Objective Information-Theoretic Algorithm for Detecting Brainstem-Evoked Responses to Complex Stimuli

DOI: 10.3766/jaaa.25.8.2

Gavin M. Bidelman*†

Abstract

Background: The scalp-recorded frequency-following response (FFR), an auditory-evoked potential with putative neural generators in the rostral brainstem, provides a robust representation of the neurophysiologic encoding of complex stimuli. The FFR is rapidly becoming a valuable tool for understanding the neural transcription of speech and music, language-related processing disorders, and brain plasticity at initial stages of the auditory pathway. Despite its potential clinical and empirical utility, determining the presence of a response is still dependent on the subjective interpretation by an experimenter/clinician.

Purpose: The purpose of the present work was to develop and validate a fully objective procedure for the automatic detection of FFRs elicited by complex auditory stimuli, including speech.

Research Design: Mutual information (MI) was computed between the spectrographic representation of neural FFRs and their evoking acoustic stimuli to quantify the amount of shared time-frequency information between electrophysiologic responses and stimulus acoustics. To remove human subjectivity associated with typical response evaluation, FFRs were first simulated at known signal-to-noise ratios using a computational model of the auditory periphery. The MI at which model FFRs contained +3 dB Signal-to-noise ratio was taken as the criterion threshold (θ_{MI}) for the presence of a response. θ_{MI} was then applied as a binary classifier on actual neurophysiologic responses recorded previously in human participants (n = 35). Sham recordings, in which no stimulus was presented to participants, allowed us to determine the receiver operating characteristics of the MI metric and the capabilities of the algorithm to segregate true evoked responses from sham recordings.

Results: Results showed high overall accuracy (93%) in the metric's ability to identify true responses from sham recordings. The metric's overall performance was considerably better than trained human observers who, on average, accurately identified only ~75% of the true neural responses. Complementary results were found in the metric's receiver operating characteristic test performance characteristics with a sensitivity and specificity of 97% and 85%, respectively. Additionally, MI increased monotonically and was asymptotic with increasing trials (i.e., sweeps) contributing to the averaged FFR and, thus, can be used as a stopping criteria for signal averaging.

Conclusions: The present results demonstrate that the mutual information between a complex acoustic stimulus and its corresponding brainstem response can provide a completely objective and robust method for automated FFR detection. Application of the MI metric to evoked potential speech audiometry testing may provide clinicians with a more robust tool to quantitatively evaluate the presence and quality of speech-evoked brainstem responses ultimately minimizing subjective interpretation and human error.

Key Words: Auditory event-related potentials, auditory nerve modeling, automated auditory brainstem response, evoked potential classification, frequency-following response

Abbreviations: ABR = auditory brainstem response; AEP = auditory-evoked potential; AN = auditory nerve; CF = characteristic frequency; EEG = encephalogram; ERP = event-related potential; FFR =

*Institute for Intelligent Systems, University of Memphis, Memphis, TN; †School of Communication Sciences & Disorders, University of Memphis, Memphis, TN

Gavin M. Bidelman, Ph.D., School of Communication Sciences & Disorders, University of Memphis, 807 Jefferson Ave., Memphis, TN 38105; E-mail: g.bidelman@memphis.edu

Portions of this work were presented at the 35th Annual Meeting of the Association for Research in Otolaryngology in San Diego, CA, 2012.

frequency-following response; MI = mutual information; PSTH = post-stimulus time histogram; ROC = receiver operating characteristic; SNR = signal-to-noise ratio

INTRODUCTION

Objective measures of audiologic function, e.g., evoked potential audiometry (Picton et al, 1977), are the backbone of current newborn hearing screening standards and neurotologic observation (e.g., intraoperative monitoring). Auditory-evoked potentials (AEPs) are neuroelectric brain responses recorded at the scalp that reflect the summed activity of various nuclei along the ascending auditory pathway. In particular, the auditory brainstem response (ABR) has become a routine measure used to noninvasively assess the functional integrity of auditory structures and estimate hearing sensitivity in both adults (Davis and Hirsh, 1976; Stapells, 2000) and newborns (Sininger et al, 1997; Berg et al, 2011). Importantly, these neurophysiologic tests offer a means to probe the hearing mechanism without an overt patient response. As such, objective audiometric techniques provide an ideal means to obtain diagnostic measures on difficult-to-test or uncooperative patients (e.g., infants). Indeed, current newborn hearing screening guidelines recommend the use of electrophysiologic measures such as the ABR to gauge auditory neural function and to identify abnormal auditory processing before infants leave the neonatal nursery (American Academy of Pediatrics, 2007). Early detection of hearing impairment and threshold estimation via evoked potential audiometry can be used to guide initial fitting of amplification in very young infants (Stapells, 2000). Having objective tools to detect and evaluate these AEPs is therefore critical in maximizing potential treatment benefits and long-term language outcomes (Carney and Moeller, 1998).

Rapid treatment intervention rests critically on the ability to detect and validate the presence of the AEP. Unfortunately, traditional recording practice requires that responses be identified by the subjective interpretation of a human observer. With use of ABRs for example, hearing thresholds are typically estimated based on visual waveform inspection whereby an operator identifies the lowest stimulus intensity, yielding a reproducible wave V complex such that the component is just distinguishable from background noise (Picton, 2010). Although visual waveform inspection remains the conventional "gold standard" in evoked potential analysis, the objectivity of AEP measures is all but lost by the subjective nature of conventional human validation (Vidler and Parkert, 2004; Bogaerts et al, 2009). To address this issue, recent methodologies have been developed in an attempt to remove the subjectivity and bias of human observers in order to provide a "hands-off" approach to response detection (for review, see Sininger, 1993; Hyde et al, 1998). Information regarding the overall waveform shape, amplitude, and morphologic pattern can be used in either template matching or cross-correlation procedures to assess response reliability (Coppola et al, 1978; Picton et al, 1983). In these approaches, the degree of similarity is computed between a running AEP average and known response template (e.g., average AEP of multiple individuals). Given the deterministic nature of the AEP, higher correlations are indicative of a present response. Other techniques (e.g., "split-buffer" or "± reference" averaging) compare the difference in running AEP averages between two recording traces and, thus, provide an estimate of the residual electroencephalogram (EEG) noise/signal-to-noise ratio (SNR) with increasing number of stimulus sweeps (Picton et al, 1983). A response is then assumed to be present when the running average exceeds a predefined stopping rule (e.g., SNR >3 dB).

The most widely known approach under these techniques is the F_{sp} ("single point F") metric, a statistic that considers the ratio between signal variance at a given sample in the AEP trace to the variance of the background noise (Elberling and Don, 1984; Sininger, 1993; Özdamar and Delgado, 1996). The rationale behind the F_{sp} is that the AEP is a deterministic signal whose amplitude is (theoretically) constant across trials. In contrast, background noise of the EEG is assumed to be stochastic. Hence, the greater the random EEG noise, the greater the sweep-to-sweep variability in a given time-point's amplitude (Sininger, 1993). The F_{sp} is appealing in that its sampling distribution can be quantified, thus allowing the operator or software to express response SNR statistically, and terminate averaging when a critical F_{sp} (and corresponding p-value) is exceeded. F_{sp} and similar automated ABR detection algorithms are currently available in most commercial clinical and research-grade AEP systems (e.g., Bio-Logic: *ABaer*; Interacoustics: *Eclipse*; IHS: *Smart Screener;* NeuroScan: *Curry 7).* Although these techniques have experienced success in the analysis of click-evoked ABRs, comparable objective algorithms have yet to be developed for other AEPs of interest to the clinical and research communities.

