Chapter 7
Individual Differences in Temporal Perception and Their Implications for Everyday Listening

Barbara Shinn-Cunningham, Leonard Varghese, Le Wang, and Hari Bharadwaj

Abstract Growing evidence shows that individual differences among listeners with normal hearing thresholds reflect underlying differences in how well the auditory system encodes temporal features of sound. In the laboratory, these differences manifest in a range of psychophysical tasks. In everyday life, however, the situations that reveal these differences are often social settings where listeners are trying to understand one talker in the presence of other competing sound sources (the "cocktail party" setting). Physiologically, the brainstem's envelope-following response (a specific form of the frequency-following response) correlates with individual differences in behavior. Motivated by both animal and human studies, this chapter reviews the evidence that behavioral and physiological differences across individual listeners with normal hearing thresholds reflect differences in the number of auditory nerve fibers responding to sound despite normal cochlear mechanical function (cochlear neuropathy). The chapter also points out some of the measurement issues that need to be considered when designing experiments trying to probe these kinds of individual differences in coding of clearly audible, supra-threshold auditory information.

Keywords Auditory brainstem · Binaural hearing · Binaural interaction component · Envelope-following response · FFR · Frequency-following response · Individual differences · Selective attention

B. Shinn-Cunningham (✉) · L. Varghese · L. Wang
Biomedical Engineering, Boston University, 677 Beacon Street, Boston, MA 02215, USA
e-mail: shinn@bu.edu

L. Varghese
e-mail: lennyv@bu.edu

L. Wang
e-mail: lwang@bu.edu

H. Bharadwaj
Martinos Center for Biomedical Imaging, Massachusetts General Hospital, 149 Thirteenth St., Boston, MA 02129, USA
e-mail: hari@nmr.mgh.harvard.edu

N. Kraus et al. (eds.), *The Frequency-Following Response*,
Springer Handbook of Auditory Research 61,
DOI 10.1007/978-3-319-47944-6_7

7.1 Introduction

Historically, the majority of psychoacoustic studies have explored how variations in perceptual ability depend on acoustic stimulus parameters. Often in such studies, individual differences across listeners confound interpretations: they are a source of noise and interfere with the differences that are the focus of study. However, a growing number of studies have started to exploit repeatable individual differences that are present across listeners with normal audiometric thresholds.

The envelope-following response (EFR), a specific form of the frequency-following response (FFR), indexes important differences in temporal coding fidelity in listeners with normal hearing thresholds. The finding that individual differences in ability are related to differences in objective physiological measurements supports the idea that sensory coding fidelity differs amongst listeners with normal audiometric thresholds and that this affects hearing in everyday settings.

After briefly describing the EFR in Sects. 7.2, 7.3 reviews evidence that the EFR reflects, in part, differences in the number of auditory nerve fibers (ANFs) responding to sound, which can be reduced (e.g., by noise exposure and by ordinary aging) without affecting detection thresholds (Kujawa and Liberman 2009; Lobarinas et al. 2013). Section 7.4 then considers why these differences manifest in everyday settings when listeners need to direct selective auditory attention. Section 7.5 discusses various factors that should be considered when designing experiments to investigate individual differences in the EFR, and Sect. 7.6 comments on some aspects of brainstem sound processing that *do not* cause consistent signatures in the EFR.

7.2 The Envelope-Following Response

7.2.1 Defining the Envelope-Following Response

The voltage measured on the scalp in response to sound reflects a mixture of brain activity, including both cortical responses (which are relatively large at the scalp) and responses from deeper, subcortical portions of the auditory pathway (roughly two orders of magnitude smaller than cortically generated potentials). The observable voltages on the scalp are the sum of electrical activity from an enormous number of individual neurons and their interconnections, each of which can produce electrical potentials. If they are in opposing directions, the sum of these electrical potentials will cancel. As a result, the only signals that survive to be observable on the scalp are those that cause synchronous potentials with similar polarity at the measuring electrodes (for a review of the physics of neuroelectric and neuromagnetic measurements, see Hamalainen et al. 1993).

Responses phase locked to periodic sounds originating from either the subcortical or cortical portions of the auditory pathway often are collectively referred to as auditory steady-state responses or ASSRs (Galambos et al. 1981; Stapells et al. 1984). The EFR is a specific form of ASSR measured by presenting a periodic input signal (typically with a periodicity in the 80–450 Hz range so that the subcortical portion of the response is emphasized; see Sect. 2.3) in opposite polarities and then averaging the responses (Goblick and Pfeiffer 1969; Aiken and Picton 2008).

To understand what the EFR encodes, one can first consider how a sound pressure wave is represented when it is transduced into neural energy. This transduction takes place in the cochlea, which turns the uni-dimensional sound pressure into a multi-dimensional representation. Specifically, the cochlea breaks the input down into a parallel representation in which each "channel" of the representation responds to a different band of input acoustic frequencies. As a result of this cochlear frequency selectivity, a narrow frequency band of sound drives each ANF. Because the "driving signal" for each ANF is narrowband, the temporal information encoded by each ANF frequency channel can be logically separated into two parts: (1) the temporal fine-structure (TFS) corresponding to nearly sinusoidal *carrier* fluctuations near the center frequency of the narrowband driving signal, and (2) slow fluctuations in the energy of the TFS known as the *envelope* of that carrier, whose frequency content is limited by the bandwidth of the corresponding cochlear filter (Boashash 1992). The (non-negative) envelope of a stimulus is also referred to as the *modulation* in the signal.

The bandwidth of the cochlear filters increases with increasing center frequency; therefore, the temporal fluctuations in the envelope of the driving functions can be more rapid with increasing center frequency (Moore 2003). This is illustrated in Fig. 7.1, which shows the signal exciting three different places along the cochlea for a click train repeating at 100 Hz. Each of the illustrated cochlear channels (100, 1000, and 4000 Hz) responds to the frequency components of the click train falling within its critical band. For the 100 Hz signal, only a single component (at 100 Hz) falls within the cochlear critical band (Fig. 7.1A). As a result, the response in this channel consists of TFS fluctuating at 100 Hz and an envelope that rises to a constant value, beginning from the start of the click train. In contrast, the higher-frequency channels have many harmonics (frequency components that are multiples of the fundamental frequency of 100 Hz) that sum together to produce the driving function at that cochlear place (see Fig. 7.1B, C). Summing harmonics with a common fundamental frequency produces a signal that is periodic with the period of the fundamental frequency. Therefore, the 1000-Hz and 4000-Hz cochlear places have periodic driving functions with a period of 1/100 Hz (10 ms), which is seen in their envelopes. In addition, because the bandwidth of the response from the 4000-Hz cochlear channel is broader than that of the 1000-Hz channel (and thus encompasses more harmonics), the envelope from the 4000-Hz channel has more rapid amplitude fluctuations (the waveform is "sharper" in time; compare envelopes in Fig. 7.1B, C). Finally, the TFS fluctuates at a rate determined by the center

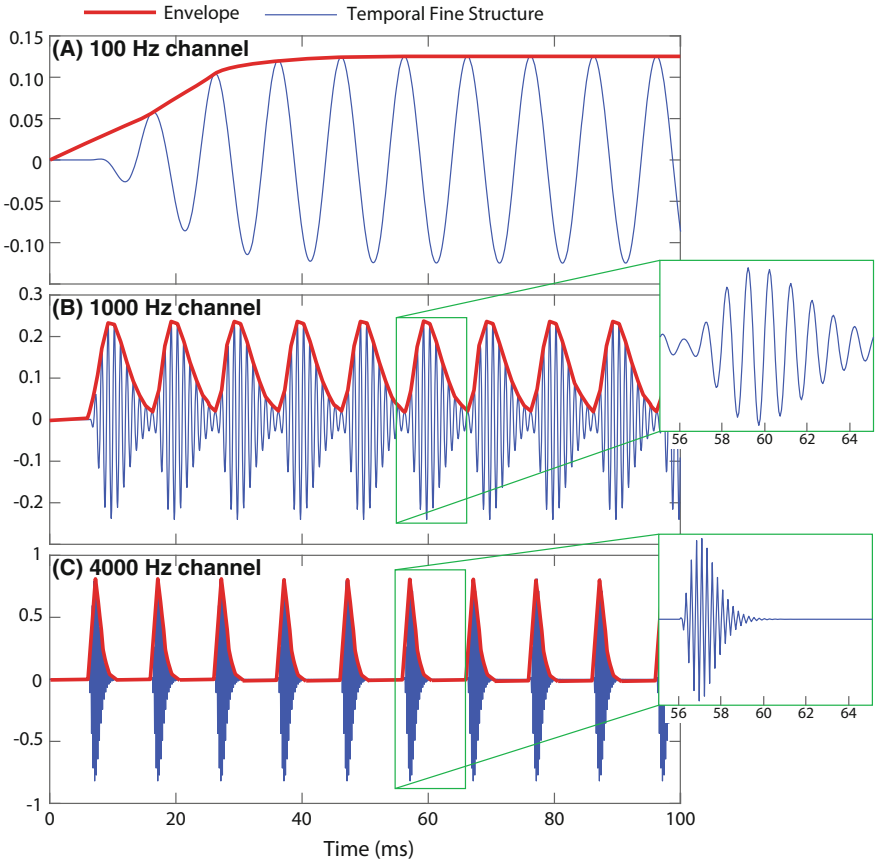

Fig. 7.1 Illustration of the "driving function" at three different places along the cochlea for a click train repeating at 100 Hz. In a low-frequency channel centered at 100 Hz (**A**), the driving signal looks like a sinusoid in the steady state with a flat envelope. In the 1000-Hz (**B**) and 4000-Hz (**C**) channels, the envelope is periodic with a repetition rate equal to the 100 Hz periodicity of the input. However, the 1000-Hz envelope varies more slowly in time than the 1000-Hz channel envelope (compare *red waveforms* in **B** and **C**). The temporal fine structure (the rapidly fluctuation oscillations in *blue*) varies with a periodicity equal to the center frequency of the channel and is four times slower for the 1000-Hz channel than for the 4000-Hz channel (compare *blue waveforms* in the insets of **B** and **C**)

frequency of each critical band (with periods of 1/1000 or 1 ms, and 1/4000 or 0.25 ms for the 1000-Hz and the 4000-Hz channels, respectively; see the insets in Fig. 7.1B, C).

While Fig. 7.1 shows the driving functions at different cochlear places, the actual response coming out of the cochlea cannot track very rapid positive and negative TFS fluctuations. Instead, ANFs at each place along the cochlea respond to some combination of the TFS and the envelope in their driving functions in a

proportion that depends on the rate of TFS fluctuations. The ANF firing pattern does not track TFS well above about 2000 Hz; thus, the envelope of the driving function tends to dominate responses of the high-frequency cochlear channels.

Because the EFR is measured by averaging responses to positive and negative polarity acoustic presentations, the EFR cancels out all portions of the measured electrical response that take on opposite values in response to the two types of presentations (see Fig. 3.3 of Krishan and Gandor, Chap. 3). This includes much of the response elicited by a signal's TFS, as well as electromagnetic contamination (e.g., from unshielded or improperly shielded audio transducers) and the cochlear microphonic (see Skoe and Kraus 2010). Conversely, the EFR enhances responses that are the same for positive and negative polarity inputs. Given this, the EFR tends to be driven by periodicities in the envelope of the input acoustic waveform, which explains the name *envelope-following response* (see Sect. 7.4.1 for further discussion). The EFR differs from other forms of FFRs in that it tends to have a higher signal-to-noise ratio (SNR) (Skoe and Kraus 2010) and, therefore, is often a more robust signature of neural activity than other types of FFRs.

One frustration in conducting research on FFRs is the lack of consistent nomenclature. The term "FFR" is now often used to denote both EFRs as well as other forms of ASSRs. Yet historically, ASSRs tracking neural envelope periodicities were known as either "amplitude modulation following responses" or EFRs (Dolphin and Mountain 1992; Kuwada et al. 2002) in order to distinguish them from ASSRs phase locked to the TFS of pure tones, which were called FFRs (Marsh et al. 1975). In the interest of specificity and based on historical precedent, this chapter uses the term EFR to refer to responses derived by summing equal numbers of positive and negative polarity presentations of a periodic auditory input.

