Published in final edited form as:
Ear Hear. 2010 June ; 31(3): 302–324. doi:10.1097/AUD.0b013e3181cdb272.

Auditory brainstem response to complex sounds: a tutorial

Erika Skoe[a,+] and **Nina Kraus**[a,b]

[a] Department of Communication Sciences, Northwestern University

[b] Departments of Neurobiology & Physiology; Otolaryngology; Northwestern University

Abstract

This tutorial provides a comprehensive overview of the methodological approach to collecting and analyzing auditory brainstem responses to complex sounds (cABRs). cABRs provide a window into how behaviorally relevant sounds such as speech and music are processed in the brain. Because temporal and spectral characteristics of sounds are preserved in this subcortical response, cABRs can be used to assess specific impairments and enhancements in auditory processing. Notably, subcortical function is neither passive nor hardwired but dynamically interacts with higher-level cognitive processes to refine how sounds are transcribed into neural code. This experience-dependent plasticity, which can occur on a number of time scales (e.g., life-long experience with speech or music, short-term auditory training, online auditory processing), helps shape sensory perception. Thus, by being an objective and non-invasive means for examining cognitive function and experience-dependent processes in sensory activity, cABRs have considerable utility in the study of populations where auditory function is of interest (e.g., auditory experts such as musicians, persons with hearing loss, auditory processing and language disorders). This tutorial is intended for clinicians and researchers seeking to integrate cABRs into their clinical and/or research programs.

Keywords

auditory brainstem response; speech; music; complex sounds; tutorial

INTRODUCTION

The human soundscape is characterized by complex sounds with rich harmonic structures, dynamic amplitude modulations, and rapid spectrotemporal fluctuations. This complexity is represented by an exceptionally precise temporal and spectral neural code within the auditory brainstem, an ensemble of nuclei belonging to the efferent and afferent auditory systems. Within the brainstem, two broad classes of time-locked responses can be defined, namely transient and sustained. As the names suggest, brief, non-periodic stimulus features evoke transient responses, whereas periodic features elicit sustained phaselocked responses (see Figure 1). Discovered nearly 40 years ago (Jewett et al. 1970; Jewett and Williston 1971; Moushegian et al. 1973), brainstem responses can be measured using scalp electrodes that pick up electrical potentials generated by the synchronous activity of neural populations in the

---

[+]**Corresponding author** Erika Skoe Northwestern University 2240 Campus Drive Evanston IL, 60208 USA eeskoe@northwestern.edu Phone: 847-491-2465 Fax: 847-491-2523.

brainstem. Because these aggregate neural responses can be recorded objectively and passively, they offer an excellent means to assess auditory function in a clinical setting. It is for this reason that the click-evoked ABR has enjoyed wide-scale clinical use as a metric for determining auditory thresholds and detecting neuropathologies (Sininger 1993; Starr et al. 1996; Hood 1998; Hall 2006).

Early studies of the auditory brainstem response (ABR) used simple stimuli such as clicks and sinusoids to tap into and maximize these transient and sustained auditory brainstem responses. While clicks and tones have been instrumental in defining these basic response patterns, they are poor approximations of the behaviorally relevant sounds we encounter outside the laboratory (e.g., speech and music, non-speech vocal sounds, and environmental sounds). Moreover, although complex sounds consist of both sustained and transient features, the response to a complex sound is not necessarily predictable from the response to click and tones (Palmer and Shamma 2004; Song et al. 2006; Johnson et al. 2008a). For these reasons, auditory neuroscience has gradually transitioned to using sounds that are more complex.

Greenberg (1980) was one of the first to adopt complex stimuli for recording ABRs. Following the seminal work by Young and Sachs (Young and Sachs 1979) showing that speech formants are preserved in the discharge pattern of the auditory nerve, Greenberg demonstrated that speech-specific information (i.e., vowel formants) is also faithfully encoded in the auditory brainstem response (Greenberg 1980). This stimulus fidelity was further supported by Galbraith and colleagues who demonstrated that cABRs to words can be heard as intelligible speech when converted from a neural signal into an audio signal (Galbraith et al. 1995). Importantly, because cABRs occur several milliseconds after the stimulus onset, they reflect a response of neural origin and not the cochlear microphonic (CM) (Galbraith et al. 1995), which occurs nearly simultaneously with the stimulus onset (See section IV for more on the CM).

A plethora of complex stimuli has now been used to examine how the temporal and spectral features of sounds are preserved in the auditory brainstem response (Table 1). By far the two most extensively studied are the consonant-vowel (CV) constructions /dɑ/ (Cunningham et al. 2001; Plyler and Ananthanarayan 2001; King et al. 2002; Russo et al. 2004; Wible et al. 2004; Kraus and Nicol 2005; Russo et al. 2005; Wible et al. 2005; Musacchia et al. 2007; Johnson et al. 2008a; Banai et al. 2009; Burns et al. 2009; Parbery-Clark et al. in press) and various Mandarin syllables with differing pitch contours (i.e., lexical tones) (Krishnan et al. 2004; Krishnan et al. 2005; Xu et al. 2006; Wong et al. 2007; Song et al. 2008; Krishnan and Gandour 2009; Krishnan et al. 2009b). The auditory brainstem response to /dɑ/ has been investigated under different recording conditions: monaural (Cunningham et al. 2001; Banai et al. 2009) and binaural (Musacchia et al. 2008; Parbery-Clark et al. in press) stimulation; left ear and right ear stimulation (Hornickel et al. 2009a); audiovisual and auditory-only stimulation (Musacchia et al. 2006; Musacchia et al. 2007); and in the presence of background noise (Cunningham et al. 2001; Russo et al. 2004; Russo et al. 2005; Russo et al. 2008a; Russo et al. 2008b; Parbery-Clark et al. in press). Moreover, in addition to manipulating stimulus parameters (e.g., duration of the stimulus, duration of the formant transition, and formant frequency settings), cABRs to /dɑ/ has been evaluated before and after auditory training (Russo et al. 2008b; Song et al. 2008), across the lifespan (Johnson et al. 2008b; Burns et al. 2009), and in a number of different populations including musicians (Musacchia et al. 2007; Musacchia et al. 2008; Kraus et al. 2009; Parbery-Clark et al. in press) and children with dyslexia, specific language impairment, and autism spectrum disorders (Cunningham et al. 2001; Banai et al. 2005; Banai and Kraus 2008; Banai et al. 2009; Hornickel et al. 2009b; Chandrasekaran et al. 2009) reviewed in Banai and Kraus 2008). Similarly, brainstem responses to syllables with Mandarin pitch contours have also been studied from numerous perspectives including: native and non-native Mandarin speakers (Krishnan et al. 2005; Xu et al. 2006); musicians and non-musicians (Wong et al. 2007); before and after auditory training (Song et al. 2008); under speech and non-speech conditions (i.e., contour embedded in iterated rippled noise) (Swaminathan et al. 2008; Krishnan et al. 2009a); and using native (curvilinear) (Krishnan et al. 2005) and non-native (linear) (Xu et al. 2006) pitch contours (Table 1).

The list of stimuli used to evoke cABRs extends beyond speech syllables and includes words and phrases (e.g., “car”, “minute”, “chair”, “rose” (Galbraith et al. 1995); “dani” (unpublished data); “chicken pot pie” (Galbraith et al. 2004)). Investigators have also started to explore the use of environmental and affective non-speech vocal sounds (e.g., a baby's cry) (Strait et al. 2009) and musical sounds as viable stimuli for brainstem-evoked recordings. Work on music-evoked ABRs has included a bowed cello note (Musacchia et al. 2007; Musacchia et al. 2008), a five-note musical melody (Skoe and Kraus 2009), and consonant and dissonant two-note intervals synthesized from an electric piano (Lee et al. 2009; tone-complexes) and tone complexes (Bidelman and Krishnan, 2009; see also Greenberg et al., 1997). Despite being a relatively new endeavor, we anticipate that this arena of research will experience a surge in the upcoming years.

The study of cABRs is a young science. There are many more stimuli, populations, and experimental paradigms yet to be explored. Although interest in this topic is growing, most of the existing cABR research has come from a handful of laboratories around the world. Each laboratory has taken a slightly different approach to collecting and analyzing cABRs and while the exact methodologies may differ, this work has led to a complete rethinking of what the human auditory brainstem response represents and how it can be used to study auditory function. Taken as a whole, this work demonstrates that cABRs are an objective and non-invasive means for studying auditory function in expert (e.g., musicians, native speakers), typically developing, and impaired populations (e.g., persons with hearing loss, auditory processing disorders and language impairments). Perhaps most crucially, this work has revealed that subcortical processing of sound is not hardwired. It is instead malleable with experience and inextricably linked to cognitive functions involving language and music. This retuning of subcortical function likely involves the corticofugal pathway, an expansive tract of efferent connections that are even more abundant than afferent connections (Galbraith et al. 1995; Krishnan et al. 2005; Banai et al. 2007; Kraus and Banai 2007; Banai and Kraus 2008; Tzounopoulos and Kraus 2009; see Banai et al. 2009 and Kraus et al. 2009 for treatments of language and music work, respectively).

It should be pointed out that we use the term “auditory brainstem response” to describe both transient and sustained responses originating in the auditory brainstem and midbrain (see Chandrasekaran and Kraus 2009 for an in depth discussion of the origins of the cABR). Because the acronym ABR has been widely adopted to refer to click-evoked auditory brainstem responses, ABRs to complex sounds are often differentiated by other names including speech-ABR or music-ABR. However, because this tutorial focuses on complex sounds in a more general sense that includes non-speech vocal sounds and environmental sounds, we adopt the phrase complex-ABR or cABR when referring to subcortical responses evoked by complex sounds of any variety.

This tutorial, which represents nearly a decade of accumulated knowledge, was written to encourage researchers and clinicians to adopt cABRs into their clinical and/or research programs. To help answer frequency asked questions and to prevent often-encountered stumbling blocks, we provide a comprehensive overview of our systematic approach to stimulus selection, data collection and analyses. While we focus primarily on the methodologies used in our laboratory, we also discuss alternative approaches that others have found successful. Because there are a number of systems on the market that can be used to collect evoked potentials (EPs), we frame this tutorial generally and do not provide instructions that are specific to a particular system. More advanced topics are covered in footnotes and figure captions.

Recommended background reading

This tutorial serves as companion to recent review articles produced by our lab (Kraus and Nicol 2005; Chandrasekaran et al. 2007; Kraus and Banai 2007; Banai and Kraus 2008; Chandrasekaran and Kraus 2009; Kraus et al. 2009; Tzounopoulos and Kraus 2009). If you are not trained in electrophysiology, we suggest that you become familiar with the literature on the electrophysiologic response to clicks and tones, as this work formed the foundation for cABR research and still guides our interpretation and analysis today. Also, because ABRs to complex sounds are recorded using essentially the same data acquisition procedures as ABRs to click and tones, and because many of the experimental considerations are also shared, we abbreviate our treatment of these topics and refer the reader to the following resources where these subjects are explored in great depth: Hood 1998, Hall 2006, Burkard et al. 2007, and Krishnan 2007.

I. CLINICAL CONSIDERATIONS

Brainstem responses to complex sounds are well suited for clinical applications. In addition to being sensitive biological markers of maturation (Johnson et al. 2008a; Burns et al. 2009) and auditory training (Russo et al. 2005; Song et al. 2008), cABRs are highly replicable across test sessions and reliably measured under passive conditions using a small number of electrodes (Russo et al. 2004; Russo et al. 2005). By providing information about the biological basis of hearing and language disorders, cABRs can also help to identify those individuals who are most likely to benefit from auditory training (Hayes et al. 2003; Russo et al. 2005). Thus in the assessment of hearing and language function, cABRs can complement existing technologies such as click-ABRs and distortion product otoacoustic emissions (Elsisy and Krishnan 2008; Dhar et al. 2009) and behavioral tests of language and auditory processing, including tests of speech in noise (SIN) perception.

Our work with clinical populations has revealed a link between cABRs and higher-level language processes such as reading and SIN perception (Banai et al. 2009; Hornickel et al. 2009b; Chandrasekaran et al. 2009; Parbery-Clark et al. in press). For children with language-based learning and reading impairments, brainstem deficits are specific to the fast spectrotemporal elements of speech (Cunningham et al. 2001; King et al. 2002; Wible et al. 2004; Banai et al. 2005; Wible et al. 2005; Johnson et al. 2007; Banai et al. 2009; Hornickel et al. 2009b). This is in contrast to the more pervasive encoding deficits seen in children with autism, which also include abnormal subcortical pitch tracking (Russo et al. 2008b; Russo et al. 2009). Subcortical responses to speech also show a longer developmental trajectory than click-ABRs (Johnson et al. 2008a), suggesting that cABRs could provide an objective index in the early identification of auditory processing deficits that lead to learning or literacy disorders (Johnson et al. 2008b). This work with clinical populations led to the development of BioMARK (Biological Marker of Auditory Processing, Natus Medical, Inc.; see http://www.brainvolts.northwestern.edu under "Clinical Technologies"), a clinical measure of speech-sound encoding.

In clinical practice, poor speech perception in noise is a commonly encountered complaint. Although everyone experiences reduced speech perception in noise, children with auditory processing disorders and language-based learning impairments, older adults, and individuals with sensorineural hearing loss often experience excessive difficulty in suboptimal listening situations (Dubno et al. 1984; Pichora-Fuller et al. 1995; Kraus et al. 1996; Bradlow et al. 2003; Ziegler et al. 2009). These perceptual findings are reflected in the electrophysiological response when noise (e.g., white noise, multi-talker babble) masks the stimulus. In addition to showing that acoustic noise disrupts cABRs in normal populations (Russo et al. 2004), our has work revealed that cABR abnormalities, which can often emerge in responses recorded with masking noise (Cunningham et al. 2001), are linked to poorer SIN perception (Hornickel et al. 2009b). This relationship between subcortical function and speech perception in noise is also evident in musicians who demonstrate better performance on clinical tests of SIN perception relative to non-musician controls, as well as more robust cABR representation of stimulus features in background noise. Thus, by counteracting the deleterious effects of masking noise (Parbery-Clark et al. 2009; Parbery-Clark et al. in press), music may provide a potential remediation strategy for individuals with SIN impairments.

Although cABRs can be used to identify a possible disorder, they do not provide the specificity needed to pinpoint the site or exact nature of the disorder. This is because an abnormal outcome on a single measure may reflect more than one underlying cause or disorder. The same can be said for click-ABRs (Hood 1998). Thus, no single cABR measure should be considered in isolation when forming a clinical interpretation of the results (See Section V). However, when interpreted collectively, cABR measures provide an objective means for corroborating suspected disorders.

II. STIMULUS SELECTION AND CREATION

Stimulus Selection

Stimulus selection should factor in (1) the population being studied, (2) the specific research questions at hand, (3) electrophysiological properties of the auditory brainstem and (4) the acoustic features that maximize transient and sustained responses (Table 2).

As a practical guide to stimulus selection, we start with our rationale for selecting /dɑ/ as our primary stimulus. We then describe the transient and sustained aspects of complex stimuli and how they manifest in the cABR. Because there is such a clear transparency between the acoustic features of the stimulus and their subcortical transcription, a basic knowledge of acoustics is necessary for both selecting the stimulus and analyzing the response. While it is beyond the scope of this tutorial to provide a comprehensive overview, we conclude this section with brief descriptions of the complex nature of speech and music.

If you are new to cABRs, you are strongly advised to start with stimuli that have been thoroughly characterized and vetted (e.g., /dɑ/, vowels, Mandarin pitch contours) to ensure that your collection system is functioning properly (Section III). When using novel stimuli, pilot studies are mandatory. During the piloting phase, several different stimulus tokens should be used to determine whether robust and reliable cABRs can be obtained. (For a general overview of the techniques used for detecting and assessing evoked potentials, refer to Elberling and Don 2007). Tips for maximizing cABRs are provided throughout this section and summarized in Table 2.

Why /dɑ/?—Although our lab currently uses a large repertoire of sounds, our early cABR work focused on the syllable /dɑ/ (Cunningham et al. 2001; Russo et al. 2004), an acoustically complex sound, which begins with a stop burst, characterized by an aharmonic and broadband frication, and is followed by a harmonically rich and spectrally dynamic formant transition. This CV syllable was chosen for a number of reasons. First, /dɑ/ is a relatively universal syllable that is included in the phonetic inventories of most European languages (UCLA Phonological Segment Inventory Database, (Maddieson 1984)). Second, the sound consists of a transient segment followed by a sustained periodic segment. It is, in a sense, much like a click followed by a tone — two acoustic signals whose brainstem response properties have been extensively characterized. Due to these acoustic similarities, the transient onset response to the stop burst is similar to the click-ABR, and the sustained response to the vowel is similar to tone-evoked frequency following response (FFR). Third, stop consonants pose great perceptual challenges to clinical populations such as the hearing- and learning-impaired (Tallal and Stark 1981; Turner et al. 1992; Kraus et al. 1996). Moreover, in noisy backgrounds, because stop-bursts are rapid and low in amplitude compared to vowels, even normal hearing adults and children can find it difficult to discriminate between contrastive stop consonants (e.g., “dare” vs. “bare”). Finally, we continue to use this syllable as our primary stimulus because it elicits clear and replicable auditory brainstem responses.

Transient Features—Transient responses, which are characterized by fast response peaks lasting fractions of milliseconds, are evoked by brief, non-sustained stimulus features such as the onset and offset of sounds. In the case of speech syllables, transient features include the onset of the sound, the onset of vocal chord vibration (i.e., voicing) and the offset of sound (Figure 1). For a simple musical stimulus, such as a bowed note of a violin, transient features include the initial burst of sound created by the bow contacting the string and the offset of sound. The morphology of the cABR onset is dictated by the attack characteristics (i.e., how quickly the sound reaches full volume) of the specific acoustic token. Stimuli with sharper rise times (i.e., abrupt onset/amplitude bursts; clicks) are more broadband (i.e., less frequency specific) and cause broader and more simultaneous activation of the cochlea, which enlists a larger population of neurons to fire synchronously (Hood 1998) and leads to more robust (i.e., larger and earlier) transient responses.