The human brainstem frequency-following response (FFR) is a sustained, scalp-recorded "neurophonic" potential that reflects phase-locked activity from a population of neural elements within the rostral brainstem (for reviews, see Krishnan, 2007; Chandrasekaran and Kraus, 2010; Skoe and Kraus, 2010). The response is characterized by a periodic waveform that follows the individual cycles of the evoking acoustic stimulus. Given its latency (~7–10 msec) and response characteristics (e.g., limit of phase-locking), putative neural generator(s) of the FFR likely include the inferior colliculus of the midbrain (Marsh et al, 1974; Smith et al, 1975;

Sohmer et al, 1977). Unlike more transient evoked AEPs (e.g., ABR, cortical P1-N1-P2), FFRs capture time-varying spectrotemporal properties of auditory stimuli (Krishnan et al, 2004; Bidelman and Krishnan, 2010) and, thus, offer a unique window into the neural encoding of complex sounds not afforded by traditional click responses. Indeed, the remarkable fidelity of the FFR is evident by the fact that the neural response to speech is intelligible when replayed as an audio stimulus (Galbraith et al, 1995). Given its level of detail, there is now considerable interest in the response as a means to probe dynamic sound processing at the level of the brainstem. Recent FFR studies, for example, have investigated subcortical representations of linguistic pitch prosody (for review, see Krishnan et al, 2012), melodic and harmonic aspects of music (for review, see Bidelman, 2013), and timbral aspects of speech and nonspeech sounds (Krishnan, 2002; Bidelman and Krishnan, 2010; Strait et al, 2012; Bidelman et al, 2013). These studies suggest that the FFR may offer a valuable tool for understanding the early neural transcription of behaviorally relevant auditory signals not afforded by traditional AEPs currently used clinically.

Recent studies have also demonstrated that brainstem FFRs are sensitive to auditory experience, as the response is differentially tuned dependent on a listener's language background (Krishnan and Gandour, 2009; Krishnan et al, 2012) and musical training (Musacchia et al, 2007; Wong et al, 2007; Bidelman et al, 2011). Experience-dependent changes in FFR (e.g., enhanced magnitudes, improved timing precision) have also been observed with short-term perceptual learning (Song et al, 2008; Carcagno and Plack, 2011). Collectively, these studies suggest that FFRs may offer an objective method to monitor brain plasticity and behavioral improvements gained through auditory training regimens (Chandrasekaran et al 2012; Song et al, 2012; Skoe et al, 2013). Moreover, speech-evoked brainstem responses are impaired in children with language and learning disorders (Banai et al, 2007; Banai et al, 2009; Basu et al, 2010; Rocha-Muniz et al, 2012) despite normal click-ABRs (Song et al, 2006). These findings imply that the FFR might offer important diagnostic function not tapped by conventional audiologic evaluation. Together, studies suggest that the FFR response is suitable for investigating the neural transcription of speech and music, auditory processing disorders, and brain plasticity at subcortical levels of processing. Yet, despite its potential clinical and empiric usefulness, determining the presence of a response is still dependent on the subjective interpretation by an experimenter or clinician.

The primary aim herein was to develop and evaluate a fully objective and automated algorithm for the detection of brainstem FFRs elicited by complex stimuli (e.g., speech). We exploit the remarkable fidelity of the FFR and its ability to encode dynamic, spectrotemporal properties of acoustic stimuli. To this end, the current detection algorithm was developed not on response time-waveform inputs (as in the majority of automated methods) but rather on response spectrograms. Spectrograms are advantageous in that they provide a three-dimensional representation of the neural activity (time, frequency, amplitude) and, thus, a higher dimensionality of detail than the two-dimensional time-waveform alone. We adopt a well-known metric from information theory, namely mutual information (MI), to quantify spectral similarity between stimulus and FFR responses. MI is a well-established measure used in image processing applications and in medical image registration (e.g., aligning magnetic resonance images) to compute the correspondence between images (for review, see Pluim et al, 2003). Here, MI is used to compute the similarity between stimulus and response time-frequency representations (i.e., spectrographic images), allowing us to quantify FFR quality knowing only the input stimulus. To identify an adequate criterion for use of MI as a detection metric, we first validate the approach on simulated responses. Model-generated FFRs allow us to fully control response SNR and the presence/absence of the response that would have to be estimated and/or subjectively determined in actual evoked potential recordings. Parameterizing the SNR of model FFRs allows us to empirically determine a criterion threshold value for the detection metric under consideration. Currently, the success of an objective algorithm is evaluated by comparing its performance with "gold standard" human ratings (e.g., Jeng et al, 2011b). In the current study, determining a detection criterion from simulated data, rather than observer judgments, removes this layer of subjective confirmation. Finally, we validate the efficacy of the MI algorithm by assessing its ability to identify recorded FFRs (true biologic responses) from sham recordings (containing no biologic response). We also examine its potential as a stopping criterion for signal averaging, with the goal of improving FFR recording efficiency.

METHODS

Participants and Overview of Original FFR Data

The evoked potential data herein represent a subsample of recordings (n = 35) reported in our previous studies (Bidelman and Krishnan, 2010; Krishnan et al, 2010). Specific participant demographics and electrophysiologic recording protocols are reported in the original reports. Briefly, participants were young adults (age range: 19–35 yr) who spoke American English. All were nonmusicians (<3 yr musical training), were strongly right handed (Oldfield, 1971), exhibited normal audiometric hearing thresholds (i.e., ≤25 dB HL at octave frequencies between 250–4000 Hz), and reported no history of neurologic or psychiatric illness.

AEPs were originally recorded differentially using a vertical montage with an electrode placed on the high-

forehead at the hairline (~Fpz; noninverting) referenced to the right mastoid (A2; inverting reference). The evoking stimulus was a 250 msec synthetic vowel token (/i/) (Bidelman and Krishnan, 2010). The vowel consisted of a time-varying fundamental frequency (F0), which rose from 103–130 Hz for the duration of the stimulus and four steady-state formant frequencies (F1 = 300 Hz, F2 = 2500 Hz, F3 = 3500 Hz, F4 = 4530 Hz), held constant. Stimuli were presented to the right ear using ER-3 insert earphones (80 dB SPL) at a rate of 2.76/sec. Control of the original experimental protocol was accomplished by the Tucker-Davis (System III) BioSig module using a sampling rate of 24,414 Hz. Each response waveform represents an average of 3000 artifact-free trials over a 280 msec acquisition window. Response time-waveforms were filtered between 70–3000 Hz (6 dB/octave rolloff) to isolate the brainstem component of the EEG (Bidelman et al, 2013). Complete details of the original FFR recordings can be found in Bidelman and Krishnan (2010).