7.2.2 *Challenges in Localizing the Source of the Envelope-Following Response*

EFRs provide a convenient, noninvasive method for measuring some aspects of subcortical neural activity. However, inferring what brain structures produce EFRs is challenging. Any form of electroencephalography (EEG) recorded at the scalp, including EFRs, measures a response that is the sum of all neural activity recorded through multiple layers of tissue and bone. On top of this, electromagnetic noise is ubiquitous, coming from line noise in the power system, electric fields generated by experimental equipment, muscle artifacts (e.g., from eye blinks), and even incidental neural activity that is not associated with the presented stimulus. Even if noise were insignificant, determining what brain region produces a particular voltage on the scalp is an ill-posed problem that is mathematically impossible to solve without independently constraining the solution (Hamalainen et al. 1993).

These challenges are particularly problematic when trying to determine what subcortical activity is reflected in the scalp voltage. The subcortical sites generating neural activity are deep in the brain and far from the recording sites on the scalp. As a result, the SNR at the scalp is low. Moreover, all of these deep, subcortical sources are at roughly the same distance to all electrodes, which means that they contribute nearly the same signal to every electrode on the scalp—even electrodes that are far away from each other. This detail, combined with the fact that activity at each subcortical stage is temporally correlated with activity at the next stage, makes it almost impossible to separate activity from different sources by combining information across multiple electrical sensors (an approach that is used to infer source locations of cortical neural activity from an EEG).

7.2.3 *Frequency Content of the Brainstem Envelope-Following Response*

Luckily, cortical activity that is observable in scalp voltages tends to be made up of relatively low frequency components, dropping off above about 80 Hz, while synchronous subcortical activity can be observed up to much higher frequencies. One reason for this is that the way temporal information is encoded changes as information ascends the auditory pathway.

In the peripheral parts of the pathway, the temporal pattern of the neural firing encodes both TFS and amplitude modulation. As information passes to higher processing centers, temporal cues are transformed from being represented in temporal structure to being encoded by a neural "place" code (coded by which neurons are firing, rather than the temporal pattern of their firing). For instance, coding of envelope modulation undergoes a transformation from phase-locked changes in firing rate (which would help to drive scalp signals phase locked to the modulation) to a more place-based code as one traverses from the ANF to the midbrain (see Joris et al. 2004). Consistent with this, cortical neurons tend to have longer lasting, slower, and more integrative responses than do subcortical neurons (e.g., Escabi and Read 2003).

Because of this frequency dependence, lower-frequency responses in the EEG signal (below 60 Hz or so) are dominated by cortical responses, while higher-frequency responses (above about 80 Hz) primarily reflect responses from subcortical regions of the brain. Both reversible inactivation studies (Kuwada et al. 2002) and irreversible lesion studies (Sohmer et al. 1977; Kiren et al. 1994) offer additional, direct evidence that a subcortical source is the dominant generator of EFRs above 80 Hz. These studies suggest that across different mammalian species the currents in inferior colliculus (IC) neurons produce the dominant response in ASSRs for frequencies above 80 Hz; lesions at higher stages of the auditory pathway, including primary auditory cortex, do not strongly influence ASSRs in this frequency range.

Practically speaking, in humans, EFR measurements are weak above 500 Hz due to SNR limitations (Cohen et al. 1991), although one study reports measuring significant EFRs up to 1000 Hz (see Purcell et al. 2004). In understanding this limit, it is helpful to recall that the ANF critical bandwidth increases with characteristic frequency (i.e., the frequency of input that causes the greatest response in a particular ANF); moreover, the critical bandwidth also increases with sound level (see Moore 2003). Since the critical bandwidth determines the fastest modulations contained in the signal driving a particular ANF, the modulation rates that might be conveyed increase with characteristic frequency.

In humans, the broadest cochlear filters can support about a 5–600 Hz modulation rate at moderate sound levels. In contrast to humans, in cats, ANFs with high characteristic frequencies (i.e., >10 kHz) phase lock to envelope fluctuations only up to about 1000 Hz; that is, the modulation bandwidth is not limited by the cochlea (Joris and Yin 1992). For the lower-frequency ANFs in cats, the envelope cutoff frequency decreases with the characteristic frequency of the fiber, suggesting that the limitation is imposed by the critical bandwidth of the cochlear filter (Joris and Yin 1992).

Figure 7.2A illustrates the low-pass nature of the EFR in human subjects measured in response to amplitude-modulated noise for modulation frequencies between 100 and 500 Hz. Because the noise in the measurements tends to be proportional to the reciprocal of the frequency, the SNR in the EFR decreases more slowly than does the absolute power. This can be seen in Fig. 7.2B, which plots the phase-locking value (PLV), a measure of response synchrony (see Sect. 7.4.2) in the EFR as a function of frequency. The PLV is a metric that directly reflects the SNR (e.g., Zhu et al. 2013). As shown in the figure, both the absolute strength of the EFR (Fig. 7.2A) and the PLV (Fig. 7.2B) decrease rapidly above 450 Hz. Thus, for all practical purposes, the EFR can be measured effectively in humans for input signals with periodicities falling in the range of roughly 80–500 Hz.

7.2.4 *Using Response Phase to Infer the Source of the Envelope-Following Response*

Analysis of the phase of the EFR as a function of input modulation frequency supports invasive animal studies in suggesting a dominant midbrain source of the EFR. The slope (derivative) of the phase versus stimulus-modulation-frequency function gives the response group delay. In frequency ranges where the group delay is constant (phase is a linear function of frequency), the group delay reflects the neural delay from the input to the dominant response source. If the measured EFR is not dominated by a single source at a given frequency but rather reflects a mixture that contains strong responses from multiple stages of the auditory pathway (each with a different delay), the group delay generally will vary with modulation frequency. By considering how the phase versus frequency slope changes with

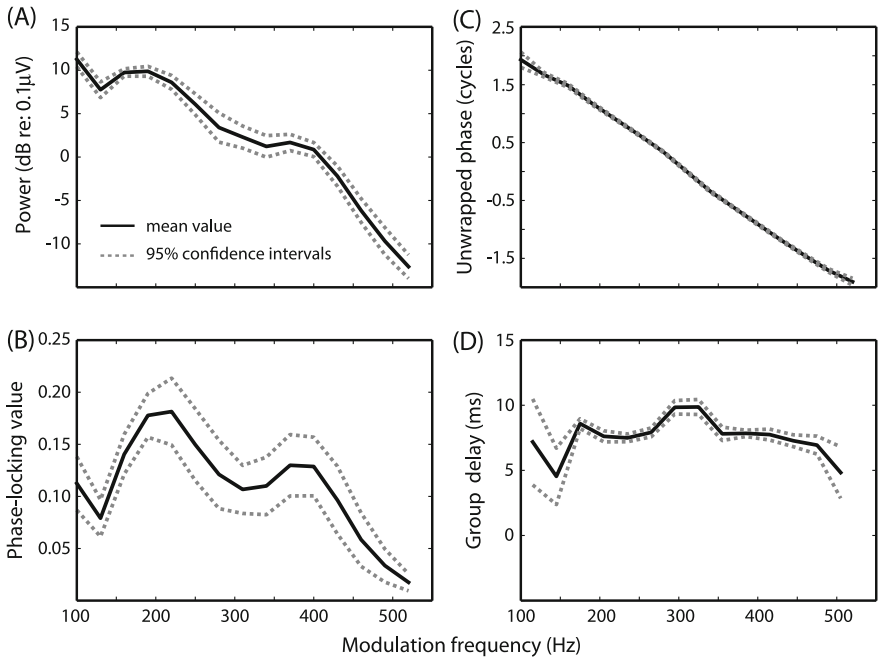

Fig. 7.2 Sample human envelope-following response (EFR) data in response to amplitude modulated noise as a function of modulation frequency. *Solid lines* show the mean response across subjects, while the *dashed lines* surrounding the mean show the 95% confidence intervals. (**A**) Power at the EFR modulation frequency. (**B**) The phase-locking value (measuring the consistency in the phase of the response at the modulation frequency across independent trials in response to the modulated noise). (**C**) The unwrapped EFR phase. (**D**) The estimated group delay (taking the difference of the EFR phase at adjacent frequencies from Fig. 7.1C)

frequency, one can begin to infer the frequency ranges over which the voltage mixture on the scalp reflects one dominant source as well as the latency of the response of this source (Kuwada et al. 2002; Shaheen et al. 2015).

Above 200 Hz, group delay varies substantially across species, probably due to anatomical differences, leading to different mixtures of responses from different parts of the pathway in the scalp-recorded voltage (e.g., Okada et al. 1997). While rabbits and mice have frequency regions of constant phase slopes out to 500 and 700 Hz, respectively (Kuwada et al. 2002; Pauli-Magnus et al. 2007), slopes are constant only up to 200 Hz in gerbils (Dolphin and Mountain 1992). In humans, EFRs exhibit a relatively constant group delay above about 80 Hz. The estimated (unwrapped) phase of the response as a function of modulation frequency is shown in Fig. 7.2C for modulation rates from 100 up to 500 Hz. Across this range, the phase is nearly linear. Figure 7.2D shows the group delay as a function of modulation frequency; the group delay is nearly constant, hovering around a value of about 8 ms, consistent with a midbrain source (likely the IC).

While the IC is likely a dominant source of the EFR, this does not mean that the recorded responses directly reflect the output firing patterns of IC neurons. Instead, the post-synaptic currents flowing in IC neurons, driven by lemniscal inputs, are likely the primary source of the EFR response. In particular, the action potentials (spikes) generated in IC are unlikely to induce observable potentials on the scalp. Spikes are brief and induce both depolarizing and hyperpolarizing currents along the neuronal axis. Consequently, they would have to be significantly better aligned in time across the IC population than the lemniscal inputs to produce an observable net signal. Spikes produce a quadrupolar current pattern, which does not produce large observable voltages beyond a very short distance; in contrast, post-synaptic currents produce a dipole current pattern, yielding voltages that can be observed at greater distances, such as between two scalp electrodes (Hamalainen et al. 1993; Milstein and Koch 2008).

7.3 Individual Differences in Listeners with Normal Cochlear Function

A number of EFR studies have found that the strength of the EFR differs across groups, such as musicians versus nonmusicians (Wong et al. 2007; Strait et al. 2011) or listeners with and without learning disabilities (Wible et al. 2005; Hornickel et al. 2011). These results demonstrate that the strength of the brainstem response is an important index of perceptual abilities. The finding that experience helps to shape subcortical neural responses (e.g., Skoe et al. 2014) helps to explain some of these group differences, such as why trained musicians tend to have stronger brainstem responses than do nonmusicians (e.g., Parbery-Clark et al. 2011). Experience is thus one potential source of individual variation in EFR strength (e.g., Anderson et al. 2013). Such results suggest that training may help ameliorate perceptual deficits indexed by the EFR (e.g., Whitton et al. 2014; Slater et al. 2015).

Other recent studies reveal robust individual differences in the subcortical responses across ordinary individuals selected randomly within a fairly homogenous subject group. Specifically, there are now a number of studies that measured EFRs as well as various perceptual abilities in ordinary listeners, all of whom had no known hearing or learning deficits, normal hearing thresholds, and normal cochlear function. These studies identified differences in EFR strength that were correlated with a range of basic perceptual abilities and the ability to selectively attend to speech in a noisy mixture (e.g., Ruggles et al. 2012; Bharadwaj et al. 2015). These studies suggest that EFRs may be useful for diagnosing subtle hearing differences that come from differences in the fidelity of auditory coding in the brainstem and that have real-world behavioral consequences (Bharadwaj et al. 2014; Shaheen et al. 2015).

This section presents evidence that at least a portion of the individual variation in EFR strength comes from differences in the number of ANFs that respond to sound. This view is motivated by the growing interest in *hidden hearing loss* (more formally known as *cochlear synaptopathy* or *cochlear neuropathy*) in both animals (Kujawa and Liberman 2009; Valero et al. 2016) and humans (Schaette and McAlpine 2011; Plack et al. 2014). Cochlear neuropathy is distinct from *auditory neuropathy*, which is characterized by a profound disruption of auditory brainstem responses (ABRs) with reduced amplitude or even absent ABR wave V responses (see Starr et al. 1996).

