Journal of Neuroscience Methods 293 (2018) 59–66

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Sonification of scalp-recorded frequency-following responses (FFRs) offers improved response detection over conventional statistical metrics

Gavin M. Bidelman[a,b,c,*]

[a] School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA
[b] Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA
[c] Univeristy of Tennessee Health Sciences Center, Department of Anatomy and Neurobiology, Memphis, TN, USA

HIGHLIGHTS

- Describe method for detecting FFRs based on sonification (neural → acoustic).
- Compared sonification to objective metrics (*MI, SNR, Corr, F-test, MSC*).
- Sonifications: 2–3 x efficiency compared to statistical approaches.
- Listening to FFRs rapid way to monitor EEG and identify stopping point for averaging.

ARTICLE INFO

Article history:
Received 9 July 2017
Received in revised form 15 August 2017
Accepted 12 September 2017
Available online 14 September 2017

Keywords:
Data auralization
EEG detection algorithms
F-test
Mean-squared coherence (MSC)
Objective audiometry

ABSTRACT

Background: The human frequency-following response (FFR) is a neurophonic potential used to examine the brain's encoding of complex sounds (e.g., speech) and monitor neuroplastic changes in auditory processing. Given the FFR's low amplitude (order of nanovolts), current conventions in literature recommend collecting several thousand trials to obtain a robust evoked response with adequate signal-to-noise ratio.
New method: By exploiting the spectrotemporal fidelity of the response, we examined whether auditory playbacks (i.e., "sonifications") of the neural FFR could be used to assess the quality of running recordings and provide a stopping rule for signal averaging.
Results and comparison with existing method: In a listening task over headphones, naïve listeners detected speech-evoked FFRs within ~500 sweeps based solely on their perception of the presence/absence of a tonal quality to the response. Moreover, response detection based on aural sonifications offered similar and in some cases a 2–3× improvement over objective statistical techniques proposed in the literature (i.e., *MI, SNR, MSC, F-test, Corr*).
Conclusions: Our findings suggest that simply listening to FFR responses (sonifications) might offer a rapid technique to monitor real-time EEG recordings and provide a stopping rule to terminate signal averaging that performs comparably or better than current approaches.

1. Introduction

Auditory event-related brain potentials (ERPs) offer a means to probe neurophysiological function of the hearing mechanism without overt subject response. As such, ERPs are applicable to diagnostic testing particularly in difficult to test patients or assessing auditory and language development in infants. In this regard, frequency-following responses (FFRs) have proven particularly useful for understanding the neurobiological coding of complex speech and musical sounds, experience-dependent plasticity (for detailed reviews, see Bidelman, 2017; Kraus, 2017; Krishnan et al., 2012), and speech-language development (e.g., Anderson et al., 2015; Jeng et al., 2011b). FFRs are neurophonic potentials that reflect sustained neural ensemble activity emitted predominately from the upper brainstem that mirror spectrotemporal properties of the acoustic stimulus (Bidelman, 2015a; Chandrasekaran and Kraus, 2010; Krishnan, 2007; Skoe and Kraus, 2010). FFRs provide a neural "fingerprint" of sound within the human EEG. Germane to the proposed experiments, we have recently shown that listeners'

* Correspondence to: School of Communication Sciences & Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN 38152, USA.
E-mail address: g.bidelman@memphis.edu

http://dx.doi.org/10.1016/j.jneumeth.2017.09.005
0165-0270/

perceptual discrimination of normal and degraded speech is predicted by the degree to which important speech cues (voice pitch, formants) are coded in their FFRs (Bidelman, 2016; Bidelman and Alain, 2015; Bidelman and Krishnan, 2010; Bidelman and Patro, 2016). These studies reveal that FFRs carry perceptual salient information of the speech signal and with high neural fidelity.

The majority of current FFR protocols recommend collecting several thousand sweeps (i.e., 2000–6000 trials) before response assessment (Bidelman, 2014, 2015b; Chandrasekaran and Kraus, 2010); (but see Skoe and Kraus, 2012; Tichko and Skoe, 2017). Although some of these recommendations are based on objective (statistical) detection techniques that assess sweep-to-sweep noise levels in the recordings (e.g., Bidelman, 2014; Jeng et al., 2011a), most are only rule-of-thumb conventions of different labs. In this regard, several statistics have been proposed in the literature to detect other classes of ERPs including the *F*-test and magnitude-squared coherence (MSC) (Champlin, 1992; Dobie and Wilson, 1996). Generally, these statistics become more powerful with increasing number of trials. As such, a stopping rule can be applied when a criterion value or significance level is achieved (e.g., $p < 0.05$). Such metrics are currently available in several commercial EEG/ERP systems. However, as we have recently argued (Bidelman, 2014; Bidelman and Bhagat, 2016), it remains unclear if these are the most optimal statistics for characterizing sustained potentials like the FFR; many of these metrics (e.g., *F*-test) are limited because they are applied to singular features of the stimulus (e.g., power at a single frequency bin of the spectrum). Consequently, they cannot be applied broadly to sustained potentials like FFRs as they are typically elicited by time-varying speech and are thus dynamic in the time-frequency plane.

A remarkable property of speech-evoked FFRs is that they are *intelligible* to human listeners when replayed as auditory stimuli (Galbraith et al., 1995). Our recent work has further shown that the intelligibility of FFRs supports the categorical perception of speech (Weiss and Bidelman, 2015). Moreover, enhancements in the neural representation for speech that result from experience-dependent plasticity (e.g., enhanced speech coding in musicians vs. nonmusicians) are also evident in aural sonifications of the FFR (Weiss and Bidelman, 2015). In our previous report, we converted neural FFRs to a speech sound continuum of audio playbacks. We then presented these "sonifications" to naïve listeners. Findings showed that individuals could accurately categorized speech sounds based on the phonetic cues that were transcribed by the neurophonic potential. This suggested that FFRs carry behaviorally relevant speech cues to the point they could be accurately identified by external observers. Although not our impetus at the time, we posited that listening to playbacks of FFRs might be a useful way to quickly evaluate recording quality. Building on the fact that human hearing is exquisitely sensitive to fine temporal and spectral acoustic cues (Moore, 2003), we reasoned that external observers simply listening to audio sonifications of the FFR might provide a novel means to detect the response and terminate signal averaging during data collection.