For both speech and music, attack characteristics are important for imparting timbre (sound quality) and they contribute to the identification of specific speech sounds (Rosen 1992) and instruments (Grey 1977; Howard and Angus 2001). Within the classes of speech sounds, obstruent stop consonants (e.g., /d/, /p/, /k/) have, by definition, sharper stimulus onsets than nasals and glides (e.g., /m/ and /y/, respectively) and produce more robust onset responses. Although fricatives and affricates have not been used to elicit cABR (to the best of our knowledge), based on stimulus characteristics we hypothesize the following continuum: earlier and larger onsets for obstruent stops, with affricates (e.g., /tʃ/ pronounced “ch”), fricatives (e.g., /z/) and sonorants (a class of sounds comprising nasals, glides and liquids (e.g., /r/ and /l/)) having increasingly smaller and later onsets. Similarly, when choosing a musical stimulus for eliciting cABRs, the specific attack properties of the instrument should be taken into consideration. For example, percussive instruments, like the piano, have fast, steep attacks and bowed string instruments have comparatively smoother attacks (Figure 2). Likewise, the mode of playing an instrument affects the attack (e.g., a plucked string has a shorter rise time than a bowed string). In addition, abrupt changes in the amplitude envelope of the sound also trigger onset-like transient responses. For example, Strait and colleagues recorded brainstem responses to the sound of a baby crying — this particular token included multiple amplitude bursts that produced a series of sharp transient responses (Strait et al., 2009) (Figure 2).

Sustained Features—Sounds containing continuous acoustic features such as sinusoidal tones, harmonically complex vowels, and musical notes elicit sustained brainstem responses reflecting synchronous population-wide neural phaselocking. Moushegian and colleagues (1973) were the first to describe the sustained response in human scalp-recorded brainstem potentials. Using tones (sinusoids ranging from 250 Hz – 2 kHz) they demonstrated that each frequency evokes a unique response, in which the pattern of neural discharge is time-locked to the temporal structure of the eliciting sound. For example, the brainstem response to a 250 Hz tone follows the periodicity of the tone such that response peaks occur at 4 ms intervals (period = 1/frequency; 4 ms = 1/250 Hz). For this reason, sustained brainstem responses are often called *frequency following responses*. Scalp-recorded FFRs can be recorded to frequencies as high as 1.5 kHz (Moushegian et al. 1973; Krishnan 2002; Aiken and Picton 2008), although phaselocking becomes weaker with increasing frequency (Greenberg 1980; Greenberg et al. 1987; Krishnan 2007), reflecting the low-pass nature of brainstem phaselocking. Thus, subcortical phaselocking provides a mechanism for representing low frequencies contributing to pitch and timbre (Greenberg 1980; Wile & Balaban 2007; Kraus et al. 2009; Bidelman and Krishnan 2009) (Figure 3), while a place code likely underlies the neural encoding of frequencies that are too high to elicit an FFR (Langner and Schreiner 1988; Chandrasekaran and Kraus 2009).

To obtain strong sustained responses, the cABR stimulus should have a low pitch with a fundamental frequency (F_0) in the range of 80-300 Hz. In speech, the F_0 ranges from ~80 Hz for a deep male voice to ~400 Hz for a child. While speech can contain spectral information up to 10 kHz, the spectral information necessary for distinguishing different consonants and vowels is largely below 3000 Hz. When selecting a speech-phoneme for cABR recording, keep in mind that some speech formants, including the second formant of many vowels, are above the range of brainstem phaselocking (Moushegian et al. 1973; Greenberg 1980) and may not be observable in the response (see however Johnson et al. 2008b). A wide range of frequencies is also encountered in music. For example, the lowest note on a standard 88-key piano occurs at 32.70 Hz and the highest at 4186 Hz (See Everest 2001 for the frequency ranges of various instruments). Because the F_0s of instruments are generally higher than speech, they often fall outside the limits of strong brainstem phaselocking. Therefore, it may be necessary to use an instrument within this cABR target range, such as the trombone.

Time Varying and Harmonically Complex Sounds—Real-life sounds, unlike sine waves, have non-stable F_0s and harmonics, and complex harmonic structures. For time-varying stimuli, such as diphthongs, consonant-vowel formant trajectories, musical glissandos, and tone language pitch contours, cABRs follow even small changes in frequency with inter-peak intervals systematically increasing or decreasing with changing frequency (Figures 11 and 13). For harmonically complex sounds, phaselocking is observed to the frequencies physically present in the stimulus, as well as the frequencies introduced via nonlinear processes within the auditory pathway. Examples include phaselocking to the amplitude envelope (Hall 1979; Aiken and Picton 2006; 2008) and distortion products (Kemp 2002; Abel and Kossl 2009; Lee et al. 2009; Elsisy and Krishnan 2008) (Figure 4).

Speech—During speech production, sound is produced when air leaving the lungs causes the vocal chords to vibrate. For speech, the F_0 is determined by the rate of this vibration. Because the vocal chords close more slowly than they open, the sound that is produced is not a sinusoid. It is instead a complex waveform, resembling a triangle or saw-tooth wave, containing harmonic energy at integer multiples of the F_0. This harmonically rich sound is then filtered (i.e., shaped) by the speech articulators (i.e., teeth, tongue, lips, etc.) to form different speech sounds. Different articulator configurations change the resonance properties of the vocal tract causing certain harmonics to be amplified and others to be attenuated. Formants, which correspond to peaks in the spectrum, arise from this filtering. Each speech sound can be uniquely identified by its characteristic formant pattern, with the first two or three formant being sufficient for identifying most speech sounds (Liberman 1954). The cABR, which synchronizes to the F_0 and harmonics of the speech waveform, contains greater energy for harmonics coinciding with formants (Krishnan 2002; Aiken and Picton 2008) because there is more energy in the signal at these frequencies. In other words, brainstem nuclei perform a type of “formant capture” (Young and Sachs 1979; Krishnan 2002), whereby harmonics adjacent to the formant regions are emphasized. Also, note that in the speech spectrum, that the F_0 has less energy than the speech-formants (Figure 5). However, because the opening and closing of the vocal folds produces a signal that is naturally amplitude-modulated, F_0 and other modulating frequencies are introduced into the neural system during nonlinear cochlear processing (Brugge et al. 1969; Regan and Regan 1988; Lins and Picton 1995; Aiken and Picton 2008).

Music—In contrast to speech, which is dominated by fast spectrotemporal transitions, music has more sustained temporal and spectral elements, slower transitions, and finer frequency spacing (Zatorre et al. 2002; Shannon 2005). In music, the mechanism of F_0 generation depends on the instrument. For example, the reed is the source of F_0 vibration for the oboe and clarinet, whereas the string is the source for the violin and guitar. In the same way that speech sounds are characterized by unique formant configurations, instruments also have characteristic harmonic structures that impart timbre. Specifically, the timbre of a musical sound is determined by the rise time of the attack (discussed above), the spectral flux (i.e., change in harmonics over time) and the spectral centroid (i.e., the distribution of the harmonics) (Grey 1977). The clarinet, for example, has a lower harmonic structure dominated by odd harmonics (the even harmonics have been attenuated). The flute, saxophone, trombone and tuba, which are all characterized by strong odd and even harmonics, can be differentiated by the distribution of the harmonics (e.g., the energy of the tuba is concentrated in the lower harmonics).

As can be seen in Figure 5, the harmonic structure of musical sounds is partially preserved in the response. Generally speaking, FFRs are more robust when there is less spectral flux (i.e., brass and woodwind families (Grey 1977)). The timbre of a musical instrument also depends on how quickly the sound decays (e.g., a piano has both a fast onset and quick decay while an electric piano has a slower onset and decay). For the purposes of eliciting an FFR, sounds with longer decays elicit responses that are more sustained (Figure 2). For more information on the rather complex and multifaceted topic of musical timbre refer to Fletcher and Rossing 1991 and Howard and Angus 2001.

Stimulus Duration—Within the speech-ABR literature, the length of vowel stimuli has varied between 60 ms to 2 seconds (Krishnan 2002; Dajani et al. 2005; Aiken and Picton 2006; 2008) and for CVs from 40 to 500 ms (Musacchia et al. 2007; Banai et al. 2009). In our experiments using musical stimuli, the duration has ranged from 170 ms for a 2-note musical interval (Lee et al., 2009; Figure 4) to 1.1 seconds for a five-note musical melody (Skoe et al., 2009; Figure 2)

Because of the sheer number of stimulus presentations required to obtain a robust response, there is an obvious tradeoff between stimulus duration and length of the recording session(Section III). For example, to record 6000 trials to a synthesized 40 ms /dɑ/ takes approximately 9 minutes, assuming an interstimulus interval (ISI) of 50 ms. Yet natural sounds generally occur on the order of seconds and not fractions of seconds, which necessarily requires much longer recording sessions. In our experience, the only factor limiting the stimulus duration is the feasibility of having a subject remain still for a long time. Thus, stimulus duration may need to be restricted in order to present the desired number of stimuli in a reasonable amount of time. For speech syllables, one tactic is to record ABRs to a stimulus containing the consonant and consonant-vowel transition (CV) without a steady-state vowel (Russo et al. 2004; Johnson et al. 2007; Johnson et al. 2008a; Banai et al. 2009; Dhar et al. 2009; Hornickel et al. 2009a) (cf. stimulus in Figure 10 vs Figure 1). Because each consonant-vowel pair has a unique formant transition, the steady-state vowel can be removed with little impact on the percept. In fact, musical timbre and vowel identity can be accurately determined from 1-4 cycles of the F_0 (Gray 1942; Robinson 1995; Robinson and Patterson 1995) but pitch identification requires at least four cycles (Robinson 1995; Robinson and Patterson 1995). Stimulus duration greatly affects pitch because lower frequencies have longer periods than higher frequencies (e.g., a 20 ms stimulus can have no meaningful frequency representation under 50 Hz).

Stimulus Creation

With modern computers, recording natural sounds is relatively simple. The process requires a sound-attenuated chamber, a microphone, a high resolution sound card and software for recording (e.g., Adobe Audition (Adobe Systems, Inc., San Jose, CA), Pratt (Boersma and Weenink 2009)) (see Aiken and Picton, 2008 and Wong et al., 2007 (supplement) for more details). To ensure that a viable token can be extracted, multiple recordings, and when possible, multiple speakers/instruments should be used. Both natural and synthetic sounds should be created with a high digitization rate (>20 kHz). However, because some stimulus presentation systems require a specific sample rate, the recordings may need to be resampled. Likewise, when comparing the stimulus and the response via cross-correlation (Section IV) the two signals must have the same sampling rate. For these reasons, it is best to sample at a high rate so that upsampling is not necessary.

Speech—Although natural speech and music tokens are ideal in the sense that they are an accurate representation of real-world sounds, they are inherently more complex, variable, and aperiodic. Consequently, with natural tokens it is difficult to study how specific physical characteristics are represented at a subcortical level. Having precise control over the stimulus parameters is especially important when multiple stimuli are compared across a single dimension. For example, /bɑ/, /dɑ/ and /gɑ/ can be distinguished based on their differing second formant trajectories (F2) (Liberman 1954); Figure 9). However, natural utterances of /bɑ/, /dɑ/ and /gɑ/ vary on more parameters than simply F2 (discussed in Johnson et al. 2008b). In these cases investigators rely on speech synthesizers such as Klatt (Klatt 1976) to create stimuli with precisely defined time-varying and sustained features[1].

In the case of the F_0, programs such as STRAIGHT (Kawahara 2009) and Praat (Boersma and Weenink 2009) can be used to remove aperiodicities, raise or lower the F_0, or apply a particular time-varying pitch contour (Wong et al. 2007; Russo et al. 2008b). To generate stimuli with pitch contours, hybrid stimuli can be made by manipulating the F_0 of a natural speech token or combining two natural speech tokens using the PSOLA method (Moulines and Charpentier 1990) in a program like Pratt (Boersma and Weenink 2009). See Wong et al., 2007 (supplement) and Russo et al., 2008 for more details.

Music—Due to the increased prevalence of computer-made music, a large number of tools are currently available to generate music stimuli. The choice of the right tool depends on the desired trade-off between acoustic control and sound naturalness. Strict acoustic control of the stimuli can be best achieved through additive synthesis in programming environments like MATLAB (The Mathworks, Natick, MA). Acoustic samples of real instruments, which can be found in some music software packages (e.g., Garritan Personnal Orchestra in Finale software, MakeMusic, Inc), are another source for music stimuli. An intermediate solution is to use synthesizers, many of which are available as plugins for music software like Cubase Studio (Steinburg Media Technologies)[2].

Other sounds—It can be difficult to construct synthetic sounds with strong affective quality. Thus, natural recordings such as those available from the Center for the Study of Emotions and Attention (University of Florida, Gainesville, FL) can be used to study paralinguistic aspects of sounds (Strait et al. 2009). Similarly, for environmental sounds, we suggest selecting a stimulus from a corpus of natural sounds (e.g., Series 6000 General

---

[1] Klatt, which can function as both a cascade and parallel synthesizer (Holmes 2001), facilitates the manipulation of dozens of features including duration, output sampling rate, the amplitude of frication, the number of formants, and the frequency and bandwidth of each formant. In our experience, the process of creating a synthetic speech sound with the desired percept requires patience and a lot of trial and error. In addition, although you do have control over many parameters, the actual output may deviate from the input because of the complex interaction among parameters. To confirm that the stimulus meets the desired specifications, the synthetic sound should be acoustically analyzed in a program like Pratt (Boersma and Weenink 2009) or Adobe Audition (Adobe Systems, Inc., San Jose, CA).

[2] Unfortunately, these synthesizers are often black boxes. Although they offer control over certain acoustic features, the sound they provide is not as good as samples of real instruments. For all methods of stimulus generation, the acoustic properties of the stimulus should be checked with a sound analyzer before proceeding with the experiment. Acoustic analyzer software dedicated to music sounds includes the MIR toolbox for MATLAB (http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox), the IPEM toolbox for MATLAB (http://www.ipem.ugent.be/Toolbox/), and PsySound3 (http://psysound.wikidot.com)

Sound Effect Library, a royalty-free CD of environmental sounds (Sound Ideas, Ontario, Canada)).

III. STIMULUS PRESENTATION

This section covers topics relating to stimulus presentation including stimulus intensity; monaural and binaural stimulation; left and right ear stimulation, stimulus polarity, stimulation rate; transducers (i.e., earphones, loudspeakers); jitter in the stimulus presentation, and multiple stimulus conditions. A summary is provided in Table 2.

Intensity

Speech, music and other complex sounds are typically delivered supra threshold within the “conversational” range of 60 to 85 dB SPL. Like the familiar click-ABR, cABRs are also intensity-dependent. This necessitates that the intensity be stable across subjects and recording sessions. Before each test session, the output intensity should be calibrated using a sound level meter with a coupler that enables the output to be measured directly from the insert earphones (see below).

The effects of increasing intensity have been examined in the cABR literature (Krishnan 2002; Akhoun et al. 2008a). Using a /ba/ syllable, Akhoun and colleagues explored how the timing of the speech-evoked onset response and speech-evoked FFR (elicited by the same stimulus) varied as function of intensity (0 to 60 dB SL, in 10 dB increments). Consistent with the click-ABR and tone-evoked FFR literature, both response components showed systematic latency shifts with increasing intensity. However, the FFR peaks showed a steeper latency-intensity function than the onset response, suggesting that the onset response and speech-evoked FFR reflect distinct neural processes. Using a similar design, Krishnan recorded cABRs to steady-state vowels between 55 dB nHL and 85 dB nHL (in 10 dB increments) and found that the harmonics in the formant range were clearly represented for each intensity. Although the amplitudes of the individual harmonics tended to increase with increasing intensity, the trajectory was not identical for all harmonics, nor was the increase always linear. Taken together, this work suggests appears that different components of the cABR are distinctively impacted by intensity level.

Monaural and Binaural Stimulation

It is well established that when a sound is heard with both ears that it is perceived to be louder than when the same sound is presented at the same intensity to just one ear (theoretically *binaural loudness summation* is estimated to be 6 dB). Because the auditory brainstem plays an integral role in binaural processing (reviewed in Moore 1991), binaural interaction effects have been widely studied in the click-ABR and tone-FFR literature (Dobie and Berlin 1979; Ballachanda et al. 1994; Krishnan and McDaniel 1998; Ballachanda and Moushegian 2000). Although similar parametric experiments have not been conducted for more complex stimuli, the same principles are assumed to apply. For practical reasons, binaural stimulation is preferable when testing adults, not only because it leads to larger and more robust responses, but also because it is more realistic in that we usually listen with both ears. However, monaural stimulation is used for individuals with asymmetric hearing thresholds, children and other populations who have difficulty sitting still during testing, or when the subject must to attend to another sound.

Left and Right Ear Stimulation

Left and right-ear stimulation produce similar, yet not identical ABRs responses, (Akhoun et al. 2008b); In fact, the well established right-ear advantage for speech is evident for discrete components of the speech-ABR, indicating that separate norms should be developed for the left and right ears (Hornickel et al. 2009a). For a review of the click-ABR and tone-FFR literature see Hornickel et al. 2009a.

Stimulus Polarity

Periodic sound waves traveling through air consist of alternating regions of compression (i.e., condensation) and decompression (i.e., rarefaction) of air molecules. In a time-amplitude plot of a sound wave, condensation and rarefaction manifest themselves as positive or negative deflections (respectively) from the baseline. Because clicks consist of a single positive or negative deflection, they are defined as either having condensation or rarefaction polarity. However, because periodic sounds oscillate between condensation and rarefaction states, the same terminology is not used. In Figures 6 and 8, we have adopted 'A' and 'B' to refer to the two different polarities. To convert a stimulus from one polarity to another, the waveform is shifted by 180° (i.e., multiplied by −1).

When collecting cABRs, two different approaches can be followed: (1) record the response to a single stimulus polarity (Krishnan 2007; Aiken and Picton 2008) or (2) record responses to both polarities and either add (Russo et al. 2004; Akhoun et al. 2008a) or subtract responses (Greenberg 1980; Greenberg et al. 1987; Wile and Balaban 2007; Krishnan 2002) to the two stimulus polarities. The process of adding will accentuate the lower-frequency components of the response including phaselocking to the amplitude envelope and minimize stimulus artifact and the cochlear microphonic (CM) (see Section V for a further discussion of artifacts). Subtracting will bias the higher-frequency components by maximizing the spectral response, although this process can also maximize artifact contamination. It should be noted that while we use the addition method for many of our published analyses, our results have been internally replicated with single-polarity stimuli. For more advanced considerations of stimulus polarity see Figure 6, footnote 3, Don et al. 1996, and Aiken and Picton 2008[3].