7.3.1 What Is Hidden Hearing Loss?

It has long been known that moderate noise exposure can lead to temporary threshold shifts (TTS) (see Quaranta et al. 1998). Immediately after noise exposure, sound detection thresholds can be elevated by as much as 40 dB but then recover back to normal over the course of days. Because clinical "hearing loss" is defined as having elevated hearing thresholds, by definition, listeners with TTS do not have hearing loss. Indeed, until relatively recently, because hearing thresholds recover and there is no loss of hair cells due to TTS (Bohne and Harding 2000), noise exposure of this type was assumed to incur no permanent hearing damage.

A growing number of animal studies have upended this assumption (Kujawa and Liberman 2015; Liberman 2015). Noise exposure that causes TTS has been shown to produce a rapid loss of as many as 40–60% of the ANF synapses driven by cochlear inner hair cells, which are the cells that generate the ascending signal conveying information in the auditory pathway (Kujawa and Liberman 2006, 2009). This loss of synapses subsequently leads to a slow death of ANF cell bodies (spiral ganglion cells) and central axons (Lin et al. 2011; Kujawa and Liberman 2015). Even in cases where the effects on synapses and spiral ganglion cells are pronounced, the effect on cochlear function can be negligible; cochlear mechanical function (including the tuning of the cochlea) can be normal in animals suffering from cochlear neuropathy (Kujawa and Liberman 2009). Most hearing screenings reveal losses associated with damage to inner and outer hair cells by looking for: (1) elevated detection thresholds, (2) reduced amplification in the cochlea, (3) wider-than-normal cochlear tuning, and (4) reduced otoacoustic emissions. Yet, with hidden hearing loss, these measures are normal, making the deficit "hidden" to typical hearing screening.

How can it be that hearing thresholds are normal even though the number of ANFs is significantly reduced? One reason is that synaptopathy causes a diffuse loss, leaving behind ANFs throughout the cochlea (Liberman et al. 1997; Lobarinas et al. 2013). In addition, some evidence suggests that synapse loss preferentially affects ANFs that have high thresholds and low spontaneous firing rates (SR) (Furman et al. 2013; Kujawa and Liberman 2015). As shown in Fig. 7.3, each healthy inner hair cell in the cochlea typically drives multiple ANFs that differ in

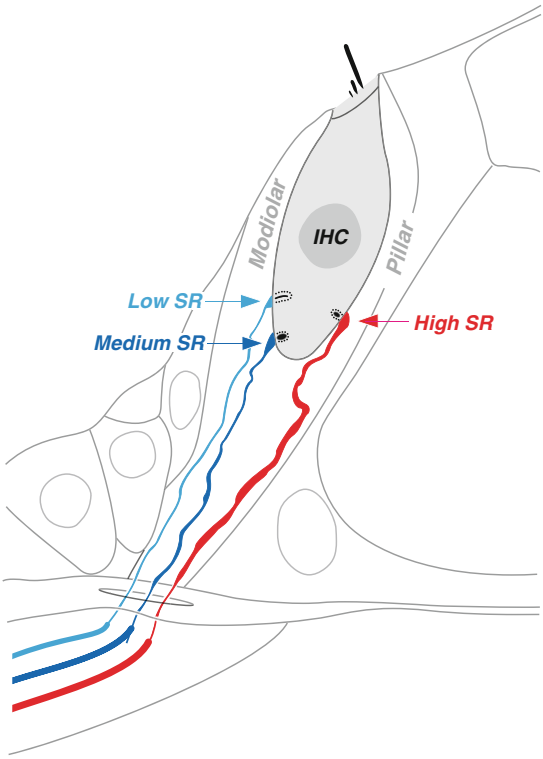

Fig. 7.3 Illustration of how terminals of the cochlear nerve innervate a single inner hair cell. Each inner hair cell typically has synaptic contacts with multiple auditory nerve fibers with *high*, *medium*, and *low spontaneous rates* (*SR*). (Figure from Bharadwaj et al. 2014)

their spontaneous firing rates and thresholds (i.e., the sound level at which the response of the ANF differs from its spontaneous firing pattern) (see Moore 2003). While low-threshold, high-SR ANFs begin to increase their firing rates when sound just exceeds perceptual detection thresholds, mid-threshold and high-threshold ANFs (or "higher-threshold ANFs" for brevity) only contribute to neural activity at supra-threshold sound levels. If noise exposure preferentially damages higher-threshold ANFs, it makes sense that a large number of ANFs may cease to respond without influencing detection thresholds. However, the effect of this loss on the encoding of acoustic temporal details in supra-threshold sound can be substantial (Plack et al. 2014).

Why might higher-threshold fibers be more susceptible to noise-exposure damage than other fibers? Pharmacological studies suggest that cochlear neuropathy is the result of a type of glutamate excitotoxicity, a process in which neurons are damaged and die off through over-activity in response to the neurotransmitter glutamate (e.g., Pujol et al. 1993; Mehta et al. 2013). In the central nervous system, glutamate excitotoxicity is mediated by an increase in intracellular calcium concentration (Szydlowska and Tymianski 2010). Mitochondria within cell bodies comprise an important intracellular calcium buffering system. In inner hair cells within the cochlea, fewer mitochondria are associated with higher-threshold fibers

(Liberman 1980). Given that higher-threshold fibers typically respond with lower firing rates than do low-threshold ANFs, the smaller number of mitochondria may be sufficient to ward off excitotoxicity in ordinary settings; however, in the face of ongoing noise that drives higher-threshold fibers at a continuous, high rate, these cells may be vulnerable to glutamate excitotoxicity (Bourien et al. 2014).

Animal studies show that cochlear neuropathy decreases the magnitude of supra-threshold, click-evoked ABR wave I responses (coming from the auditory nerve) but not the magnitude of wave V (coming from the midbrain; see Hickox and Liberman 2014). In animals with extreme neuropathy (with a loss of >95% of cochlear nerve afferent synapses), plasticity leads to an enhanced neural gain in the brainstem and cortex that compensates for the weak ANF response, producing detection thresholds that are near normal (Chambers et al. 2016). While these changes ameliorate some of the effects of a weak ANF drive, they cannot compensate fully for ANF loss; temporal coding in the denervated animals is poorer than in control animals. These findings further help explain why more subtle cochlear neuropathy may have a big impact on the representation of temporal features of supra-threshold sounds without affecting detection thresholds and why ABR wave V may have a normal magnitude even when ABR wave I amplitude is reduced.

7.3.2 Hidden Hearing Loss in Humans

While there are no data yet to directly support the idea that cochlear neuropathy occurs in humans, a growing number of studies hint that it accounts for some of the individual variability seen in listeners with normal cochlear mechanical function. As noted above, listeners with normal hearing thresholds vary significantly in their ability to utilize precise temporal information (Ruggles et al. 2011; Bharadwaj et al. 2015). This variability correlates with difficulties in using spatial-selective attention to focus on and understand speech in a noisy background (Ruggles and Shinn-Cunningham 2011), underscoring the clinical relevance of these differences.

In one such study, young adult subjects were recruited with no special criteria except that they had normal hearing thresholds and no known auditory deficits (Bharadwaj et al. 2015). Individual differences amongst this cohort were nonetheless large. Perceptual abilities correlated with EFR strength, especially at high sound levels and shallow modulation depths when higher-threshold ANFs are important for coding temporal features. This is illustrated in Fig. 7.4, which demonstrates the consistent relationships between the EFR strength (plotted along the x axes) and perceptual thresholds for amplitude modulation detection (Fig. 7.4A) and for envelope interaural time difference (ITD) discrimination (Fig. 7.4B). Both of these perceptual measures rely on fine temporal information, and both are significantly correlated with the strength of the EFR when a shallow modulation drives the brainstem response.

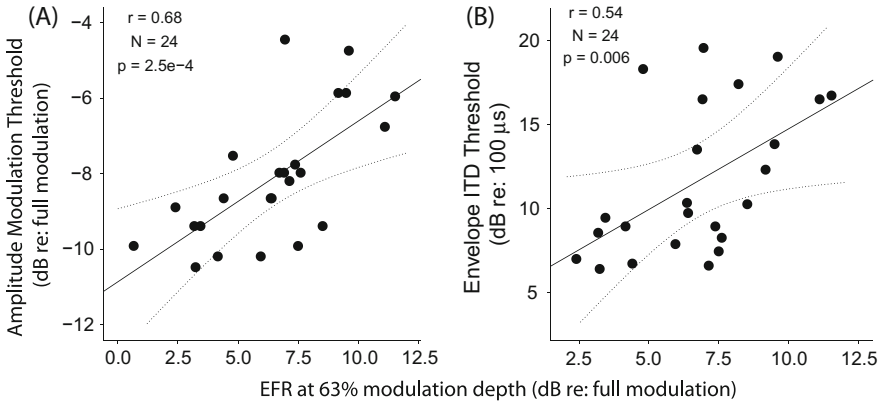

Fig. 7.4 Relationship between envelope-following response (EFR) strength and perceptual thresholds in young adult listeners with normal cochlear function. (**A**) Amplitude modulation detection thresholds are correlated with the EFR strength. (**B**) Discrimination thresholds for envelope ITD are correlated with EFR strength. *r*, correlation coefficient (Data are from the study described in Bharadwaj et al. 2014)

Crucially, listeners had normal compressive growth of cochlear responses (measured by distortion product otoacoustic emissions), normal frequency tuning (measured by psychoacoustic estimation), and pure-tone audiometric thresholds of 15 dB hearing level (HL) or better at octave frequencies between 250 Hz and 8 kHz. In other words, although perceptual differences were correlated with the EFR (an objective measure of the precision of brainstem temporal coding), these differences could not be explained by cochlear mechanical function. These findings suggest that cochlear neuropathy may be quite common, affecting a large percentage of the population, including relatively young listeners.

Other studies in humans also support the view that human listeners with normal cochlear function may suffer from different degrees of cochlear neuropathy. For instance, listeners can vary significantly in their ability to discriminate both frequency modulation and ITDs (see Strelcyk and Dau 2009; Grose and Mamo 2010). The computation of ITDs depends directly on temporal precision in ANF responses and subsequent processing centers (such as neurons in the superior olivary complex). Indeed, sensitivity to ITD cues was one of the perceptual abilities that correlated with EFR strength (Bharadwaj et al. 2015).

On the physiological side, listeners with normal hearing thresholds show large inter-subject variability in the magnitude of ABR wave I (Schaette and McAlpine 2011; Stamper and Johnson 2015) again supporting the view that listeners with normal audiograms suffer from neuropathy to varying degrees. As in animal studies, while ABR wave I amplitude varies significantly across individuals, the magnitude of ABR wave V does not (Schaette and McAlpine 2011; Stamper and Johnson 2015).

One study has shown that perceptual differences correlate with differences in human ABRs: in young adults with no known hearing deficits, wave I magnitude was related to ITD sensitivity (Mehraei et al. 2016). Consistent with previous animal studies, wave V magnitude was unrelated to wave I magnitude or perceptual ability (although effects of noise on wave V timing were correlated with wave I amplitude). Taken together, these results suggest that cochlear neuropathy is common amongst human listeners who have normal audiograms, many of whom do not even realize that they may have communication difficulties.

7.3.3 *Effects of Aging on Hidden Hearing Loss*

In animal models, natural aging produces cochlear neuropathy (see Anderson, Chap. 11, for a discussion of the effects of aging and hearing loss). Aging mice raised without exposure to any loud sound (and without significant hair cell loss) exhibit a loss of 30–40% of inner hair cell synapses by roughly 3/4 of their lifespan, an age at which thresholds are elevated by less than 10 dB (Sergeyenko et al. 2013). This kind of neurodegeneration may selectively affect higher-threshold fibers (Schmiedt et al. 1996). Counts of spiral ganglion cells in an age-graded series of human temporal bones show degeneration of 30%, on average, from birth to death, even in cases with no hair cell loss (Makary et al. 2011). These anatomical results support the idea that aging alone can produce hidden hearing loss.