Simulated FFR Responses from a Computational Model

Model Description

We used a neurobiologically plausible computational model of the auditory nerve (AN) (Zilany et al, 2009; Zilany and Carney, 2010) to simulate brainstem FFRs (Dau, 2003). The details of this phenomenologic model are beyond the scope of the present report and are reviewed only briefly here. Essentially, the model takes an acoustic stimulus time-waveform at its input and for a given AN fiber with characteristic frequency (CF), outputs a realistic pattern of action potentials (i.e., spike trains). The current model represents the latest generation of a well-established model rigorously tested against actual physiologic AN responses to both simple and complex stimuli, including tones, broadband noise, and speech-like sounds (Zilany and Bruce, 2007). The model incorporates several important nonlinearities observed in the auditory periphery, including cochlear filtering, level-dependent gain (i.e., compression) and bandwidth control, and long-term adaptation, as well as two-tone suppression. Model tuning curves were fit to the CF-dependent variation in threshold and bandwidth for high-spontaneous rate fibers in normal-hearing cats (Miller et al, 1997). The stochastic nature of AN responses is accounted for by a modified nonhomogenous Poisson process, which includes effects of both absolute and relative refractory periods and captures the major stochastic properties of single-unit AN responses (e.g., Young and Barta, 1986).

Model FFR Generation

The AN model was used to simulate the scalp-recorded FFR using methodology described by Dau (2003) (Fig. 1A). This approach is based on the assumption that far-field event-related potentials (ERPs) recorded via scalp electrodes can theoretically be estimated as the convolution of an elementary unit waveform (i.e., impulse response) with the instantaneous discharge rate from a given auditory nucleus (Goldstein and Kiang, 1958; Dau, 2003). In the present implementation, 10 repetitions of the vowel /i/ (Bidelman and Krishnan, 2010) were used to evoke AN spike-trains. To realize the evoked potential response, spikes were generated from 300 model fibers (CFs: 80–11,000 Hz) to simulate neural activity distributed across the cochlear partition. Activity from the entire ensemble was then summed to form a population discharge pattern (i.e., post-stimulus time histogram [PSTH]) for the complete AN array. The population PSTH was then convolved with a unitary response function, simulating the impulse response of nuclei within the auditory brainstem (for details, see Dau, 2003). Pink noise (i.e., $1/f$ distribution) was then added to simulate the spectral density of the inherent random fluctuations in scalp-recorded EEG (Granzow et al, 2001; Dau, 2003). The resulting model waveform provides a mirror approximation of the spectrotemporal characteristics of true FFR potentials recorded in human listeners (Fig. 1B). Model FFRs provide an idealized waveform to validate the performance of a detection metric, as both the presence and quality (i.e., SNR) of the responses are known *a priori* and can be parametrically manipulated.

MI Detection Metric

For arbitrary inputs, MI is a dimensionless quantity (measured in bits) that measures the degree of shared information (i.e., mutual dependence) between two random variables. Said differently, it reflects the amount of information—or reduction in uncertainty—that knowing either variable provides about the other. More generally, for two random variables (A and B), MI is computed according to Equation 1:

$$MI(A,B) = \sum_{a\in A}\sum_{b\in B} p(a,b)\log\left(\frac{p(a,b)}{p(a)p(b)}\right) \qquad (1)$$

where $p(a,b)$ is the joint probability of A and B, and $p(a)$ and $p(b)$ are the marginal probabilities of A and B, respectively. In the specific case where A and B are two images (e.g., spectrograms), MI can be interpreted as the distance between the joint distribution of the images' grayscale pixel values $p(a, b)$ and the joint distribution for two independent images, $p(a)p(b)$. In other words, MI computes the *dependence* or *similarity* between the two images (Pluim et al, 2003).

Basic properties of MI are worth noting: (i) MI = 0 if and only if A and B are independent, i.e., they share no commonality; (ii) MI is strictly positive (i.e., MI $\geq$0); and (iii) MI is symmetric [i.e., $MI(A,B) = MI(B,A)$].

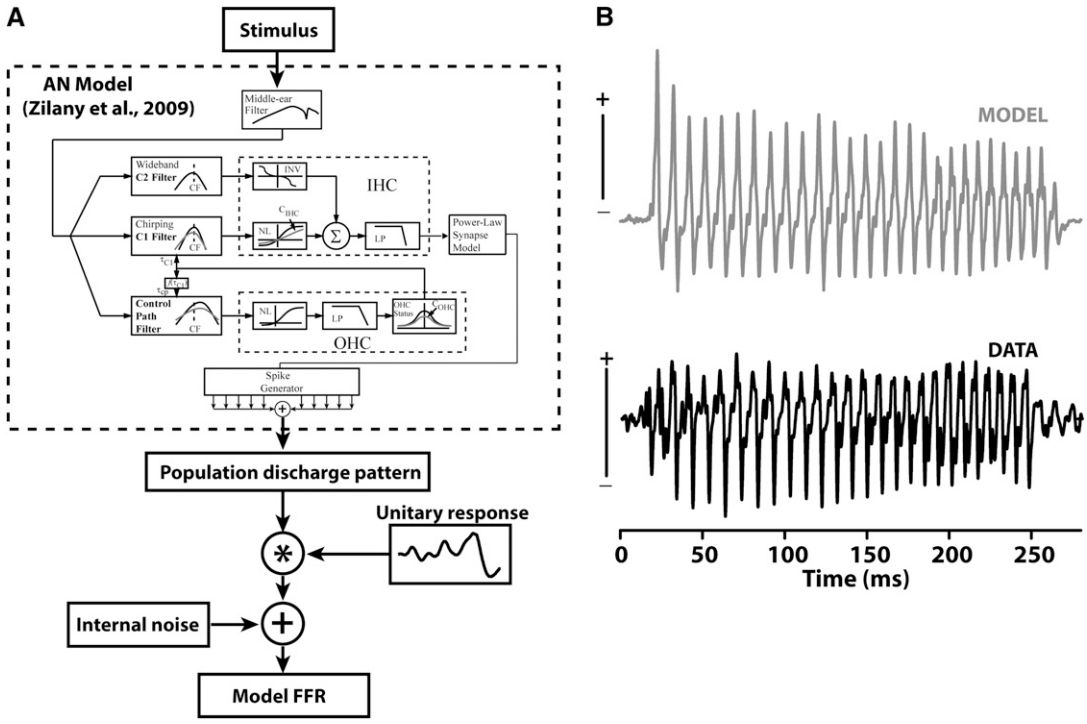

Figure 1. Computational model architecture used to simulate scalp-recorded FFRs. (*A*) The acoustic stimulus is input to a biologically plausible model of the auditory periphery (Zilany et al, 2009). The model provides a simulated realization of the neural discharge pattern for single AN fibers. After middle-ear filtering and hair cell transduction and filtering, action potentials are generated according to a nonhomogeneous Poisson process. Spikes were generated from 300 model fibers (CFs: 80–11,000 Hz) to simulate neural activity across the cochlear partition and summed to form a population discharge pattern (i.e., PSTH) for the entire AN array. Population PSTHs were then convolved with a unitary response function that simulates the impulse response of nuclei within the auditory brainstem (Dau, 2003). Additive noise then simulates the inherent random fluctuations in the scalp-recorded evoked potentials. (*B*) Comparison of model and actual FFRs recorded in human listeners (n = 35 ears; Bidelman and Krishnan, 2010; Krishnan et al, 2010). Note the remarkable similarity in the temporal phase-locking characteristics between actual and model data. Amplitude scale bars = 0.5 μV.