Presentation Rate

Presentation rate depends on the length of the stimulus and the ISI (defined as the period of silence between the offset of one stimulus and the onset of the next). A second way to express the interval of presentation is stimulus onset asynchrony (SOA), which is measured from the onset of one stimulus to the onset of the next. The two measures are essentially the same for click stimulation, as a click has virtually no duration, but ISI and SOA are very different for cABRs. While there is no specific formula for determining the optimal ISI for a

---

[3] Phaselocking to spectral energy in the stimulus follows the phase of the stimulus; thus, it inverts when the polarity is inverted. In Figure 6, this inversion is evident in the regions between the large peaks of the A and B responses. Because of this inversion, adding A and B will theoretically cancel out the spectral response. Subtraction, in contrast, enhances the spectral response (including formant frequencies and temporal fine structure) and attenuates the envelope response. Phaselocking to the amplitude envelope is independent of phase because the energy of the envelope is not present in the speech signal but is introduced into the auditory pathway during nonlinear cochlear processing. Thus, phaselocking to the amplitude envelope does not invert between A and B and it is, therefore, maximized when the responses to A and B are added. Although the two polarities of an auditory stimulus do not sound different, the two polarities do not elicit identical responses, especially in the case where the stimulus waveform is asymmetric (Greenberg 1980). An advantage to using the addition and subtraction methods is that they represent the average of two polarities, thus relieving the necessity to choose which polarity to use. Finally, one criticism of the addition method is that while it will minimize artifacts, it will also create a response at twice the frequency of the actual response (Aiken and Picton 2008). While this is the case for simple sinusoids, the polarity effects are more complicated for time-varying spectrally complex stimuli. For a complex stimulus, involving multiple frequency components, a polarity reversal will theoretically affect the response latency to a given frequency. This is because the 180° inversion shifts the phase of the frequency component by one half-period, resulting in different single unit responses for each polarity. However, because, of the complex interaction of temporally-dynamic frequency components within a population-wide response, the doubling of frequency for any particular frequency (such as F_0) is not evident in the cABR, and frequency-dependent latency shifts may be obscured or washed out (Don et al., 1996). Consequently, while the responses to the two polarities are not strictly identical, the differences are not always as extensive as might be predicted. Moreover, although the effects of stimulus polarity have been extensively explored to simple stimuli, the observations based on click stimuli do not generalize to low-frequency tone bursts (Don et al., 1996), nor to the polarity effects for either simple stimulus generalize to complex sounds. For a more in depth description and discussion of these cochlear processes and the effect of stimulus polarity on the click-evoked response see Hall (2006) and Starr and Don (1988).

given stimulus length, in surveying the cABR literature, the ISI has varied from ~30% of the stimulus length to more than double the length.

When choosing an ISI the following considerations should be made: First, changing the ISI can alter the perception of a complex sound. Second, if the ISI is not sufficiently long, the response to one stimulus may not fully conclude before the next stimulus is presented. Thus, the ISI and the duration of the averaging window should be long enough to allow for the response to return to baseline. The ISI should also allow for an ample sample of the baseline (i.e., non-stimulus-related) activity so that signal-to-noise ratios (SNRs) can be evaluated (Section V). Third, latencies and amplitudes, particularly of onset responses, are affected by rate of presentation (Hall 2006). Although rate effects have not been extensively explored in FFRs to tones or complex stimuli, in adults speech-evoked FFR latencies seem to be less susceptible to rate changes than onset responses in adults (Parthasarathy and Moushegian 1993; Krizman et al. in press). Fourth, to avoid contamination from the AC line frequency (60 Hz in North America, 50 Hz elsewhere), a presentation rate should be chosen such that the rate divided by the line frequency is not an integer (e.g., for both 50 and 60 Hz line noise, 10/sec is a bad choice, but 10.3 or 11 is okay). Alternatively, a variable ISI should be used. Fifth, when conducting simultaneous cABR-cortical EP recordings (Section IV) longer ISIs may be required to obtain robust cortical auditory EPs.

An alternative approach is to record cABRs in several blocks of continuous stimulation (i.e., no silence between stimuli) using the same procedure used to record auditory steady-state responses (Dajani et al. 2005; Aiken and Picton 2006; 2008). This technique maximizes spectral resolution at the expense of temporal resolution.

Transducer

Because circumaural headphones increase the chances for stimulus artifact contamination (Section IV), we strongly advise against using them and instead recommend electromagnetically shielded insert earphones (e.g., E•A•RTONE3A (Aearo Technologies, Minneapolis, MN), ER-3a (Etymotic Research, Inc., Elk Grove Village, IL), Bio-logic insert-earphones (Bio-logic Systems Corp., Mundelein, IL)). When testing persons with hearing aids or other populations not suited for inserts (e.g., cochlear implant wearers), loudspeakers can be used to deliver the stimulus. However, sound field delivery, causes the latency of the response to be more variable because the sound intensity changes subtly with head movements. To minimize head movements, we have the subject focus on a movie or another visual image positioned directly in front of him/her. Also, because the latency is dependent on the distance between the loudspeakers and the subject, we carefully measure and mark the location of the chair and speakers, and position the left and right speakers equidistantly.

Detecting Stimulus Jitter

One of the defining characteristics of the auditory brainstem response is that it reflects extremely fast neural activity synchronized across populations of neurons, with minute disruptions in neural precision being indicative of brainstem pathologies (Hood 1998; Hall 2006). For this reason, the delivery and recording units must be precisely time-locked to each other[4]. Even a small amount of jitter in this synchronization can ruin an ABR recording. If the timing of the stimulus does not always occur at the same time with respect to the triggering of the recording system, the response is canceled, or at the very least become

---

[4] Depending on the recording system, the software used for stimulus delivery is either integrated into the collection system or installed on a separate computer from the collection computer. In the latter case, in order to synchronize the presentation of stimulus and the recording of the response, the delivery computer sends a digital trigger to the recording computer every time a sound is played.

distorted, when trials are averaged. Thus, when a new recording system is acquired, it is important to confirm that the delivery system is properly calibrated (see below) to ensure that there is not an unexpected stimulus delay or jitter. A system that has been optimized for collecting cortical responses should also undergo testing before it can be *cleared* for brainstem testing. Because of timing and duration differences between brainstem and cortical responses, jitter may only be evident when recording brainstem responses.

To determine whether the stimulus presentation is jittered, couple the output of the delivery system into the electrode box, as if recording cABRs from a subject. Next, play a click stimulus and record the output into the recording system in continuous (non-averaging) mode. Adjust the output intensity, if the waveform is clipped in the recording. It is important to record a sizeable number of sweeps (100+) to ensure that the jitter does not creep in over time. After the recording is complete, check that each click occurs at the same time relative to the trigger across the recording. For a properly functioning system, the deviation should not exceed 0.1 ms. This is also an opportunity to determine whether the stimulus is actually simultaneous with the trigger or whether there is a delay that needs to be taken into account when processing and analyzing cABRs[5].

Multiple Stimulus Conditions

When an experiment includes multiple stimulus conditions, a block or interleaved paradigm can be used. In a block paradigm, each condition is presented separately (i.e., block 1: P P P; block 2: Q Q Q) (Johnson et al. 2008b) and in an interleaved paradigm the stimulus conditions are intermixed (i.e., P Q P Q P Q or P Q P P Q P, etc.) (Wong et al. 2007). In the blocked design, state (i.e., alertness) or expectancy effects may confound comparisons across stimulus conditions. However, if the delivery system is not designed to play multiple stimulus tokens, interleaving stimulus conditions may not be possible. While the presentation software might limit the number of stimuli that can be interleaved, there does not seem to be a corresponding neurophysiologic limit (e.g., in one experiment we interleave eight different stimulus conditions, two polarities for each, for a total of 16 different sounds). In the case where multiple stimuli are to be directly compared, it may be desirable to normalize the duration and amplitude across the stimulus set. This can be carried out in programs such as Level 16 (Tice and Carrell 1998) and Praat (Boersma and Weenink 2009).

Recent work suggests that block and interleaved designs may invoke different online context-dependent subcortical encoding mechanisms. Chandrasekaran et al. 2009 compared the response to [da] collected in a block condition and the response to the same stimulus when it was presented with a pseudo-random probability within a mix seven other speech stimuli. The response to the interleaved condition was found to have smaller spectral amplitudes compared to the blocked condition, which the authors interpret to an indication of weaker stimulus ‘tagging’ when the stimulus is presented less frequently.

IV. cABR COLLECTION

Issues relating to electrodes, filtering, sampling rate, signal averaging, simultaneous ABR-cortical EP recording, artifact reduction, and recording conditions are reviewed below and summarized in Table 3.

---

[5] Please note that this testing procedure may not be feasible for every recording system. Because the source of the jitter can come from one or a combination of several sources and because each system will have its own peculiarities, we advise contacting the manufacturer before adjusting any settings.

Electrodes and Electrode Montage

For cABRs, a vertical one-channel montage is the most common configuration. This configuration requires only three electrodes corresponding to the active (non-inverting), reference (inverting) and ground electrodes. In our lab, the preferred electrode placements are Cz (active), ipsilateral earlobe (reference), and forehead or contralateral earlobe (ground). We prefer to use the earlobe rather than the mastoid because it is a non-cephalic site that causes fewer artifacts from bone vibration (Hall 2006). For researchers who intend to record subcortical and cortical potentials simultaneously (see below), or who wish to collect them within the same session, cABRs can be recorded with an electrode cap.

Filters

Filtering is used to isolate subcortical activity from cortical potentials and to increase the SNR of the response. For cABRs, the bandpass filters match the range of settings used for click-ABRs and typically fall in the range of 100-3000 Hz. This frequency range has been found to maximize the detection of the high-frequency transient peaks, such as the click-evoked peaks I-V, which have very sharp slopes. For stimuli containing frequencies below 100 Hz (or which produce distortion products below 100; cf. Lee et al., 2009), the high pass cutoff should be lowered to ensure that none of these lower-frequency features are lost. Another approach is to record with more open filters such as 30-3000 Hz (Galbraith et al. 1995; Galbraith et al. 2004)[6].

Sampling Rate

Sampling rate (Fs), also referred to as the digitization rate, determines how many times per second the analog neural signal is digitally sampled by the recording system. In cases where only low-frequency components of the response are of interest, a low Fs (1000-2000 Hz) may be appropriate (Dajani et al. 2005; Aiken and Picton 2006). However, many researchers opt to over-sample cABR recordings (rates range from 7 kHz to 50 kHz) by sampling well above the Nyquist frequency (i.e., twice the highest frequency in the stimulus) (Krishnan et al. 2005; Musacchia et al. 2007; Akhoun et al. 2008a; Banai et al. 2009). In addition to reducing sample errors, a higher Fs, by definition, increases the temporal precision of the recording and allow for finer differentiation of response peaks. Because cABR disruptions and enhancements occur on the order of tenths of milliseconds, fine-grained temporal precision is essential. Although a higher Fs is desirable, the choice may be limited by the particular recording system. For example, some recording systems utilize a fixed number of sample points. In this case, the Fs is dependent on the duration of the recording window (Fs = sample points/duration).

Signal Averaging

Number of Sweeps—An age-old question in the EP literature is how many sweeps must be averaged to obtain a robust and reliable response. It is well established that for higher intensity stimuli roughly 1000-2000 sweeps are needed to collect click-ABRs and FFRs to tones (Hood 1998; Krishnan 2007; Thornton 2007). For cABRs a comparable but sometimes greater

---

[6] A more advanced recording technique uses a two-channel (or more) montage either to simultaneously record horizontal and vertical montages (Galbraith 1994) or to record multiple recording parameters from a single site (e.g., different filter bandwidths; see below). We commonly record in continuous (i.e., non-averaged) mode using broad filters (e.g. 0.1 to 3000 Hz) and then re-filter offline using more narrowly defined digital bandpasses. Analog filters, generally employed at the time of data collection, are more likely to introduce distortions (e.g., phase distortions) in the response especially when cutoff frequencies are near the frequency range of the response. In addition, filter choice may be restricted by the recording system to only a handful of preset values. Thus for recording systems that include the option to record with open filters, subsequent digital filtering is preferred. With offline digital filtering you have the capability to set cutoff values more precisely, and to optimize the filter settings for a particular stimulus by systematically adjusting the bandpass to be more restrictive or more encompassing. However, open filters, because of their susceptibility to cortical activity and noise contamination, can be unsatisfactory for monitoring the quality of the response during acquisition. A two-channel solution, with a second channel using a more restricted bandpass (e.g., 100-2000 Hz) for online viewing, solves this problem.

number of sweeps are obtained (1000 - 6000). However, if analyses are only carried out in the frequency domain, then spectral maxima may be detected (i.e., statistically above the noise floor) with fewer sweeps (Dajani et al. 2005; Aiken and Picton 2006; 2008).

Our lab takes a conservative approach by collecting more stimulus trials than less, typically ~2000-3000 per polarity (i.e., 4000 to 6000 total sweeps). There are several reasons for this strategy. First, this allows for the creation of sub-averages that can be used to determine response repeatability and/or track how the response evolves over time. Second, we are often interested in subtle response characteristics and small group differences that may not be apparent until additional sweeps are collected and/or repeatability is confirmed. A general principle of EP signal averaging is that the SNR is proportional to the square root of the number of sweeps (Hood 1998; Hall 2006; Thornton 2007). Thus, the overall SNR increases quickly at first and then begin to plateau with more sweeps. However, an individual component of the cABR (for example, a specific peak in the time domain or a spectral peak that is near the phaselocking limits of the brainstem) may show its own SNR progression with different response components requiring greater or fewer sweeps. While it may not be possible to determine the “optimal” number of sweeps for a given stimulus and population before the start of an experiment, the optimal range can be deduced *a posteriori* using an iterative off-line averaging technique based on a handful of subjects from whom a large number of sweeps have been collected (i.e., compare subaverages of 1000 sweeps, 1500 sweeps, 2000 sweeps, … 6000 sweeps, etc.). In the future, we envision that better characterization of the cABR will enable the number of sweeps to be reduced while still maintaining spectral and temporal precision.

Averaging Window—In the time-domain, the averaging window should be long enough to include a pre-stimulus baseline period, the response period, and a post-stimulus period. The length of the post-stimulus time window needs to account for the stimulus transmission delay and neural conduction time. A post-stimulus period between 10 ms and 50 ms is recommended to ensure that the response returns to baseline. The pre-stimulus baseline reflects the ambient EEG before the response, thereby assisting in the interpretation of the response. For example, when identifying prominent peaks in the response waveform, peak amplitudes are compared to the amplitude of the pre-stimulus period. For a given peak, if the amplitude does not exceed the baseline amplitude, it is not considered a valid (i.e., reliable) peak. The baseline period can also be used to determine the SNR (in the time and frequency domains) (Section V). For running window analyses (Section V), it is helpful to have a pre-stimulus period that is long enough to include one full analysis window. Because we typically perform running window analyses on 40 ms bins, we use a pre-stimulus window of at least 40 ms.

Simultaneous cABR - Cortical Recordings

Although we have had success simultaneously recording cABR and cortical responses (Musacchia et al. 2008) there are a number of practical limitations to this practice that arise from cABRs and cortical EPs having different optimal recording parameters. First, cABRs require a much higher Fs than cortical responses (often a tenfold or more difference) (Burkard et al. 2007). Second, because cortical responses are optimally obtained using slower stimulation rates than ABR (Burkard et al. 2007), the presentation rate must be slow for simultaneous recordings. Yet because cABRs are much smaller in amplitude (typically less than 1 microvolt), many more trials must be collected for a robust cABR than for a cortical response, often leading to long recording sessions. These factors aggregate to create extremely large files, especially when high-density electrode caps are used, leading to concerns about both computer processing power and data storage. For these reasons, we usually opt to collect brainstem and cortical-evoked responses in separate recording sessions, optimizing recording lengths, numbers of channels, and sampling rates for each.

Avoiding, Detecting and Eliminating Artifact

There are four types of artifact that can distort the auditory brainstem response recordings: external (i.e., non-biological) electrical noise, myogenic (muscular) artifact, cochlear microphonic (CM), and stimulus artifact. While artifacts can be minimized, it is best to remove the contamination at its source.

Electrical—When combating electrical artifact such as line noise (60 or 50 Hz), the best tactic is to record within an electrically shielded booth and remove all electrical sources from the booth including televisions, CRT and LCD computer monitors. Light dimmers are another serious source of noise. If the experimenter wishes to play a movie or another visual stimulus during the experiment, two different approaches can be taken. The cheaper option is to use a portable battery-powered DVD player that is placed on a table in front of the subject. The second option, and the one we employ most often, is to use an LCD projector located outside the booth that projects the visual stimulus through a booth window onto a screen inside the booth.

Another type of electrical artifact comes from the electrical trigger pulse that is used to synchronize stimulus presentation and response averaging. This artifact appears at time zero. If a long trigger is used a second artifact may appear when the trigger turns off. If the duration of the trigger pulse can be manually set within the stimulus presentation software, this type of artifact can be reduced by either shortening the trigger pulse so it occurs before the onset of the response (e.g., < 5 ms), or by elongating it to be longer than the stimulus itself.

Myogenic—Given that cABRs are typically recorded with wide bandpass filters (see above), myogenic artifacts are often not filtered out. Because myogenic artifacts produce potentials that can be many times larger than the brainstem response, trials for which the amplitude exceeds a specific threshold should be excluded from the final average (either on-line or off-line). In the cABR literature this threshold ranges from +/− 20 to +/− 75 μV (Galbraith et al. 1995; Akhoun et al. 2008b). Although this technique removes large artifacts, it does not completely expunge all myogenic contamination from the recording. For this reason, it is important to keep the subject relaxed and still during the recording session.

Cochlear Microphonic—The CM is a potential generated by the cochlear hair cells which, like the FFR, mimics the temporal waveform of the acoustic stimulus. Because of its similarity to the neural response, care must be taken to prevent or remove the CM from the recordings. The CM can be distinguished from the brainstem response in a number of ways. Unlike the cABR, which occurs at approximately 6-10 ms post-stimulus onset, the onset of the CM is nearly coincident with the stimulus. The cochlear microphonic and cABR are also differentially affected by rate, intensity and noise. For example, while cABRs *breakdown* with increases in presentation rate and simultaneous masking intensity, the CM remains unaffected. Furthermore, in contrast to cABR amplitude, which plateaus at supra threshold levels, the size of the CM usually increases linearly with moderate increases in intensity. For more information on the CM and how it can be isolated from the ABR, we refer the reader to Starr and Don 1988, Aiken and Picton 2008, and Chandrasekaran and Kraus 2009.

Stimulus artifact—Given that the cABR occurs within a matter of milliseconds after stimulation, and the fact that the cABR closely mimics the stimulating waveform, stimulus artifact is a major concern. Fortunately, this type of artifact is easy to detect (Figure 7) and with the right recording techniques can be minimized (Figure 8).