Older human listeners with normal hearing thresholds exhibit basic temporal processing deficits (see Fitzgibbons and Gordon-Salant 2010 for a review). Aging degrades temporal modulation sensitivity (Purcell et al. 2004; He et al. 2008) and leads to weaker brainstem responses (Anderson et al. 2012). Temporal deficits correlate with the strength of the EFR in older listeners with normal thresholds (Purcell et al. 2004). The highest modulation frequency to which EFRs exhibit phase locking decreases with age (Purcell et al. 2004; Grose et al. 2009), and temporal processing of both monaural and binaural sound features degrades with age (e.g., Grose and Mamo 2012; Grose et al. 2015). Indeed, even after factoring out effects of elevated hearing thresholds, aging causes degradations in temporal processing that appear well before there is evidence of speech processing deficits (Snell and Frisina 2000; Snell et al. 2002). Aging also interferes with the ability to understand speech in the presence of competing sound (Fullgrabe et al. 2014; Helfer 2015). All of these symptoms implicate deficits in temporal coding in aging listeners. Indeed, a number of researchers have concluded that in older listeners with normal thresholds, difficulties with understanding speech in noise arise because of temporal processing deficits (Helfer and Vargo 2009; Jin et al. 2014). These studies, like anatomical studies, support the view that aging leads to cochlear neuropathy in human listeners.

7.3.4 *Hidden Hearing Loss and Individual Differences in the Envelope-Following Response*

When considering the idea that cochlear neuropathy produces differences in the EFR, one question is how to reconcile the view that individual differences in the EFR come from an irreversible loss of ANFs with the many studies showing that appropriate experience and training can increase the strength of the brainstem response (e.g., Carcagno and Plack 2011; Strait and Kraus 2014). One intriguing possibility is that long-term training and experience can partially, but only partially, compensate for cochlear neuropathy. For instance, training could increase the efficiency with which the information in remaining ANFs is extracted by higher centers of the auditory pathway. In line with this, as noted previously, neuropathy decreases the magnitude of ABR wave I but not in the magnitude of wave V (see Stamper and Johnson 2015; Mehraei et al. 2016). In other words, the gain of the auditory pathway between the ANFs and the IC seems to compensate for a weaker than normal ANF response when there is cochlear neuropathy. Such compensation likely helps to ensure that detection thresholds are normal and helps to keep the overall average firing rate of brainstem neurons at the proper level. However, an increase in gain cannot fully restore coding of temporal cues, which rely on the convergence of the noisy, stochastic responses of many ANFs (e.g., Oertel et al. 2000).

Consistent with this hypothesis, following profound cochlear denervation, central compensatory processes restore responses in both cortex and midbrain; however, this compensation cannot overcome deficits in "features encoded by precise spike timing" (Chambers et al. 2016). This kind of thinking helps to resolve the counterintuitive idea that even though experience influences the strength of the EFR, some of the differences in the strength of the response reflect irreversible differences in the number of ANFs encoding sound. Furthermore, the deficits that cannot be overcome by compensatory gain changes in the midbrain and above are those in temporal processing, which explains the pattern of deficits seen in human listeners who are suspected to have cochlear neuropathy.

7.4 Why Hidden Hearing Loss Affects Daily Function

Roughly 5–10% of listeners seeking treatment at audiological clinics have normal hearing thresholds (Kumar et al. 2007; Hind et al. 2011). Typically, these patients are driven to seek help because of difficulty communicating in situations requiring them to focus selective attention. Historically, such listeners were said to have "central auditory processing disorder" (Rosen et al. 2010), a catchall diagnosis that testifies to the fact that underlying causes were not well understood; however, some of these listeners likely are suffering from cochlear neuropathy.

The fact that listeners first notice the effects of cochlear neuropathy when trying to communicate in social settings makes sense, given how neuropathy degrades

auditory temporal coding. Spectrotemporal details in a sound mixture are important for grouping of acoustic elements into perceptual objects (Shamma et al. 2011; Christiansen and Oxenham 2014), discrimination of perceptual features like pitch (Smith et al. 2002) and source location (Blauert 1997; Smith et al. 2002), as well as speech perception itself (Zeng et al. 2005). Importantly, subtle hearing deficits may not disrupt speech perception in quiet, yet they still have a debilitating effect on selective auditory attention.

7.4.1 Source Segregation

In order to selectively attend, listeners must be able to segregate sounds making up the acoustic mixture entering the ears. Source segregation depends on harmonic structure, interaural time differences, and other cues computed from acoustic features that are degraded when temporal coding is poor (Bregman 1990; Carlyon 2004). If temporal features are degraded and the target source cannot properly be segregated from the scene, selective attention will fail (Shinn-Cunningham 2008; Shinn-Cunningham and Best 2008). This idea is illustrated by visual analogy in the cartoons shown in Fig. 7.5. In people with good coding fidelity, fine details in the scene ensure that each source is distinct. In the visual analogy, features of each word in the scene are clear: words differ in their color, so are easy to perceive as distinct and separate objects (Fig. 7.5A). In contrast, even if the representation of the scene is weak, an observer may have no difficulty detecting that there are elements present in the scene: they may have normal detection thresholds (Fig. 7.5B). Yet observers may have problems understanding the supra-threshold information in the scene. Elements making up the scene are fuzzy, letter edges are blurry, and colors of different words in the scene are similar, so that the words seem to run together perceptually.

In the auditory domain, when listening to a complex scene, spectrotemporal details (e.g., periodicity, ITD, and amplitude and frequency modulation) are analogous to the letter edges and colors of a visual scene. These features are less clearly represented when a listener suffers from hidden hearing loss, so that the structural elements critical for parsing the acoustic scene are perceptually indistinct.

7.4.2 Source Selection

Successfully listening in a complex setting depends on more than simply segregating the sources from one another; it also requires selecting the desired source from the mixture by focusing selective attention. Selective auditory attention enhances the representation of the auditory object with a desired perceptual feature or attribute (Lee et al. 2012; Maddox and Shinn-Cunningham 2012). The low-level acoustic spectrotemporal structure is what enables a listener to compute perceptual

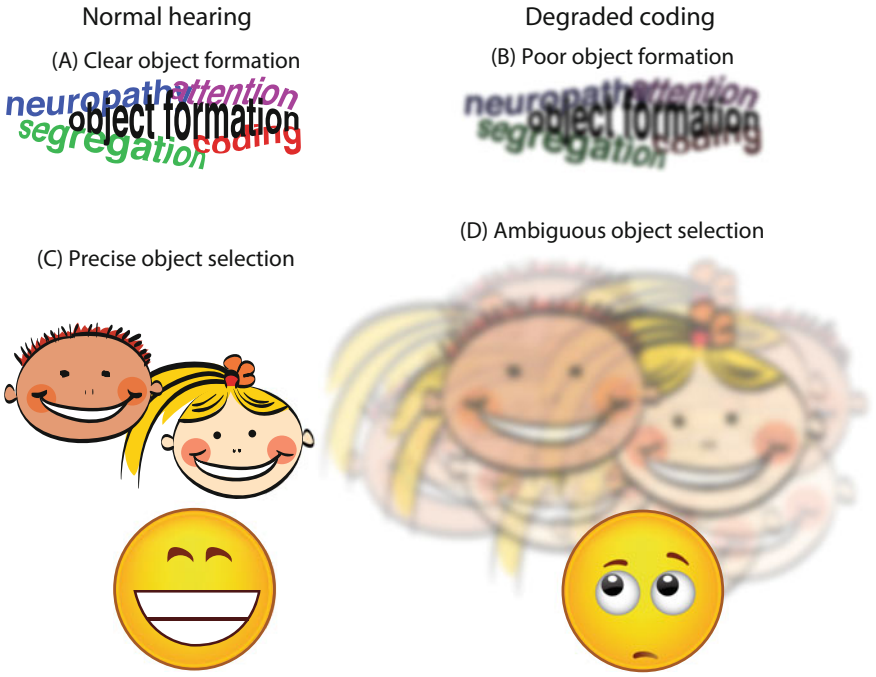

Fig. 7.5 Visual analogy illustrating the effects of poor brainstem coding fidelity on segregating and selecting a target object from a complex scene for a "good" listener (*left*) and a "bad" listener (*right*). (**A**) For a listener with a good sensory representation, each edge of each letter in the scene is represented clearly, and the similarity of the color of the letters making up each word (as well as the dissimilarity of the colors across words) allows each word to be perceived automatically as a distinct unit. In this representation, words are automatically segregated, based on the clear features of the letters and words. (**B**) With a poor representation, individual letters blur together, making the structure of each letter difficult to perceive. The poor quality of the representation also degrades the features that distinguish words from each other, further blurring together the elements of the scene and making it more challenging to separate the words in the scene. (**C**) When the peripheral representation is clear, each object is both distinct and has a clear perceived location, making it easy to deploy spatial selective attention and focus on a target from a particular direction. (**D**) When the peripheral representation is weak, spatial cues are blurry and ambiguous and can be inconsistent over time, making it difficult to focus attention and select out the target object. (Original figure by Shinn-Cunningham)

features of objects in a scene that can be used to focus attention. Specifically, low-level features such as periodicity, ITD, and amplitude and frequency modulation support computation of higher-level perceptual quantities such as pitch, location, and timbre. These attributes can be used to listen to "the high-pitched source," or "the source on the left," or "Sally, not Jim."

One clear example of a high-level feature that is degraded when temporal cues are weak (e.g., due to hidden hearing loss) is spatial location. When temporal cues are weak, the perceived location of a sound source can be broad and diffuse. Listeners with a weak temporal code can fail to select the correct source in the scene based on its less perceptually precise location. For instance, one study found large individual differences in performance on a spatial selective-attention task (Ruggles and Shinn-Cunningham 2011). In this study, when listeners failed, they did not fail to understand speech present in the sound mixture. Instead, they reported the wrong word, coming from the wrong location; that is, perceptual deficits were not severe enough to interfere with understanding the speech that was present in the mixture. The failures happened because listeners could not select the correct talker based on spatial cues. Individual variations in performance on the selective attention task correlated with differences in EFR strength, which is consistent with the idea that spatial-selective attention fails when listeners suffer from hidden hearing loss and poor temporal coding (Ruggles and Shinn-Cunningham 2011; Ruggles et al. 2012). Reverberation, a natural form of temporal degradation in the signals reaching the ears, exacerbated the selective attention errors. In other words, both external noise in the temporal acoustic features important for conveying location (from reverberation) and internal noise in the computation of ITDs (from differences in temporal coding fidelity in the brainstem) had similar, additive effects in disrupting selective auditory attention.

Figure 7.5C–D shows a visual analogy to this kind of problem. In this scene, a listener with a good peripheral representation can focus attention unambiguously to a talker to the left (Fig. 7.5C). If the spatial cues are weakly represented, however, the perceived talker locations overlap and smear into each other (Fig. 7.5D). Even if a listener can parse the scene into a male and a female talker, they may focus on the wrong talker when trying to focus on "the talker on the left" because of the spatial ambiguity in the scene. Such problems can produce communication difficulties in settings where there are multiple sources competing for attention that would not show up on a test of speech perception in quiet or even if there were nonspeech sounds present (i.e., in conditions where competing sound objects are so perceptually dissimilar that failures of selection will not occur).

7.4.3 *Understanding Speech in Noisy Settings Reveals Subtle Deficits*

The previous examples demonstrate why even modest degradations in temporal processing may lead to communication dysfunction in everyday settings (Shinn-Cunningham and Best 2008). Temporal coding problems interfere with the sound features that support both segregation and selection of the desired source from the mixture. In other words, listening to a talker amid similar, competing talkers reveals deficits that may be too subtle to be observed in other listening situations.

7.5 Interpreting Individual Differences in the Envelope-Following Response

Interpreting EFRs is complicated. The responses that are measured on the scalp are a sum of the electrical activity from different populations of neurons across different stages of processing. While evidence suggests that the EFR strength reflects true differences in the strength of temporal coding in subcortical portions of the auditory pathway, there are a number of issues that arise when interpreting EFRs. This section highlights a few of the issues of which one should be aware when trying to interpret individual differences in EFRs.

7.5.1 Encoding of Modulation in the Auditory Nerve Responses

Any modulation information that drives the EFR must be encoded in the firing patterns of the population of ANFs ascending from the cochlea. Therefore, in order to understand how the EFR is generated, one important thing to understand is how modulation information is first coded in this ANF population.