The goal of sonification itself is to reveal deeper insights into data through its translation into sound (Beans, 2017; Fitch and Kramer, 1994). The Geiger counter represents a classic example of sonification. One cannot see radiation. Yet, the conversion of radio activity into sound allows for immediate readout of radiological dose (Neuhoff, 2011). Data sonification is becoming increasingly common in the physical and medical sciences (for review, see Beans, 2017). Sonifying proteins has been achieved by assigning a musical chord to different amino acids (Carey, 2016) and analyzing long-term climate data by associating pitch changes with temperature measurements (Hansman, 2015). Sonification also has enjoyed several clinical applications. For example, Brown et al. (2015) studied the auralizations of pulse oximeter readings using pitch scaling to optimize clinical judgments of absolute levels of oxygen saturation. The detection and real-time correction of gait abnormalities via biofeedback has also been investigated in the context of Parkinson's disease where abnormal movement is easily heard via changes in the pitch and rhythmicity of sonified walking data (Schedel et al., 2016). Sonifications have also been used to identify abnormal brain tissue by listening to acoustic analogs of magnetic resonance images (MRIs) (Barrass and Kramer, 1999). More recently, sonifications of electrical brain activity (i.e., EEG) have also been used for rapid seizure detection in epilepsy patients (Baier et al., 2007; Loui et al., 2014).

With these motivations in mind, the current study investigated the use of sonification as a novel approach to detect and monitor FFR recordings. We recorded neural FFRs to speech sounds in human listeners and then presented sonifications (aural playbacks) of these responses to naïve observers as auditory stimuli. We then compared human judgments with objective detection methods (i.e., *SNR, MI, F-test, MSC, Corr.*) to determine whether listeners could detect sonified FFRs more efficiently during online averaging than established statistical approaches. We hypothesized that simply listening to FFRs might provide a more rapid decision criterion to terminate signal averaging, thereby saving valuable recording time in clinical or laboratory settings.

2. Methods and materials

2.1. Participants

FFRs were recorded in sixteen young adults (age: 22–30 years; $M \pm SD = 24.7 \pm 3.1$ years; 8 male, 8 female). All participants were students recruited at from the University of Memphis student body. All were right-handed ($72.4 \pm 49.7\%$ laterality; Oldfield, 1971), had normal hearing bilaterally (i.e., audiometric thresholds ≤ 25 dB HL from 250 to 8000 Hz; PTA = 5.9 ± 4.3 dB HL), and reported no previous history of neuropsychiatric illness. All were paid and gave written informed consent in compliance with a protocol approved by the Institutional Review Board at The University of Memphis.

2.2. Stimuli

A synthetic, steady-state vowel token/a/was created using a Klatt synthesizer (Klatt, 1980) implemented in MATLAB® 2015b (The MathWorks, Inc.) (cf. Bidelman and Yellamsetty, 2017). The stimulus duration was 100 ms including a 5 ms of $\cos^2$ ramping to minimize spectral splatter in the acoustic stimulus and onset components in the response. Vowel fundamental frequency (F0) was 150 Hz and formant frequencies (F1-F4) were 730, 1090, and 2350 Hz, respectively. We wanted to ensure that FFRs would be generated by brainstem (Bidelman, 2015a) rather than cortical sources (cf. Coffey et al., 2016). Because "cortical FFRs" are not recordable >100 Hz (Brugge et al., 2009), the choice of a high voice pitch (F0 = 150 Hz) in our stimuli helped ensure that FFRs were of subcortical origin.

Listeners heard 4000 repetitions of the speech token presented with fixed rarefaction polarity, delivered binaurally through ER-30 insert earphones (Etymotic Research) at 80 dB SPL (ISI = 50 ms). Extended acoustic tubing of these headphones (20 ft) eliminated stimulus artifact from overlapping neural responses (Bidelman and Howell, 2016; Campbell et al., 2012). We accounted for the acoustic delay of the headphones (~17.8 ms) in all analyses. The frequency response of the headphone coupled to the acoustic tubing resulted in a lowpass characteristic to the acoustic output with a cutoff frequency of ~2 kHz. This rolloff was corrected with a dual channel 15 band graphical equalizer (dbx EQ Model 215s; Harman) to achieve a flat frequency response through 4 kHz (Bidelman and Howell, 2016). It should be noted that this bandwidth was sufficient to transmit the most salient cues of our speech stimulus (F0, F1, and F2 frequencies) and cover the effective bandwidth of FFRs (∼1–1.5 Hz) (Chandrasekaran and Kraus, 2010). Stimulus presentation was controlled by a MATLAB® routed to a TDT RP2 interface (Tucker-Davis Technologies).

2.3. Electrophysiological recordings

Electrophysiological protocols were similar to our recent published reports on speech-FFRs (Bidelman, 2014, 2015b). Neuroelectric activity was recorded differentially between Ag/AgCl disc electrodes placed on the scalp at the high forehead (∼Fpz) referenced to linked mastoids (A1/A2) (mid-forehead electrode = ground). This montage highlights vertically oriented dipole sources from the brainstem (Bidelman, 2015a; Galbraith, 1994). Interelectrode impedance was ≤3 kΩ. EEGs were digitized at 10 kHz (SynAmps RT amplifiers; Compumedics Neuroscan) using an online passband of DC – 4000 Hz. EEGs were then epoched (0–150 ms window) and averaged in the time domain to derive FFRs to the speech stimulus. Sweeps exceeding ±50 μV were rejected as artifacts prior to averaging. FFRs were then bandpass filtered (80–2000 Hz) for response visualization and quantification.

2.4. FFR analysis

2.4.1. Sonification of FFRs

Auditory sonification stimuli were created by converting single-trial FFRs (neural responses) into digital audio files (Galbraith et al., 1995; Weiss and Bidelman, 2015). Specifically, WAV files were created from subaverages of the FFR where each subaverage contained the accumulation of all prior trials plus 15 additional sweeps. This parcellation resulted in 267 audio files (=4000/15 sweeps) per FFR recording with increasing clarity ranging from total noise (sonification #1 containing only 15 sweeps) to a clear sounding response (sonification #267 containing all 4000 sweeps). Note that this mimics the online averaging of FFR recordings and increasing response fidelity with accumulating sweeps (cf. Fig. 1). Each sonification was RMS normalized and filtered between 80 and 2000 Hz to equate the overall intensity and bandwidth and between sonifications (e.g., Weiss and Bidelman, 2015). This guaranteed listeners would judge responses based on the inherent quality of the signal (e.g., SNR) rather than trivial acoustic differences that might fluctuate sweep-to-sweep (e.g., level cues). The sonification procedure was repeated for each of the original FFR recordings (see Section 2.3) to create a collection of 16 different sonification conditions (i.e., one trial set for each participant's recordings).

