Taylor & Francis
Taylor & Francis Group

International Journal of Audiology 2016; 55: 313–319

International Journal of Audiology

Technical Report

Objective detection of auditory steady-state evoked potentials based on mutual information

Gavin M. Bidelman[1,2] & Shaum P. Bhagat[2]

[1]Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA and [2]School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA

The British Society of Audiology

The International Society of Audiology

NAS NORDIC AUDIOLOGICAL SOCIETY

Abstract

Objective: Recently, we developed a metric to objectively detect human auditory evoked potentials based on the mutual information (MI) between neural responses and stimulus spectrograms. Here, the MI algorithm is evaluated further for validity in testing the auditory steady-state response (ASSR), a sustained potential used in objective audiometry. *Design*: MI was computed between spectrograms of ASSRs and their evoking stimuli to quantify the shared time-frequency information between neuroelectric activity and stimulus acoustics. MI was compared against two traditional ASSR detection metrics: *F-test* and magnitude-squared coherence (*MSC*). *Study Sample*: Using an empirically derived threshold ($\ominus_{MI}$=1.45), MI was applied as a binary classifier to distinguish actual biological responses recorded in human participants (n=11) from sham recordings, containing only EEG noise (i.e., non-stimulus-control condition). *Results*: MI achieved high overall accuracy (>90%) in identifying true ASSRs from sham recordings, with true positive/true negative rates of 82/100%. During online averaging, comparison with two other indices (*F-test, MSC*) indicated that MI could detect ASSRs in roughly half the number of trials (i.e., ~400 sweeps) as the *MSC* and performed comparably to the *F-test,* but showed slightly better signal detection performance. *Conclusions*: MI provides an alternative, more flexible metric for efficient and automated ASSR detection.

Keywords: Auditory evoked potentials (AEPs); automated auditory brainstem response (AABR); evoked potential classification; F-test. Magnitude-squared coherence (MSC); objective audiometry

Introduction

Conventional auditory evoked potential (AEP) recording practice requires that responses be identified by the subjective interpretation of human observers. This subjectivity leads to biases in the detection and interpretation of evoked responses (Vidler & Parker, 2004; Bogaerts et al, 2009). Recently, auditory steady-state responses (ASSRs) have gained popularity as they offer more rapid physiological assessment by allowing evaluation of both ears simultaneously at multiple audiometric frequencies (Lins et al, 1995; Picton et al, 1998; John & Picton, 2000; Cone-Wesson et al, 2002; Stroebel et al, 2007; Bhagat, 2008). ASSRs are also desirable because response detection is based on a statistical comparison between signal and noise power in the evoked potential average rather than human waveform inspection. This offers a fully objective means to assess ASSR signal quality and remove human subjectivity from electrophysiological assessment.

Objective evaluation of the ASSR is typically performed in the frequency domain where a test statistic [e.g., *F*-test and magnitude-squared coherence (MSC)] is applied to the ASSR spectrum to determine the probability of whether a response is present at the stimulus frequency relative to the surrounding noise floor (Dobie & Wilson, 1996; John & Picton, 2000; Vidler & Parker, 2004; Sturzebecher & Cebulla, 2013). These statistics become more powerful with increasing number of sweeps. Consequently, a stopping rule can be applied when a criterion significance level is achieved (typically $p < 0.01$). Both the *F*-test and MSC provide robust statistical power for detecting the presence of the ASSR (Champlin, 1992; Dobie & Wilson, 1996) and can be used ''online'' as an automated stopping rule for signal averaging—saving valuable time in clinical testing. Yet, these current metrics are somewhat limited in that they are specific to the evoking stimulus (e.g., sinusoidally amplitude modulated (SAM) tones). With more complex stimuli (e.g., multi-frequency SAM tones, time-varying speech), response detection would require incorporating multiple test statistics applied to each stimulus frequency of interest. Alternate objective indices might be more sensitive and flexible to detect the ASSR and other sustained AEPs.