In most modern EP collection systems, the stimulus waveform is sent as an electrical signal to a transducer where it is converted to an acoustic signal. If the transducer is not properly shielded, the electrical signal can "leak" and get picked up by the electrodes and recorded by the EP system along with the response (Figure 8). In addition to using electromagnetically shielded earphones (Akhoun et al. 2008a; Akhoun et al. 2008b), it is also good practice to double-check that the electrode leads and transducer cables are not touching and to also position the electrodes and transducer as far apart as possible. This can be achieved with insert earphones that use a plastic tube to separate the transducer and foam earplug. Using longer tubes and/or positioning the transducer outside the test booth can further minimize stimulus artifact. For an in depth discussion of stimulus artifacts in cABR recordings, refer to Aiken and Picton 2008; Akhoun et al. 2008a; Akhoun et al. 2008b.

Minimizing the Effects of CM and Stimulus Artifact—Given that both artifacts follow the phase of the stimulus exactly, stimulus artifact and CM can be minimized from the response by adding responses to alternating stimulus polarities (Figure 8, Section III).

Active vs. Passive Test Conditions

Because ABRs are not greatly affected by sleep, click-ABRs are often collected while the patient is asleep or sedated. Similarly, to reduce muscular artifact, many cABR researchers allow or even encourage their subjects to fall asleep on a cot or to recline comfortably in a chair (Dajani et al. 2005; Krishnan et al. 2005; Aiken and Picton 2008). In order to rule out differences in state as a potential confound, our subjects stay awake during testing. To promote relaxation and stillness, subjects watch a movie or read a book. For monaural stimulation, the movie soundtrack is played at a low level (~40 dB SPL) so that it can be heard in the non-test ear without masking the auditory stimulation. Subtitles are displayed for binaural recordings.

cABRs can also be recorded under active test conditions in which the subject performs a task (e.g., detecting/counting oddball stimulus tokens) (Musacchia et al. 2007). For example, using an audiovisual paradigm, Musacchia and colleagues (2007) revealed that active multi-sensory integration can shape how subcortical sensory processes respond to speech and music (Musacchia et al. 2006; Musacchia et al. 2007). While there is some disagreement in the literature as to whether attention modulates the click-ABR (Picton and Hillyard 1974; reviewed in Rinne et al. 2008), Galbraith and colleagues have demonstrated that attentional state can govern the FFR to tones and speech (Galbraith and Arroyo 1993; Galbraith and Kane 1993; Galbraith and Doan 1995; Galbraith et al. 1998; Galbraith et al. 2003; Hoormann et al. 2004). This is consistent with recent functional MRI work showing that selective auditory attention tasks can modulate the activation of subcortical structures (Rinne et al., 2008).

Notably, to study the dynamic nature of auditory processing, the subject need not be performing an active task during data collection. A growing body of research supports the use of passive recording conditions to study how brainstem function is fine-tuned by experience. Although the subject is not actively processing the sounds evoking the response, cABRs tap into how previous active engagement with sound has shaped brainstem processes over the course of lifelong or short-term auditory experiences. This refinement of the sensory system likely results from an interplay between subcortical structures and high-order cognitive processes via the corticofugal system (Banai et al. 2005; Krishnan et al. 2005; Russo et al. 2005; Musacchia et al. 2007; Wong et al. 2007; Song et al. 2008; Banai et al. 2009; Lee et al. 2009; Strait et al. 2009; reviewed in Banai and Kraus 2008; Kraus et al. 2009; Tzounopoulos and Kraus 2009).

V. DATA ANALYSIS

To analyze the transient and sustained aspects of cABRs, our lab employs a battery of measures to appraise the timing and magnitude of neural synchrony, as well as the strength and precision of phaselocking. Because cABRs are rich in temporal and spectral information, the use of multiple measures allows us to (1) dissect individual components of the response and how they reflect distinct aspects of processing and (2) describe brainstem encoding in a holistic manner. Due to transparency between the temporal and spectral features of the stimulus and the response, our analyses are largely stimulus-driven. That is to say, we base our analyses and interpretation on the acoustic make-up of the stimulus. Because of this stimulus-response fidelity, commonly employed digital signal processing tools such as cross-correlation and Fourier analysis can be used to analyze both the stimulus and response. Each of these techniques comes in many variants and each belongs to a large family of analysis methods; however, we generally use each in its most basic form. For more information on digital signal processing, we refer the reader to van Drongelen 2007; Wallisch 2009 and Porat 1997.

This section includes an overview and illustration of the most common signal processing techniques used to evaluate cABRs, namely peak latency and amplitude measurements, root-mean-square (RMS) amplitude, cross-correlation and Fourier analysis. A summary is provided in Table 4. The analyses described below are performed offline on the averaged time-domain response or sub-averages. Although some of these measurements can be made directly by the EP collection unit, others require the use of computational software packages such as MATLAB (The Mathworks, Natick, MA). For researchers and clinicians who are not in a position to code their own algorithms, we have developed an open source MATLAB-based toolbox (The Brainstem Toolbox) that is available for free upon request under the GNU General Public License (contact eeskoe@northwestern.edu for more information).

Analyzing Transient Responses

Peak Latency and Amplitude—To characterize the transient features of the response, individual response peaks relating to major acoustic landmarks in the stimulus are identified (Figure 1). For each peak, latency (time relative to stimulus onset) and amplitude measurements are obtained. Inter-peak measurements are also calculated; these include inter-peak amplitude, duration, slope and area (Russo et al. 2004). In general, transient peaks occur within 6-10 ms after the corresponding stimulus landmark. Automated peak-picking algorithms can be used to objectively identify maxima (peaks) or minima (troughs) known to occur within a given latency range. To be considered a reliable peak, the absolute amplitude must be larger than the baseline activity recorded prior to the onset of the stimulus. Confidence in selection of ambiguous peaks is aided by referring to subaverages. Once the peaks have been identified, they are visually confirmed or rejected by multiple raters who are blind to subject group and/or stimulus contrast(s). When the raters disagree, the selection is determined by the most experienced rater. Bear in mind, however, that agreement among raters may reflect common training in peak identification. Consequently, if peaks cannot be identified by objective methods, an external rater should also be consulted whenever feasible.

A number of techniques have been developed to aid in the identification of difficult to identify/low-amplitude peaks. These include wavelet denoising (Quian Quiroga et al. 2001; Russo et al. 2004) and high-pass filtering (Johnson et al. 2008b; Hornickel et al. 2009b). When determining which peaks to pick in the cABR a novel stimulus, start by generating a grand average response of all subjects, and then compare the grand average to the stimulus waveform to determine where the two waveforms match up. Once this has been performed, the individual waveforms should be reviewed to determine which peaks have high replicability across subjects (i.e., < 1 ms deviation across subjects/groups).

Differences in Latency Over Time—When multiple stimulus conditions are compared, a more advanced technique involves calculating the latency changes between/among conditions as a function of time. For example, recent work from our lab (Johnson et al. 2008b; Hornickel et al. 2009b) showed that the formant frequency differences differentiating the stop consonants /bɑ/, /dɑ/ and /gɑ/ are represented by systematic and progressive latency differences in the cABR with /gɑ/ responses occurring first, followed by /dɑ/ then /bɑ/ (i.e., higher stimulus frequencies yield earlier response latencies). These latency differences can be visualized using a latency-tracking plot (Figure 9).

Analyzing Sustained Responses

Static and Sliding Window Analyses—The response to periodic features (e.g., steady-state vowels, formant transitions, pitch contours, steady-state musical notes, glissandos) can be analyzed using RMS, cross-correlation and Fourier analysis. Each of the analysis techniques described below can be used to perform “static” window or “sliding” window (also called running window) analyses. A single region of the time-amplitude waveform is evaluated in a static window analysis. For sliding-window analyses, small time bins (i.e., windows) of the signal are analyzed in succession. The technique captures how the signal changes over time and it is often used to create a three dimensional representation of the signal, such as spectrograms (Figures 5 and 13) and correlograms (Figure 12). For time-frequency-varying stimuli such as Mandarin pitch contours, diphthongs and glissandos, “frequency tracks” are generated using sliding window analysis (Krishnan et al. 2005; Wong et al. 2007) to capture how the changing F_0 or harmonic is tracked in the response over time.

Root Mean Square (RMS) Amplitude—For cABRs, RMS amplitude represents the magnitude of neural activation over a given time period (Russo et al. 2004; Akhoun et al. 2008a). RMS is calculated by (1) squaring each point, (2) finding the mean of the squared values and then (3) taking the square root of the mean. The quotient of response RMS amplitude (i.e., signal) and pre-stimulus baseline RMS amplitude (i.e., noise) can serve as a measure of signal to noise ratio (Russo et al. 2004). If the SNR is less than 1, the pre-stimulus activity is larger than the “response” activity. In cases where the SNR of the cABR is less than 1.5, we recollect the cABR or exclude the subject. A typical cABR has an SNR is in the range of 2.5 to 3, although SNRs as high as 6 are not uncommon.

Cross-Correlation—Correlation is a useful tool for comparing the overall morphology and timing of two signals (e.g., stimulus vs. response; Figure 10) (Russo et al. 2004; Akhoun et al. 2008a). In general terms, cross-correlation determines the extent to which two signals are correlated, as a function of the time shift between them. At a given time displacement, if two signals are identical, the cross-correlation coefficient (r) is 1. If the signals are identical but 180° out of phase, it is −1. However, if the signals are completely dissimilar, $r = 0$. In addition to using cross-correlation to determine the degree of similarity, it can also be used to quantify the time delay between two signals (i.e., time displacement that produces the greatest r-value) (Figure 10). The onset of the response can be objectively determined in this manner (Akhoun et al. 2008a) by correlating the stimulus and the response. In addition, two responses can be cross-correlated to determine how much the response has been degraded in noise (Russo et al. 2004; Russo et al. 2005) or how the response changes for different stimulus conditions, e.g., left vs. right ear stimulation (Hornickel et al. 2009a). When performing stimulus-to-response correlations, the stimulus is low pass filtered to remove the high frequencies that are not present in the response (Russo et al. 2004; Akhoun et al. 2008a) (Figure 10) and then resampled to match the sample rate of the response (Section III).

Autocorrelation—Cross-correlation can also be used to find repeating patterns (i.e., periodicities) within a signal, such as the fundamental periodicity (Krishnan et al. 2005; Wong et al. 2007) and the temporal envelope (Lee et al. 2009). This class of cross-correlations is called autocorrelation because a signal is correlated with itself. Autocorrelations are performed by making a copy of a signal and then shifting the copy forward in time.

The fundamental frequency is represented in the stimulus and cABR by peaks occurring at the period of the F_0 (period = 1/frequency). The interpeak interval (period) can be found by calculating the time shift at which the signal best correlates with itself. Thus, autocorrelation is an objective way for determining interpeak intervals and can be used to estimate the F_0 of the response (calculated as 1/d, where d is the time shift needed to obtain the maximum autocorrelation). The strength of phaselocking to the F_0 can also be estimated by this maximal correlation coefficient value. In addition, autocorrelation functions that have broader morphology reflect less robust responses, and steeper functions reflect sharper and more robust phaselocking (Krishnan et al. 2005; Lee et al. 2009).

An autocorrelogram, created via sliding window analysis, is a visual representation of how well the signal correlates with itself across time. In the three dimensional graph (Figure 11, bottom), the degree of correlation is represented in color, with the vertical axis representing the time shift and the horizontal axis representing time. Autocorrelograms can be used to evaluate frequency-tracking to the F_0 and amplitude envelope. Phaselocking can then be described in terms of consistency (i.e., how much the maximum r-value deviates over time) and strength over time (i.e., the average maximum r-value over time) (Figure 11) (Krishnan et al. 2005; Wong et al. 2007; Lee et al. 2009).

Fourier Analysis—A frequency domain representation of the cABR can be generated using Fourier analysis. This method can be used to measure the precision and magnitude of neural phaselocking at specific frequencies or frequency ranges.

One of the basic properties of periodic waveforms is that when two or more waves interact, the resulting waveform is the sum of the individual components (assuming a linear system). This is the principle underlying Fourier analysis. Using Fourier analysis, a complex waveform consisting of many frequency components is decomposed into a set of sine waves. The magnitude of each sine wave corresponds to the amount of energy contained in the complex waveform at that frequency. The spectral composition of a complex wave can then be represented by plotting the frequency of the sine wave on the x-axis and the magnitude on the y-axis (Figures 4 and 12). The fast Fourier transform (FFT) is the most common algorithm for performing spectral analysis (Cooley and Tukey 1965) although other Fourier-based methods have been used by cABR researchers (Dajani et al. 2005; Aiken and Picton 2008). The FFT is most efficient (i.e., faster) when the signal N (defined as the number of points) is a power of two. However, software such as MATLAB (The Mathworks, Natick, MA) and Mathematica (Wolfram Research, Inc., Champaign, IL) do not require the input to be a set length. Fourier analyses can also be used to generate a frequency-domain average. An alternative, less computationally demanding technique, is to perform an FFT on the time-amplitude average.

When dealing with finite signals such as cABRs, the frequency resolution is dependent on the duration of the sample being analyzed (resolution = 1/T, where T = duration in sec). For a 50 ms signal, the frequency resolution is 1 sample per 0.05 sec or 20 Hz. The resulting frequency spectrum contains information only at integer multiples of 20 Hz (i.e., 0, 20, 40, 60…Nyquist frequency; 0 Hz = DC component). If the signal contains a frequency component at 130 Hz, the amplitude of the 130 Hz component, which is not an integer multiple of the sampling period, “leaks” into the neighboring components (i.e., 120, 140, etc.). This leakage can be reduced by increasing T. As a general rule of thumb, T should be long enough to include a minimum of 2 to 4 cycles of the lowest frequency of interest. For example, if you wish to characterize a 100 Hz frequency component, the length of the signal should be at least 20 ms (i.e., (1/100) × 2). One trick for "increasing" T, without actually taking a longer sample of the signal, is to add a series of zeros to the end of the original sample (often called zero padding) (Dajani et al. 2005; Wong et al. 2007). For example, if the 50 ms sample has a 20 kHz sample rate (1000 point sample), to increase the resolution from 20 Hz to 1 Hz, 19,000 zeros are added onto the end of the sample before performing the FFT.

Another thing to consider when performing Fourier analyses is that the FFT treats the sample as if it were a continuous loop, in which the first and last samples are contiguous. Thus, if the starting and ending amplitudes are the not same, the amplitude difference gets reflected in the FFT output. When the discontinuity is large, it creates a click-like feature in the response. Because clicks are broadband, this discontinuity results in *frequency splattering* that contaminates the accuracy of the spectral analysis. In order to prevent this splatter, a common countermeasure is to multiply the signal by a windowing function, which tapers the amplitudes on both ends, so that the sample begins and ends at zero with zero amplitude. Window functions come in many different shapes, although we typically use a Hanning window, which has a bell-shaped function.

For cABRs, frequency spectra are analyzed with respect to the frequency composition of the stimulus. Because stimulus and response amplitudes occur on different scales, the amplitudes must be normalized in order to plot the two spectra on the same plot. This can be achieved by converting both spectra to decibels (Aiken and Picton 2008) or by dividing each spectral amplitude by the corresponding spectral maximum (Figure 4) (Lee et al. 2009). When analyzing the response in the frequency domain, spectral maxima corresponding to the stimulus F_0 and its harmonics are identified, and the phase and amplitude (modulus of the FFT) of the maxima are recorded. Fourier analysis is also useful for calculating the amplitude over a range of frequencies, especially in cases when the stimulus has time-varying features such as formant transitions (Figure 12) (Banai et al. 2009). By performing an FFT on the prestimulus time window, the spectral noise floor can be estimated (Figure 12) and used to calculated spectral SNRs.

If performed as part of a sliding-window analysis, the FFT can be used to generate a spectrogram, a three dimensional graph of the frequency spectrum as a function of time. This type of analysis is often referred to as Short-Term Fourier Transform (STFT). In these plots, the horizontal axis (x-axis) represents time, the vertical axis (y-axis) represents frequency and the third dimension represents the amplitude at a given time-frequency point. The third dimension is usually represented using a color continuum. Frequency-tracks can be derived from STFTs of the response (Figures 5 and 13) (Musacchia et al. 2007; Song et al. 2008). Wavelets provide an alternative method for performing time-frequency analyses (see Addison et al. 2009 for an overview of the emerging role of wavelets in biological signal analysis)

VI. SUMMARY

1. Auditory brainstem responses provide an objective and non-invasive means for examining how behaviorally relevant sounds such as speech and music are transcribed into neural code. The brainstem response is ideal for studying this process because stimulus features are preserved in the response. Notably, this process is not "hard coded". Brainstem encoding of speech and other complex signals is shaped by short-term and lifelong musical and language experience and is thereby tightly coupled with higher-order cognitive processes. Aspects of the response are selectively impaired or enhanced in impaired and expert populations

(e.g., children with reading impairments and musicians) facilitating the delineation of specific underlying processes.

2. Like click-evoked auditory brainstem responses, cABRs are well suited for clinical applications because they can be meaningfully applied to individuals
3. A variety of speech and musical stimuli have been used to evoke auditory brainstem responses. When choosing a stimulus, the acoustic properties of the stimulus matter. To maximize transient responses, the sound should have sharp onsets/amplitude bursts. A low pitch stimulus (<300 Hz), or a stimulus with its fundamental periodicity in this range, is needed to obtain strong phaselocked (i.e., sustained) responses to the F_0 and harmonics.
4. cABRs are generally elicited at supra threshold levels (60-80 dB SPL) using monaural or binaural stimulation via electromagnetically shielded insert earphones. If the stimulus presentation is jittered by even a small amount, cABRs are canceled when trials are averaged.
5. cABRs can be recorded using the same data acquisition procedures as click-ABRs and tone-FFRs. Additionally, manipulations of stimulus polarity can be used to enhance different aspects of the response and to minimize stimulus artifacts and the CM.
6. Because of the transparency between the stimulus and response, digital signal processing tools (e.g., cross-correlation, Fourier analysis) can be used to analyze both the stimulus and response. Sliding-window analysis is used to track how the response changes over time.