We presented sonified FFRs to an additional $n = 8$ young, normal-hearing listeners who did not participate in the original electrophysical recordings. Although all had some experience with auditory ERPs, they were naïve to the purpose of the study and none were familiar with aural data sonification nor FFR recordings and their interpretation. On average, listeners had 4.1 ± 4.5 years of formal musical training. FFRs were presented over Sennheiser HD 280 circumaural headphones via a custom coded MATLAB® GUI. Listeners were told they would repeatedly hear auditory stimuli that sounded like noise with or without a faint tonal quality (i.e., pitch percept). FFR sonifications were presented continuously with an ISI = 50 ms to match the rate of the original FFR recordings. As successive sweeps were presented back to back, listeners were instructed to terminate the program when they clearly perceived (were confident they heard) a pitch. The sweep number at which the listener terminated playback (i.e., between 1 and 4000 trials) was logged for that FFR recording. Listeners then self-initiated the next block to begin playback of FFR sonifications from another set of recordings. This procedure was repeated for each of the 16 FFRs sonification sets. Logging the termination sweep number for each sonification condition per participant allowed us to compare the number of trials in neural FFRs needed to elicit an aural sensation of pitch.

2.4.2. Objective evoked potential detection metrics

To compare the efficacy of FFR sonification as a stopping rule for neural recordings, we compared it with four established time- and frequency-domain measures for objectively detecting auditory evoked potentials. Each metric was computed on a sweep-to-sweep basis to track its growth with an increasing number of FFR trials. Of interest was a comparison of the stopping point between approaches and whether FFR sonification would detect neural responses in a fewer number of averages.

2.4.2.1. MI. We used mutual information (MI) to compare neural FFRs to their evoking speech stimulus on a sweep-to-sweep basis (Bidelman, 2014; Bidelman and Bhagat, 2016). Briefly, MI is computed between the *spectrogram* (i.e., grayscale time-frequency images) of the stimulus and neural response. Identical parameters were used to compute both the stimulus and response spectrograms. MI is a dimensionless quantity (measured in bits), that provides a measure of the *similarity* between the two images that captures both linear and nonlinear dependencies between the stimulus and response (Bidelman, 2014; Pluim et al., 2003). For full details of the MI algorithm for detecting FFRs, the reader is referred to Bidelman (2014). We used a criterion MI value of MI = 1 bit, which corresponds to a response SNR of +3 dB (Bidelman, 2014).

2.4.2.2. Signal-to-noise ratio (SNR). Sweep-to-sweep SNR was computed as $10\log(\mathrm{RMS}_{\mathrm{FFR}}/\mathrm{RMS}_{\mathrm{noise}})$ (cf. Hu et al., 2010), where $\mathrm{RMS}_{\mathrm{FFR}}$ and $\mathrm{RMS}_{\mathrm{noise}}$ are the RMS amplitudes of the FFR (signal) portion of the epoch window and pre-response baseline period (i.e., 0–10 ms) containing only neural noise, respectively. [*Note*: the acoustic delay of the ER-30 meant that FFRs were initiated ∼17.8 ms after the stimulus trigger at $t = 0$; the initial 10 ms pre-response period contained only EEG noise]. The criterion SNR value for FFR detection was set at SNR = + 3 dB to parallel MI.

2.4.2.3. Mean-squared coherence (MSC). Sweep-to-sweep MSC was computed according the statistics described in Dobie and Wilson (1989). MSC is computed on subaverages of the response using the Fourier phase (φ) and amplitude (A) at the Fourier bin or interest (Champlin, 1992; Dobie and Wilson, 1989). Intuitively, MSC is equal to the power of the grand mean neural response divided by the mean power of the individual subaverages (Dobie and Wilson, 1989) and ranges from 0 (no response, total noise) to 1 (no noise, total response). We computed MSC at the Fourier bin corresponding to the F0 (150 Hz) of the response, since that was the dominant spectral component of the speech FFR (see Fig. 1). We used the conventional criterion value for the MSC = 0.632, corresponding to a significance level of $p = 0.05$ (Dobie and Wilson, 1989).

2.4.2.4. F-test. The *F*-test is another well-known statistical approach based on frequency domain analysis that provides a probability of whether a response is present at the stimulus frequency (Dobie and Wilson, 1996; John and Picton, 2000; Zurek, 1992). The underlying assumption is that in the spectral domain, response energy is localized to a frequency bin near the stimulus frequency; activity in adjacent bins contain only random noise with zero mean and equal variance distributed equally across the noise bins (John and Picton, 2000). As in *MSC*, we computed the *F*-test at the response F0. The ratio of signal power to the powers in N adjacent frequency bins is distributed according to a F distribution with 2 and $2N$–1 degrees of freedom (John and Picton, 2000). In the current study, we used $N = 12$ frequency