*Address for editorial *Correspondence*: Gavin M. Bidelman, PhD, School of Communication Sciences & Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN, 38152, TEL: (901) 678-5826, FAX: (901) 525-1282, Email: g.bidelman@memphis.edu

(Received 15 April 2015; revised 12 December 2015; accepted 4 January 2016)

ISSN 1499-2027 print/ISSN 1708-8186 online
DOI: 10.3109/14992027.2016.1141246

Abbreviations

AEP	auditory evoked potential
ABR	auditory brainstem response
ASSR	auditory steady-state response
FFR	frequency-following response
FP	false positive (i.e., ''false alarm''); FPR, false positive rate
FN	false negative (i.e., ''miss''); MI, mutual information
MSC	magnitude-squared coherence
SAM	sinusoidal amplitude modulation
TP	true positive (i.e., ''hit''); TPR, true positive rate
TN	true negative (i.e., ''correct rejection'')

Recently, we have developed a fully objective and automated algorithm for the detecting the speech-evoked brainstem frequency-following response (FFR) (Bidelman, 2014). Our detection metric is based on the statistical comparison between the spectrographic representations of the stimulus signal and neural response which we adopted from information theory and image processing: mutual information (MI). This metric quantifies the spectral similarity between stimulus and neural response spectrograms. Spectrograms are advantageous for assessing AEP signal quality because they provide a three-dimensional representation of the neural activity (time, frequency, amplitude) and thus, a higher dimensionality of detail than a 2D time-waveform alone. In our previous study, we showed that MI could be used as a robust means to objectively detect not only the presence but also signal quality of the speech-evoked FFR with 97% true positive rate and 85% true negative rate (Bidelman, 2014). The metric was also useful as a stopping rule for signal averaging and demonstrated that ~1500 sweeps were adequate to detect the speech-FFR with a signal-to-noise ratio (SNR) of +3 dB (Bidelman, 2014). Moreover, the metric can be applied to time-varying signals and thus generalizes well across AEP classes and eliciting stimuli.

Given the increased popularity and clinical utility of the ASSR (e.g., Picton et al, 1998; John & Picton, 2000; Cone-Wesson et al, 2002; Stroebel et al, 2007; Bhagat, 2008), the aim of the present study was to further validate MI (Bidelman, 2014) as an objective metric for broader AEP evaluation and use in detecting ASSR responses to SAM-tone stimuli. We endeavored to assess the MI metric's ability to identify recorded ASSRs (true biological responses) from sham recordings (containing no biological response). We then compared the proposed MI metric for use as an ''online'' stopping criterion for signal averaging against two other well-established frequency-domain ASSR detection indices: magnitude-squared coherence (MSC) (Dobie & Wilson, 1989; Champlin, 1992) and the *F-test* (Dobie & Wilson, 1996; John & Picton, 2000). These three metrics were tracked on a sweep-by-sweep basis to determine differences in their stopping rule (i.e., efficiency) for signal averaging. Comparable or better performance of the proposed MI measure compared to current detection metrics would support broad efficacy of the MI across AEPs and a potential alternative to other objective ASSR detection approaches.

Experiment 1: MI for ASSR detection

Methods

Participants

Eleven normal-hearing young adults (11 female; age: 20–28 yrs) participated in Experiment 1. All participants had normal hearing thresholds ($\leq$25 dBHL, 250–8000 Hz) bilaterally. Each gave written-informed consent in compliance with a protocol approved by the Louisiana State University Institutional Review Board.

Stimuli

Stimuli were SAM tones with a carrier frequency (f_c) of 1000 Hz and modulation frequency (f_m) of 40 Hz (100% modulation depth). Eight cycles of the 40-Hz modulation were recorded. Stimuli were delivered monaurally to the right ear using Eartone-3A inserts at 70 dB SPL.

ASSR recordings and analysis

ASSRs were recorded differentially between electrodes placed on the vertex (Cz) and right earlobe (A2) (mid forehead = ground). Interelectrode impedances were $\leq$3 kΩ. Electrophysiological recordings were collected with BioSig (Tucker Davis Technologies). Recordings were amplified and bandpass filtered (10–200 Hz, −6 dB/octave rolloff. ASSRs were digitized at 10 kHz over a 204.8 ms epoch (window). Sweeps containing voltages >90% of the A/D converter's dynamic range were rejected prior to averaging. Final average ASSR waveforms contained 512 sweeps. In addition to SAM tone stimulation, sham recordings were also obtained from each subject. Sham runs were identical to ASSR recordings with the exception that the insert headphone was removed from the ear canal, thus preventing stimulus delivery to the participant but allowing the continued recording of EEG noise (e.g., Aiken & Picton, 2008; Bidelman, 2014).

ASSR and sham traces were analyzed using the Fast Fourier Transform (FFT) to index spectral content of the scalp potentials. Magnitudes were measured from response spectra, relative to the noise floor, at the fundamental and harmonics of the modulation response (i.e., 40, 80, 120, 160 Hz). Comparison of spectral magnitudes allowed us to verify the presence (ASSR) and absence (sham) of a neural response. Post-processing and analyses were performed using custom routines coded in MATLAB® 2014b (v. 8.4).