VII. CONCLUSION

Neural transcription of sound in the auditory brainstem and midbrain is arguably a measure of auditory processing and as such can be applied to research and clinical assessment whenever auditory processing is of interest. This includes the investigation of auditory specialization (e.g., musicians, native language speakers) and the management of auditory disorders (e.g., auditory processing disorders, language-based learning impairments such as dyslexia, specific language impairment and autism, hearing loss and age-related hearing decline) that result in pervasive difficulties with speech perception in noise. Auditory brainstem responses to complex sounds provide an objective neural metric for determining the effectiveness of remediation strategies, providing the outcome measures that clinicians need to strengthen their role in advocating for auditory training and remediation across the lifespan. Together with other converging lines of research (Fritz et al. 2007; Weinberger 2007; Luo et al. 2008; Atiani et al. 2009), the cABR has reinforced the notion that a contemporary view of the auditory system must include its cognitive and sensory functions. That is, subcortical function inherently reflects a confluence of sensory and cognitive processes that likely operate in a reciprocally interactive manner. This view can help the field of audiology more effectively address socially and clinically meaningful aspects of human communication. It is hoped that the methodological information in this tutorial will help move forward our knowledge and clinical management of auditory processing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Mary Ann Cheatham, Trent Nicol, Bharath Chandrasekaran, and the members of the Auditory Neuroscience Lab for their careful review of this manuscript. We extend our gratitude to Alexandra Parbery-Clark and Frederic Marmel for their contributions to the music-related aspects of Section II. Co-author Nina Kraus currently serves as consultant to Natus Corp. for BioMARK, a clinical measure of auditory function, and receives royalties through Northwestern University. This work was supported by NSF 0842376 and NIH R01 DC01510, F32 DC008052.

Abbreviations

ABR	auditory brainstem response
cABR	auditory brainstem response to complex sounds
CV	consonant-vowel
CM	cochlear microphonic
EP	evoked potential
EEG	electroencephalogram
FFR	frequency following response
F1	first formant
FFT	fast Fourier Transform
F_0	fundamental frequency
Fs	sampling rate
ISI	interstimulus interval
RMS	root-mean-square
SNR	signal to noise ratio
SIN	Speech in Noise
SOA	stimulus onset asynchrony
STFT	Short-term Fourier Transform

REFERENCES

Abel C, Kossl M. Sensitive response to low-frequency cochlear distortion products in the auditory midbrain. J Neurophysiol 2009;101:1560–1574. [PubMed: 19036870]

Addison P, Walker J, Guido R. Time-frequency analysis of biosignals. IEEE Eng Med Biol Mag 2009;28:14–29. [PubMed: 19775952]

Aiken SJ, Picton TW. Envelope following responses to natural vowels. Audiol Neurootol 2006;11:213–232. [PubMed: 16612051]

Aiken SJ, Picton TW. Envelope and spectral frequency-following responses to vowel sounds. Hear Res 2008;245:35–47. [PubMed: 18765275]

Akhoun I, Gallego S, Moulin A, et al. The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. Clin Neurophysiol 2008a;119:922–933. [PubMed: 18291717]

Akhoun I, Moulin A, Jeanvoine A, et al. Speech auditory brainstem response (speech abr) characteristics depending on recording conditions, and hearing status: An experimental parametric study. J Neurosci Methods 2008b;175:196–205. [PubMed: 18789971]

Atiani S, Elhilali M, David SV, et al. Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. Neuron 2009;61:467–480. [PubMed: 19217382]

Ballachanda BB, Moushegian G. Frequency-following response: Effects of interaural time and intensity differences. J Am Acad Audiol 2000;11:1–11. [PubMed: 10741352]

Ballachanda BB, Rupert A, Moushegian G. Asymmetric frequency-following responses. J Am Acad Audiol 1994;5:133–137. [PubMed: 8180429]

Banai, K.; Kraus, N. The dynamic brainstem: Implications for APD. Plural Publishing Inc; San Diego, CA: 2008.

Banai K, Abrams D, Kraus N. Sensory-based learning disability: Insights from brainstem processing of speech sounds. Int J Audiol 2007;46:524–532. [PubMed: 17828668]

Banai K, Nicol T, Zecker SG, et al. Brainstem timing: Implications for cortical processing and literacy. J Neurosci 2005;25:9850–9857. [PubMed: 16251432]

Banai K, Hornickel J, Skoe E, et al. Reading and subcortical auditory function. Cereb Cortex 2009;19:2699–2707. [PubMed: 19293398]

Bidelman G, Krishnan K. Neural Correlates of Consonance, Dissonance, and the Hierarchy of Musical Pitch in the Human Brainstem. J Neurosci 2009;29(42):13165–13171. [PubMed: 19846704]

Boersma P, Weenink D. Praat: Doing phonetics by computer (version 5.1.05). 2009

Bradlow AR, Kraus N, Hayes E. Speaking clearly for children with learning disabilities: Sentence perception in noise. J Speech Lang Hear Res 2003;46:80–97. [PubMed: 12647890]

Brugge JF, Anderson DJ, Hind JE, et al. Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. J Neurophysiol 1969;32:386–401. [PubMed: 4977838]

Burkard, RF.; Eggermont, JJ.; Don, M., editors. Auditory evoked potentials: Basic principles and clinical application. Lippincott Williams and Wilkins; Philadelphia: 2007.

Burns, K.; Soll, A.; Vander Werff, K. Brainstem responses to speech in younger and older adults. American Auditory Society Annual Meeting; Scottsdale AZ. March 2009; 2009.

Chandrasekaran B, Kraus N. The scalp-recorded brainstem response to speech: Neural origins and plasticity. Psychophysiology. 2009

Chandrasekaran B, Gandour JT, Krishnan A. Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. Restor Neurol Neurosci 2007;25:195–210. [PubMed: 17942999]

Chandrasekaran B, Hornickel J, Skoe E, et al. Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: Implications for developmental dyslexia. Neuron. 2009 doi 10.1016/j.neuron.2009.10.006.

Cooley JW, Tukey JW. An algorithm for the machine calculation of complex fourier series. Mathematics of Computation 1965;19:297–301.

Cunningham J, Nicol T, Zecker SG, et al. Neurobiologic responses to speech in noise in children with learning problems: Deficits and strategies for improvement. Clin Neurophysiol 2001;112:758–767. [PubMed: 11336890]

Dajani HR, Purcell D, Wong W, et al. Recording human evoked potentials that follow the pitch contour of a natural vowel. IEEE Trans Biomed Eng 2005;52:1614–1618. [PubMed: 16189976]

Dhar S, Abel R, Hornickel J, et al. Exploring the relationship between physiological measures of cochlear and brainstem function. Clin Neurophysiol 2009;120:959–966. [PubMed: 19346159]

Dobie RA, Berlin CI. Binaural interaction in brainstem-evoked responses. Arch Otolaryngol 1979;105:391–398. [PubMed: 454297]

Don M, Vermiglio AJ, Ponton CW, et al. Variable effects of click polarity on auditory brain-stem response latencies: Analyses of narrow-band abrs suggest possible explanations. J Acoust Soc Am 1996;100:458–472. [PubMed: 8675840]

Dubno JR, Dirks DD, Morgan DE. Effects of age and mild hearing loss on speech recognition in noise. The Journal of the Acoustical Society of America 1984;76:87–96. [PubMed: 6747116]

Elberling, C.; Don, M. Detecting and assessing synchronous neural activity in the temporal domain (snr, response detection). In: Burkard, RF.; Eggermont, JJ.; Don, M., editors. Auditory evoked potentials : Basic principles and clinical application. Lippincott Williams & Wilkins; Philadelphia: 2007. p. 102-123.

Elsisy H, Krishnan A. Comparison of the acoustic and neural distortion product at 2f1-f2 in normal-hearing adults. Int J Audiol 2008;47:431–438. [PubMed: 18574781]

Everest, FA.; ebrary Inc.. The master handbook of acoustics. 4th ed.. McGraw-Hill; New York: 2001. p. xix, 615

Fletcher, NH.; Rossing, TD. The physics of musical instruments. Springer-Verlag; New York: 1991.

Fritz JB, Elhilali M, David SV, et al. Auditory attention--focusing the searchlight on sound. Curr Opin Neurobiol 2007;17:437–455. [PubMed: 17714933]

Galbraith GC. Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. Electroencephalogr Clin Neurophysiol 1994;92:321–330. [PubMed: 7517854]

Galbraith GC, Arroyo C. Selective attention and brainstem frequency-following responses. Biol Psychol 1993;37:3–22. [PubMed: 8110920]

Galbraith GC, Kane JM. Brainstem frequency-following responses and cortical event-related potentials during attention. Percept Mot Skills 1993;76:1231–1241. [PubMed: 8337070]

Galbraith GC, Doan BQ. Brainstem frequency-following and behavioral responses during selective attention to pure tone and missing fundamental stimuli. Int J Psychophysiol 1995;19:203–214. [PubMed: 7558987]

Galbraith GC, Olfman DM, Huffman TM. Selective attention affects human brain stem frequency-following response. Neuroreport 2003;14:735–738. [PubMed: 12692473]

Galbraith GC, Arbagey PW, Branski R, et al. Intelligible speech encoded in the human brain stem frequency-following response. Neuroreport 1995;6:2363–2367. [PubMed: 8747154]

Galbraith GC, Bhuta SM, Choate AK, et al. Brain stem frequency-following response to dichotic vowels during attention. Neuroreport 1998;9:1889–1893. [PubMed: 9665621]

Galbraith GC, Amaya EM, de Rivera JM, et al. Brain stem evoked response to forward and reversed speech in humans. Neuroreport 2004;15:2057–2060. [PubMed: 15486481]

Gray GW. Phonemic microtomy: The minimum duration of perceptible speech sounds. Speech Monographs 1942;9:75.

Greenberg S. Wpp, no. 52: Temporal neural coding of pitch and vowel quality. Department of Linguistics, UCLA. Working Papers in Phonetics. 1980

Greenberg S, Marsh JT, Brown WS, et al. Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. Hear Res 1987;25:91–114. [PubMed: 3558136]

Grey JM. Multidimensional perceptual scaling of musical timbres. The Journal of the Acoustical Society of America 1977;61:1270–1277. [PubMed: 560400]

Hall, JW. New handbook of auditory evoked responses. Allyn and Bacon; Boston: 2006.

Hall JW 3rd. Auditory brainstem frequency following responses to waveform envelope periodicity. Science 1979;205:1297–1299. [PubMed: 472748]

Hayes EA, Warrier CM, Nicol TG, et al. Neural plasticity following auditory training in children with learning problems. Clin Neurophysiol 2003;114:673–684. [PubMed: 12686276]

Holmes, J.; Holmes, W. Speech synthesis and recognition. Taylor & Francis, Inc; 2002.

Hood, LJ. Clinical applications of the auditory brainstem response. Singular Pub. Group; San Diego: 1998.

Hoormann J, Falkenstein M, Hohnsbein J. Effects of spatial attention on the brain stem frequency-following potential. Neuroreport 2004;15:1539–1542. [PubMed: 15232279]

Hornickel J, Skoe E, Kraus N. Subcortical laterality of speech encoding. Audiol Neurootol 2009a; 14:198–207. [PubMed: 19122453]

Hornickel J, Skoe E, Nicol T, et al. Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. Proc Natl Acad Sci U S A 2009b;106:13022–13027. [PubMed: 19617560]

Howard, DM.; Angus, JAS. Acoustics and psychoacoustics. 2nd ed.. Focal Press; Oxford; Boston: 2001.

Jewett DL, Williston JS. Auditory-evoked far fields averaged from the scalp of humans. Brain 1971;94:681–696. [PubMed: 5132966]

Jewett DL, Romano MN, Williston JS. Human auditory evoked potentials: Possible brain stem components detected on the scalp. Science 1970;167:1517–1518. [PubMed: 5415287]

Johnson KL, Nicol TG, Zecker SG, et al. Auditory brainstem correlates of perceptual timing deficits. J Cogn Neurosci 2007;19:376–385. [PubMed: 17335387]

Johnson KL, Nicol T, Zecker SG, et al. Developmental plasticity in the human auditory brainstem. J Neurosci 2008a;28:4000–4007. [PubMed: 18400899]

Johnson KL, Nicol T, Zecker SG, et al. Brainstem encoding of voiced consonant--vowel stop syllables. Clin Neurophysiol 2008b;119:2623–2635. [PubMed: 18818121]

Kawahara, H. Straight. 2009. http://www.Wakayama-u.Ac.Jp/~kawahara/straighttrial/

Kemp DT. Otoacoustic emissions, their origin in cochlear function, and use. Br Med Bull 2002;63:223–241. [PubMed: 12324396]

King C, Warrier CM, Hayes E, et al. Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. Neurosci Lett 2002;319:111–115. [PubMed: 11825683]

Klatt D. Software for cascade/parallel formant synthesizer. Journal for the Acoustic Society of America 1976;67:971–975.

Kraus N, Nicol T. Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. Trends Neurosci 2005;28:176–181. [PubMed: 15808351]

Kraus N, Banai K. Auditory processing malleability: Focus on language and music. Current Directions in Psychological Science 2007;16:105–109.

Kraus N, Skoe E, Parbery-Clark A, et al. Experience-induced malleability in neural encoding of pitch, timbre, and timing. Ann N Y Acad Sci 2009;1169:543–557. [PubMed: 19673837]

Kraus N, McGee TJ, Carrell TD, et al. Auditory neurophysiologic responses and discrimination deficits in children with learning problems. Science 1996;273:971–973. [PubMed: 8688085]

Krishnan A. Human frequency-following responses: Representation of steady-state synthetic vowels. Hearing Research 2002;166:192–201. [PubMed: 12062771]

Krishnan, A. Frequency-following response. In: Burkard, RF.; Eggermont, JJ.; Don, M., editors. Auditory evoked potentials: Basic principles and clinical application. Lippincott Williams & Wilkins; Philadelphia: 2007. p. 313-335.

Krishnan A, McDaniel SS. Binaural interaction in the human frequency-following response: Effects of interaural intensity difference. Audiol Neurootol 1998;3:291–299. [PubMed: 9705526]

Krishnan A, Gandour JT. The role of the auditory brainstem in processing linguistically-relevant pitch patterns. Brain Lang 2009;110(3):135–148. [PubMed: 19366639]

Krishnan A, Swaminathan J, Gandour JT. Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. J Cogn Neurosci 2009a; 21:1092–1105. [PubMed: 18702588]

Krishnan A, Xu Y, Gandour JT, et al. Human frequency-following response: Representation of pitch contours in chinese tones. Hear Res 2004;189:1–12. [PubMed: 14987747]

Krishnan A, Xu Y, Gandour J, et al. Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res Cogn Brain Res 2005;25:161–168. [PubMed: 15935624]

Krishnan A, Gandour JT, Bidelman GM, et al. Experience-dependent neural representation of dynamic pitch in the brainstem. Neuroreport 2009b;20:408–413. [PubMed: 19223793]

Krizman J, Skoe E, Kraus N. Stimulus rate and subcortical auditory processing of speech. Audiology Neurotology. in press.

Langner G, Schreiner CE. Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J Neurophysiol 1988;60:1799–1822. [PubMed: 3236052]

Lee KM, Skoe E, Kraus N, et al. Selective subcortical enhancement of musical intervals in musicians. J Neurosci 2009;29:5832–5840. [PubMed: 19420250]

Liberman, AM. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. American Psychological Association; Washington: 1954.

Lins OG, Picton TW. Auditory steady-state responses to multiple simultaneous stimuli. Electroencephalogr Clin Neurophysiol 1995;96:420–432. [PubMed: 7555916]

Luo F, Wang Q, Kashani A, et al. Corticofugal modulation of initial sound processing in the brain. J Neurosci 2008;28:11615–11621. [PubMed: 18987197]

Maddieson, I. Patterns of sounds. Cambridge University Press; Cambridge [Cambridgeshire] ; New York: 1984.

Moore DR. Anatomy and physiology of binaural hearing. Audiology 1991;30:125–134. [PubMed: 1953442]

Moulines E, Charpentier F. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Commun 1990;9:453–467.

Moushegian G, Rupert AL, Stillman RD. Laboratory note. Scalp-recorded early responses in man to frequencies in the speech range. Electroencephalogr Clin Neurophysiol 1973;35:665–667. [PubMed: 4128165]

Musacchia G, Strait D, Kraus N. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hear Res 2008;241:34–42. [PubMed: 18562137]

Musacchia G, Sams M, Nicol T, et al. Seeing speech affects acoustic information processing in the human brainstem. Exp Brain Res 2006;168:1–10. [PubMed: 16217645]

Musacchia G, Sams M, Skoe E, et al. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc Natl Acad Sci U S A 2007;104:15894–15898. [PubMed: 17898180]

Nunez, PL.; Srinivasan, R. Electric fields of the brain : The neurophysics of eeg. 2nd ed.. Oxford University Press; Oxford ; New York: 2006.

Palmer, A.; Shamma, S. Physiological representations of speech. In: Greenberg, S.; Ainsworth, WA.; Popper, AN.; Fay, RR., editors. Speech processing in the auditory system, Vol. Springer handbook of auditory research. Vol. 18. Springer; New York: 2004. p. xiv, 476

Parbery-Clark A, Skoe E, Kraus N. Musical experience limits the degradative effects of background noise on the neural processing of sound. Journal of Neuroscience. in press.

Parbery-Clark A, Skoe E, Lam C, et al. Musician enhancement for speech-in-noise. Ear Hear. 2009

Pichora-Fuller MK, Schneider BA, Daneman M. How young and old adults listen to and remember speech in noise. The Journal of the Acoustical Society of America 1995;97:593–608. [PubMed: 7860836]

Picton TW, Hillyard SA. Human auditory evoked potentials. Ii. Effects of attention. Electroencephalogr Clin Neurophysiol 1974;36:191–199. [PubMed: 4129631]

Plyler PN, Ananthanarayan AK. Human frequency-following responses: Representation of second formant transitions in normal-hearing and hearing-impaired listeners. J Am Acad Audiol 2001;12:523–533. [PubMed: 11791939]

Porat, B. A course in digital signal processing. John Wiley; New York: 1997.

Quian Quiroga R, Sakowitz OW, Basar E, et al. Wavelet transform in the analysis of the frequency composition of evoked potentials. Brain Res Brain Res Protoc 2001;8:16–24. [PubMed: 11522524]

Regan D, Regan MP. The transducer characteristic of hair cells in the human ear: A possible objective measure. Brain Res 1988;438:363–365. [PubMed: 3345444]

Rinne T, Balk MH, Koistinen S, et al. Auditory selective attention modulates activation of human inferior colliculus. J Neurophysiol 2008;100:3323–3327. [PubMed: 18922948]

Robinson K. The duration required to identify the instrument, the octave, or the pitch chroma of a musical note. Music Perception 1995;13:1–15.

Robinson K, Patterson RD. The stimulus duration required to identify vowels, their octave, and their pitch chroma. The Journal of the Acoustical Society of America 1995;98:1858–1865.