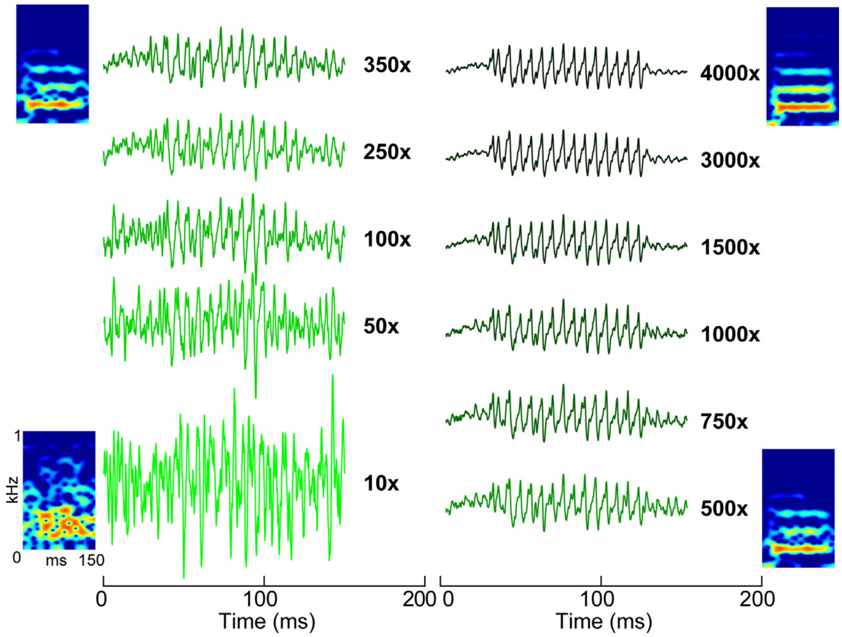

Fig. 1. Sweep-by-sweep changes in FFR time waveforms and spectrograms of grand average data. Initially, with only 10 trials contributing to the FFR average, temporal waveforms are dominated by EEG noise and spectrographic details of the response are unclear. FFRs converge to their characteristic periodicity with robust spectral harmonics (bands in the spectrogram) by ~350 sweeps and are fully stable by ~1000 sweeps as EEG noise diminishes in the running average according to $\sqrt{N}$.

bins (6 on either side of the signal). The *F*-ratio is then computed as in John and Picton (2000). Using $N = 12$ bins of the FFT, the *F*-ratio is compared to a critical *F*-value with 2 and 23 degrees of freedom. We used the conventional *F*-test criterion value of $p = 0.05$ (John and Picton, 2000) to evaluate the detection of FFRs on a sweep-by-sweep basis.

2.4.2.5. Correlation (Corr). We computed the sweep-to-sweep Pearson correlation between each listener's FFR temporal waveform and their averaged evoked response (i.e., 4000 trial average). This approach provided a running index of the temporal similarity between the developing response average and a template of the listener's FFR based on the final average. This correlation metric is akin to template matching techniques that assess response reliability (e.g., Coppola et al., 1978; Picton et al., 1983) and stimulus-to-response correlations (Bidelman, 2016; Bidelman and Krishnan, 2010; Lau et al., 2017; Parbery-Clark et al., 2009) or response consistency measures (Bidelman, 2016; Skoe et al., 2013a) reported in the literature. As there is no established criterion for correlation, we used a value of $r = 0.8$ which is considered a high (strong) positive association (Cohen, 1988; Hinkle et al., 2003).

3. Results

Fig. 1 illustrates grand averaged FFR response time waveforms and spectrograms with increasing number of sweeps in the running average. Initially, with few (e.g., 10 trials) contributing to the evoked potential average, FFR waveforms lack their characteristic periodicity and are dominated by physiological EEG noise. Similarly, with low trial count, spectrographic details of the response are unclear. FFRs converge to their characteristic periodicity and robust spectral harmonic structure (integer related bands in the spectrogram of the F0 = 150 Hz) by ~250 sweeps and are fully stabilized by ~1000 sweeps as EEG noise diminishes in the running average according to $\sqrt{N}$. By several thousand sweeps (where most FFR studies terminate averaging) the response has long fully developed and the salience of the harmonic structure is clear.

Sweep-by-sweep growth in the five objective (i.e., statistical) detection metrics and subjective sonification judgments by human listeners are shown in Fig. 2. Criterion values for each objective metric were based on published norms: *MSC*: 0.632 ($p = 0.05$) (Dobie and Wilson, 1989); *SNR*: +3 dB; *F*-test: $p = 0.05$ (John and Picton, 2000); *MI*: 1 (Bidelman, 2014); *Corr*: 0.8 (Cohen, 1988; Hinkle et al., 2003). The criterion for subjective sonifications was based on human judgments of the presence/absence of the neural response based on the pitch quality of the FFR when replayed as an audio stimulus. In each case, individual subjects' responses reach the stopping criterion at a different number of sweeps. Objective measures including *MI*, *SNR*, and *Corr* require over 1000 sweeps for terminating averaging and declaring the presence of the response. In contrast, based on sonifications, FFR averaging for the same responses can be terminated in only ~500 sweeps. These observations suggest that the stopping rule varied with test metric.

Fig. 3 shows the number of sweeps required to achieve a specified stopping criterion for signal averaging. A mixed model ANOVA (GLIMMIX; SAS® Institute, Inc) with subjects as a random factor and detection metric as the fixed, within-subject factor was conducted on the stopping sweep number. The dependent variable was sqrt-transformed to improve homogeneity of variance assumptions necessary for parametric statistics. Many participants' FFRs approached, but did not fully converge to the criterion threshold for the *F*-test. Consequently, this metric was not considered in the statistical analysis. The ANOVA revealed a significant effect of metric type on the stopping rule ($F_{4,53} = 6.74$, $p = 0.0002$). A planned contrast confirmed that sonification enabled the termination of averaging in upwards of half the number of sweeps compared to the objective statistical measures ($t_{53} = 2.79$, $p = 0.0073$). Post hoc Tukey-Kramer adjusted multiple comparisons further revealed that sonification was roughly 2–3 x as efficient as MI ($p = 0.0047$) and SNR ($p = 0.024$). Sonification was equally as efficient as the *Corr* ($p = 0.24$) and *MSC* metrics ($p = 0.99$). These findings confirm that detecting FFRs via sonification of the neurophonic response is as good or in some cases better than using objective statistical techniques.