Mutual information (MI) detection metric

We computed the mutual information (MI) between spectrographic representations of the stimulus and ASSR to index the degree to which neural responses captured spectrotemporal details of the acoustic input (Bidelman, 2014). MI is a dimensionless quantity (measured in bits), which quantifies the degree of shared information (i.e., mutual dependence) between two variables. Said differently, it reflects the reduction in uncertainty that knowing either signal provides about the other. For two random variables (A and B), MI is computed as Eq. 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 distributions for two independent images, $p(a)p(b)$ (see *SI material* for further details). Applied here, MI quantifies the degree to which neural responses capture the collective time-frequency characteristics of the stimulus based on comparison between stimulus and response spectrograms (Bidelman, 2014).

MI was computed between the stimulus and each response spectrogram allowing us to assess the degree to which neural responses reflected spectrotemporal properties of the evoking stimulus. In response to a SAM tone, rectification in the cochlear haircell transduction process produces a response at the modulation/envelope frequency (here 40 Hz) that is captured in the scalp-recorded potential but is not present in the stimulus itself (Lins et al, 1995; John & Picton, 2000). Hence, comparisons between raw stimulus and response spectrograms would show large discrepancy near the frequency of interest (40 Hz), leading to artificially low MI. To circumvent this natural disparity between stimulus and response spectra, we half-wave rectified the stimulus waveform prior to spectrographic analysis—a common operator to model nonlinear haircell transduction bias (e.g., Oxenham et al, 2004). This ensured that both the stimulus and response time-frequency representations contained prominent energy at the modulation frequency (here 40 Hz). Note however, that we do not further scale this operator to avoid manipulating the amount of simulated rectification and potential false optimization of MI. While stimulus rectification is necessary here given the use of highly stereotyped SAM stimuli, for more complex stimuli (e.g., speech tokens), we have found that in practice, half-wave rectification is unnecessary to compute a valid MI (Bidelman, 2014).

In our previous study (Bidelman, 2014), we made no *a priori* assumptions regarding an acceptable threshold for MI in use as a detection metric for AEPs. A criterion threshold was determined empirically using computational modeling and responses at known SNRs. For the detection of speech-evoked FFRs (e.g., Bidelman & Krishnan, 2010), we found that a threshold θ_{MI}=1 (corresponding to an SNR of +3 dB) provided robust classification characteristics (97% true positive rate, 85% true negative rate) for discriminating true biological FFRs from noise-sham recordings. While this criterion was appropriate for sustained FFRs to speech stimuli, it was unclear in the present study if the same threshold would be appropriate for the non-speech stimuli used here for traditional ASSRs (i.e., SAM tones). Thus, a criterion threshold was determined empirically by comparing the *d*-prime between ASSR and sham responses as a function of the lowpass filter cutoff frequency applied to the response. Maximal segregation of true from sham responses was observed with a cutoff of 200 Hz. This filter setting is optimal as it corresponds with the upper cutoff of the original recordings (10-200 Hz) but still passes the most salient harmonics of the ASSR response (e.g., 40, 80 120 Hz). The resulting criterion MI (θ_{MI}=1.45) was applied to all subsequent analyses.

MI classifier performance metrics

Once determined empirically, θ_{MI} was applied as a binary classifier to ASSR and sham recordings. Traces 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 (Bidelman, 2014). Classifier performance was evaluated by computing standard metrics used in signal detection theory (*d*-prime) and receiver operating characteristics (ROC) including true and false positive rates. True positive rate [i.e., TPR = 100*TP/(TP + FN)] was computed as the percentage of actual ASSR recordings correctly identified; false-positive rate as the percentage of sham recordings erroneously classified as a biological response [i.e., FPR = 100*FP/(FP + TN)] (see Figure 2; Jeng et al, 2011). True negative rate (TNR) was computed as 1-FPR.

Results

ASSR spectrotemporal properties

ASSR time waveforms and response spectra are shown for actual and sham recordings in Figure 1. Response spectra illustrate robust phase-locked neural activity at the 40 Hz modulation rate and its upper harmonics (e.g., 80 and 120 Hz) for ASSR but not sham recordings. Bonferroni adjusted paired *t*-tests (adjusted for four comparisons) revealed significant ASSR responses at the 40 Hz [t_{10} = 5.76, p< 0.0001], 80 Hz [t_{10} = 3.62, p=0.0024], and 120 Hz [t_{10} = 2.61, p=0.0129] components relative to the noise floor amplitude as measured in response FFTs; responses at 160 Hz were indistinguishable between actual and sham recordings [t_{10} = -3.41, p =0.99] as both appeared lower than the noise floor. These findings confirm that the ASSRs contain robust phase-locked neural activity whereas sham recordings contain no biological response (nor stimulus artifact) and are thus suitable for use as ''catch trials'' in validating the proposed MI detection metric.