As described in Sect. 7.2.1, ANF neural spikes are phase locked to a mixture of both TFS and envelope modulation. The degree of phase locking to the TFS rolls off with frequency, with a knee point (in humans) near 2000 Hz. Many EFRs are measured using broadband, periodic inputs, such as a broadband complex tone (Zhu et al. 2013) or a speech syllable that has a fixed fundamental frequency (Russo et al. 2004). With such broadband stimuli, even though most of the cochlea is excited, the EFR is dominated by responses from middle to high frequency regions of the cochlea (Zhu et al. 2013). This can be explained by considering what acoustic energy drives the low-frequency and high-frequency ANFs.

In the lowest frequencies, individual harmonics excite different ANFs, leading to "resolved harmonics" in the excitation pattern. For a steady-state input, a single, resolved sinusoidal harmonic will cause a constant drive, with no envelope fluctuations (recall Fig. 7.1A). In contrast, in high-frequency channels, multiple harmonics fall within a single ANF critical band (unresolved harmonics). For channels responding to unresolved harmonics, the driving signal is periodic with a period corresponding to the fundamental frequency of the input (recall Figs. 7.1B, C). For these channels, envelope fluctuations at the fundamental frequency of the input signal dominate the periodicity in the neural firing pattern, strongly contributing to the EFR. This suggests that for a broadband, periodic input, EFRs come primarily from responses in frequency channels responding to unresolved harmonics.

For narrowband, low-level sounds, phase-locked ANF activity (to both TFS and envelope) is limited to a small region at the tonotopic place tuned to the input (Ananthanarayan and Durrant 1992; Herdman et al. 2002). This specificity led some researchers to propose using ASSRs for objective audiometry (Gardi et al. 1979;

Lins et al. 1996). However, narrowband supra-threshold sounds that are at a comfortable listening level (or louder) cause activity that spreads out from the best place on the cochlea; this spread of excitation can be quite pronounced, especially toward the basal (higher frequency) end of the cochlea. When EFRs are measured for supra-threshold acoustic inputs, the ANFs that drive higher auditory centers are typically spread over a large swath of the cochlea, even if the acoustic input is band limited (e.g., John et al. 1998; Herdman et al. 2002). Therefore it is difficult to deduce how activity from a specific place along the cochlea contributes to EFRs. One approach to minimizing the spread of excitation is to use notched noise maskers, so that contributions from off-frequency channels are attenuated (e.g., Bharadwaj et al. 2015).

The measured EFR is a sum of all neural activity; it therefore depends on the phase alignment of the responses in different frequency channels. If responses in two distinct neural populations are both large, but out of phase with each other, they can cancel one another. While models can predict phase disparities in the responses across the population of ANFs (e.g., Shinn-Cunningham et al. 2013; Verhulst et al. 2015), it is more difficult to predict what phase differences are present in envelope-modulation driven responses in the neural population at the level of the IC or how this impacts the final EFR.

7.5.2 Metrics to Quantify the Envelope-Following Response

Time domain methods make sense for analyzing transient events, such as the waves composing the traditional click-evoked ABR. In contrast, frequency-domain analyses efficiently characterize periodic neural activity such as EFRs. Typically, to measure the EFR, a periodic signal is presented on multiple trials. The response to each trial includes multiple cycles of the underlying periodic signal. Frequency-domain analysis focuses on the degree to which the voltages on the scalp align to the periodic input signal by analyzing the response at the fundamental frequency of the input and possibly at its harmonics by combining the responses across trials. Alternatively, the EFR can also be measured by presenting a long-duration input and then breaking the measured output into equal epochs (ensuring that each epoch has the same starting phase with respect to the periodic input signal; e.g., Schoof 2014). This approach effectively treats each epoch as a separate trial.

Conceptually, the signal that one is trying to measure, $s(t)$, is identical across different trials. The variation in responses from one trial to another is due to noise, $n(t)$, which can be thought of as a random process. This noise comes from various sources, including physiological activity unrelated to the input (electrical activity associated with nonauditory sensory activity, cognitive functions, or electromyographic activity) and/or from the environment (noise from the recording environment or devices, including harmonics of 60 Hz line noise). The noise $n(t)$ is typically assumed to be zero mean and uncorrelated with $s(t)$. Under these assumptions, $n(t)$ adds to the variability in measurements, but does not change the

expected mean across trials, which equals the signal $s(t)$. However, the metrics used to quantify the EFR are inherently affected by $n(t)$. This means that when comparing EFRs, one must be aware of the effects of noise on the EFR metric.

A number of frequency-domain measures have been utilized to quantify the EFR. Two that have been applied often are the power spectral density (PSD) and the phase-locking value (PLV), each of which is a function of frequency. To calculate the PSD, the waveforms from each of M trials or epochs are averaged together. $PSD(f)$ is then computed as the square of the absolute value of the Fourier transform of this average. The expected value of the PSD equals the sum of the expected signal power and the expected noise power after averaging:

$$PSD(f) = \left|S^2(f)\right| + \frac{1}{M}\left|N^2(f)\right| \tag{7.1}$$

where $S(f)$ and $N(f)$ are the Fourier transforms of $s(t)$ and $n(t)$, respectively.

The PSD is easy to interpret when the noise floor is the same across conditions and/or subjects. If the noise characteristics differ, however, interpretation of the PSD can be problematic: if the PSD at a particular frequency varies significantly across conditions or subjects, it could either be due to differences in the signal or differences in the noise. Subtracting off an estimate of the noise at each frequency can normalize the PSD and mitigate this problem. For instance, for EFRs, the signal is assumed to be zero for all frequencies except the fundamental frequency and its harmonics. The PSD at these nonsignal frequencies provides a direct estimate of the noise. Typically, the noise floor varies relatively smoothly with frequency (often proportional to the reciprocal of the frequency), allowing the noise to be estimated from neighboring frequency bins.

The PLV measures the phase consistency of the response across individual trials (or epochs), ignoring the magnitude of each trial (see Dobie and Wilson 1993). At low SNRs, the PLV is better able to detect the presence of the signal than is the PSD (Dobie and Wilson 1993; Lachaux et al. 1999). Because it ignores the magnitude of the response, chance performance depends only on the number of trials (epochs) being combined to form the estimate of the PLV, making it easy to determine whether or not there is a significant signal in the measurement (Zhu et al. 2013). The magnitude of the PLV depends on the SNR; thus, just as with the PSD, comparisons of PLVs across conditions or across listeners depends upon appropriately characterizing the noise in the measurements (e.g., using resampling methods).

The choice of what kind of metric to use to quantify the EFR should take into account both the SNR of the measurements and the goal of the study. Because the PLV takes on values between zero (response phases are randomly distributed from $-\pi$ to π) and one (response phases are equal on each trial), it is a compressive function of signal level when the SNR is high. That is, the same amount of change in SNR produces increasingly smaller changes in the PLV as SNR increases. As a result, conditions that differ in the strength of the signal in measurements with a high SNR may be difficult to distinguish using the PLV. Compared to the PLV, the

PSD is more sensitive to changes in signal power when SNR is high; across the range of SNRs, the PSD increases linearly with signal strength (see Eq. 7.1). At low SNRs, the PLV scales approximately linearly with SNR. In this SNR regime, the PLV is more likely to differentiate differences in signal level that might be unobservable using the PSD. Taking these issues together, if the goal is simply to detect the presence of a significant signal rather than to estimate differences in the strength of the EFR, the PLV either equals or outperforms the PSD across all SNRs. However, if the goal is to quantify the magnitude of differences in signal strength across individuals, groups, or conditions, either the PSD or the PLV may be better, depending on the SNR.

This effect of noise on the metrics quantifying the EFR can be especially problematic when comparing different behavioral conditions and trying to conclude whether or not the listener state has an influence on the brainstem response. Cortical activity is one of the main sources of noise in EFR recordings. Moreover, cortical activity depends strongly on task demands. Imagine an experiment exploring the question of whether the EFR strength differs when a subject is attending to an auditory source versus attending to a visual source. The different tasks of listening versus watching will change the distribution of cortical activity on the scalp and thus change the amount of noise in the EFR measurement. Alternatively, imagine a "blocked" experimental design where different listening conditions are presented without sufficient randomization. If a subject's focus varies slowly through time (e.g., due to fatigue or inattentiveness), cortical activity will reflect this shift, and different noise levels will bias EFR measures differently in different conditions. Care should be taken to tease apart changes in noise levels from changes in the signal to avoid misinterpreting differences in the estimates of the EFR strength.

7.5.3 *Effects of Stimulus Characteristics*

Responses in the auditory system exhibit a host of nonlinear effects, including forward masking, adaptation, and the like. EFR measurements often implicitly assume that the response that is being measured is constant across trials and, within each trial, the response has settled into a constant, steady-state response. This is not a fair assumption. Indeed, the one study that explored adaptation effects demonstrated that the ASSR to a periodic stimulus is stronger at stimulus onset compared to the later portion of the stimulus (Gockel et al. 2015).

Adaptation effects will be weaker and EFRs will be stronger when each trial consists of a short stimulus and when inter-trial intervals are long. However, a shorter duration stimulus contains fewer cycles of the periodic input, so that the neural response to the input is more affected by onset and offset transients and is in its pseudo-steady-state for a proportionately briefer portion of time. An alternative approach is to present an ongoing stimulus and to analyze epochs of the output response. With this kind of approach, adaptation effects will be maximal, but the

neural response should be more stable (asymptoting toward a true steady-state response).

In considering how to design EFR stimuli and the effects of adaptation, it is worth mentioning that lower-threshold and higher-threshold ANFs differ in their adaptation time constants. Specifically, high-threshold ANFs (i.e., the fibers that may be most susceptible to cochlear neuropathy) have a longer recovery time than do low-threshold fibers (Relkin and Doucet 1991; Furman et al. 2013). Thus, differences in the proportion of high-threshold ANFs versus low-threshold ANFs are likely to affect how adaptation influences EFRs.

7.5.4 Electrode Configuration

The placement of recording electrodes on the scalp and the choice of reference site influence EFR measurements strongly (Stillman et al. 1978; Galbraith 1994). Most EFR studies use a vertical one-channel montage, which emphasizes sustained phase-locked neural activity from the rostral generators in the brainstem (Smith et al. 1975; Stillman et al. 1978). This configuration requires an active lead (usually the vertex channel CZ), reference electrode(s) (usually the earlobes or mastoids), and a ground electrode. Often, the earlobe is the preferred reference (rather than the mastoid) for auditory subcortical recordings because it is a noncephalic site and results in smaller bone vibration artifacts (Hall 2007).

Multiple electrode recordings can be combined to estimate brainstem responses; however, in estimating the EFR, simple time-domain averaging or application of principal component analyses can decrease, rather than increase, the effective SNR of the recorded signal. Specifically, small phase differences in the total signal reaching different recording channels can lead to cancellation of responses. Benefits of multi-channel recording can be realized by averaging frequency-domain amplitudes at the modulation frequency of interest or by using a complex principal components analysis (Bharadwaj and Shinn-Cunningham 2014).

7.6 What the Envelope-Following Response May not Reveal

Although this chapter focuses on why the EFR reflects differences in the precision of subcortical temporal coding, it is also important to mention what kinds of neural processing the EFR response does not index strongly. In particular, there are aspects of auditory processing that are unassailably present in the brainstem (e.g., from electrophysiological animal recordings, neuroimaging techniques, or other approaches) but that do not cause robust effects on the EFR. Two examples are

mentioned here as a reminder that the EFR may be insensitive to a manipulation that has a clear effect on subcortical neural processing.

7.6.1 Binaural Processing

The ITD, computed by comparing the timing of inputs reaching the left and right ears through a coincidence detector, is arguably the perceptual feature that relies most heavily on precise temporal coding in the brainstem (for a review, see Joris et al. 1998). Indeed, ITD sensitivity correlates with EFR strength and both correlate with individual differences in the strength of temporal coding in the brainstem. Many neurons in the IC are sensitive to ITDs, responding preferentially to some ITD values over others (e.g., see Kuwada and Yin 1987). Click ABRs reflect binaural processing in the *binaural interaction component* (BIC), which is defined as the difference between the ABR when sound is presented to both ears and the sum of the monaural ABR responses for clicks presented separately to the left and right ears (Wrege and Starr 1981). The ABR BIC has a relatively low SNR but is generally consistent with a response generated at the level of the lateral lemniscus or IC (e.g., Brantberg et al. 1999). In addition to click responses, other transient brainstem responses are sensitive to binaural cues (Parbery-Clark et al. 2013). Since the EFR itself is dominated by responses coming from the level of IC and is a good index of temporal acuity in individual listeners, it stands to reason that binaural processing might also be reflected in EFRs.