Intuitively, properties (i–ii) suggest that if a response spectrogram is entirely noise, MI will approach 0 whereas a robust response will tend to converge with the stimulus spectrogram, producing large positive values for MI. Property (iii) ensures that the MI computed between stimulus and response spectrograms is identical to that computed between the response and stimulus spectrograms (i.e., commutative property).

MI threshold Determination and Validation

Conceptually, MI was used in the present study to quantify the similarity between spectrograms of brainstem responses and that of the evoking acoustic stimulus. Thus, we compute the degree to which neural responses capture the collective dynamic, time-frequency characteristics of the stimulus. Model FFRs (see Figure 1B) were generated with SNRs between +25 and –25 dB (5 dB steps) by systematically adding broadband noise to the model waveform (Fig. 2). Additionally, clean (+∞ SNR) and 100% noise (–∞ SNR) responses were included to compute the upper and lower bound of the MI metric, respectively. Using model responses allowed us to simulate an evoked FFR at known SNRs, which would be impossible to accomplish with actual neurophysiologic recordings. Spectrograms were then computed for the stimulus and a given response and converted to grayscale images. Spectrograms were computed using the "spectrogram" routine in MATLAB 2013 (The MathWorks). This routine computed a 2^{14} point FFT in consecutive 50 msec segments (Hamming windowed) computed every 3 msec. Time waveforms were zero-padded to minimize edge effects and ensure that spectrograms ran to the end of the signal's duration. Identical parameters were used to compute both the stimulus and response spectrograms.

The information captured in the time-frequency characteristics of the FFR is limited (biologically) only by the rolloff of phase-locking at the level of the brainstem (<2000 Hz) (Liu et al, 2006). Thus, spectrogram images were limited to a frequency span of 2000 Hz in order to mimic this neurobiologic limit; spectral content beyond this range does not contain stimulus-related, time-locked neural energy and, thus, only adds noise to the spectrographic representation. MI was computed according to Equation 1 between the stimulus and each response spectrogram, allowing us to assess the change in the metric as a function of response SNR.

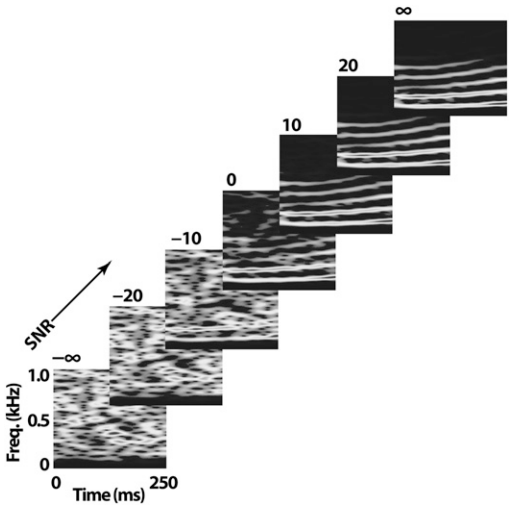

Figure 2. Model responses as a function of SNR. Varying-response SNR was used to determine an acceptable threshold for the detection metric empirically. Panels show model FFR response spectrograms evoked by a 250 msec synthetic version of the vowel /i/. Note the clear band of response energy near the voice F0 frequency (~100 Hz) and its integer-related harmonics. The decay in spectral amplitude above ~500 Hz reflects the rolloff of phase-locking within the brainstem, which acts as a quasi-lowpass filter. Broadband noise was systematically added to the clean-model FFR to yield a continuum of identical responses that differed only by a known SNR. SNRs of $+\infty$ and $-\infty$ represent tokens with 0% (i.e., clean) and 100% noise, respectively.

Importantly, we made no *a priori* assumptions regarding the resulting values of MI or an acceptable threshold for its use as a detection metric. To determine one empirically, an MI corresponding to a response of +3 dB SNR was taken as a criterion value (θ_{MI}). Similar objective criteria (i.e., response power ≥ twice power of surrounding noise) have been applied to other physiologic responses used in audiologic testing (e.g., otoacoustic emissions; Janssen and Müller, 2008). All processing and analyses were performed in MATLAB.

θ_{MI} Validation and Classifier Performance on Human FFR Data

θ_{MI} for Response Classification

Once determined empirically, θ_{MI} was applied as a binary classifier to previously recorded FFR responses obtained in n = 35 listeners (Bidelman and Krishnan, 2010; Krishnan et al, 2010). Additional sham recordings (n = 20) were used to assess the classifier's false-positive and misclassification rates. Sham recordings were identical to FFR recordings (e.g., Bidelman and Krishnan, 2010) with the exception that in those runs, the headphone was removed from the ear, thus preventing stimulus delivery to the participant but allowing the continued recording of EEG noise. Recordings yielding MI $\geq\theta_{MI}$ were classified as neural responses, whereas recordings with MI $<\theta_{MI}$ were considered to be noise (i.e., no response). Classifier performance was evaluated by computing standard metrics used in signal detection theory (d') and receiver operating characteristics (ROC), e.g., sensitivity and specificity. True-positive rate (i.e., sensitivity) was computed as the percentage of actual FFR recordings correctly identified by the metric; false-positive rate was computed as the percentage of sham recordings erroneously classified as a response. Test specificity was computed as 1 – false positive rate.

θ_{MI} as a Criterion for Determining an Adequate Number of ERP Trials

In addition to detection, an objective metric should be suitable as a stopping criterion for signal averaging. To test the efficacy of θ_{MI} in this application, we applied the metric to a second set of FFR data containing 2,000–3,000 individual trial epochs (Bidelman et al, 2013). MI was computed, as above, between the running ERP average response and stimulus spectrograms as each additional sweep was included in the average evoked potential. This allowed us to trace the growth of MI as a function of the trials contributing to the measured response and determine an appropriate stopping criterion to detect the brainstem FFR.

Classification Based on Human Observers

Current practice in evoked potential audiometry typically involves that the presence or absence of a neural response be determined by the subjective interpretation of a human observer. To compare this "gold standard" with the proposed objective MI algorithm, we asked three human observers to identify true biologic FFR responses in a closed set identification task. Each observer was well trained in the recording and analysis of AEPs and had extensive experience with the identification and interpretation of brainstem FFRs. Observers were shown a set of 55 spectrograms (e.g., Figure 2), 35 of which reflected time-frequency plots of true FFR responses (Bidelman and Krishnan, 2010; Krishnan et al, 2010) and 25 sham recordings, presented in random order. Although more information is usually available to judge responses (e.g., time-waveforms, replications, etc.), observers were required to make their decision based solely on spectrograms. This step ensured that human raters were given identical information as supplied to the MI algorithm. On each trial, observers viewed an image of a single spectrogram and were asked to record whether they believed that image represented a true response. As with the objective method, sensitivity and specificity (1 – false positive rate) were computed for each observer by calculating the percentage of FFR spectrograms correctly identified as responses (hits = FFR spectrogram classified as a neural response) and the number of

sham noise spectrograms incorrectly labeled as an FFR response (false alarms).

RESULTS

MI computed from model FFRs are shown as a function of response SNR in Figure 3. MI decreased monotonically from ~1.3 to ~0.3 bits with progressively noisier responses. Given the idealized nature of the simulated model waveforms, these values can be taken to represent the ceiling and noise floor of the metric, respectively, with an overall dynamic range of ~1 bit. As is shown in Figure 3, the point along the function corresponding to a response with +3 dB SNR corresponds to an MI of ~1. Whereas lower values still enable the detectability of a response, a +3 dB criterion represents a conservative threshold and one commonly applied in the literature to other physiologic data. For convenience, we adopt an MI = 1 as the criterion threshold for detecting the presence of a neural response. This empirically determined threshold (i.e., $\theta_{MI} = 1$) was subsequently used as a binary classifier for identifying true biologic responses from sham recordings.