MI classifier performance for identifying true ASSRs from sham recordings

Figure 2 shows MI computed from actual ASSR recordings as well as sham traces—containing no biological response. Applying a threshold of θ_{MI}=1.45, only two ASSR recordings (9.1%) are misclassified from the n=22 total observations (n=11 ASSR, n=11 shams). Responses were highly distinguishable from sham traces based on their MI [t_{10}= 6.28, p<0.0001] (Figure 2B). This was confirmed by bootstrap resampling (Efron & Tibshirani, 1993) which showed no overlap in the 95% confidence intervals between the MIs of true and sham recordings [ASSR: 1.52-1.67; sham: 1.26-1.36].

Classifier performance for common signal detection and classification metrics are shown for the MI and *F-test* (Dobie & Wilson, 1996; John & Picton, 2000) metrics in Table 1. See *SI Material* for definition of the *F-test.* [Note that MSC could not be applied on the waveforms in Exp 1 as this metric requires computing subaverages from single-trial data and only grand averages were originally collected for this dataset.] Overall, MI yielded 91% accuracy (9% misclassification) with a corresponding true positive and true negative rate of 82% and 100%, respectively. These values are corroborated by the metric's overall true positive rate (d' = 3.23) and minimal bias (c= 0.71). These operating characteristics are comparable to our previous report applying the MI metric to detect speech-evoked FFRs (Bidelman, 2014). These performance characteristics demonstrate that the mutual information between a stimulus and neural response provides an objective means for distinguishing the ASSR generated from the auditory system from background EEG noise.

Experiment 2: MI as a stopping rule for ASSR averaging

In Experiment 2, we investigated the application of using MI as an online stopping rule for ASSR signal averaging.

Materials & Methods

Participants

Six additional normal-hearing adults (4 female; age: 28.2 ± 4.6 years) participated in Experiment 2. All participants had normal hearing thresholds ($\leq$ 25 dBHL, 250–8000 Hz) bilaterally.

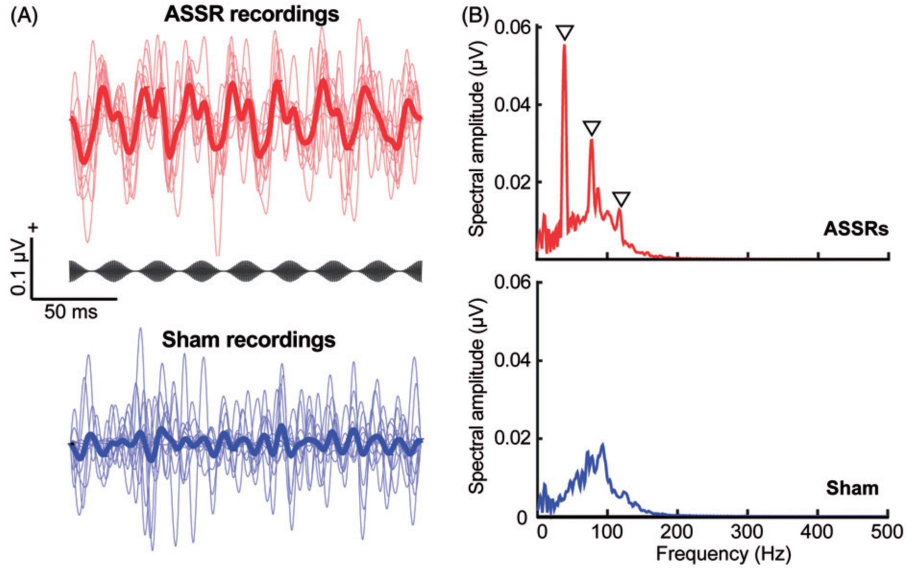

Figure 1. Auditory steady-state responses (ASSR) elicited by a 40 Hz sinusoidally modulated 1 kHz tone. (A) ASSR time-waveforms for true neurobiological (top) responses and sham (bottom) recordings in which the earphone was removed from the ear canal. Light traces, overlay of responses from individual subjects; thick dark trace, grand average. The stimulus waveform is shown in gray in the middle. (B) Average response spectra for true ASSR (top) vs. sham (bottom) recordings. True ASSRs show dominant energy at the modulation frequency (40 Hz) and integer-related harmonics (i.e., 80, 120 Hz) well above the noise floor (▽). In contrast, sham recordings show no definable or harmonically-related response peaks (only EEG noise), confirming the absence of a neural response (or stimulus artifact) in the non-stimulus-control condition.