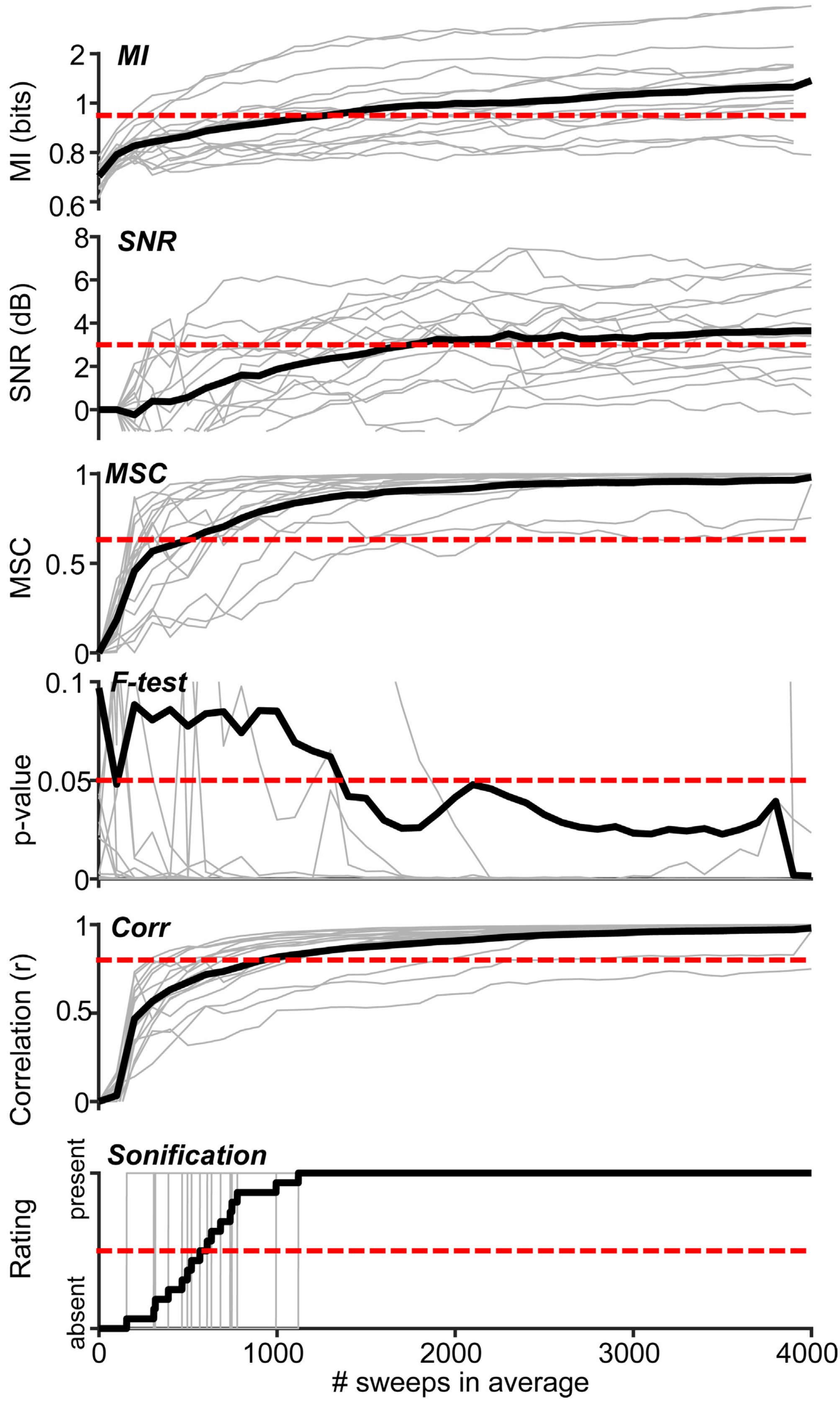

Fig. 2. Comparison of the growth in objective statistical metrics vs. sonification for detecting FFRs during online recording. (*top three panels*) Sweep-by-sweep detection based on objective statistical measures (*MI*, *SNR*, *MSC*, *F*-test, and *Corr*.). Dashed lines, significance thresholds for response detection [*MSC*: 0.632 (p = 0.05) (Dobie and Wilson, 1989); *SNR*: +3 dB; *F*-test: p = 0.05 (John and Picton, 2000); *MI*: 1 (Bidelman, 2014); *Corr*: 0.8 (Cohen, 1988; Hinkle et al., 2003)]. Light gray traces track the growth of individual subjects' FFRs; dark traces, grand averaged data. (*bottom panel*) Sonification results for detecting FFRs based on listeners' perception of audio playbacks of the neural responses. By ~500 sweeps, listeners perceive a clear tonal quality (pitch) to FFRs and declare the response is present.

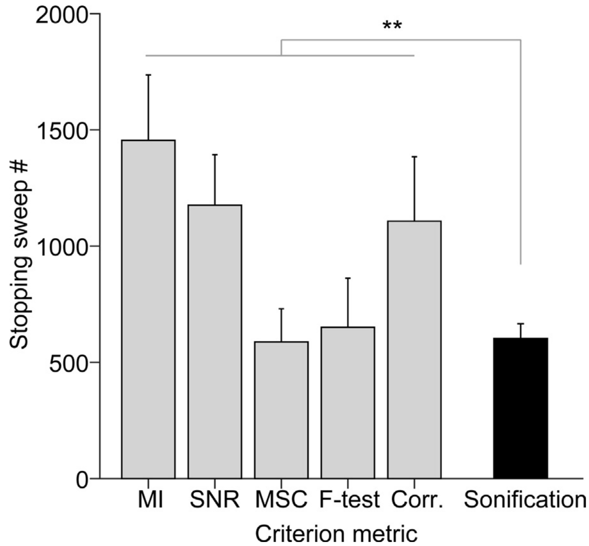

Fig. 3. Sonification yields an earlier stopping rule than objective statistical techniques during FFR signal averaging. Comparison of each detection metrics' stopping rule for signal averaging and identifying the presence of FFRs. Stopping criteria for objective statistics (gray) were based on a $p < 0.05$ (*MSC* and *F*-test), $r = 0.8$ (*Corr.*), and $MI = 1$ (MI); the criterion for sonification was based on the sweep at which human listeners subjectively perceived a tonal quality to acoustic playbacks of the FFR. Sonification is as efficient or in some cases more efficient (cf. *SNR*, *MI*) than statistical response detection approaches. error bars = ±1 s.e.m.; $^{**}p < 0.01$.

We used the intraclass correlation coefficient (ICC) to compare the absolute agreement of sonification judgments among (i) listeners and (ii) sonification stimuli (e.g., Bidelman et al., 2017b). The ICC is a normalized statistic akin to a correlation coefficient that extends to multiple (i.e., >2) observations of the same unit or group (Koch, 1982; McGraw and Wong, 1996). The ICC between listeners was 0.55 ($F_{15,105} = 2.44$, $p = 0.004$), whereas ICC between stimuli was 0.69 ($F_{7,105} = 3.66$, $p = 0.001$). These results indicate moderate agreement in listeners sonification ratings (Koch, 1982; McGraw and Wong, 1996) and even stronger consistency in the ratings of individual stimuli.

Intense auditory experiences (e.g., musicianship, tone-languages) are thought to enhance the detection of signals in noise (for review, see Bidelman, 2017). However, we did not find a correlation between raters' years of musical training and their detection of FFRs via sonification (Pearson's $r = -0.19$, $p = 0.65$). However, this might be expected given the minimal musical training of listeners in our cohort. Nevertheless, the fact that naïve listeners can use sound to detect neural responses is consistent with our previous report showing untrained listeners can exploit auditory cues within FFR sonifications (Weiss and Bidelman, 2015).