A few studies have reported statistically significant differences between the sum of FFRs from left and right monaural inputs and FFRs to binaural stimuli (e.g., Clark et al. 1997; Ballachanda and Moushegian 2000). However, these studies used pure tones as the acoustic stimuli and thus only assessed the subcortical responses that are phase locked to TFS. Moreover, other studies failed to find any signature of binaural processing in FFRs, reporting that the binaural response was roughly equal to the sum of the two monaural responses (Gerken et al. 1975; Zhang and Boettcher 2008).

Conflicting results are seen for other binaural phenomena, such as evidence of physiological correlates of the binaural masking level difference (BMLD: the difference in the detection threshold for a tone in noise when the tone is presented with an ITD that differs from the ITD of the noise, compared to when both are diotic). One study measured the ASSR to diotic 500 Hz tones in the presence of simultaneous noise and found larger ASSR amplitudes when the noises at the two ears were in phase than when either the tone or the noise was 180° out of phase (Wilson and Krishnan 2005). However, another study concluded that the only correlate of the BMLD was in cortical responses, for slow modulations (7 or 13 Hz), with no significant response from brainstem sources for 80 Hz modulations (Wong and Stapells 2004).

Taken together, the results of these various studies suggest that subcortical FFR signatures of binaural processing are weak. Data from an example experiment lends

support to the idea that the binaural EFR does not reflect spatial-dependent processing. Figure 7.6 shows EFR responses (quantified by the PLV) to broadband click trains presented at a repetition rate of 100 Hz. Figure 7.6A shows the PLV as a function of frequency for one typical subject for a binaural, diotic input, and for the sum of the left and right ear monaural presentations. Figure 7.5B shows the mean PLV at the 100 Hz fundamental frequency of the input (averaged across subjects and plotted as a function of the binaural stimulus ITD) for binaural responses, the sum of the left and right monaural responses, and "corrected" binaural responses (described below).

For both the individual example subject (Fig. 7.6A) and the mean PLV (Fig. 7.6B), the EFR in the binaural condition is greater than the summed monaural

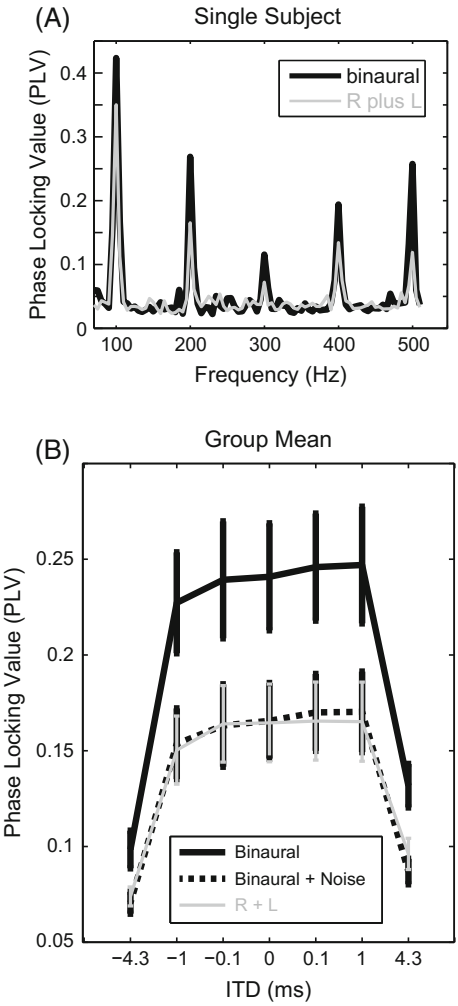

Fig. 7.6 Binaural and monaural envelope-following responses (EFRs) in response to 100 Hz click trains. (**A**) The phase-locking value, plotted as a function of frequency, for one sample subject. The *solid black line* shows the response to a binaural (diotic) presentation, and the *gray line* shows the response to the sum of the monaural responses. (**B**) The across-listener average phase-locking value at 100 Hz, plotted as a function of interaural time difference. *Error bars* show the across-subject standard deviation. The *solid black line* shows the response to binaural stimuli; the *gray line* shows the response to the sum of the monaural responses; the *dashed black line* shows the response to binaural stimuli after compensating for the difference in the noise levels of the summed response and the binaural response. (See text for additional explanation)

responses. This result seems to hint that some component of the response reflects binaural processing. However, this comparison does not take into account the noise in the conditions being compared. Since the presentation order was randomized, the noise floor should be identical across measurement conditions. This means that when the two monaural signals are added, the total noise in the sum has twice the noise power (3 dB more) than the binaural recording. This difference in noise floor actually accounts for the apparent difference between the binaural EFR and the sum of the monaural responses. Once a compensatory level of noise is added to the binaural condition, the binaural response is essentially identical to the sum of the monaural responses (Fig. 7.6B).

Figure 7.6B also shows that the ITD influences the binaural FFR, but only when the ITD is extremely large or extremely small. Importantly, this effect of ITD is also explained by the monaural responses. This reduction at large magnitude ITDs occurs because the left and right ear responses cancel each other; for artificially large ITDs with a magnitude of 4.3 ms, the monaural responses are delayed relative to one another by roughly one-half of the repetition period.

There is no question that binaural cues affect responses in the brainstem, and specifically modulate the synaptic inputs driving the responses of individual neurons in IC. Despite this, there is not a robust, consistent signature of binaural processing in the EFR. This could be due to any number of reasons. For instance, the binaural-specific electrical response may be small compared to responses to monaural stimuli. Alternatively, depending on the ITD, there may be differences in what subpopulation of neurons responds, yet the sum of the responses across the IC population may be roughly constant, independent of ITD. Regardless, the fact that a fundamental feature such as ITD does not have a robust effect on the EFR highlights the limitations of this kind of measure.

7.6.2 *Modulation of Subcortical Responses Due to Selective Attention*

IC receives many descending projections originating in cortex. These efferent projections create a dynamic feedback loop spanning cortical and subcortical auditory processing stages (for a review, see Kraus and White-Schwoch 2015). Such feedback likely guides long-term learning and plasticity and allows cortical feedback to alter the subcortical sound based on task goals (e.g., Chandrasekaran et al. 2012, 2014).

Experience clearly tunes responses in IC. Direct electrical stimulation of auditory cortex shifts the frequency tuning of IC neurons with changes persisting for hours or longer (Suga and Ma 2003). Long-term learning shapes responses in the midbrain to enhance sound features important for perception and behavior (e.g., Chandrasekaran et al. 2007; Chambers et al. 2016). Moreover, a number of

studies show that experience has an impact on EFRs (e.g., Carcagno and Plack 2011; Strait and Kraus 2014). Such effects are considered in other chapters of this volume, including language experience (Krishnan and Gandour, Chap. 3), perceptual learning (Carcagno and Plack, Chap. 4), and musical training (White-Schwoch and Kraus, Chap. 6).

While long-term effects may influence steady-state brainstem responses, the immediate effects of task demands do not show consistent EFR effects. Despite this, it is well established that task demands change physiological responses measured in other ways. In ferrets, spectrotemporal receptive fields of IC neurons change depending on whether the ferrets are actively attending to sounds and performing a listening task compared to when they are passively hearing the same sounds (Slee and David 2015). In humans, selectively attending to a sound in one ear gives rise to higher fMRI activation in the contralateral IC compared to when attention is directed to the opposing ear (Rinne et al. 2008).

Projections from auditory cortex may modulate CN responses (Luo et al. 2008) and underlie changes in CN responses during periods of visual attention (Oatman 1976; Oatman and Anderson 1980). Visually directed attention can even alter responses at the level of the auditory nerve (Oatman 1976). Yet, despite the vast evidence for online modulatory changes in subcortical responses based on subject goals, efforts to demonstrate changes in EFRs due to selective focus of attention have produced mixed results. Although a few studies argued that EFR strength is influenced by exactly which of multiple competing sounds a listener attends (Galbraith et al. 2003; Lehmann and Schonwiesner 2014), the effect sizes are small, the effect directions are inconsistent, and efforts to replicate the effects have failed (see the discussion in Varghese et al. 2015).

7.7 Summary and Conclusions

The EFR provides a window into individual differences in the fidelity of temporal coding in subcortical portions of the auditory pathway. A portion of this variation across listeners reflects compensatory changes and experience-dependent plasticity in brainstem processing. However, a significant portion derives from sensory differences that likely reflect differences in the number of ANFs encoding sound.

Together, noise exposure and aging cause cochlear neuropathy, or death of ANFs. Such cochlear neuropathy reduces the fidelity with which temporal modulation in supra-threshold sounds, such as speech, are encoded in the auditory nerve. This subtle “hidden hearing loss” manifests primarily as perceptual deficits in temporal processing and is especially noticeable when listeners are trying to communicate in noisy social settings (conditions in which listeners must selectively attend in order to understand speech). Individual differences in EFRs quantify these differences in sensory coding and correlate with the resulting differences in perceptual ability. Still, the EFR can be difficult to interpret. When considering how

the EFR varies across listeners, across groups, or across experimental conditions, it is important to understand how the EFR is generated and measured and how measurement noise influences EFR measures. Moreover, there are many subcortical aspects of sound processing that do not influence the EFR.

Compliance with Ethics Barbara Shinn-Cunningham, Leonard Varghese, Le Wang, and Hari Bharadwaj declared that they had no conflicts of interest.

References

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research, 245*(1–2), 35–47. Doi:10.1016/j.heares.2008.08.004

Ananthanarayan, A. K., & Durrant, J. D. (1992). The frequency-following response and the onset response: Evaluation of frequency specificity using a forward-masking paradigm. *Ear and Hearing, 13*(4), 228–232.

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. Doi:10.1523/JNEUROSCI.2176-12.2012

Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013). Reversal of age-related neural timing delays with training. *Proceedings of the National Academy of Sciences of the USA, 110*(11), 4357–4362. Doi:10.1073/pnas.1213555110

Ballachanda, B. B., & Moushegian, G. (2000). Frequency-following response: Effects of interaural time and intensity differences. *Journal of the American Academy of Audiology, 11*(1), 1–11.

Bharadwaj, H. M., Masud, S., Mehraei, G., Verhulst, S., & Shinn-Cunningham, B. G. (2015). Individual differences reveal correlates of hidden hearing deficits. *The Journal of Neuroscience, 35*, 2161–2172.

Bharadwaj, H. M., & Shinn-Cunningham, B. G. (2014). Rapid acquisition of auditory subcortical steady state responses using multichannel recordings. *Clinical Neurophysiology, 125*(9), 1878–1888. Doi:10.1016/j.clinph.2014.01.011

Bharadwaj, H. M., Verhulst, S., Shaheen, L., Liberman, M. C., & Shinn-Cunningham, B. G. (2014). Cochlear neuropathy and the coding of supra-threshold sound. *Frontiers in Systems Neuroscience*. Doi:10.3389/fnsys.2014.00026

Blauert, J. (1997). *Spatial hearing* (2nd ed.). Cambridge, MA: MIT Press.

Boashash, B. (1992). Estimating and interpreting the instantaneous frequency of a signal I. Fundamentals. *Proceedings of the IEEE, 80*(4), 520–538.

Bohne, B. A., & Harding, G. W. (2000). Degeneration in the cochlea after noise damage: Primary versus secondary events. *American Journal of Otology, 21*(4), 505–509.

Bourien, J., Tang, Y., Batrel, C., Huet, A., et al. (2014). Contribution of auditory nerve fibers to compound action potential of the auditory nerve. *Journal of Neurophysiology, 112*(5), 1025–1039. Doi:10.1152/jn.00738.2013

Brantberg, K., Fransson, P. A., Hansson, H., & Rosenhall, U. (1999). Measures of the binaural interaction component in human auditory brainstem response using objective detection criteria. *Scandinavian Audiology, 28*(1), 15–26.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Carcagno, S., & Plack, C. J. (2011). Subcortical plasticity following perceptual learning in a pitch discrimination task. *Journal of the Association for Research in Otolaryngology, 12*(1), 89–100. Doi:10.1007/s10162-010-0236-1

Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences, 8*(10), 465–471.