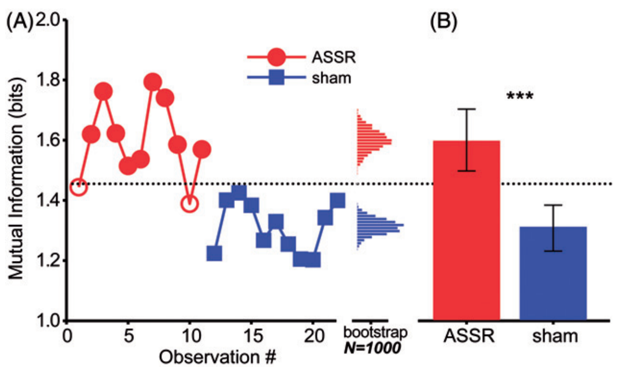

Figure 2. MI classifier performance in distinguishing true ASSRs from sham recordings. (A) MI computed from ASSRs (circles) and sham traces (squares)—where the earphone was removed from the ear canal. MI values below the empirically derived threshold value (θ_{MI} =1.45), are classified as noise (i.e., no response); values exceeding θ_{MI} =1.45 are identified as true biological responses. Misclassifications are shown as open symbols. Bootstrap resampling (right) shows the distribution of MI for neural and sham recordings for N= 1000 data resamples. (B) Average MI value computed from true vs. sham recordings. As denoted by the clear separation of recordings, the MI metric is able to distinguish actual evoked responses from EEG noise. Error bars = 95% CIs computed via bootstrapping, ***p<0.001.

Participants gave written-informed consent in compliance with a protocol approved by the University of Memphis Institutional Review Board.

Table 1. Classifier performance characteristics for MI and F-test detection metrics[†].

	MI	*F-test*
Overall performance		
Accuracy	90.91%[a]	95.45%
Misclassification rate	9.09%	4.54%
Signal detection metrics		
d-prime[b]	3.23	2.67
bias[c]	0.71	0
ROC		
True positive rate	81.8%	90.9%
True negative rate	100%	90.9%

[†]n=22 total observations (n=11 neural ASSR vs. n=11 sham noise recording; 512 sweeps).
[a]Based on MI = 1.45
[b]Computed as d' = z(H) − z(FA) from mean hit (H) and false alarm (FA) rates.
[c]Computed as bias = −[z(H) + z(FA)]/2 from mean H and FA rates.

ASSR recordings

ASSR recording procedures and stimuli were similar to Experiment 1. Briefly, EEGs were recorded between Ag/AgCl disc electrodes placed on the scalp at the high forehead at the hairline referenced to linked mastoids (A1/A2) (mid-forehead = ground). Continuous EEGs were digitized at 10 kHz (SynAmps RT amplifiers; Compumedics Neuroscan). EEGs were then windowed [−50–250 ms], filtered (10–200 Hz), and averaged in the time domain to obtain ASSR waveforms. SAM tone stimuli (see Experiment 1) were delivered binaurally at an intensity of 80 dB SPL through insert earphones (ER-2, Etymotic Research). Listeners heard 2500 exemplars of the stimulus token presented an ISI interval of 50 ms.

Comparison of MI to other objective metrics (F-test, MSC)

To test the efficiency of θ_{MI} as stopping criterion for signal averaging, we computed MI on a sweep-by-sweep basis as accumulating trials were added to the ongoing ASSR average. Similarly, we compared the ''online'' development of MI against two other frequency-domain detection metrics: (1) θ_{MI} (current study; Bidelman, 2014), (2) *F-test* (Dobie & Wilson, 1996; John & Picton, 2000), and (3) *MSC* (Dobie & Wilson, 1989; Champlin, 1992). See *SI Material* for definitions. Comparison between the three metrics allowed us to relate their performance and determine differences in their stopping rule for signal averaging, i.e., the number of trials where each index detected the presence of the ASSR response. Some participants' ASSRs approached, but did not fully converge to the criterion threshold (treated as missing values in statistical analysis).