Figure 4 shows MI computed from actual FFR recordings (Bidelman and Krishnan, 2010; Krishnan et al, 2010) as well as sham recordings—containing no biologic response. The $\theta_{MI} = 1$ criterion is denoted by the shaded region. Values exceeding this threshold are predicted to be true responses. Of the n = 55 total observations (n = 35 FFR versus n = 20 shams), only four recordings (7.3%) are misclassified. Responses are highly distinguishable from their sham counterparts based on MI (Fig. 4B). Average MI for true responses was 1.12 ± 0.06 bits, whereas MI of sham recordings was 0.93 ± 0.05 bits. An independent-samples t-test (two-tailed) revealed a significant difference in MI computed from responses relative to sham recordings [$t_{53} = 11.21$, $p < 0.001$].

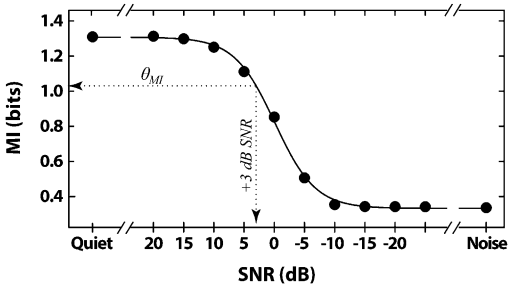

Figure 3. MI as a function of response SNR. MI was computed between the spectrograms of model FFRs (see Fig. 1) and their corresponding input stimulus to quantify the similarity in time-frequency representations. MI measures the degree to which evoked responses preserve the spectrotemporal properties of the evoking acoustic stimulus. Higher values indicate that brain responses preserve more acoustic features of the evoking stimulus. MI increases monotonically with increasing response SNR. The MI (θ_{MI}) at which responses contained +3 dB SNR (dotted lines) was taken as a criterion threshold for reliably detecting the presence of an FFR response.

Classifier performance for common signal detection and classification metrics is given in Table 1 and ROC, in Figure 5. Overall, the MI metric yielded 93% accuracy (7% misclassification) with a corresponding sensitivity and specificity of 97% and 85%, respectively. In comparison, the subjective judgments of trained human observers achieved, on average, only 75% accuracy (25% misclassification). These values are corroborated by the metric's overall sensitivity, which was notably larger ($d' = 2.94$) compared with that of human observers ($d' = 1.39$). Response bias for the MI metric was slightly larger ($c = –0.43$) relative to human judgments, who were largely unbiased in their decisions ($c = –0.03$).

Figure 6 shows the time-course of MI with increasing number of trials contributing to the ERP signal average. As denoted by the exponential fits, MI tends to grow monotonically with additional trials and asymptotes as the running average stabilizes. The function's break point (i.e., time-constant) provides an estimate of the rate of growth in MI and corresponds roughly to where the function exceeds the criterion threshold ($\theta_{MI} = 1$). The probability density function for MI rate of growth is shown in the figure's inset. Across recordings, θ_{MI} is reached by ~1500–2000 sweeps, indicating that the brainstem FFR can be adequately detected in this number of trials.

DISCUSSION

The current work presents a novel algorithm for an objective identification of brainstem FFR responses. The metric, based on the MI between neural responses and stimulus spectrograms, assesses the degree to which the evoked potential carries information of the eliciting acoustic signal. A threshold criterion for MI ($\theta_{MI} = 1$)—corresponding to a signal SNR of +3 dB—was determined empirically from a training set of model (simulated) data and then was validated on a larger test set of actual data recorded from normal-hearing listeners. The metric's performance was determined by how well it segregated true biologic responses from sham noise recordings containing only EEG noise. Overall, the proposed metric achieved superior performance, yielding >90% accuracy and equally impressive sensitivity and specificity (97% and 85%, respectively). This level of performance represents a considerable improvement from typical "gold standard" visual waveform inspection because trained human observers achieved, on average, only 75% accuracy in identifying responses based on their subjective interpretation of FFR spectrograms (Table 1). Lastly, it was shown that MI metric increases monotonically with an increasing number of stimulus presentations (i.e., trials). θ_{MI} was achieved in ~1500

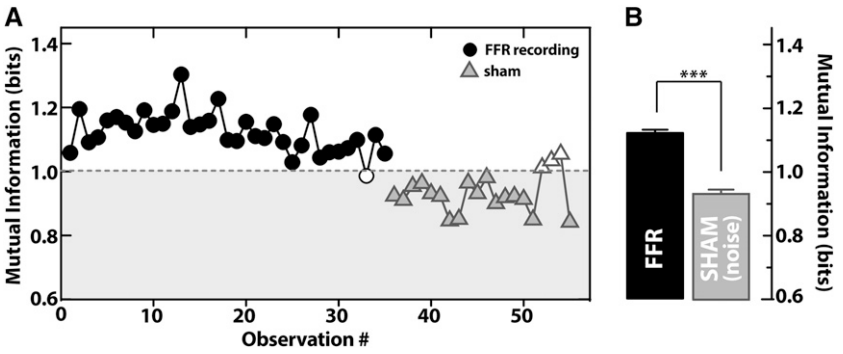

Figure 4. MI classifier performance distinguishing true brainstem FFRs from sham recordings. (*A*) MI computed from FFRs recorded from n = 35 listeners (Bidelman and Krishnan, 2010; Krishnan et al, 2010) (circles) and n = 20 sham recordings (triangles), in which the earphone was removed from the ear canal (i.e., no stimulus was presented). MI values below the empirically derived threshold value (θ_{MI} = 1; shaded region), are classified as noise (i.e., no response); values exceeding θ_{MI} = 1 are identified as true biologic responses. Misclassifications are shown as open symbols. (*B*) Average MI computed from true versus sham recordings. As denoted by the clear separation of recordings, the MI metric is robust in distinguishing actual evoked responses from EEG noise. Error bars = 1 s.e.m.; ***$p < 0.001$.

trials, indicating that brainstem FFR responses are detectable with adequate SNR within a few thousand stimulus sweeps. Thus, the metric may provide an objective criterion to terminate signal averaging, thereby saving valuable time in research and clinical EEG recordings.

It is worth noting potential limitations of the current metric. The negative, albeit small, bias in the objective metric (Table 1) indicates that it had a slight tendency toward false alarms. That is, sham recordings were sometimes classified as true responses (e.g., Figure 4A). However, we note that this bias is likely negligible in light of the metric's overall low misclassification rate (7.3%) and high accuracy (93%). The considerable improvement in detection performance of the objective method versus human observers suggests that some bias is tolerable in favor of its superior, automated detection ability.