Rosen S. Temporal information in speech: Acoustic, auditory and linguistic aspects. Philosophical Transactions: Biological Sciences 1992;336:367–373. [PubMed: 1354376]

Russo N, Larson C, Kraus N. Audio-vocal system regulation in children with autism spectrum disorders. Exp Brain Res 2008a;188:111–124. [PubMed: 18347784]

Russo N, Nicol T, Musacchia G, et al. Brainstem responses to speech syllables. Clin Neurophysiol 2004;115:2021–2030. [PubMed: 15294204]

Russo N, Nicol T, Trommer B, et al. Brainstem transcription of speech is disrupted in children with autism spectrum disorders. Developmental Science 2009;12:557–567. [PubMed: 19635083]

Russo NM, Nicol TG, Zecker SG, et al. Auditory training improves neural timing in the human brainstem. Behav Brain Res 2005;156:95–103. [PubMed: 15474654]

Russo NM, Skoe E, Trommer B, et al. Deficient brainstem encoding of pitch in children with autism spectrum disorders. Clin Neurophysiol 2008b;119:1720–1731. [PubMed: 18558508]

Shannon R. Speech and music have different requirements for spectral resolution. International Review of Neurology 2005:121–134.

Sininger YS. Auditory brain stem response for objective measures of hearing. Ear Hear 1993;14:23–30. [PubMed: 8444334]

Skoe, E.; Kraus, N. Society for Neuroscience. Chicago, IL: 2009. The effects of musical training on subcortical processing of a missing fundamental piano melody.

Song JH, Banai K, Russo NM, et al. On the relationship between speech- and nonspeech-evoked auditory brainstem responses. Audiol Neurootol 2006;11:233–241. [PubMed: 16675898]

Song JH, Skoe E, Wong PC, et al. Plasticity in the adult human auditory brainstem following short-term linguistic training. J Cogn Neurosci 2008;20:1892–1902. [PubMed: 18370594]

Starr, A.; Don, M. Brain potentials evoked by acoustic stimuli. In: Picton, TW., editor. Handbook of electroencephalography and clinical neurophysiology (revised volume 3). Human event-related potentials. Elsevier; Amsterdam: 1988. p. 97-157.

Starr A, Picton TW, Sininger Y, et al. Auditory neuropathy. Brain 1996;119(Pt 3):741–753. [PubMed: 8673487]

Strait DL, Kraus N, Skoe E, et al. Musical experience and neural efficiency: Effects of training on subcortical processing of vocal expressions of emotion. Eur J Neurosci 2009;29:661–668. [PubMed: 19222564]

Swaminathan J, Krishnan A, Gandour JT, et al. Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem. IEEE Trans Biomed Eng 2008;55:281–287. [PubMed: 18232372]

Tallal P, Stark RE. Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. The Journal of the Acoustical Society of America 1981;69:568–574. [PubMed: 7462479]

Thornton, ARD. Instrumentation and recording parameters. In: Burkard, RF.; Eggermont, JJ.; Don, M., editors. Auditory evoked potentials : Basic principles and clinical application. Lippincott Williams & Wilkins; Philadelphia: 2007. p. 42-72.

Turner CW, Fabry DA, Barrett S, et al. Detection and recognition of stop consonants by normal.Hearing and hearing impaired listeners. J Speech Hear Res 1992;35:942–949. [PubMed: 1405550]

Tzounopoulos T, Kraus N. Learning to encode timing: Mechanisms of plasticity in the auditory brainstem. Neuron 2009;62:463–469. [PubMed: 19477149]

van Drongelen, W. Signal processing for neuroscientists : Introduction to the analysis of physiological signals. Academic Press; Burlington, Mass.: 2007.

Wallisch, P. Matlab for neuroscientists : An introduction to scientific computing in matlab. Elsevier/ Academic Press; Amsterdam; Boston: 2009.

Weinberger NM. Associative representational plasticity in the auditory cortex: A synthesis of two disciplines. Learn Mem 2007;14:1–16. [PubMed: 17202426]

Wible B, Nicol T, Kraus N. Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. Biol Psychol 2004;67:299–317. [PubMed: 15294388]

Wible B, Nicol T, Kraus N. Correlation between brainstem and cortical auditory processes in normal and language-impaired children. Brain 2005;128:417–423. [PubMed: 15634732]

Wile D, Balaban E. An auditory neural correlate suggests a mechanism underlying holistic pitch perception. PLoS One 2007;2(4):e369. [PubMed: 17426817]

Wong PC, Skoe E, Russo NM, et al. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat Neurosci 2007;10:420–422. [PubMed: 17351633]

Xu Y, Krishnan A, Gandour JT. Specificity of experience-dependent pitch representation in the brainstem. Neuroreport 2006;17:1601–1605. [PubMed: 17001276]

Young ED, Sachs MB. Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. The Journal of the Acoustical Society of America 1979;66:1381–1403. [PubMed: 500976]

Zatorre RJ, Belin P, Penhune VB. Structure and function of auditory cortex: Music and speech. Trends in Cognitive Sciences 2002;6:37–46. [PubMed: 11849614]

Ziegler J, Pech-Georgel C, Florence G, et al. Speech-perception-in-noise deficits in dyslexia. Developmental Science. 2009

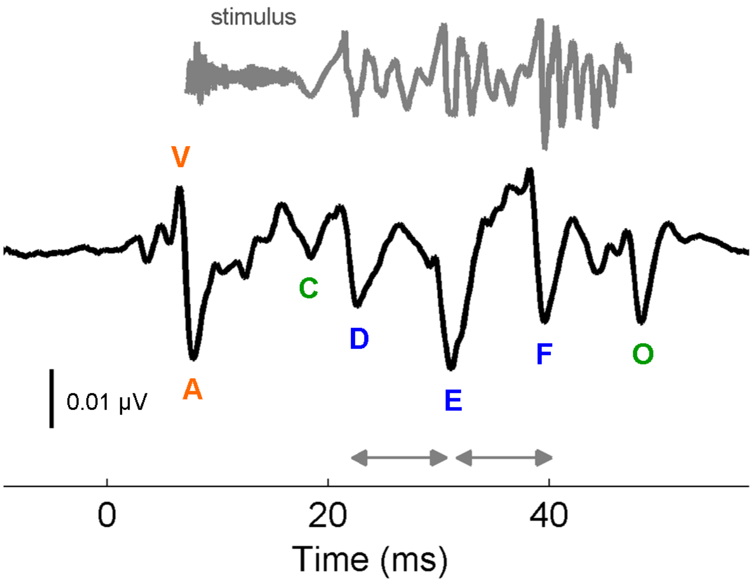

Figure 1. Transient and sustained features in the cABR to /dɑ/

Time-domain representation of a 40 ms stimulus /da/ (gray) and response (black). The cABR to /dɑ/ includes both transient and sustained response features. This stimulus evokes seven characteristic response peaks that we have termed V, A, C, D, E, F and O. As can be seen in this figure, these peaks relate to major acoustic landmarks in the stimulus. Peaks occur approximately 7 to 8 ms after the corresponding stimulus landmark, which is consistent with neural transmission time between the cochlea and rostral brainstem. In this figure, the stimulus waveform is shifted in time to account for this transmission time and maximize the visual coherence between the two signals. Along with V and A, C and O are considered transient responses in that they correspond to transient stimulus features, the beginning and end of voicing, respectively. The V-A complex, often together referred to as the onset response, is analogous to the click-evoked wave V-Vn complex. This sharp onset response arises from the broadband stop burst associated with /d/. The region between D and F forms the frequency-following response (FFR). Peaks D, E, F and the small voltage fluctuations between them correspond to sustained stimulus features, namely the fundamental frequency (F_0) and harmonics within the consonant-vowel formant transition. The D-E-F inter-peak interval (~8-9 ms duration, arrows) occurs at the period of the F_0 of the stimulus, which ramps from 103-125 Hz. We have developed a systematic approach for identifying these peaks and have established normative data for 3-4 year olds, 5-12 year olds and young adults (Johnson et al., 2008b; Dhar et al., 2009). Here, and in all figures showing a stimulus waveform, the stimulus plot is scaled to match the size of the response. Hence, the microvolt bar refers only to the response.

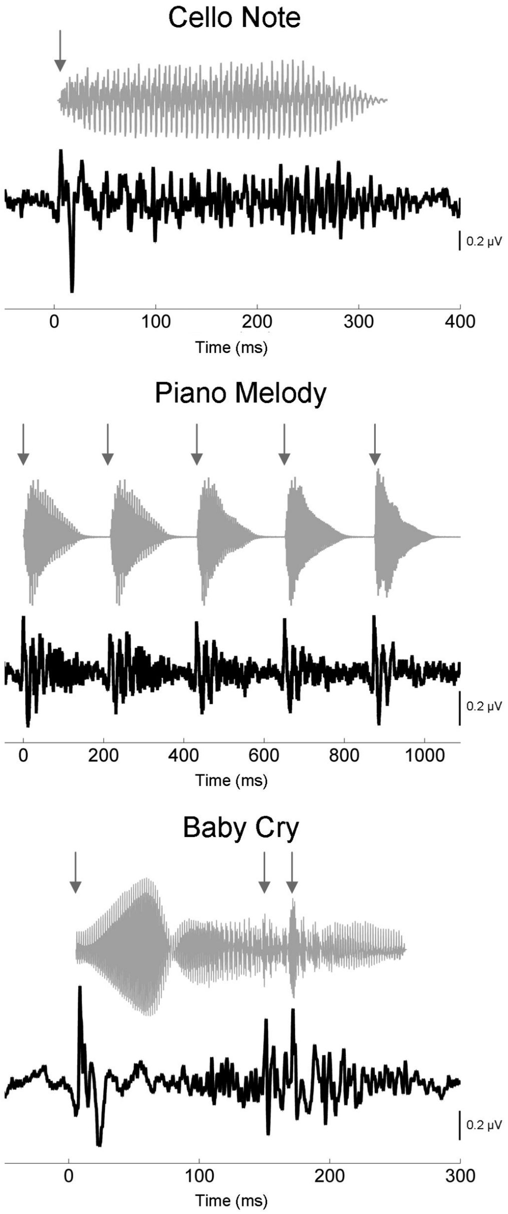

Figure 2. Transient responses

To maximize the visual coherence between the stimulus (gray) and response (black), stimulus waveforms are shifted in time to align the stimulus with the response onset. Arrows indicate major transient features. In the response, these transient features are represented as large peaks. **Top:** The brainstem response to a cello note with a low pitch (G2, 100 Hz). The sound onset occurs when the bow contacts the string and causes a brief transience before the string starts to vibrate in a periodic manner. This leads to a strong onset, followed by a more sustained response. Because of the gradual decay of this sound, a strong offset response is not apparent. Adapted from [Musacchia, Sams, Skoe and Kraus, 2007]. [to listen to stimulus see Audio file, Supplemental Digital Content 1.wav] [to listen to response see Audio file, Supplemental Digital Content 2.wav] [**Middle:** Percussive instruments, like the piano, have fast attacks and rapid decays. These features are evident in this 5-note piano melody. Large response peaks coincide with the onset of each piano note. The stimulus amplitude envelope is also preserved in the response. [to listen to stimulus see Audio file, Supplemental Digital Content 3.wav] [to listen to response see Audio file, Supplemental Digital Content 4.wav] **Bottom:** Sounds with abrupt changes in the amplitude envelope also trigger multiple onset-like transient responses. This is illustrated here using the sound of a crying baby. Adapted from [Strait, Skoe, Ashley, and Kraus, 2009]. In the top and bottom plots, the stimulus was presented binaurally; in the middle plot, it was presented monaurally.

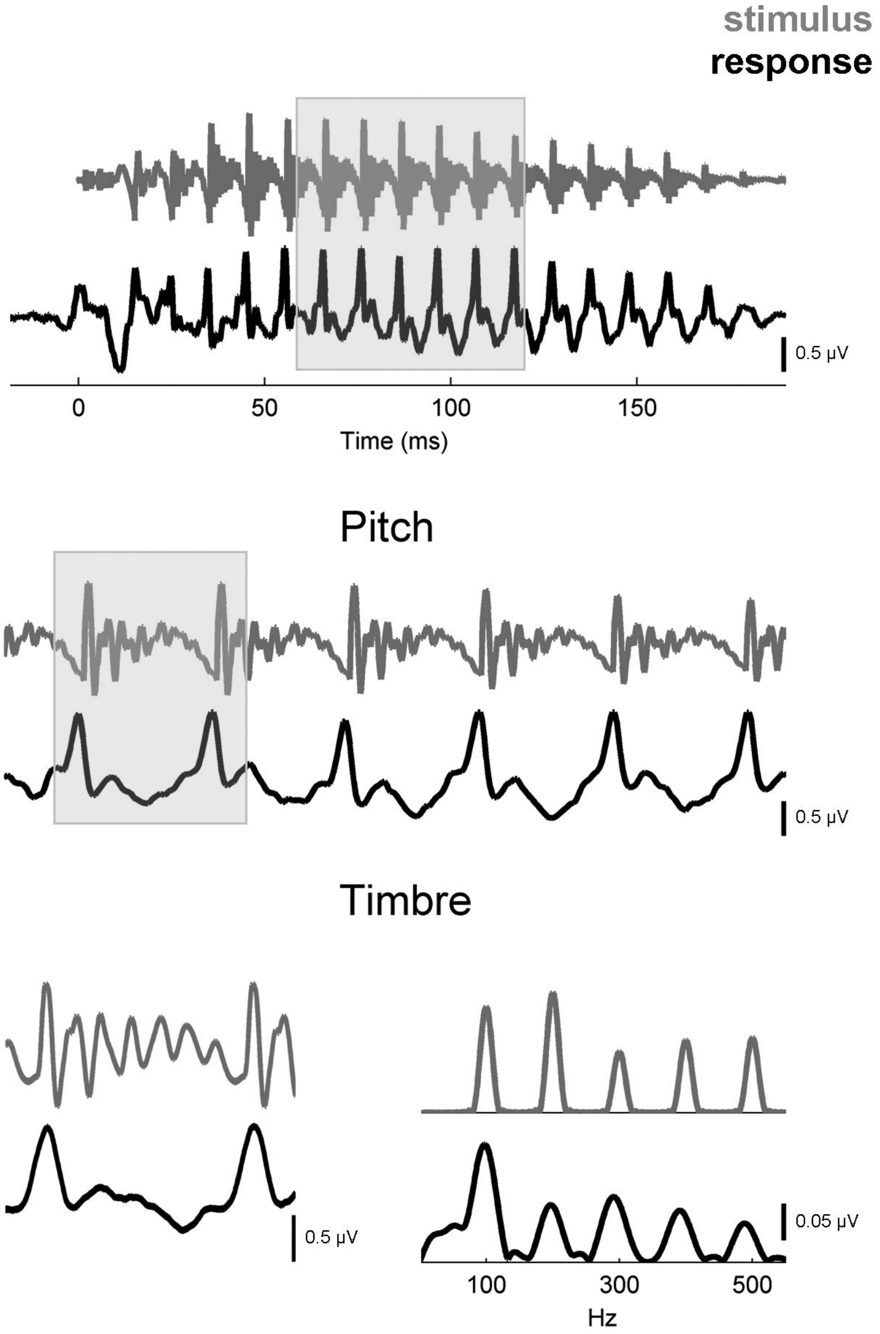

Figure 3. Sustained phaselocked responses

Low frequencies, including those associated with pitch and timbre perception, are preserved in the cABR. For complex sounds the pitch corresponds (in large part) to the lowest resonant frequency, also known as the fundamental frequency (F_0). Timbre enables us to differentiate two sounds with the same pitch. Timbre is a multidimensional property resulting from timing cues of attack and decay, and the interaction of spectral and temporal properties associated with the harmonics of the F_0. Together these timbral features give rise to the characteristic sound quality associated with a given instrument or voice. **Top:** The full view of the time-domain stimulus /dɑ/ (gray) and its cABR (black). The spectrotemporal features of the stimulus, including the F_0 and harmonics, are evident in the response. The gray box demarcates six cycles of the F_0. This section is magnified in the middle panel. **Middle:** The smallest repeating unit of the stimulus has a duration of 10 ms (i.e., the periodicity of the 100 Hz F_0). **Bottom:** The left panel shows a close-up of a single F_0 cycle. The harmonics of the F_0 (frequencies at multiples of 100 Hz) are represented as small fluctuations between the major F_0 peaks in both the stimulus and response. In the right panel, the stimulus and cABR are plotted in the frequency domain. The frequencies important for the perception of pitch and timbre are maintained in the response.