Results

Figure 3 shows the time-course for *MSC*, *F*-test, and proposed *MI* detection metrics with increasing sweeps. Each metric improves with additional trials and asymptotes as the running average stabilizes.

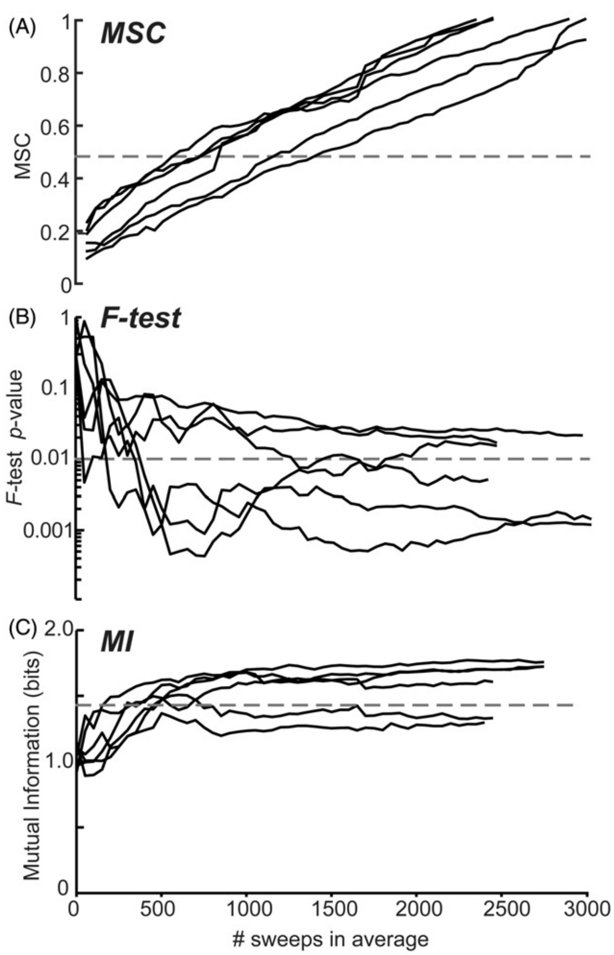

Figure 3. Comparison of the growth in three detection metrics during online ASSR recording. Sweep-by-sweep ASSR detection based on (A) *MSC*, (B) *F*-test, and (C) MI. Dashed lines, significance thresholds for response detection [dashed lines; MSC: 0.483 (p=0.01) (Dobie & Wilson, 1989); *F*-test: p=0.01 (John & Picton, 2000); MI: θ_{MI} >1.45].

Figure 4 shows the number of sweeps required to achieve a specified stopping criterion for signal averaging. Stopping criteria were based on a p<0.01 (*MSC*=0.483 and *F*-test) and θ_{MI} =1.45 (MI), respectively (Dobie & Wilson, 1989; John & Picton, 2000; Bidelman, 2014). A Kruskal-Wallace nonparametric ANOVA (used given small sample sizes) revealed a difference in terminating sweep between the three metrics [$\chi^2(2)$= 6.90, p=0.032]. Follow-up corrected multiple comparisons (Dunn procedure; Elliott & Hynan, 2011) revealed that the MI metric was superior to the *MSC* and was able to detect the ASSR in roughly half the trials (Z=-2.66, p=0.0078; *MSC*: ~850 trials; *MI*: ~400 trials). No difference was observed between the *F*-test and *MI* (Z=0.0, p=1.0), suggesting these metrics performed comparably in their efficiency.

Discussion

Results of the current study demonstrate potential advantages of a novel algorithm for objective identification of scalp-recorded auditory steady-state responses (ASSRs) based on the mutual information (MI) between neural responses and stimulus spectrograms (i.e., time-frequency representations) (Bidelman, 2014). Overall, the proposed MI metric achieved robust performance, yielding >90% overall accuracy and equally impressive true positive and true negative rates of 82% and 100%, respectively. Lastly, it was shown that MI increases monotonically with increasing number of stimulus presentations (i.e., trials) and can detect the ASSR in nearly half the number of trials using the *MSC* and comparable time efficiency as the *F*-test. Taken together, the objectivity and larger generalizability of the proposed MI metric to other AEP classes (Bidelman, 2014) make it a useful method for ASSR evaluation and as a criterion to terminate signal averaging. The current study demonstrates that MI, previously validated for human speech-evoked FFR recordings (Bidelman, 2014), is also