MI as a Monitor of ERP Signal Quality

Normal intraparticipant variability in scalp-recording potentials, recording parameters (e.g., electrode impedance and placement), and subject factors (e.g., attentional state, body temperature) all contribute to a variation in brainstem response morphology. Thus, one of the fundamental issues in interpreting evoked potential recordings is that one can never truly be sure, with absolute certainty, whether the average waveform represents an actual response or merely some portion of residual background noise (Picton et al, 1983). SNR can be estimated by comparing signal amplitude after the time-locking event with amplitude within the prestimulus baseline, or via the residual between two averaging buffers (i.e., split-buffer or ± averaging). However, even these objective measures of SNR are difficult to measure and are plagued by the requirement of separately estimating "signal" and "noise" components, which can be difficult to partition. The proposed MI metric helps circumvent these issues in that it does not make *a priori* assumptions regarding whether poststimulus activity is "signal" or "noise" per se. Instead, the mutual correspondence between all of the stimulus and response areas are considered and signal quality is determined statistically.

Variation in MI among participants was well below the variation between recording classes, allowing adequate segregation of true responses from sham recordings (Fig. 4B). Yet, although the metric was used here as a successful all-or-nothing binary classifier, it should be noted that the measure is not entirely homogeneous across all listeners (e.g., Figure 4); the response MI ranged from ~1.0–1.3 bits. Slight disparity is expected in any classifier or diagnostic test and is likely attributable to inherent differences in ERP SNR. These differences could, for example, manifest with slight discrepancies in the number of trials, sensation

Table 1. Classifier Performance Characteristics*

	MI Algorithm†	Human Observers‡
Overall Performance		
Accuracy (%)	92.7	75.7 (32.5)
Misclassification Rate (%)	7.3	24.3 (32.5)
Signal Detection Metrics		
d'§	2.94	1.39
Bias**	−0.43	−0.03
ROC		
Sensitivity (%)	97.1	76.2 (28.9)
Specificity (%)	85	75 (39.1)

*n = 55 total observations (n = 35 neural FFR versus n = 20 sham noise recordings).

†Based on MI = 1.

‡Mean (SD) of n = 3 trained observers.

§Computed as d' = z(H) - z(FA) from mean hit (H) and false alarm (FA) rates.

**Computed as bias = –[z(H) + z(FA)]/2 from mean H and FA rates.

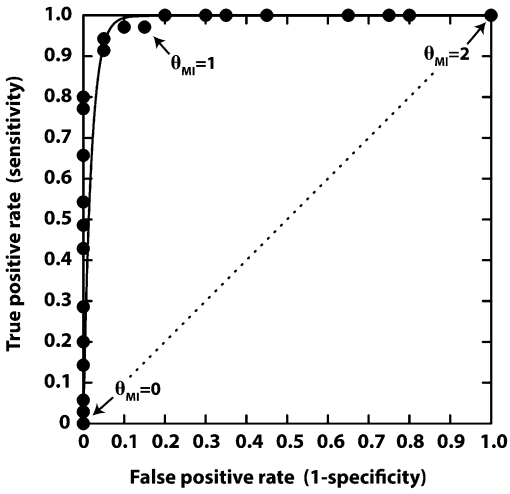

Figure 5. ROC space for the MI classifier. Individual points denote the true-positive (i.e., sensitivity) and false-positive (i.e., 1 – specificity) rates for various values of θ_{MI}. The dotted line corresponds with chance performance (i.e., $d' = 0$). With the empirically determined value of $\theta_{MI} = 1$ adopted here, the MI classifier achieves a $d' = 2.94$ and a sensitivity and specificity of 97.1% and 85%, respectively.

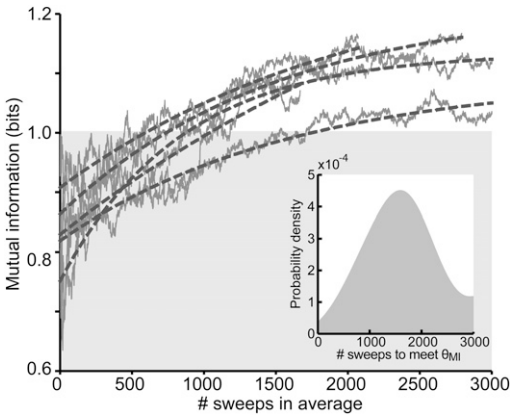

Figure 6. MI increases with the number of sweeps contributing to the evoked potential average. MI was computed between the evoking stimulus and corresponding FFR response for each new epoch added to the running average (thin solid lines). FFR data represent responses evoked by a 100 msec vowel /u/ presented at 83 dB SPL as originally reported in Bidelman et al (2013). MI increases monotonically, modeled as exponential growth, with additional averaging (dotted lines), consistent with the increase in response SNR with additional trials. Across a sample of $n = 5$ listeners, MI exceeds the criterion detection threshold ($\theta_{MI} = 1$) in ~1,500–2,000 sweeps, as denoted by the probability density function (inset). The monotonic increase of MI and the asymptotic nature of the metric indicate it can be used as stopping criteria for ERP averaging.

level of the stimulus, electrode contact impedance, and artifact specific to individual recordings. Nevertheless, the well-defined relationship between stimulus-to-response MI and signal SNR (Fig. 3) suggests that it may serve not only as a detection criterion, but also as a means to quantify signal quality for individual subjects. Moreover, applying the metric as an "online" threshold (Fig. 6), we demonstrate its potential as a stopping criterion for signal averaging. Brainstem FFRs elicited by dynamic speech sounds were detected in ~1500 sweeps with sufficient SNR (i.e., +3dB). Whereas brainstem responses evoked by other types of stimuli and presentation levels may require additional stimulus sweeps and signal averaging, using MI as a real-time monitor of signal quality could be used to maximize the efficiency of recording time.

Comparison with Other Objective Metrics

It is worth comparing the current measure with other automated detection algorithms used in electrophysiologic testing. Our MI metric bears resemblance to the F_{sp} statistic first introduced by Elberling and Don (1984) and applied to the detection of click-ABRs. However, unlike F_{sp}, which is a statistical ratio between signal and noise variance, MI is a measure of probability (i.e., commonality) between the stimulus and response spectrographic images. In this regard, our metric is somewhat akin to a stimulus-to-response correlation, in which Pearson's r is computed between ERP averages held in multiple response buffers or between subaverages taken from blocks of trials. The higher the r, the more likely the presence of a response as only the deterministic signal (i.e., ERP) replicates across trials; EEG noise is presumed to not correlate from trial to trial, although this later assumption can be easily violated (e.g., blinks time-locked to the stimulus presentation). These strict correlations are typically inappropriate in the analysis of sustained neurophysiologic responses such as the FFR. Given its periodic nature, small amounts of latency jitter between the neural response and the evoking stimulus can yield artificially low correlations[1]. Hence, we deliberately avoided correlational measures in the present work. Moreover, MI is preferred versus (cross-) correlation as it captures *all* dependencies between the stimulus and response rather than only second-order ones, as with correlation/covariance (Wang et al, 2010).

The difficulty in finding an appropriate detection metric for FFRs is complicated by the fact that there is no uniform evoking stimulus for the response. The potential can be elicited by any sustained, complex sound (for examples, see Krishnan, 2007; Skoe and

[1]The faulty nature of simple stimulus-to-response correlations is illustrated in the following exercise. Consider, two sinusoids (y_1 and y_2) with identical amplitude (A) and frequency (f) differing only by a small phase discrepancy (φ) [i.e., $y1 = A\sin(2\pi f)$ and $y2 = A\sin(2\pi f + \varphi)$]. For $f =$ 100 Hz, if we set the phase φ to equal to the approximate transmission delay of the brainstem FFR (~8 msec), the corresponding phase shift is 252°. Assuming $y1$ and $y2$ represent the stimulus and elicited FFR waveforms, the resulting Pearson's correlation (r) between these two signals amounts to a mere $r = 0.3$. Thus, despite the fact that the response identically mirrors the stimulus (i.e., a "perfect response"), correlation proves to be an inadequate measure of this correspondence. To mitigate such discrepancies, a cross-correlation approach can be used where stimulus-to-response similarity is computed as a running function of the time-lag between the two waveforms (Galbraith and Brown, 1990).