Previous studies have also shown that individual differences in FFR amplitude provide important clinical information that reveals supra-threshold temporal coding deficits in cases of learning, language, and cognitive impairments (Bidelman, 2017; Cunningham et al., 2001; Kraus, 2017; Rocha-Muniz et al., 2012). Thus, in addition to demonstrating that sonifications offer more time-efficient recordings, we aimed to ground our approach in these conventional FFR measures of individual differences. Correlational analyses revealed an association between the SNR of each FFR responses (i.e., 2000 sweep average) and the stopping sweep number determined via sonifications. These measures were negatively correlated ($r = -0.57$, $p = 0.022$), indicating that subjects with higher fidelity FFR recordings (i.e., larger SNR) sounded cleaner to external raters and were identified in fewer trials. No correlation was observed between sonification sweep number and F0 amplitude of the FFR ($r = -0.21$, $p = 0.42$). These results indicate that raters in this study most likely judged the quality of FFR playbacks based on their noise rather than signal level, *per se*. It should be noted however, that FFR quality can also be accurately gauged from sonifications even when overall response SNR is fixed (Weiss and Bidelman, 2015).

4. Discussion

By sonifying and presenting neural FFRs to human listeners, the present study shows that simply listening to these neurophonic potentials offers a straightforward way to monitor the quality of neurophysiological signals and determine the stopping point for online ERP averaging. Moreover, our results show that aural sonification offers similar and in some cases improvement over objective statistical techniques for ERP identification (i.e., *MI, SNR, MSC, F-test, Corr*). Sonification is probably most akin to the *MI* metric—which considers the joint time-frequency information contained in the spectrogram of the response (Bidelman, 2014)—because listeners presumably use all spectrotemporal cues available to them when judging the quality of FFR sonifications. Yet, FFRs were detected via sonification in ~3× fewer trials than *MI* and roughly half the number of trials as *SNR* and *Corr* metrics. Current FFR protocols recommend collecting several thousand sweeps (i.e., 2000–6000 trials) before response assessment (Bidelman, 2014, 2015b; Chandrasekaran and Kraus, 2010; Skoe and Kraus, 2010, 2012). Our findings thus indicate that FFRs might be reliably detected in far fewer trials (500 sweeps). Pragmatically, fewer trials may offer new avenues for stimulus paradigms that have otherwise precluded the use of FFRs to study attention or active behavioral responses during listening tasks (cf. Bidelman, 2015b). Translationally, monitoring the sound of FFRs might also offer more rapid collection of electrophysiological recordings in patient populations, thereby saving valuable time in clinical settings.

Sonification of FFRs is currently only achievable post hoc, after responses are recorded and translated to auditory stimuli. However, this process is relatively simple and would be straightforward to implement in commercial systems as an additional tool to monitor the ongoing EEG. Our findings indicate that sonification might offer an advantage over some well-known, "hands-off" statistical techniques currently used in ERP detection. While each metric has a different unit and measurement scale, we attempted to match their equivalence in terms of the quality of the FFR recording (i.e., matched signal to noise ratio in the case of *MI* and *SNR*). In this regard, the pattern of sweep-by-sweep growth (Fig. 2) offers a way to compare equivalencies between different approaches. In the present study, we used published criterion for each metric (e.g., $p < 0.05$ for *F*-test; +3 dB for SNR). However, changing this criterion value would offer a different receiver operating characteristic and detection of responses. Fig. 2 therefore offers a way to directly compare across measures and assess how different values map between metrics at the same "significance level."

We found that naïve listeners unexperienced with FFR sonifications accurately judged the sound quality of neural responses and detected them in only several hundred sweeps This was comparable to or better than the efficiency of some objective signal detection metrics. The use of untrained listeners demonstrates an important strength of our sonification approach: no formal training is needed to use sound playbacks for response assessment. This would be relevant in both research and clinical applications as experimenters and clinicians would not require extensive acclimatization to properly use auditory displays of online neural data. The fact that unskilled listeners are implicitly good at judging sonifications is consist with our previous studies demonstrating that the fidelity of speech-FFRs can be accurately judged by naïve listeners to the point they can be identified categorically (e.g., /u/vs./a/continuum; Weiss and Bidelman, 2015). Nevertheless, perceptual training can improve the identification of noise-degraded speech over time (e.g., Anderson et al., 2013). Previous studies also demonstrate listeners rapidly improve in differentiating seizures from non-seizure based on EEG sonifications within several trials of practice (Loui et al., 2014). In the present study, we did not find a correlation between raters' auditory experience (i.e., musical training) and their FFR judgments nor did we note within-rater improvements across the 16 sonification exemplars. However, our task was not designed to assess learning effects as stimuli were drawn from different recordings and raters only heard each sonification set once (i.e., no repetitions). Nevertheless, future studies could assess how short-term learning (explicit or implicit) and more "experienced" listeners hear and make use of sonification data. Presumably, raters might become more sensitive and less biased in their judgments (e.g., Loui et al., 2014), leading to even more efficient detection of FFRs than observed here.

FFRs have become popular for assessing neurophysiological function related to real-world listening skills in normal and impaired systems (for review, see Bidelman, 2017). For example, recent studies have also demonstrated aberrant FFR encoding of speech with language learning problems (Basu et al., 2010; Rocha-Muniz et al., 2012; White-Schwoch et al., 2015), increasing age and hearing loss (Anderson et al., 2012; Bidelman et al., 2014; Clinard and Tremblay, 2013), and mild cognitive impairments (Bidelman et al., 2017a)—among other factors. Collectively, previous studies suggest that the FFR might be used as an objective "biomarker" for identifying auditory-based impairments, monitoring disease state (Bidelman and Alain, 2017; Johnson et al., 2005), and tracking behavioral gains in learning tasks (Skoe et al., 2013b). We infer that auditory EEG displays, like the sonifications used here, might aid the early detection of speech-language impairments due to neurodegenerative disorders (Barrass and Kramer, 1999) and provide an objective index of speech perception abilities based solely on listening to individuals' brain activity (e.g., Weiss and Bidelman, 2015). In future work, it would be interesting to investigate whether sonifications can be used to identify normal vs. aberrant speech processing as is the case with nascent forms of cognitive impairment (Bidelman et al., 2017a) and children at risk for language-learning problems (White-Schwoch et al., 2015).