Chambers, A. R., Resnik, J., Yuan, Y., Whitton, J. P., et al. (2016). Central gain restores auditory processing following near-complete cochlear denervation. *Neuron, 89*(4), 867–879. Doi:10.1016/j.neuron.2015.12.041

Chandrasekaran, B., Kraus, N., & Wong, P. C. (2012). Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology, 107*(5), 1325–1336. Doi:10.1152/jn.00923.2011

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Experience-dependent neural plasticity is sensitive to shape of pitch contours. *NeuroReport, 18*(18), 1963–1967. Doi:10.1097/WNR.0b013e3282f213c5

Chandrasekaran, B., Skoe, E., & Kraus, N. (2014). An integrative model of subcortical auditory plasticity. *Brain Topography, 27*(4), 539–552. Doi:10.1007/s10548-013-0323-9

Christiansen, S. K., & Oxenham, A. J. (2014). Assessing the effects of temporal coherence on auditory stream formation through comodulation masking release. *The Journal of the Acoustical Society of America, 135*(6), 3520–3529. Doi:10.1121/1.4872300

Clark, J. L., Moushegian, G., & Rupert, A. L. (1997). Interaural time effects on the frequency-following response. *Journal of the American Academy of Audiology, 8*(5), 308–313.

Cohen, L. T., Rickards, F. W., & Clark, G. M. (1991). A comparison of steady-state evoked potentials to modulated tones in awake and sleeping humans. *Journal of the Acoustical Society of America, 90*(5), 2467–2479.

Dobie, R. A., & Wilson, M. J. (1993). Objective response detection in the frequency domain. *Electroencephalography and Clinical Neurophysiology, 88*(6), 516–524.

Dolphin, W. F., & Mountain, D. C. (1992). The envelope-following response: Scalp potentials elicited in the Mongolian gerbil using sinusoidally AM acoustic signals. *Hearing Research, 58*(1), 70–78.

Escabi, M. A., & Read, H. L. (2003). Representation of spectrotemporal sound information in the ascending auditory pathway. *Biological Cybernetics, 89*(5), 350–362.

Fitzgibbons, P. J., & Gordon-Salant, S. (2010). Age-related differences in discrimination of temporal intervals in accented tone sequences. *Hearing Research, 264*(1–2), 41–47. Doi:10.1016/j.heares.2009.11.008

Fullgrabe, C., Moore, B. C., & Stone, M. A. (2014). Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. *Frontiers in Aging Neuroscience, 6*, 347. Doi:10.3389/fnagi.2014.00347

Furman, A. C., Kujawa, S. G., & Liberman, M. C. (2013). Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates. *Journal of Neurophysiology*. Doi:10.1152/jn.00164.2013

Galambos, R., Makeig, S., & Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *Proceedings of the National Academy of Sciences of the USA, 78*(4), 2643–2647.

Galbraith, G. C. (1994). Two-channel brainstem frequency-following responses to pure tone and missing fundamental stimuli. *Electroencephalography and Clinical Neurophysiology, 92*(4), 321–330.

Galbraith, G. C., Olfman, D. M., & Huffman, T. M. (2003). Selective attention affects human brainstem frequency-following response. *NeuroReport, 14*(5), 735–738. Doi:10.1097/01.wnr.0000064983.96259.49

Gardi, J., Merzenich, M., & McKean, C. (1979). Origins of the scalp recorded frequency-following response in the cat. *Audiology, 18*(5), 358–381.

Gerken, G. M., Moushegian, G., Stillman, R. D., & Rupert, A. L. (1975). Human frequency-following responses to monaural and binaural stimuli. *Electroencephalography and Clinical Neurophysiology, 38*(4), 379–386.

Goblick, T. J., Jr., & Pfeiffer, R. R. (1969). Time-domain measurements of cochlear nonlinearities using combination click stimuli. *The Journal of the Acoustical Society of America, 46*(4), 924–938.

Gockel, H. E., Krugliak, A., Plack, C. J., & Carlyon, R. P. (2015). Specificity of the human frequency-following response for carrier and modulation frequency assessed using adaptation. *Journal of the Association for Research in Otolaryngology, 16*(6), 747–762. Doi:10.1007/s10162-015-0533-9

Grose, J. H., & Mamo, S. K. (2010). Processing of temporal fine structure as a function of age. *Ear and Hearing, 31*, 755–760. Doi:10.1097/AUD.0b013e3181e627e7

Grose, J. H., & Mamo, S. K. (2012). Frequency modulation detection as a measure of temporal processing: Age-related monaural and binaural effects. *Hearing Research, 294*(1–2), 49–54. Doi:10.1016/j.heares.2012.09.007

Grose, J. H., Mamo, S. K., Buss, E., & Hall, J. W., III. (2015). Temporal processing deficits in middle age. *American Journal of Audiology, 24*(2), 91–93. Doi:10.1044/2015_AJA-14-0053

Grose, J. H., Mamo, S. K., & Hall, J. W., III. (2009). Age effects in temporal envelope processing: Speech unmasking and auditory steady state responses. *Ear and Hearing, 30*(5), 568–575. Doi:10.1097/AUD.0b013e3181ac128f

Hall, J. (2007). *New handbook of auditory evoked responses*. Boston: Pearson.

Hamalainen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencepahlography: Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics, 65*(2), 413–497. Doi:10.1103/RevModPhys.65.413

He, N. J., Mills, J. H., Ahlstrom, J. B., & Dubno, J. R. (2008). Age-related differences in the temporal modulation transfer function with pure-tone carriers. *Journal of the Acoustical Society of America, 124*(6), 3841–3849. Doi:10.1121/1.2998779

Helfer, K. S. (2015). Competing speech perception in middle age. *American Journal of Audiology, 24*(2), 80–83. Doi:10.1044/2015_AJA-14-0056

Helfer, K. S., & Vargo, M. (2009). Speech recognition and temporal processing in middle-aged women. *Journal of the American Academy of Audiology, 20*(4), 264–271.

Herdman, A. T., Picton, T. W., & Stapells, D. R. (2002). Place specificity of multiple auditory steady-state responses. *The Journal of the Acoustical Society of America, 112*(4), 1569–1582.

Hickox, A. E., & Liberman, M. C. (2014). Is noise-induced cochlear neuropathy key to the generation of hyperacusis or tinnitus? *Journal of Neurophysiology, 111*(3), 552–564. Doi:10.1152/jn.00184.2013

Hind, S. E., Haines-Bazrafshan, R., Benton, C. L., Brassington, W., et al. (2011). Prevalence of clinical referrals having hearing thresholds within normal limits. *International Journal of Audiology, 50*(10), 708–716. Doi:10.3109/14992027.2011.582049

Hornickel, J., Chandrasekaran, B., Zecker, S., & Kraus, N. (2011). Auditory brainstem measures predict reading and speech-in-noise perception in school-aged children. *Behavioral and Brain Research, 216*(2), 597–605. Doi:10.1016/j.bbr.2010.08.051

Jin, S. H., Liu, C., & Sladen, D. P. (2014). The effects of aging on speech perception in noise: Comparison between normal-hearing and cochlear-implant listeners. *Journal of the American Academy of Audiology, 25*(7), 656–665. Doi:10.3766/jaaa.25.7.4

John, M. S., Lins, O. G., Boucher, B. L., & Picton, T. W. (1998). Multiple auditory steady-state responses (MASTER): Stimulus and recording parameters. *Audiology, 37*(2), 59–82.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiology Review, 84*(2), 541–577. Doi:10.1152/physrev.00029.2003

Joris, P. X., Smith, P. H., & Yin, T. C. (1998). Coincidence detection in the auditory system: 50 years after Jeffress. *Neuron, 21*(6), 1235–1238. S0896-6273(00)80643-1 [pii].

Joris, P. X., & Yin, T. C. (1992). Responses to amplitude-modulated tones in the auditory nerve of the cat. *The Journal of the Acoustical Society of America, 91*(1), 215–232.

Kiren, T., Aoyagi, M., Furuse, H., & Koike, Y. (1994). An experimental study on the generator of amplitude-modulation following response. *Acta Otolaryngolica Supplement, 511*, 28–33.

Kraus, N., & White-Schwoch, T. (2015). Unraveling the biology of auditory learning: A cognitive-sensorimotor-reward framework. *Trends in Cognitive Sciences, 19*(11), 642–654.

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. Doi:10.1523/JNEUROSCI.4985-05.2006

Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: Cochlear nerve degeneration after "temporary" noise-induced hearing loss. *The Journal of Neuroscience, 29*(45), 14077–14085. Doi:10.1523/JNEUROSCI.2845-09.2009

Kujawa, S. G., & Liberman, M. C. (2015). Synaptopathy in the noise-exposed and aging cochlea: Primary neural degeneration in acquired sensorineural hearing loss. *Hearing Research, 330*(Pt B), 191–199. Doi:10.1016/j.heares.2015.02.009

Kumar, G., Amen, F., & Roy, D. (2007). Normal hearing tests: Is a further appointment really necessary? *Journal of the Royal Society of Medicine, 100*(2), 66. Doi:10.1258/jrsm.100.2.66-a

Kuwada, S., Anderson, J. S., Batra, R., Fitzpatrick, D. C., et al. (2002). Sources of the scalp-recorded amplitude-modulation following response. *Journal of the American Academy of Audiology, 13*(4), 188–204.

Kuwada, S., & Yin, T. C. T. (1987). Physiological studies of directional hearing. In W. A. Yost & G. Gourevitch (Eds.), *Directional hearing* (pp. 146–176). New York: Springer.

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping, 8*(4), 194–208. Doi:10.1002/(SICI)1097-0193(1999)8:4<194:AID-HBM4>3.0.CO;2-C [pii]

Lee, A. K. C., Larson, E., & Maddox, R. K. (2012). Mapping cortical dynamics using simultaneous MEG/EEG and anatomically-constrained minimum-norm estimates: An auditory attention example. *Journal of Visual Experiments, 68*, e4262. Doi:10.3791/4262

Lehmann, A., & Schonwiesner, M. (2014). Selective attention modulates human auditory brainstem responses: Relative contributions of frequency and spatial cues. *PLoS ONE, 9*(1), e85442. Doi:10.1371/journal.pone.0085442

Liberman, M. C. (1980). Morphological differences among radial afferent fibers in the cat cochlea: An electron-microscopic study of serial sections. *Hearing Research, 3*(1), 45–63.

Liberman, M. C. (2015). Hidden hearing loss. *Scientific American, 313*(2), 48–53.

Liberman, M. C., Chesney, C., & Kujawa, S. (1997). Effects of selective inner hair cell loss on DPOAE and CAP in carboplatin-treated chinchillas. *Auditory Neuroscience, 3*(3), 255–268.

Lin, H. W., Furman, A. C., Kujawa, S. G., & Liberman, M. C. (2011). Primary neural degeneration in the guinea pig cochlea after reversible noise-induced threshold shift. *Journal of the Association for Research in Otolaryngology, 12*(5), 605–616. Doi:10.1007/s10162-011-0277-0

Lins, O. G., Picton, T. W., Boucher, B. L., Durieux-Smith, A., et al. (1996). Frequency-specific audiometry using steady-state responses. *Ear and Hearing, 17*(2), 81–96.

Lobarinas, E., Salvi, R., & Ding, D. (2013). Insensitivity of the audiogram to carboplatin induced inner hair cell loss in chinchillas. *Hearing Research, 302*, 113–120. Doi:10.1016/j.heares.2013.03.012

Luo, F., Wang, Q., Kashani, A., & Yan, J. (2008). Corticofugal modulation of initial sound processing in the brain. *The Journal of Neuroscience, 28*(45), 11615–11621. Doi:10.1523/JNEUROSCI.3972-08.2008

Maddox, R. K., & Shinn-Cunningham, B. G. (2012). Influence of task-relevant and task-irrelevant feature continuity on selective auditory attention. *Journal of the Association for Research in Otolaryngology, 13*(1), 119–129. Doi:10.1007/s10162-011-0299-7

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. Doi:10.1007/s10162-011-0283-2

Marsh, J. T., Brown, W. S., & Smith, J. C. (1975). Far-field recorded frequency-following responses: Correlates of low pitch auditory perception in humans. *Electroencephalography and Clinical Neurophysiology, 38*(2), 113–119.