Kraus, 2010). To date, the majority of FFR studies have used the response as a neural index of voice pitch processing. The overwhelming focus of much of this work has investigated brainstem encoding of linguistic pitch patterns as found in tonal languages of the world (e.g., Mandarin Chinese and Thai; Wong et al, 2007; Krishnan and Gandour, 2009; Krishnan et al, 2010; Bidelman et al, 2011; Krishnan et al, 2012). Indeed, the "following" nature of the response can provide an objective measure of how well the brainstem tracks time-varying linguistic and musical pitch patterns (Bidelman et al, 2011). With this mind, Jeng et al (2011b) recently developed and evaluated two novel algorithms for detecting human FFRs to voice pitch. Stimuli consisted of the four lexical tones of Mandarin Chinese that comprise level (Tone 1), rising (Tone 2), dipping (Tone 3), and falling (Tone 4) pitch contours. F0 was extracted from short-term autocorrelation and narrow-band response spectrograms to index temporal and spectral measures of pitch encoding, respectively (e.g., see Bidelman and Krishnan, 2010). A series of metrics (e.g., *pitch-tracking accuracy, pitch-noise ratio*) were used to quantify the magnitude and degree to which brainstem responses accurately captured the detail of each token's F0 information. Aggregated across stimuli and the four measures explored, the authors reported an average sensitivity/specificity of 72/88% and 82/78% for temporal and spectral pitch measures, respectively (see Table 1 and 2 of Jeng et al, 2011b).

These metrics may be promising, but they are limited by two factors. First, they assess the neural response to only a single stimulus feature, namely, F0 (i.e., "pitch"). Although applicable to lexical tones, such measures (e.g., "pitch tracking accuracy") are not as relevant for responses evoked by speech stimuli of nontonal languages (e.g., English) where pitch plays only a suprasegmental role. Thus, a detection metric that only evaluates pitch processing alone is likely to be of less diagnostic value than a metric that captures more global response characteristics, as with our MI procedure. The inadequacy of measuring only a single response feature is also evident when considering the metrics use as an online tool for signal averaging. In consideration of a voice pitch detection metric alone, roughly 3,000–4,000 stimulus trials are needed to reach asymptotic levels in performance and detect the speech-evoked FFR (Jeng et al, 2011a). In contrast, our proposed MI metric, which considers a multidimensional set of spectral features (including those related to pitch), provides robust detection of the speech-FFR in roughly half as many trials (Fig. 6). Secondly, previous "objective" metrics have been validated by comparing an algorithm's classification to human judgments (Jeng et al, 2011b). Unfortunately, associating a test metric with a subjective interpretation reduces the impartiality desired in a fully objective procedure (as noted by Jeng et al, 2011b). Motivated by these inconsistencies, we pursued a fully objective approach to the development and validation of our MI detection metric. We empirically determined a threshold criterion from simulated model data with known response and SNR, thereby removing the subjective confirmation of observer judgments. Taken together, the larger generality and higher objectivity of the proposed MI metric make it a robust method for AEP evaluation.

CONCLUSIONS

A novel metric was developed to objectively detect brainstem FFRs elicited by complex auditory stimuli. The measure is based on the MI between the spectrograms of the neural response and its eliciting acoustic stimulus. An adequate threshold for response detection was determined empirically based on simulated model responses produced at known SNRs. Applying the criterion ($\theta_{MI} = 1$) as a decision metric, we show that the algorithm achieves 93% accuracy in classifying true FFR responses from sham recordings. ROC test performance characteristics indicated a sensitivity and specificity of 97% and 85%, respectively. We infer that the MI between an acoustic stimulus and its corresponding brainstem representation can provide a completely objective and robust method for automated FFR detection. Application of the current metric to evoked potential audiometry testing may provide clinicians with a more robust tool to objectively evaluate the presence and quality of speech-evoked electrophysiologic responses. The metric and its computation could be easily incorporated into most commercially available ABR/AEP systems similar to other statistical detection metrics (e.g., F_{sp}). This objective index can also be used to decrease valuable recording time in both research and clinical settings by providing a termination rule for ERP signal averaging.

Acknowledgments. The author thanks Drs. Chris Smalt and Saradha Ananthakrishnan for their assistance in manual waveform inspection. A MATLAB implementation of the MI algorithm is available from the author upon request.

REFERENCES

American Academy of Pediatrics, Joint Committee on Infant Hearing. (2007) Year 2007 position statement: Principles and guidelines for early hearing detection and intervention programs. *Pediatrics* 120(4):898–921.

Banai K, Abrams D, Kraus N. (2007) Sensory-based learning disability: Insights from brainstem processing of speech sounds. *Int J Audiol* 46(9):524–532.

Banai K, Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N. (2009) Reading and subcortical auditory function. *Cereb Cortex* 19(11): 2699–2707.

Basu M, Krishnan A, Weber-Fox C. (2010) Brainstem correlates of temporal auditory processing in children with specific language impairment. *Dev Sci* 13(1):77–91.

Berg AL, Prieve BA, Serpanos YC, Wheaton MA. (2011) Hearing screening in a well-infant nursery: profile of automated ABR-fail/OAE-pass. *Pediatrics* 127(2):269–275.

Bidelman GM, Krishnan A. (2010) Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res* 1355:112–125.

Bidelman GM, Gandour JT, Krishnan A. (2011) Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J Cogn Neurosci* 23(2):425–434.

Bidelman GM. (2013) The role of the auditory brainstem in processing musically relevant pitch. *Front Psychol* 4:264.

Bidelman GM, Moreno S, Alain C. (2013) Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage* 79:201–212.

Bogaerts S, Clements JD, Sullivan JM, Oleskevich S. (2009) Automated threshold detection for auditory brainstem responses: comparison with visual estimation in a stem cell transplantation study. *BMC Neurosci* 10:104.

Carcagno S, Plack CJ. (2011) Subcortical plasticity following perceptual learning in a pitch discrimination task. *J Assoc Res Otolaryngol* 12(1):89–100.

Carney AE, Moeller MP. (1998) Treatment efficacy: hearing loss in children. *J Speech Lang Hear Res* 41(1):S61–S84.

Chandrasekaran B, Kraus N. (2010) The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47(2):236–246.

Chandrasekaran B, Kraus N, Wong PC. (2012) Human inferior colliculus activity relates to individual differences in spoken language learning. *J Neurophysiol* 107(5):1325–1336.

Coppola R, Tabor R, Buchsbaum MS. (1978) Signal to noise ratio and response variability measurements in single trial evoked potentials. *Electroencephalogr Clin Neurophysiol* 44(2):214–222.

Dau T. (2003) The importance of cochlear processing for the formation of auditory brainstem and frequency following responses. *J Acoust Soc Am* 113(2):936–950.