Acknowledgement

The author thanks Jessany Knapp and Bonnie Brown for their assistance in data collection.

References

Anderson, S., Parbery-Clark, A., White-Schwoch, T., Kraus, N., 2012. Aging affects neural precision of speech encoding. J. Neurosci. 32, 14156–14164.
Anderson, S., White-Schwoch, T., Parbery-Clark, A., Kraus, N., 2013. Reversal of age-related neural timing delays with training. Proc. Natl. Acad. Sci. U. S. A. 110, 4357–4362.
Anderson, S., Parbery-Clark, A., White-Schwoch, T., Kraus, N., 2015. Development of subcortical speech representation in human infants. J. Acoust. Soc. Am. 137, 3346–3355.
Baier, G., Hermann, T., Stephani, U., 2007. Event-based sonification of EEG rhythms in real time. Clin. Neurophysiol. 118, 1377–1386.
Barrass, S., Kramer, G., 1999. Using sonification. Multimedia Syst. 7, 23–31.
Basu, M., Krishnan, A., Weber-Fox, C., 2010. Brainstem correlates of temporal auditory processing in children with specific language impairment. Dev. Sci. 13, 77–91.
Beans, C., 2017. Science and Culture Musicians join scientists to explore data through sound. Proc. Natl. Acad. Sci. 114, 4563–4565.
Bidelman, G.M., Alain, C., 2015. Hierarchical neurocomputations underlying concurrent sound segregation: connecting periphery to percept. Neuropsychologia 68, 38–50.
Bidelman, G.M., Alain, C., 2017. Auditory biomarker identified for early cognitive impairment. Hear. J. 70, 18–20.
Bidelman, G.M., Bhagat, S.P., 2016. Objective detection of auditory steady-state evoked potentials based on mutual information. Int. J. Audiol. 55, 313–319.
Bidelman, G.M., Howell, M., 2016. Functional changes in inter- and intra-hemispheric auditory cortical processing underlying degraded speech perception. Neuroimage 124, 581–590.
Bidelman, G.M., Krishnan, A., 2010. Effects of reverberation on brainstem representation of speech in musicians and non-musicians. Brain Res. 1355, 112–125.
Bidelman, G.M., Patro, C., 2016. Auditory perceptual restoration and illusory continuity correlates in the human brainstem. Brain Res. 1646, 84–90.
Bidelman, G.M., Yellamsetty, A., 2017. Noise and pitch interact during the cortical segregation of concurrent speech. Hear. Res. 351, 34–44.
Bidelman, G.M., Villafuerte, J.W., Moreno, S., Alain, C., 2014. Age-related changes in the subcortical-cortical encoding and categorical perception of speech. Neurobiol. Aging 35, 2526–2540.
Bidelman, G.M., Lowther, J.E., Tak, S.H., Alain, C., 2017a. Mild cognitive impairment is characterized by deficient hierarchical speech coding between auditory brainstem and cortex. J. Neurosci. 37, 3610–3620.
Bidelman, G.M., Pousson, M., Dugas, C., Fehrenbach, A., 2017b. Test-retest reliability of dual-recorded brainstem vs. cortical auditory evoked potentials to speech. J. Am. Acad. Audiol., http://dx.doi.org/10.3766/jaaa.16167.
Bidelman, G.M., 2014. Objective information-theoretic algorithm for detecting brainstem evoked responses to complex stimuli. J. Am. Acad. Audiol. 25, 711–722.
Bidelman, G.M., 2015a. Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient ABR. Hear. Res. 323, 68–80.
Bidelman, G.M., 2015b. Towards an optimal paradigm for simultaneously recording cortical and brainstem auditory evoked potentials. J. Neurosci. Meth. 241, 94–100.
Bidelman, G.M., 2016. Relative contribution of envelope and fine structure to the subcortical encoding of noise-degraded speech. J. Acoust. Soc. Am. 140, EL358–63.
Bidelman, G.M., 2017. Communicating in challenging environments: noise and reverberation. In: Kraus, N., Anderson, S., White-Schwoch, T., Fay, R.R., Popper, A.N. (Eds.), Springer Handbook of Auditory Research: The Frequency-following Response: A Window into Human Communication. Springer Nature, New York, N.Y.
Brown, Z., Edworthy, J., Sneyd, J.R., Schlesinger, J., 2015. A comparison of linear and logarithmic auditory tones in pulse oximeters. Appl. Ergon. 51, 350–357.
Brugge, J.F., Nourski, K.V., Oya, H., Reale, R.A., Kawasaki, H., Steinschneider, M., Howard 3rd, M.A., 2009. Coding of repetitive transients by auditory cortex on Heschl's gyrus. J. Neurophysiol. 102, 2358–2374.
Campbell, T., Kerlin, J.R., Bishop, C.W., Miller, L.M., 2012. Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. Ear Hear. 33, 144–150.
Carey, J., 2016. Science and culture musical genes. Proc. Natl. Acad. Sci. 113, 1958–1959.
Champlin, C.A., 1992. Methods for detecting auditory steady-state potentials recorded from humans. Hear. Res. 58, 63–69.
Chandrasekaran, B., Kraus, N., 2010. The scalp-recorded brainstem response to speech: neural origins and plasticity. Psychophysiology 47, 236–246.
Clinard, C.G., Tremblay, K.L., 2013. Aging degrades the neural encoding of simple and complex sounds in the human brainstem. J. Am. Acad. Audiol. 24, 590–599.
Coffey, E.B., Herholz, S.C., Chepesiuk, A.M., Baillet, S., Zatorre, R.J., 2016. Cortical contributions to the auditory frequency-following response revealed by MEG. Nat. Commun. 7, 11070.
Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. L. Erlbaum Associates, Hillsdale, N.J.
Coppola, R., Tabor, R., Buchsbaum, M.S., 1978. Signal to noise ratio and response variability measurments in single trial evoked potentials. Electroencephalogr. Clin. Neurophysiol. 44, 214–222.
Cunningham, J., Nicol, T., Zecker, S.G., Bradlow, A., Kraus, N., 2001. Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. Clin. Neurophysiol. 112, 758–767.
Dobie, R.A., Wilson, M.J., 1989. Analysis of auditory evoked potentials by magnitude-squared coherence. Ear Hear. 10, 2–13.
Dobie, R.A., Wilson, M.J., 1996. A comparison of *t*-test, F test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids. J. Acoust. Soc. Am. 100, 2236–2246.
Fitch, W.T., Kramer, G., 1994. Sonifying the body electric: superiority of an auditory over a visual display in a complex,multivariate system. In: Kramer, G. (Ed.), Auditory Display: Sonification, Audification, and Auditory Interfaces. Addison-Wesley, Reading, MA, pp. 307–325.
Galbraith, G., Arbagey, P.W., Branski, R., Comerci, N., Rector, P.M., 1995. Intelligible speech encoded in the human brain stem frequency-following response. Neuroreport 6, 2363–2367.
Galbraith, G.C., 1994. Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. Electroencephalogr. Clin. Neurophysiol. 92, 321–330.
Hansman, H., 2015. This Song Is Composed From 133 Years of Climate Change Data. Smithsonian.com.
Hinkle, D.E., Wiersma, W., Jurs, S.G., 2003. Applied Statistics for the Behavioral Sciences, 5th ed. Houghton Mifflin, Boston, MA.
Hu, L., Mouraux, A., Hu, Y., Iannetti, G.D., 2010. A novel approach for enhancing the signal-to-noise ratio and detecting automatically event-related potentials (ERPs) in single trials. Neuroimage 50, 99–111.
Jeng, F.C., Hu, J., Dickman, B., Lin, C.Y., Lin, C.D., Wang, C.Y., Chung, H.K., Li, X., 2011a. Evaluation of two algorithms for detecting human frequency-following responses to voice pitch. Int. J. Audiol. 50, 14–26.