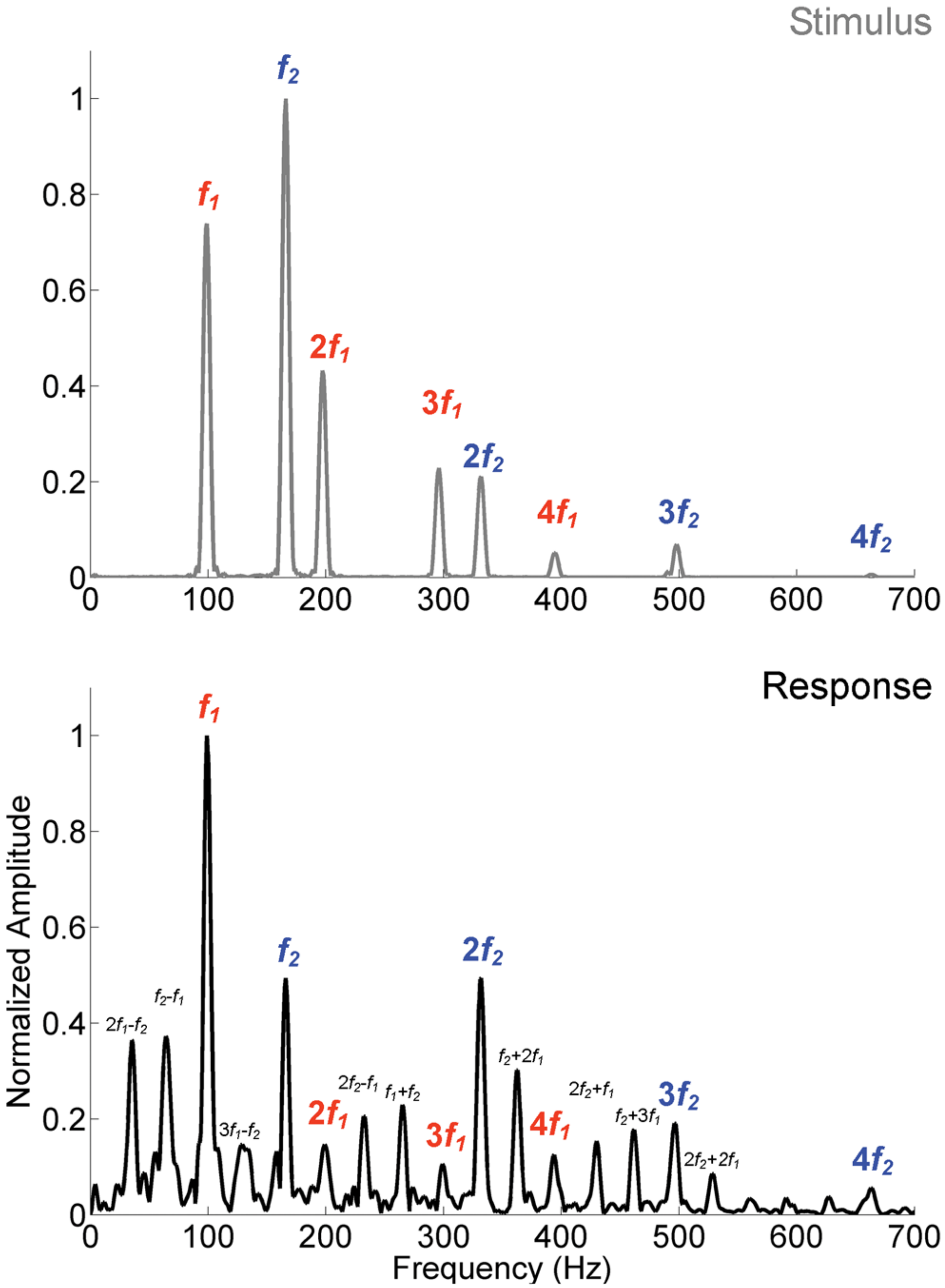

Figure 4. Distortion products (DPs) in the cABR

Stimulus (top) and response (bottom) spectra for a consonant musical interval (major 6th). This musical stimulus was created from G2 and E3 notes produced by an electric piano. When two harmonically complex notes are played simultaneously, the F_0s and harmonics interact via nonlinear auditory processes to create DPs that are measurable in the response but not present in the stimulus. In this figure, parentheses denote the DPs, f_1 denotes the lower tone (G2, red) and f_2 denotes the upper tone (E3, blue). Adapted from [Lee, Skoe, Kraus and Ashley, 2009] [to listen to stimulus see Audio file, Supplemental Digital Content 5.wav] [to listen to response see Audio file, Supplemental Digital Content 6.wav].

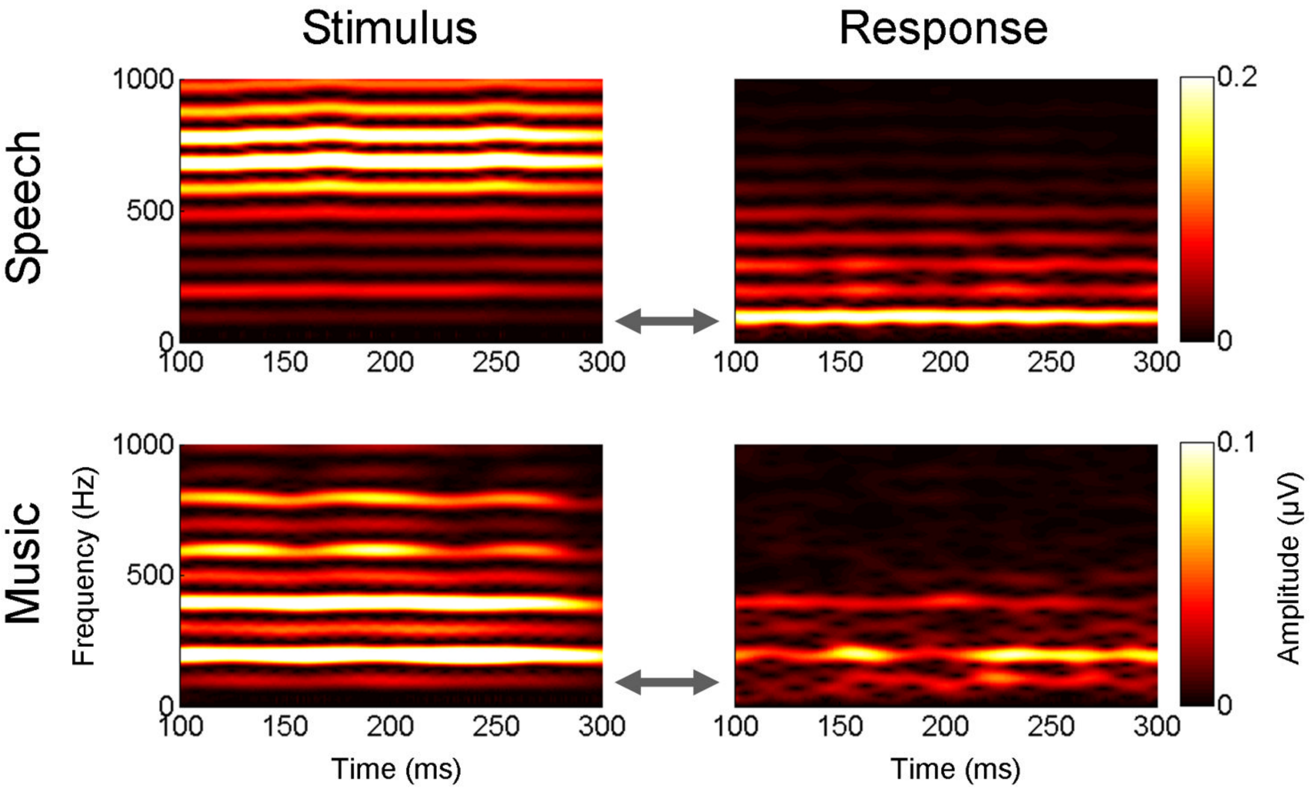

Figure 5. cABRs to harmonically complex signals

The sustained aspects of brainstem responses (right) and their evoking stimuli (left) can be visualized using spectrograms (Section V and Figure 13). These graphs represent a 200 ms steady-state (unchanging) segment of the vowel /α/ (top) and the cello note (bottom, see also Figure 2) used in [Musacchia et al. 2007]. In this example, the speech (top) and musical stimulus (bottom) have the same pitch (F_0 = 100 Hz; arrows), yet have very different harmonic structures and consequently different timbres. These acoustic differences account for the different response patterns. For the cello (bottom), the dominant frequency bands occur at 200 and 400 Hz in both the stimulus and response. For the speech signal (top), the harmonics around the first formant (700 Hz) have much more energy than the F_0. Yet, lower frequencies dominate the response. This reflects the low-pass nature of brainstem phaselocking and the nonlinear processes that amplify the energy of the F_0 and the lower-harmonics. Adapted from [Musacchia, Sams, Skoe and Kraus, 2007].

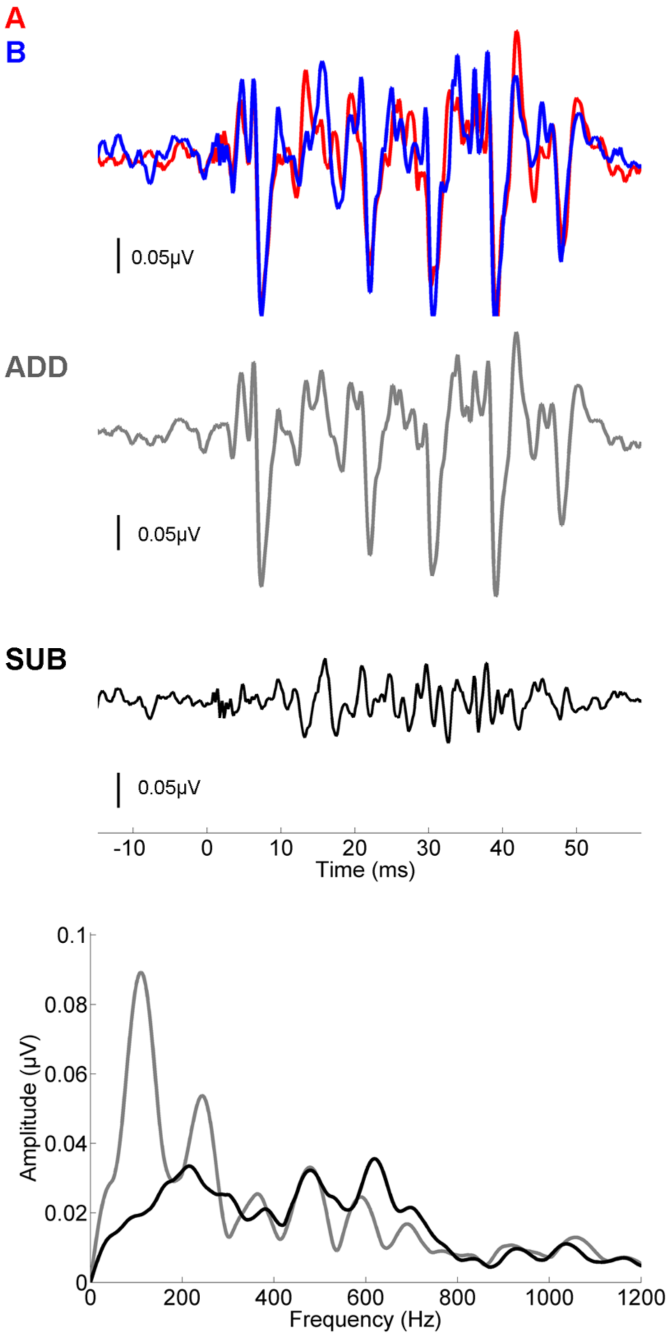

Figure 6. Stimulus polarities

cABRs the two polarities of the /dɑ/ stimulus from Figure 1. For shorthand, they are referred to as polarity A (red) and B (blue). The cABRs to A and B are quite similar, especially for the prominent negative going peaks corresponding to the F_0 (top) (Akhoun et al., 2008b). By adding or subtracting A and B, spectral and envelope components of the response, respectively, can be separated (see footnote 3). Adding (gray) accentuates the lower-frequency components of the response, including the temporal envelope, and minimizes stimulus artifact and the cochlear microphonic (see Figure 8, and Section V for a discussion of artifacts). Subtracting (black) emphasizes the higher-frequency components by maximizing the spectral response; however, this process can also maximize artifact contamination. In the bottom panel, the ADD and SUB responses are plotted in the frequency domain. In contrast to the ADD response, which has peaks occurring at F_0 (~100 Hz) and the harmonics of the F_0, the SUB response has well-defined peaks in the 200-700 Hz range. This range corresponds to the first formant trajectory of this stimulus. In this figure, ADD = (A+B)/2; SUB = (A−B)/2).

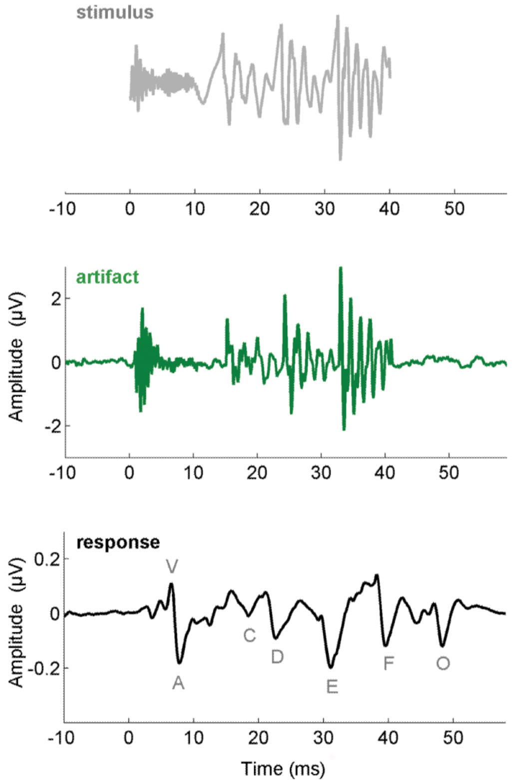

Figure 7. Detecting stimulus artifact

Stimulus artifacts are easily discernable in cABR recordings. Unlike the response (bottom), the artifact (middle) contains frequencies that are higher than the phaselocking capabilities of the brainstem (Moushegian *et al.*, 1973). In contrast to the cABR response, which occurs within 6-10 ms after the stimulus (top) is played, the stimulus artifact also exhibits no delay. In addition, the artifact is often larger than a typical cABR. In this example, the artifact to a 40 ms /dɑ/ (Figure 1) is 10 times larger than the response. Stimulus artifact can be minimized by using electromagnetically shielded insert earphones and adding the responses to alternating polarities (Figure 8).

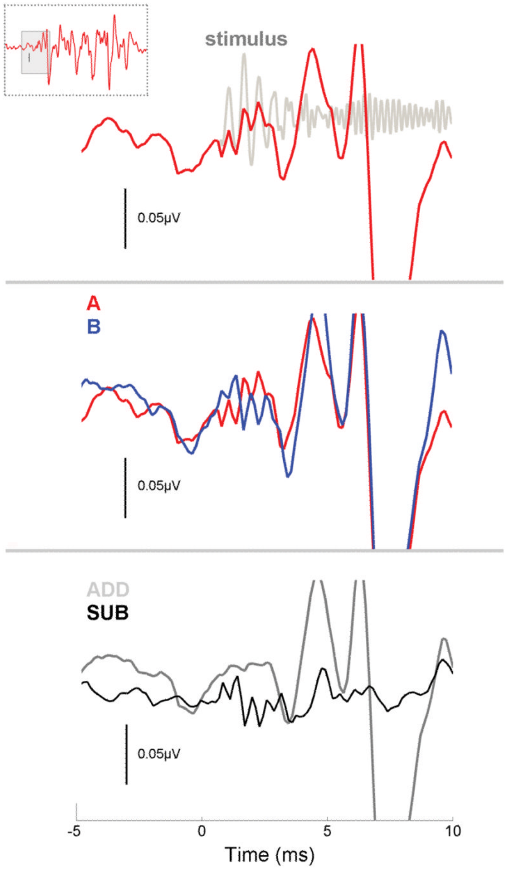

Figure 8. Adding polarities (A and B) minimizes stimulus artifact and cochlear microphonic

cABRs to the A and B polarities of the 40-ms /dɑ/ from Figure 1. **Top:** The response to polarity A (inset) is magnified (−5 to 10 ms) to illustrate the stimulus artifact. When the A response (red) is compared to the stimulus (light gray), the two waveforms align in phase for ~4 ms. This is because the stimulus artifact (and CM) follow the temporal pattern of the stimulus waveform. **Middle:** The response to polarity B (blue) is inverted with respect to the A response in this region. **Bottom:** By adding together A and B responses (gray), the artifact is canceled. In contrast, the artifact is accentuated when the two polarities are subtracted (black). Thus, while the analysis of the subtracted waveform or the single polarity response can be complicated by unwanted artifacts, the added response ensures a response of neurogenic origins (Aiken et al., 2008). In this figure, ADD = (A+B)/2; SUB = (A−B)/2.

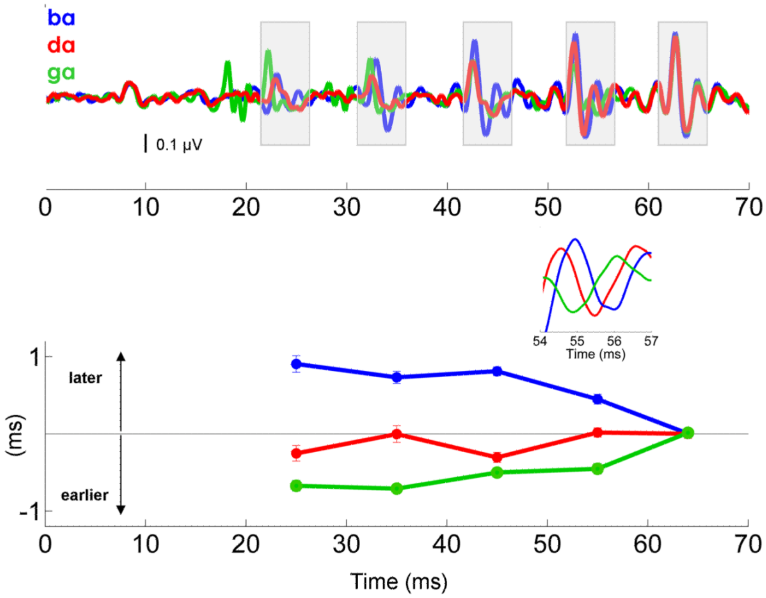

Figure 9. Tracking latency differences over time

The frequency trajectories that differentiate the CV stop syllables /ba/, /da/, /ga/ are represented in the cABR by latency differences, with /ga/ responses occurring first, followed by /da/ and then /ba/, i.e., higher frequencies yield earlier peak latencies than lower frequencies. In the stimulus, the frequency differences diminish over the course of the 50-ms formant transition. **Top:** This pattern is reflected in the timing of the cABR (/ga/ < /da/ < /ba/) and is most apparent at four discrete response peaks. Peaks ~55 are magnified in the inset. **Bottom:** The normalized latency difference between responses is plotted as a function of time (see Johnson et al., 2008b; Hornickel et al. 2009b, for details).

