55*(2), 620–630.

Killion, M., & Niquette, P. (2000). What can the pure-tone audiogram tell us about a patient's SNR loss? *The Hearing Journal, 53*(3), 46–53.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 116*(4 Pt 1), 2395–2405.

Kinsler, L. E., Frey, A. R., Coppens, A. B., & Sanders, J. V. (2000). *Fundamentals of acoustics*. New York: Wiley.

Knight, R. T., Staines, W. R., Swick, D., & Chao, L. L. (1999). Prefrontal cortex regulates inhibition and excitation in distributed neural networks. *Acta Psychologica, 101*(2–3), 159–178.

Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., et al. (2014). Music enrichment programs improve the neural encoding of speech in at-risk children. *The Journal of Neuroscience, 34* (36), 11913–11918.

Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language, 110*(3), 135–148.

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. *Brain Research, 1313*, 124–133.

Krishnan, A., Gandour, J. T., Bidelman, G. M., & Swaminathan, J. (2009). Experience-dependent neural representation of dynamic pitch in the brainstem. *NeuroReport, 20*(4), 408–413.

Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25*(1), 161–168.

Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences of the USA, 109*(20), 7877–7881.

Krizman, J., Skoe, E., & Kraus, N. (2010). Stimulus rate and subcortical auditory processing of speech. *Audiology and Neuro-Otology, 15*, 332–342.

Kujawa, S. G., & Liberman, M. C. (2006). Acceleration of age-related hearing loss by early noise exposure: Evidence of a misspent youth. *The Journal of Neuroscience, 26*(7), 2115–2123.

Laroche, M., Dajani, H. R., Prevost, F., & Marcoux, A. M. (2013). Brainstem auditory responses to resolved and unresolved harmonics of a synthetic vowel in quiet and noise. *Ear and Hearing, 34*(1), 63–74.

Larsen, E., Iyer, N., Lansing, C. R., & Feng, A. S. (2008). On the minimum audible difference in direct-to-reverberant energy ratio. *The Journal of the Acoustical Society of America, 124*(1), 450–461.

Li, X., & Jeng, F.-C. (2011). Noise tolerance in human frequency-following responses to voice pitch. *The Journal of the Acoustical Society of America, 129*, 21–26.

Lifshitz, S. (1925). Optimum reverberation for an auditorium. *Physical Review, 25*(3), 391–394.

Liu, C., & Kewley-Port, D. (2004). Formant discrimination in noise for isolated vowels. *The Journal of the Acoustical Society of America, 116*(5), 3119–3129.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences of the USA, 103*(49), 18866–18869.

Makary, C. A., Shin, J., Kujawa, S. G., Liberman, M. C., & Merchant, S. N. (2011). Age-related primary cochlear neuronal degeneration in human temporal bones. *Journal of the Association for Research in Otolaryngology, 12*(6), 711–717.

Marmel, F., Linley, D., Carlyon, R. P., Gockel, H. E., et al. (2013). Subcortical neural synchrony and absolute thresholds predict frequency discrimination independently. *Journal of the Association for Research in Otolaryngology, 14*, 755–766.

McShefferty, D., Whitmer, W. M., & Akeroyd, M. A. (2015). The just-noticeable difference in speech-to-noise ratio. *Trends in Hearing, 19*, 1–9.

Middelweerd, M. J., Festen, J. M., & Plomp, R. (1990). Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. *Audiology, 29*(1), 1–7.

Munte, T. F., Kohlmetz, C., Nager, W., & Altenmuller, E. (2001). Superior auditory spatial tuning in conductors. *Nature, 409*(6820), 580.

Nabelek, A. K. (1988). Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss. *The Journal of the Acoustical Society of America, 84*(2), 476–484.

Nabelek, A. K., & Dagenais, P. A. (1986). Vowel errors in noise and in reverberation by hearing-impaired listeners. *The Journal of the Acoustical Society of America, 80*(3), 741–748.

Nabelek, A. K., & Letowski, T. R. (1985). Vowel confusions of hearing-impaired listeners under reverberant and nonreverberant conditions. *Journal of Speech and Hearing Disorders, 50*(2), 126–131.

Nabelek, A. K., & Letowski, T. R. (1988). Similarities of vowels in nonreverberant and reverberant fields. *The Journal of the Acoustical Society of America, 83*(5), 1891–1899.

Nabelek, A. K., Letowski, T. R., & Tucker, F. M. (1989). Reverberant overlap- and self-masking in consonant identification. *The Journal of the Acoustical Society of America, 86*(4), 1259–1265.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America, 95*(2), 1085–1099.

Oxenham, A. J., Fligor, B. J., Mason, C. R., & Kidd, G., Jr. (2003). Informational masking and musical training. *The Journal of the Acoustical Society of America, 114*(3), 1543–1549.

Parbery-Clark, A., Marmel, F., Bair, J., & Kraus, N. (2011). What subcortical-cortical relationships tell us about processing speech in noise. *European Journal of Neuroscience, 33*(3), 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*(45), 14100–14107.

Picton, T. W., Alain, C., Woods, D. L., John, M. S., et al. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiology and Neuro-Otology, 4*(2), 64–79.

Picton, T. W., Woods, D. L., Baribaeu-Braun, J., & Healy, T. M. G. (1977). Evoked potential audiometry. *Journal of Otolaryngology, 6*(2), 90–119.