Mehraei, G., Hickox, A. E., Bharadwaj, H. M., Goldberg, H., et al. (2016). Auditory brainstem response latency in noise as a marker of cochlear synaptopathy. *The Journal of Neuroscience, 36*(13), 3755–3764. Doi:10.1523/JNEUROSCI.4460-15.2016

Mehta, A., Prabhakar, M., Kumar, P., Deshmukh, R., & Sharma, P. L. (2013). Excitotoxicity: Bridge to various triggers in neurodegenerative disorders. *European Journal of Pharmacology, 698*(1–3), 6–18. Doi:10.1016/j.ejphar.2012.10.032

Milstein, J. N., & Koch, C. (2008). Dynamic moment analysis of the extracellular electric field of a biologically realistic spiking neuron. *Neural Computation, 20*(8), 2070–2084. Doi:10.1162/neco.2008.06-07-537

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). San Diego, CA: Academic Press.

Oatman, L. C. (1976). Effects of visual attention on the intensity of auditory evoked potentials. *Experimental Neurology, 51*(1), 41–53.

Oatman, L. C., & Anderson, B. W. (1980). Suppression of the auditory frequency-following response during visual attention. *Electroencephalography and Clinical Neurophysiology, 49*(3–4), 314–322.

Oertel, D., Bal, R., Gardner, S. M., Smith, P. H., & Joris, P. X. (2000). Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus. *Proceedings of the National Academy of Sciences of the USA, 97*(22), 11773–11779. Doi:10.1073/pnas.97.22.11773

Okada, Y. C., Wu, J., & Kyuhou, S. (1997). Genesis of MEG signals in a mammalian CNS structure. *Electroencephalography and Clinical Neurophysiology, 103*(4), 474–485.

Parbery-Clark, A., Strait, D. L., Hittner, E., & Kraus, N. (2013). Musical training enhances neural processing of binaural sounds. *The Journal of Neuroscience, 33*(42), 16741–16747. Doi:10.1523/JNEUROSCI.5700-12.2013

Parbery-Clark, A., Strait, D. L., & Kraus, N. (2011). Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia, 49*(12), 3338–3345. Doi:10.1016/j.neuropsychologia.2011.08.007

Pauli-Magnus, D., Hoch, G., Strenzke, N., Anderson, S., et al. (2007). Detection and differentiation of sensorineural hearing loss in mice using auditory steady-state responses and transient auditory brainstem responses. *Neuroscience, 149*(3), 673–684. Doi:10.1016/j.neuroscience.2007.08.010

Plack, C. J., Barker, D., & Prendergast, G. (2014). Perceptual consequences of "hidden" hearing loss. *Trends in Hearing,* 18. Doi:10.1177/2331216514550621

Pujol, R., Puel, J. L., Gervais d'Aldin, C., & Eybalin, M. (1993). Pathophysiology of the glutamatergic synapses in the cochlea. *Acta Otolaryngolica, 113*(3), 330–334.

Purcell, D. W., John, S. M., Schneider, B. A., & Picton, T. W. (2004). Human temporal auditory acuity as assessed by envelope-following responses. *Journal of the Acoustical Society of America, 116*(6), 3581–3593.

Quaranta, A., Portalatini, P., & Henderson, D. (1998). Temporary and permanent threshold shift: An overview. *Scandinavian Audiology Supplement, 48*, 75–86.

Relkin, E. M., & Doucet, J. R. (1991). Recovery from prior stimulation. I: Relationship to spontaneous firing rates of primary auditory neurons. *Hearing Research, 55*(2), 215–222.

Rinne, T., Balk, M. H., Koistinen, S., Autti, T., et al. (2008). Auditory selective attention modulates activation of human inferior colliculus. *Journal of Neurophysiology, 100*(6), 3323–3327. Doi:10.1152/jn.90607.2008

Rosen, S., Cohen, M., & Vanniasegaram, I. (2010). Auditory and cognitive abilities of children suspected of auditory processing disorder (APD). *International Journal of Pediatric Otorhinolaryngology, 74*(6), 594–600. Doi:10.1016/j.ijporl.2010.02.021

Ruggles, D., Bharadwaj, H., & Shinn-Cunningham, B. G. (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. Doi:10.1073/pnas.1108912108

Ruggles, D., Bharadwaj, H., & Shinn-Cunningham, B. G. (2012). Why middle-aged listeners have trouble hearing in everyday settings. *Current Biology, 22*(15), 1417–1422. Doi:10.1016/j.cub.2012.05.025

Ruggles, D., & Shinn-Cunningham, B. (2011). Spatial selective auditory attention in the presence of reverberant energy: Individual differences in normal-hearing listeners. *Journal of the Association for Research in Otolaryngology, 12*(3), 395–405. Doi:10.1007/s10162-010-0254-z

Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology, 115*(9), 2021–2030. Doi:10.1016/j.clinph.2004.04.003

Schaette, R., & McAlpine, D. (2011). Tinnitus with a normal audiogram: Physiological evidence for hidden hearing loss and computational model. *The Journal of Neuroscience, 31*(38), 13452–13457. Doi:10.1523/JNEUROSCI.2156-11.2011

Schmiedt, R. A., Mills, J. H., & Boettcher, F. A. (1996). Age-related loss of activity of auditory-nerve fibers. *Journal of Neurophysiology, 76*(4), 2799–2803.

Schoof, T. (2014). *The effects of ageing on the perception of speech in noise.* Dissertation, University College London, London, UK.

Sergeyenko, Y., Lall, K., Liberman, M. C., & Kujawa, S. G. (2013). Age-related cochlear synaptopathy: An early-onset contributor to auditory functional decline. *Journal of Neuroscience, 33*(34), 13686–13694. Doi:10.1523/JNEUROSCI.1783-13.2013

Shaheen, L. A., Valero, M. D., & Liberman, M. C. (2015). Towards a diagnosis of cochlear neuropathy with envelope-following responses. *Journal of the Association for Research in Otolaryngology, 16*(6), 727–745. Doi:10.1007/s10162-015-0539-3

Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences, 34*(3), 114–123. Doi:10.1016/j.tins.2010.11.002. S0166-2236(10)00167-0 [pii]

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12*(5), 182–186. Doi:10.1016/j.tics.2008.02.003

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification, 12*(4), 283–299. Doi:10.1177/1084713808325306

Shinn-Cunningham, B., Ruggles, D. R., & Bharadwaj, H. (2013). How early aging and environment interact in everyday listening: From brainstem to behavior through modeling. *Basic Aspects of Hearing: Physiology and Perception, 787*, 501–510. Doi:10.1007/978-1-4614-1590-9_55

Skoe, E., Chandrasekaran, B., Spitzer, E. R., Wong, P. C., & Kraus, N. (2014). Human brainstem plasticity: The interaction of stimulus probability and auditory learning. *Neurobiological Learning and Memory, 109*, 82–93. Doi:10.1016/j.nlm.2013.11.011

Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: A tutorial. *Ear and Hearing, 31*(3), 302–324. Doi:10.1097/AUD.0b013e3181cdb272

Slater, J., Skoe, E., Strait, D. L., O'Connell, S., et al. (2015). Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. *Behavioral and Brain Research, 291*, 244–252. Doi:10.1016/j.bbr.2015.05.026

Slee, S. J., & David, S. V. (2015). Rapid task-related plasticity of spectrotemporal receptive fields in the auditory midbrain. *The Journal of Neuroscience, 35*(38), 13090–13102. Doi:10.1523/JNEUROSCI.1671-15.2015

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature, 416*(6876), 87–90.

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.

Snell, K., & Frisina, D. R. (2000). Relationship among age-related differences in gap detection and word recognition. *The Journal of the Acoustical Society of America, 107*(3), 1615–1626.

Snell, K. B., Mapes, F. M., Hickman, E. D., & Frisina, D. R. (2002). Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity. *The Journal of the Acoustical Society of America, 112*(2), 720–727.

Sohmer, H., Pratt, H., & Kinarti, R. (1977). Sources of frequency-following responses (FFR) in man. *Electroencephalography and Clinical Neurophysiology, 42*(5), 656–664.

Stamper, G. C., & Johnson, T. A. (2015). Auditory function in normal-hearing, noise-exposed human ears. *Ear and Hearing, 36*(2), 172–184. Doi:10.1097/AUD.0000000000000107

Stapells, D. R., Linden, D., Suffield, J. B., Hamel, G., & Picton, T. W. (1984). Human auditory steady state potentials. *Ear and Hearing, 5*(2), 105–113.

Starr, A., Picton, T. W., Sininger, Y., Hood, L. J., & Berlin, C. I. (1996). Auditory neuropathy. *Brain, 119*(Pt 3), 741–753.

Stillman, R. D., Crow, G., & Moushegian, G. (1978). Components of the frequency-following potential in man. *Electroencephalography and Clinical Neurophysiology, 44*(4), 438–446.

Strait, D. L., Hornickel, J., & Kraus, N. (2011). Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behavioral Brain Function, 7*(1), 44. Doi:10.1186/17449081-7-44

Strait, D. L., & Kraus, N. (2014). Biological impact of auditory expertise across the life span: Musicians as a model of auditory learning. *Hearing Research, 308*, 109–121. Doi:10.1016/j.heares.2013.08.004

Strelcyk, O., & Dau, T. (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *Journal of the Acoustical Society of America, 125*(5), 3328–3345. Doi:10.1121/1.3097469

Suga, N., & Ma, X. (2003). Multiparametric corticofugal modulation and plasticity in the auditory system. *Nature Review Neuroscience, 4*(10), 783–794. Doi:10.1038/nrn1222

Szydlowska, K., & Tymianski, M. (2010). Calcium, ischemia and excitotoxicity. *Cell Calcium, 47*(2), 122–129. Doi:10.1016/j.ceca.2010.01.003

Valero, M. D., Hancock, K. E., & Liberman, M. C. (2016). The middle ear muscle reflex in the diagnosis of cochlear neuropathy. *Hearing Research, 332*, 29–38. Doi:10.1016/j.heares.2015.11.005

Varghese, L., Bharadwaj, H. M., & Shinn-Cunningham, B. G. (2015). Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. *Brain Research, 1626*, 146–164. Doi:10.1016/j.brainres.2015.06.038

Verhulst, S., Bharadwaj, H. M., Mehraei, G., Shera, C. A., & Shinn-Cunningham, B. G. (2015). Functional modeling of the human auditory brainstem response to broadband stimulation. *The Journal of the Acoustical Society of America, 138*(3), 1637–1659. Doi:10.1121/1.4928305

Whitton, J. P., Hancock, K. E., & Polley, D. B. (2014). Immersive audiomotor game play enhances neural and perceptual salience of weak signals in noise. *Proceedings of the National Academy of Sciences of the USA, 111*(25), E2606–E2615. Doi:10.1073/pnas.1322184111

Wible, B., Nicol, T., & Kraus, N. (2005). Correlation between brainstem and cortical auditory processes in normal and language-impaired children. *Brain, 128*(Pt 2), 417–423. Doi:10.1093/brain/awh367

Wilson, J. R., & Krishnan, A. (2005). Human frequency-following responses to binaural masking level difference stimuli. *Journal of the American Academy of Audiology, 16*(3), 184–195.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*(4), 420–422. Doi:10.1038/nn1872

Wong, W. Y., & Stapells, D. R. (2004). Brainstem and cortical mechanisms underlying the binaural masking level difference in humans: An auditory steady-state response study. *Ear and Hearing, 25*(1), 57–67. Doi:10.1097/01.AUD.0000111257.11898.64

Wrege, K. S., & Starr, A. (1981). Binaural interaction in human auditory brainstem evoked potentials. *Archives of Neurology, 38*(9), 572–580.

Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., et al. (2005). Speech recognition with amplitude and frequency modulations. *Proceedings of the National Academy of Sciences of the USA, 102*(7), 2293–2298. Doi:10.1073/pnas.0406460102

Zhang, F., & Boettcher, F. A. (2008). Effects of interaural time and level differences on the binaural interaction component of the 80 Hz auditory steady-state response. *Journal of the American Academy of Audiology, 19*(1), 82–94.

Zhu, L., Bharadwaj, H., Xia, J., & Shinn-Cunningham, B. (2013). A comparison of spectral magnitude and phase-locking value analyses of the frequency-following response to complex tones. *The Journal of the Acoustical Society of America, 134*(1), 384–395. Doi:10.1121/1.4807498