Jeng, F.C., Hu, J., Dickman, B., Montgomery-Reagan, K., Tong, M., Wu, G., Lin, C.D., 2011b. Cross-linguistic comparison of frequency-following responses to voice pitch in American and Chinese neonates and adults. Ear Hear. 32, 699–707.
John, M.S., Picton, T.W., 2000. MASTER: a Windows program for recording multiple auditory steady-state responses. Comput. Methods Programs Biomed. 61, 125–150.
Johnson, K.L., Nicol, T.G., Kraus, N., 2005. Brain stem response to speech: a biological marker of auditory processing. Ear Hear. 26, 424–434.
Klatt, D.H., 1980. Software for a cascade/parallel formant synthesizer. J. Acoust. Soc. Am. 67, 971–995.
Koch, G.G., 1982. Intraclass correlation coefficient. In: Kotz, S., Johnson, N.L. (Eds.), Encyclopedia of Statistical Sciences. John Wiley & Sons, New York, pp. 213–217.
Kraus, N., 2017. Springer Handbook of Auditory Research: The Frequency-following Response: A Window into Human Communication. Springer Nature, New York, N.Y.
Krishnan, A., Gandour, J.T., Bidelman, G.M., 2012. Experience-dependent plasticity in pitch encoding: from brainstem to auditory cortex. Neuroreport 23, 498–502.
Krishnan, A., 2007. Human frequency following response. In: Burkard, R.F., Don, M., Eggermont, J.J. (Eds.), Auditory Evoked Potentials: Basic Principles and Clinical Application. Lippincott Williams & Wilkins, Baltimore, pp. 313–335.
Lau, J.C.Y., Wong, P.C.M., Chandrasekaran, B., 2017. Context-dependent plasticity in the subcortical encoding of linguistic pitch patterns. J. Neurophysiol. 117, 594.
Loui, P., Koplin-Green, M., Frick, M., Massone, M., 2014. Rapidly learned identification of epileptic seizures from sonified EEG. Front. Hum. Neurosci. 8, 820.
McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. Psychol. Methods 1, 30–46.
Moore, B.C.J., 2003. Introduction to the Psychology of Hearing, 5th ed. Academic Press, San Diego.
Neuhoff, J.G., 2011. The Sonification Handbook. Logos Publishing House, Berlin.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.
Parbery-Clark, A., Skoe, E., Kraus, N., 2009. Musical experience limits the degradative effects of background noise on the neural processing of sound. J. Neurosci. 29, 14100–14107.
Picton, T.W., Linden, R.D., Hamel, G., Maru, J.T., 1983. Aspects of averaging. Semin. Hear. 4, 327–341.
Pluim, J.P., Maintz, J.B., Viergever, M.A., 2003. Mutual-information-based registration of medical images: a survey. IEEE Trans. Med. Imaging 22, 986–1004.
Rocha-Muniz, C.N., Befi-Lopes, D.M., Schochat, E., 2012. Investigation of auditory processing disorder and language impairment using the speech-evoked auditory brainstem response. Hear. Res. 294, 143–152.
Schedel, M., et al., 2016. Interactive sonification of gait: realtime biofeedback for people with Parkinson's disease. In: Schedel, M. (Ed.), Proceedings of ISon 2016, Fifth Interactive Sonification Workshop. Bielefeld, Germany.
Skoe, E., Kraus, N., 2010. Auditory brain stem response to complex sounds: a tutorial. Ear Hear. 31, 302–324.
Skoe, E., Kraus, N., 2012. A little goes a long way: how the adult brain is shaped by musical training in childhood. J. Neurosci. 32, 11507–11510.
Skoe, E., Krizman, J., Kraus, N., 2013a. The impoverished brain: disparities in maternal education affect the neural response to sound. J. Neurosci. 33, 17221–17231.
Skoe, E., Krizman, J., Spitzer, E., Kraus, N., 2013b. The auditory brainstem is a barometer of rapid auditory learning. Neuroscience 243, 104–114.
Tichko, P., Skoe, E., 2017. Frequency-dependent fine structure in the frequency-following response: the byproduct of multiple generators. Hear. Res. 348, 1–15.
Weiss, M.W., Bidelman, G.M., 2015. Listening to the brainstem: musicianship enhances intelligibility of subcortical representations for speech. J. Neurosci. 35, 1687–1691.
White-Schwoch, T., Woodruff Carr, K., Thompson, E.C., Anderson, S., Nicol, T., Bradlow, A.R., Zecker, S.G., Kraus, N., 2015. Auditory processing in noise: a preschool biomarker for literacy. PLoS Biol. 13, e1002196.
Zurek, P.M., 1992. Detectability of transient and sinusoidal otoacoustic emissions. Ear Hear. 13, 307–310.