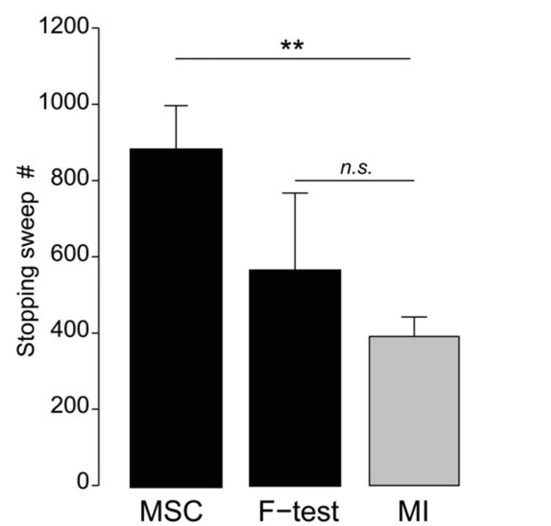

Figure 4. Comparison of three detection metrics' stopping rule for signal averaging and objectively identifying the presence of ASSRs. Stopping criteria were based on a p<0.01 (*MSC* and *F*-test) and θ_{MI} =1.45 (MI), respectively. Of the three indices, MI allows detection in the fewest number of stimulus sweeps. **p < 0.01. Error bars = ±1 s.e.m.

applicable to the detection of the ASSR, an evoked potential with increasing use in clinical audiology (e.g., Picton et al, 1998; John & Picton, 2000; Cone-Wesson et al, 2002; Stroebel et al, 2007; Bhagat, 2008; Sturzebecher & Cebulla, 2013)

The positive, albeit small, bias in the metric (Table 1) indicates it had a slight tendency toward false negatives, i.e., true ASSRs being erroneously classified as sham responses. However, this bias is likely negligible in light of the metric's relatively low misclassification rate (9%) and high accuracy (91%). More importantly, comparisons with other detection metrics including the *F-test* and *MSC* revealed largely comparable performance in the MI index. Its primary advantage seems to be in time efficiency. Compared to *MSC*, MI was able to detect the ASSR in roughly half the number of stimulus sweeps during online averaging. Direct comparisons with the *F*-test, perhaps the most widely used metric for ASSR detection (Dobie & Wilson, 1996; John & Picton, 2000), revealed that MI performed comparably in time efficiency; both the *F-test* and MI identified the presence of the ASSR in ~400 trials (Figure 4) and showed similar ROC detection performance (Table 1), overall accuracy (*MI*: 91%, *F*-test: 95%), but slightly higher misclassification error (*MI*: 9%, *F*-test: 4.5%). However, signal detection metrics indicated higher *d-prime* for the MI compared to the *F-test* suggesting slightly better discrimination of true ASSR responses from noise.

In our previous report, we also demonstrated the MI metric's superiority over ''gold standard'' judgments of human observers, whose detection is inherently subjective in nature (Bidelman, 2014). We have previously shown that the MI metric can be applied to other classes of AEPs including the speech-evoked FFRs elicited by dynamic, spectrotemporally complex stimuli (Bidelman, 2014). Moreover, unlike the *F*-test and *MSC* which requires knowledge of stimulus frequency content to ''tag'' the correct response (i.e., frequency bin) from the noise floor, MI does not require this *a priori* knowledge or the investigator to specify stimulus and noise components. In addition, MI is an information-theoretic measure that is ''distribution free'' and therefore requires fewer assumptions than the *F*-test and *MSC,* which both utilize parametric (distribution-based) statistics. Unlike these other metrics, MI can also be easily applied to time-varying signals (Bidelman, 2014). Thus, in addition to potentially broader application, MI seems to offer a useful alternative to other objective detection approaches that performs as well or some cases better than current ASSR analysis techniques.

Study limitations and directions for future work

One application of the ASSR is objective determination of hearing thresholds (Johnson & Brown, 2005; Sturzebecher & Cebulla, 2013). As a first step toward developing a new ASSR detection metric, we investigated the efficacy of the MI metric for detecting *suprathreshold* stimuli (70-80 dB SPL). While suprathreshold stimuli do find some use clinically in objective auditory applications, e.g., newborn hearing screenings (American Academy of Pediatrics, 2007), they are more commonplace in research applications, e.g., assessing stimulus-related changes in neural responses (e.g., Bidelman et al, 2013) or differences in auditory function between populations (e.g., Kraus & Banai, 2007).