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *The Journal of the Acoustical Society of America, 63*, 533–549.

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech and Hearing Research, 29*(2), 146–154.

Plyler, P. N., & Ananthanarayan, A. K. (2001). Human frequency-following responses: Representation of second formant transitions in normal-hearing and hearing-impaired listeners. *Journal of the American Academy of Audiology, 12*(10), 523–533.

Popelka, G. R., Moore, B. C. J., Popper, A. N., & Fay, R. R. (2016). *Hearing Aids*. New York: Springer Science+Business Media.

Prevost, F., Laroche, M., Marcoux, A. M., & Dajani, H. R. (2013). Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel. *Clinical Neurophysiology, 124*(1), 52–60.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics, 27*(03), 465–485.

Ruggles, D. R., Bharadwaj, H., & Shinn-Cunningham, B. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proceedings of the National Academy of Sciences of the USA, 108*(37), 15516–15521.

Russo, N. M., Bradlow, A. R., Skoe, E., Trommer, B. L., et al. (2008). Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clinical Neurophysiology, 119*(8), 1720–1731.

Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology, 115*(9), 2021–2030.

Sabine, W. (1962). *Collected Papers on Acoustics*. Los Altos Hills, CA: Dover.

Sayles, M., & Winter, I. M. (2008). Reverberation challenges the temporal representation of the pitch of complex sounds. *Neuron, 58*(5), 789–801.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303–304.

Slater, J., Skoe, E., Strait, D., O'Connell, S., et al. (2015). Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. *Behavioural Brain Research, 291*, 244–252.

Smalt, C. J., Krishnan, A., Bidelman, G. M., Ananthakrishnan, S., & Gandour, J. T. (2012). Distortion products and their influence on representation of pitch-relevant information in the human brainstem for unresolved harmonic complex tones. *Hearing Research, 292*(1–2), 26–34.

Smith, J. C., Marsh, J. T., & Brown, W. S. (1975). Far-field recorded frequency-following responses: Evidence for the locus of brainstem sources. *Electroencephalography and Clinical Neurophysiology, 39*(5), 465–472.

Sohmer, H., & Pratt, H. (1977). Identification and separation of acoustic frequency following responses (FFRs) in man. *Electroencephalography and Clinical Neurophysiology, 42*(4), 493–500.

Sohmer, H., Pratt, H., & Kinarti, R. (1977). Sources of frequency-following responses (FFR) in man. *Electroencephalography and Clinical Neurophysiology, 42*, 656–664.

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2011). Perception of speech in noise: Neural correlates. *Journal of Cognitive Neuroscience, 23*(9), 2268–2279.

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2012). Training to improve hearing speech in noise: Biological mechanisms. *Cerebral Cortex, 22*(5), 1180–1190.

Stiles, J. (2000). Neural plasticity and cognitive development. *Developmental Neuropsychology, 18*(2), 237–272.

Swaminathan, J., & Heinz, M. G. (2012). Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. *The Journal of Neuroscience, 32*(5), 1747–1756.

Sweetow, R. W., & Sabes, J. H. (2006). The need for and development of an adaptive listening and communication enhancement (LACE) program. *Journal of the American Academy of Audiology, 17*(8), 538–558.

von Békésy, G. (1938). Über die Enstehung der Entfernungsempfindung beim Hören (On the origin of distance perception in hearing). *Akustische Zeitschrift, 3*, 21–31.

Wang, D. L., & Brown, G. J. (2006). *Computational auditory scene analysis*. Hoboken, NJ: Wiley.

Watkins, A. J. (2005). Perceptual compensation for effects of reverberation in speech identification. *The Journal of the Acoustical Society of America, 118*(1), 249–262.

Weiss, M. W., & Bidelman, G. M. (2015). Listening to the brainstem: Musicianship enhances intelligibility of subcortical representations for speech. *The Journal of Neuroscience, 35*(4), 1687–1691.

Wilson, R. H., & McArdle, R. (2005). Speech signals used to evaluate functional status of the auditory system. *Journal of Rehabilitation Research and Development, 42*(4 Suppl 2), 79–94.

Wong, P. C. M., Uppunda, A. K., Parrish, T. B., & Dhar, S. (2008). Cortical mechanisms of speech perception in noise. *Journal of Speech, Language, and Hearing Research, 51*, 1026–1041.

Woods, D. L., & Hillyard, S. A. (1978). Attention at the cocktail party: Brainstem evoked responses reveal no peripheral gating. In D. A. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (pp. 230–233). Washington, DC: U.S. Government Printing Office. (EPA 600/9-77-043).

Yang, W., & Bradley, J. S. (2009). Effects of room acoustics on the intelligibility of speech in classrooms for young children. *The Journal of the Acoustical Society of America, 125*(2), 922–933.

Yost, W. A. (1996). Pitch of iterated rippled noise. *The Journal of the Acoustical Society of America, 100*(1), 511–518.

Yund, E. W., & Woods, D. L. (2010). Content and procedural learning in repeated sentence tests of speech perception. *Ear and Hearing, 31*, 769–778.

Zahorik, P. (2002). Direct-to-reverberant energy ratio sensitivity. *The Journal of the Acoustical Society of America, 112*(5 Pt 1), 2110–2117.

Zendel, B. R., & Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience, 21*(8), 1488–1498.

Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and Aging, 27*(2), 410–417.