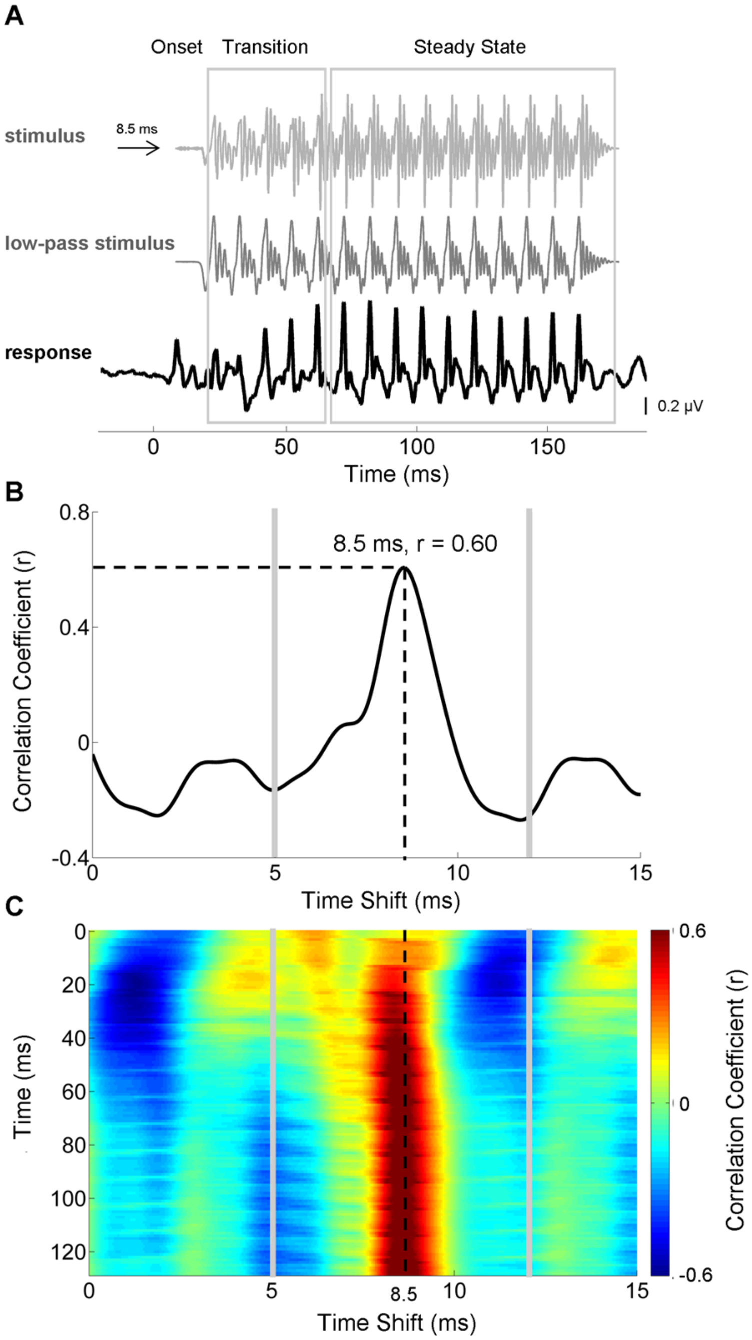

Figure 10. Stimulus-to-response cross-correlation

Cross-correlation is used to compare the timing and morphology of two signals **A.** The cABR (black, bottom) to a 170 ms /dɑ/ (gay, top) is compared to a low-pass filtered version (gray, middle) of the evoking stimulus. The stimulus consists of an onset stop-burst, CV formant transition and a steady-state (i.e., unchanging) vowel. **B:** This plot represents the degree to which the low-pass stimulus and response are correlated as a function of the time shift. The maximal correlation is reached at an 8.5 ms time displacement, an indication of the neural transmission delay (r_{max} = 0.60; r_{max} = 0.32 for unfiltered stimulus (correlogram not shown)). An alternative approach is to cross-correlate the response with the stimulus envelope (Akhoun et al., 2008a), which can often lead to higher correlation values. **C:** Running-window analyses can be used to visualize and quantify the similarity of two signals across time. In this example, when the same low-pass stimulus and response are compared in this manner (40 ms windows), the two signals are more similar during the steady-state portion, although the time displacement is consistent across time. [to listen to stimulus see Audio file, Supplemental Digital Content 7.wav] [to listen to response see Audio file, Supplemental Digital Content 8.wav]

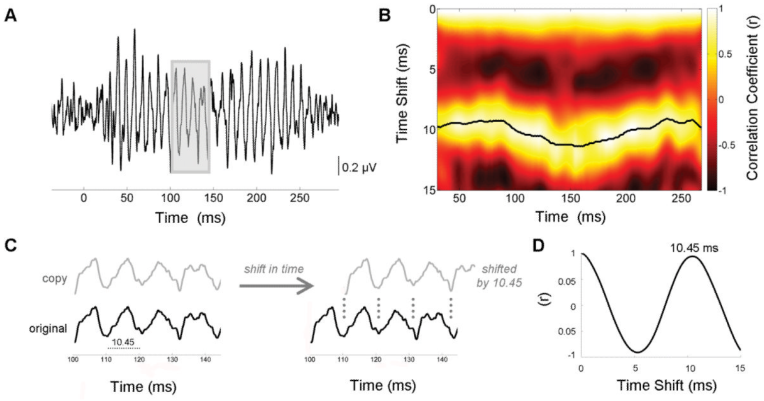

Figure 11. An illustration of frequency tracking by autocorrelation

By cross-correlating a cABR waveform with itself, the time interval between peaks can be determined. The frequency of the F_0 and other periodic aspects of the response, including the temporal envelope (Krishnan et al., 2004; Lee et al., 2009), can be derived from an autocorrelogram. This technique can also be used to calculate the strength of phaselocking to these features. **A.** In this example, a cABR to a syllable /mi/ with a dipping F_0 contour (Mandarin Tone 3; black line) is plotted. **B.** By applying the autocorrelation technique on 40 ms sliding-windows, a frequency contour can be tracked over time. Colors represent strength of correlation; white is highest. **C and D.** An illustration of cross-correlation performed of a single time window (100-140 ms; demarcated in A). When a copy of this window is shifted 10.4 ms, the first peak of the copy lines up with the second peak of the original (C). A correlogram (D) represents the degree of correlation as a function of the time shift. The highest correlation occurs at 10.4 ms; thus, the fundamental periodicity of this window is 1/10.4 ms or 96 Hz. The strength of correlation at 10.4 ms is r = 0.98, indicating a strong phaselocking to 96 Hz in this time window.

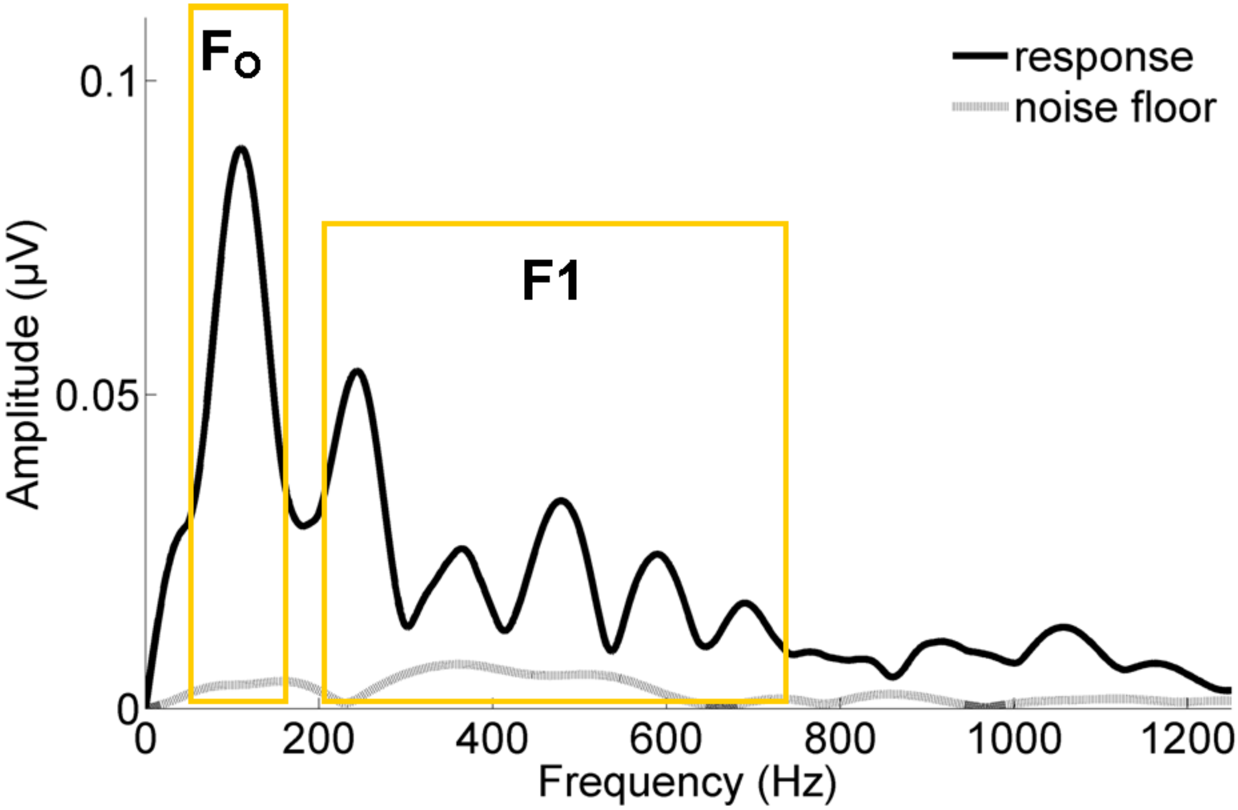

Figure 12. Fast Fourier analysis of a complex signal with time-varying features

This cABR was evoked by a 40-ms /dɑ/ sound comprising an onset stop-burst followed by a CV formant transition period. A frequency-domain representation of the FFR was generated using the fast Fourier transform (FFT). As a measure of phaselocking, spectral amplitudes are calculated over a range of frequencies corresponding to F_0 (103-125 Hz) and first formant (F1; 220 – 720 Hz). The noise floor is plotted in gray. The time-domain representation of this response is plotted in Figure 1.

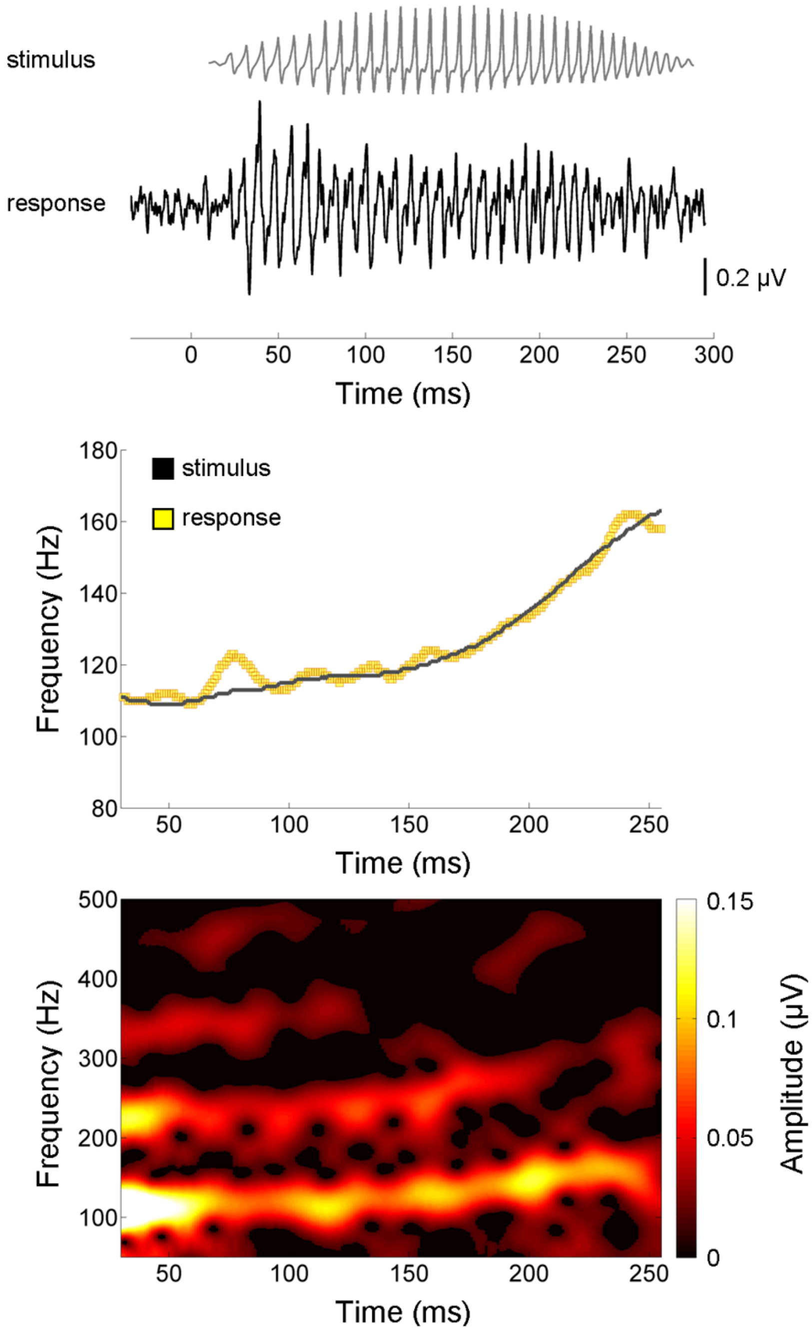

Figure 13. An illustration of frequency-tracking by STFT method

STFT is a method for examining frequency-tracking, that enables the tracking of the stimulus F_0 and harmonics **Top:** A cABR to a /mi/ syllable with a rising pitch contour (Mandarin Tone 2). The rising pitch is evident in the increasingly smaller inter-peak intervals in the stimulus and response over time. **Middle:** The estimated response F_0 (yellow) contour is plotted against the known F_0 of the stimulus (black). Each point represents the spectral maximum within a single 40-ms window of a sliding-window STFT analysis. The precision of phaselocking can be measured by calculating the frequency error between the stimulus and response trajectories ($^{\text{Wong et al. 2007}}$; Russo et al. 2008). **Bottom:** Plotting the resulting spectrogram of the STFT procedure enables a visualization of the response's tracking of F_0 and its harmonics. [to listen to stimulus see Audio file, Supplemental Digital Content 9.wav] [to listen to response see Audio file, Supplemental Digital Content 10.wav]

Table 1

Survey of speech stimuli used to evoke cABRs

Speech syllable	Examples	Publications
VOWELS		
Synthetic	/ɑ/, /**u**/	Krishnan 2002
Natural	/ ɛ/, /I/, / i/, /ɑ/, /æ/, /Λ/, /u/	Greenburg et al. 1980; Dajani et al. 2005, Aiken and Picton 2006; Aiken and Picton 2008
CONSONANT VOWELS SYLLABLES (CVs)		
Synthetic	/dɑ /	Cunningham et al., 2001; Plyler and Ananthanarayan, 2001; King et al. 2002; Wible et al. 2004; Russo et al. 2004; Wible et al. 2005; Russo et al. 2005; Kraus and Nicol 2005; Johnson et al. 2007; 2008; Banai et al. 2005; 2009; Burns et al. 2009; Chandarasekaran et al. 2009; Parbery-Clark et al. in press;
	/bɑ /	Akhoun et al. 2008a; Akhoun et al. 2008b
	bɑ -dɑ -gɑ continuum	Plyer and Ananthanarayan 2001; Johnson et al. 2008; Hornickel et al. 2009
Natural	Mandarin pitch contours	
	/yi/	Krishnan et al. 2005; Xu et al. 2006
	/mi/	Wong et al. 2008; Song et al. 2008
Hybrid	/yɑ/ with linearly rising and falling pitch contours	Russo et al. 2008

Table 2

Recommended stimulus and presentation parameters for cABRs

Parameter	Recommendation	Rationale/Comments
STIMULUS		
Type	speech, music, non-speech vocal sounds, environmental sounds, etc.	examine how behaviorally-relevant sounds are turned into neural code
Characteristics		
Transient	well-defined temporal features such as strong attacks and amplitude bursts	maximize transient responses
Sustained	$F_0 < 500$ Hz	maximize sustained responses
Creation	natural, synthetic, or hybrid	cABR stimuli can be created with many different software packages
Duration	short: 40 ms to 100 ms long: 100 ms to 500 ms	minimizes recording time maximizes naturalness
STIMULUS PRESENTATION		
Intensity	well above threshold: 60-80 dB SPL	stimuli should be precisely calibrated before each test session using a sound level meter
Monaural Stimulation	separate norms should be collected for each ear	monaural is preferred for children
Binaural Stimulation	maximizes response characteristics	binaural is more realistic than monaural
Transducer	magnetically shielded ear inserts	minimizes stimulus artifact
Rate and ISI	rate: dependent on stimulus duration ISI: >= 30% of stimulus duration	see table 3 for recording-based issues that impact rate and ISI decisions
Presentation Software	perform thorough testing to ensure precise, non-jittered stimulus presentation	because of the temporal sensitivity of the cABR, a small amount of jitter will spoil the response

Table 3

Recommended cABR recording parameters

Parameter	**Recommendation**	**Rationale/Comments**
Electrodes	vertical montage (active Channel: Cz; reference: earlobe(s); ground: forehead)	for rostral brainstem recordings; a horizontal montage is used for recording more peripheral structures
Sampling Rate	6000-20000 Hz	better temporal precision with higher sampling rates
Filtering	low pass cutoff: 2000-3000 Hz	more defined transient peaks
	high pass cutoff: 30-100 Hz	depends on spectral characteristics of stimulus
	if possible, collect cABR with open filters (1-3000 Hz) and bandpass filter offline using digital filters	digital filters minimize temporal phase shifts
Signal Averaging	2 or more sub-averages of 2000-3000 sweeps	determine response replicability spectral-domain averaging will increase spectral estimates and require fewer sweeps.
Averaging Window	begin 10-50 ms before stimulus onset	an adequate sample of the baseline is needed to determine whether a particular response peak is above the noise floor for running window analyses: the pre-stimulus time window should be > or = to the duration of the analysis window
	extend 10-50 ms after stimulus onset	neural activity should return to baseline
Simultaneous cABR-cortical response recording	recommended only if large files can be accommodated and longer sessions are appropriate	
Minimizing artifacts	passive collection protocol	minimizes myogenic artifacts
	electromagnetically shielded insert ear phones	minimizes stimulus artifact
	alternating stimulus polarity	enables adding of responses to minimize both stimulus artifact and cochlear microphonic.
	use electrically shielded test booth project movie into test booth	minimize electrical artifact
	artifact rejection criterion: > 20 μV	exclude trials exceeding typical neural response size.

Table 4

Methods for analyzing cABRs

Method	**Description**	**Comment**
TRANSIENT FEATURES		
peak latency and amplitude	delineation of transient response peaks	
SUSTAINED FEATURES		
RMS amplitude	global measure of magnitude	used to calculate SNRs
Fourier analysis	frequency domain representation	used to measure precision and magnitude of neural phaselocking at specific frequencies and frequency ranges amplitude and phase are recorded
Cross-Correlation	compares the timing and morphology of two signals examples: (1) stimulus-to-response (2) response-to-response	signal 1 is shifted in time relative to signal 2 to find the shift that produces the strongest correlation if the correlation coefficient (r) = 1, the signals are identical. If r = 0, the signals are completely dissimilar
Autocorrelation	a signal is cross-correlated with itself	can be used to calculate the time-displacement between two signals used to find (1) repeating patterns in signals such as phaselocked activity to the F_0 and the amplitude envelope (2) strength of phaselocking
Sliding window analyses	small time bins (i.e., windows) of the signal are analyzed in succession to create a three-dimensional representation of the response (e.g. spectrograms and autocorrelograms)	used to evaluate and visualize how cABRs changes over time.