Davis H, Hirsh SK. (1976) The audiometric utility of brain stem responses to low-frequency sounds. *Audiology* 15(3):181–195.

Elberling C, Don M. (1984) Quality estimation of averaged auditory brainstem responses. *Scand Audiol* 13(3):187–197.

Galbraith GC, Arbagey PW, Branski R, Comerci N, Rector PM. (1995) Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport* 6(17):2363–2367.

Galbraith GC, Brown WS. (1990) Cross-correlation and latency compensation analysis of click-evoked and frequency-following brain-stem responses in man. *Electroencephalogr Clin Neurophysiol* 77(4):295–308.

Goldstein MH, Kiang NYS. (1958) Synchrony of neural activity in electric responses evoked by transient acoustic stimuli. *J Acoust Soc Am* 30:107–114.

Granzow M, Riedel H, Kollmeier B. (2001) Single-sweep-based methods to improve the quality of auditory brain stem responses Part I: Optimized linear filtering. *Z Audiol.* 40(1):32–44.

Hyde M, Sininger YS, Don M. (1998) Objective detection and analysis of auditory brainstem response: An historical perspective. *Semin Hear* 19:97–113.

Janssen T, Müller J. (2008) Otoacoustic emissions as a diagnostic tool in a clinical context.Ineds In: Manley GA, Fay RR, Popper AN, eds. *Springer Handbook of Auditory Research: Active Processes and Otoacoustic Emissions*. New York, NY: Springer.

Jeng FC, Chung HK, Lin CD, Dickman B, Hu J. (2011a) Exponential modeling of human frequency-following responses to voice pitch. *Int J Audiol* 50(9):582–593.

Jeng FC, Hu J, Dickman B, et al. (2011b) Evaluation of two algorithms for detecting human frequency-following responses to voice pitch. *Int J Audiol* 50(1):14–26.

Krishnan A. (2002) Human frequency-following responses: representation of steady-state synthetic vowels. *Hear Res* 166(1-2): 192–201.

Krishnan A, Xu Y, Gandour JT, Cariani PA. (2004) Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res* 189(1-2):1–12.

Krishnan A. (2007) Human frequency following response. In: Burkard RF, Don M, Eggermont JJ, eds. *Auditory Evoked Potentials: Basic Principles and Clinical Application*. Baltimore, MD: Lippincott Williams & Wilkins.

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

Krishnan A, Gandour JT, Bidelman GM. (2010) The effects of tone language experience on pitch processing in the brainstem. *J Neurolinguist* 23(1):81–95.

Krishnan A, Gandour JT, Bidelman GM. (2012) Experience-dependent plasticity in pitch encoding: from brainstem to auditory cortex. *Neuroreport* 23(8):498–502.

Liu LF, Palmer AR, Wallace MN. (2006) Phase-locked responses to pure tones in the inferior colliculus. *J Neurophysiol* 95(3): 1926–1935.

Marsh JT, Brown WS, Smith JC. (1974) Differential brainstem pathways for the conduction of auditory frequency-following responses. *Electroencephalogr Clin Neurophysiol* 36(4):415–424.

Miller RL, Schilling JR, Franck KR, Young ED. (1997) Effects of acoustic trauma on the representation of the vowel "eh" in cat auditory nerve fibers. *J Acoust Soc Am* 101(6):3602–3616.

Musacchia G, Sams M, Skoe E, Kraus N. (2007) Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci USA* 104(40):15894–15898.

Oldfield RC. (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9(1):97–113.

Özdamar O, Delgado RE. (1996) Measurement of signal and noise characteristics in ongoing auditory brainstem response averaging. *Ann Biomed Eng* 24(6):702–715.

Picton TW, Woods DL, Baribeau-Braun J, Healey TM. (1977) Evoked potential audiometry. *J Otolaryngol* 6(2):90–119.

Picton TW, Linden RD, Hamel G, Maru JT. (1983) Aspects of averaging. *Semin Hear* 4:327–341.

Picton TW. (2010) *Human Auditory Evoked Potentials, 1 edn*. San Diego, CA: Plural Publishing.

Pluim JP, Maintz JB, Viergever MA. (2003) Mutual-information-based registration of medical images: a survey. *IEEE Trans Med Imaging* 22(8):986–1004.

Rocha-Muniz CN, Befi-Lopes DM, Schochat E. (2012) Investigation of auditory processing disorder and language impairment using the speech-evoked auditory brainstem response. *Hear Res* 294(1-2):143–152.

Sininger YS. (1993) Auditory brain stem response for objective measures of hearing. *Ear Hear* 14(1):23–30.

Sininger YS, Abdala C, Cone-Wesson B. (1997) Auditory threshold sensitivity of the human neonate as measured by the auditory brainstem response. *Hear Res* 104(1-2):27–38.

Skoe E, Kraus N. (2010) Auditory brain stem response to complex sounds: a tutorial. *Ear Hear* 31(3):302–324.

Skoe E, Krizman J, Spitzer E, Kraus N. (2013) The auditory brainstem is a barometer of rapid auditory learning. *Neuroscience* 243:104–114.

Smith JC, Marsh JT, Brown WS. (1975) Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalogr Clin Neurophysiol* 39(5):465–472.

Sohmer H, Pratt H, Kinarti R. (1977) Sources of frequency following responses (FFR) in man. *Electroencephalogr Clin Neurophysiol* 42(5):656–664.

Song JH, Banai K, Russo NM, Kraus N. (2006) On the relationship between speech- and nonspeech-evoked auditory brainstem responses. *Audiol Neurootol* 11(4):233–241.

Song JH, Skoe E, Wong PC, Kraus N. (2008) Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci* 20(10):1892–1902.

Song JH, Skoe E, Banai K, Kraus N. (2012) Training to improve hearing speech in noise: biological mechanisms. *Cereb Cortex* 22(5):1180–1190.

Stapells DR. (2000) Threshold estimation by the tone-evoked auditory brainstem response: A literature meta-analysis. *J Speech Lang Pathol Audiol* 42:74–83.

Strait DL, Chan K, Ashley R, Kraus N. (2012) Specialization among the specialized: auditory brainstem function is tuned in to timbre. *Cortex* 48(3):360–362.

Vidler M, Parkert D. (2004) Auditory brainstem response threshold estimation: subjective threshold estimation by experienced clinicians in a computer simulation of the clinical test. *Int J Audiol* 43(7):417–429.

Wang T, Bebbington MS, Harte DS. (2010) A comparative study of coherence, mutual information and cross-intensity models. *Int J Info Sys Sci* 6:49–60.

Wong PC, Skoe E, Russo NM, Dees T, Kraus N. (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 10(4):420–422.

Young ED, Barta PE. (1986) Rate responses of auditory nerve fibers to tones in noise near masked threshold. *J Acoust Soc Am* 79(2):426–442.

Zilany MS, Bruce IC. (2007) Representation of the vowel /ɛ/ in normal and impaired auditory nerve fibers: model predictions of responses in cats. *J Acoust Soc Am* 122(1):402–417.

Zilany MS, Bruce IC, Nelson PC, Carney LH. (2009) A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. *J Acoust Soc Am* 126(5):2390–2412.

Zilany MS, Carney LH. (2010) Power-law dynamics in an auditory-nerve model can account for neural adaptation to sound-level statistics. *J Neurosci* 30(31):10380–10390.