In our previous report examining brainstem FFRs, we showed that MI decreases according to a sigmoidal function with decreasing SNR (Bidelman, 2014). Hence, it is possible that the criterion MI detailed here ($\theta_{MI} = 1.45$), may not be appropriate for stimuli of lower level. It should be noted however, that pure scaling to the signal or response will not change the computed MI; based on the definition of mutual information (Equation 1), multiplying by a constant will not change information content between signals *A* and *B*. Practically speaking, this means that MI is invariant to changes in purely the signal level or response level. MI is however, sensitive to the *SNR* between the signal and response. Hence, lower level stimuli (e.g., near threshold) would tend to lower MI given the reduction in SNR of the resulting ASSR (i.e., decreased response amplitude but constant EEG noise). Presumably, MI would tend to weaken with lower stimulus intensities which would then close the gap between ASSR and noise floor spectra, making response detection more difficult. While we have no reason to believe that the proposed MI metric would not increase testing-efficacy for less intense stimuli (and therefore noisier AEPs), future studies are needed to validate the metric and its generalizability across a wider range of stimulus parameters including intensity and different modulation rates.

Acknowledgements

This work was supported by grants from the American Hearing Research Foundation (AHRF) and American Academy of Audiology (AAA) Foundation awarded to G.M.B. (2014) and S.P.B. (2004). Correspondence for reprints and materials should be addressed to G.M.B. [g.bidelman@memphis.edu].

Declaration of interest: The authors report no declarations of interest.

References

Aiken S.J. & Picton T.W. 2008. Envelope and spectral frequency-following responses to vowel sounds. *Hear Res*, 245, 35–47.

American Academy of Pediatrics J.C.o.I.H. 2007. Position Statement: Principles and guidelines for early hearing detection and identification. *Pediatrics*, 120, 898–921.

Bhagat S.P. 2008. The effects of monotic and dichotic interference tones on 40 Hz auditory steady-state responses in normal-hearing adults. *J Am Acad Audiol*, 19, 101–119.

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

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

Bogaerts S., Clements J.D., Sullivan J.M. & Oleskevich S. 2009. Automated threshold detection for auditory brainstem responses: Comparison with visual estimation in a stem cell transplantation study. *BMC Neurosci*, 10, 1–7.

Champlin C.A. 1992. Methods for detecting auditory steady-state potentials recorded from humans. *Hear Res*, 58, 63–69.

Cone-Wesson B., Dowell R.C., Tomlin D., Rance G. & Ming W.J. 2002. The auditory steady-state response: comparisons with the auditory brainstem response. *J Am Acad Audiol*, 13, 173–187.

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.

Efron B. & Tibshirani R. 1993. *An introduction to the bootstrap*. New York: Chapman & Hall.

Elliott A.C. & Hynan L.S. 2011. A SAS® macro implementation of a multiple comparison post hoc test for a Kruskal–Wallis analysis. *Comput Methods Programs Biomed*, 102, 75–80.

Jeng F.C., Hu J., Dickman B., Lin C.Y., Lin C.D., et al. 2011. Evaluation of two algorithms for detecting human frequency-following responses to voice pitch. *Int J Audiol*, 50, 14–26.

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 T.A. & Brown C.J. 2005. Threshold prediction using the auditory steady-state response and the tone burst auditory brain stem response: a within-subject comparison. *Ear Hear*, 26, 559–576.

Kraus N. & Banai K. 2007. Auditory-processing malleability: Focus on language and music. *Curr Dir Psychol Sci*, 16, 105–110.

Lins O.G., Picton P.E., Picton T.W., Champagn S.C. & Durieux-Smith A. 1995. Auditory steady-state responses to tones amplitude-modulated at 80-110 Hz. *J Acoust Soc Am*, 97, 3051–3063.

Oxenham A.J., Bernstein J.G.W. & Penagos H. 2004. Correct tonotopic representation is necessary for complex pitch perception. *Proc Natl Acad Sci USA*, 101, 1421–1425.

Picton T.W., Durieux-Smith A., Champagne S.C., Whittingham J., Moran L.M., et al. 1998. Objective evaluation of aided threhsolds using auditory steady-state responses. *Journal of American Academy of Audiology*, 9, 315–331.

Stroebel D., Swanepoel W. & Groenewald E. 2007. Aided auditory steady-state responses in infants. *Int J Audiol*, 46, 287–292.

Sturzebecher E. & Cebulla M. 2013. Automated auditory response detection: Improvement of the statistical test strategy. *Int J Audiol*, 52, 861–864.

Vidler M. & Parker D. 2004. Auditory brainstem response threshold estimation: Subjective threshold estimation by experienced clinicians in a computer simulation of a clinical test. *Int J Audiol*, 43, 417–429.