Ear and Hearing

Improved detection of vowel envelope frequency-following responses using Hotelling's T2 analysis

--Manuscript Draft--

Manuscript Number:	EANDH-D-17-00092R2
Full Title:	Improved detection of vowel envelope frequency-following responses using Hotelling's T2 analysis
Article Type:	Research Article
Section/Category:	Auditory Neuroscience: Central
Keywords:	Objective response detection; Envelope Frequency Following Responses; EEG; Auditory Steady-State Responses; Simulation; Natural speech response detection
Corresponding Author:	steven bell, PhD University of Southampton Southampton, Hampshire UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Southampton
Corresponding Author's Secondary Institution:	
First Author:	Frederique Vanheusden
First Author Secondary Information:	
Order of Authors:	Frederique Vanheusden
	Steven Bell
	Michael Chesnaye
	David Simpson
Order of Authors Secondary Information:	
Abstract:	Objectives: Objective detection of brainstem responses to natural speech stimuli is an important tool for the evaluation of hearing aid fitting, especially in people who may not be able to respond reliably in behavioral tests. Of particular interest is the envelope Frequency Following Response (eFFR), which refers to the EEG response at the stimulus' fundamental frequency (and its harmonics), and here in particular to the response to natural spoken vowel sounds. This paper introduces the frequency-domain Hotelling's T2 (HT2) method for eFFR detection. This method was compared, in terms of sensitivity in detecting eFFRs at the fundamental frequency (HT2_F0), to two different single channel frequency domain methods (F-test on Fourier Analyzer amplitude spectra - FA-F-Test and Magnitude Squared Coherence - MSC) in detecting envelope following responses to natural vowel stimuli in simulated data and EEG data from normal hearing subjects. Sensitivity was assessed based on the number of detections and the time needed to detect a response for a false-positive rate of 5%. The study also explored the whether a single-channel, multi-frequency HT2 (HT2_3F) and a multichannel, multi-frequency HT2 (HT2_MC) could further improve response detection. Design: Four repeated words were presented sequentially at 70dB SPL LAeq through ER-2 insert earphones. The stimuli consisted of a prolonged vowel in a /hVd/ structure (where V represents different vowel sounds). Each stimulus was presented over 440 sweeps (220 condensation, 220 rarefaction). EEG data were collected from 12 normal hearing adult participants. After pre-processing and artefact removal, eFFR detection was compared between the algorithms. For the simulation study, simulated EEG signals were generated by adding random noise at multiple signal-to-noise ratios (SNR - 0dB to -60dB) to the auditory stimuli as well as to a single sinusoid at the fluctuating and flattened fundamental frequency (f0). For each SNR, 1,000 sets of 440 simulated epochs were generated. Performance of the algorithms was assessed based on the

	number of sets for which a response could be detected at each SNR. Results: In simulation studies, HT2_3F significantly outperformed the other algorithms when detecting a vowel stimulus in noise. For simulations containing responses only at a single frequency, HT2_3F performs worse compared to other approaches applied in this study as the additional frequencies included do not contain additional information. For recorded EEG data, HT2_MC showed a significantly higher response detection rate compared to MSC and FA-F-Test. Both HT2_MC and HT2_F0 also showed a significant reduction in detection time compared to the FA-F-Test algorithm. Comparisons between different electrode locations confirmed a higher number of detections for electrodes close to Cz compared to more peripheral locations. Conclusion: The HT2 method is more sensitive than FA-F-Test and MSC in detecting responses to complex stimuli, as it allows detection of multiple frequencies (HT2_F3) and multiple EEG channels (HT2_MC) simultaneously. This effect was shown in simulation studies for HT2_3F and in EEG data for the HT2_MC algorithm. The spread in detection time across subjects is also lower for the HT2 algorithm, with decision on the presence of an eFFR possible within 5 minutes.
Additional Information:	
Question	**Response**
COMPLIANCE WITH NIH AND OTHER RESEARCH FUNDING AGENCY ACCESSIBILITY REQUIREMENTS A number of research funding agencies now require or request the submission of the post-acceptance / pre-copyedit (the article after peer review and acceptance but not the final published article) to a repository that is accessible online by all without charge. Lippincott Williams & Wilkins will make this submission on your behalf. Please disclose below if you have received funding for research on which your article is based from any of the following organizations: National Institutes of Health (NIH), Wellcome Trust, Howard Hughes Medical Institute (HHMI), or Research Council UK (RCUK).	Research Council UK (RCUK)
"Public Access Policy" Funding Disclosure Please disclose below if you have received funding for research on which your article is based from any of the following organizations:	Research Councils UK (RCUK)

Improved detection of vowel envelope frequency-following responses using Hotelling's T2 analysis

Frederique J. Vanheusden[1], Steven L. Bell[1], Michael A. Chesnaye[1] and David M. Simpson[1]

[1]Institute of Sound and Vibration Research, University of Southampton, Southampton, United Kingdom

Financial Disclosures/Conflicts of Interest:

This study was funded by the Engineering and Physical Sciences Research Council (EPSRC), UK (Grant number: EP/M026728/1)

Address for correspondence: Steven L. Bell, Institute of Sound and Vibration Research, University of Southampton, Highfield Campus, Tizard Building 13/4015, University Road, Southampton, SO17 1BJ, United Kingdom. E-mail: S.L.Bell@soton.ac.uk. Tel: +44 (0)23 8059 2277. Fax: +44 (0)23 8059 3190.

ABSTRACT

Objectives: Objective detection of brainstem responses to natural speech stimuli is an important tool for the evaluation of hearing aid fitting, especially in people who may not be able to respond reliably in behavioral tests. Of particular interest is the envelope Frequency Following Response (eFFR), which refers to the EEG response at the stimulus' fundamental frequency (and its harmonics), and here in particular to the response to natural spoken vowel sounds. This paper introduces the frequency-domain Hotelling's T2 (HT2) method for eFFR detection. This method was compared, in terms of sensitivity in detecting eFFRs at the fundamental frequency (HT2_F0), to two different single channel frequency domain methods (F-test on Fourier Analyzer amplitude spectra – FA-F-Test and Magnitude Squared Coherence – MSC) in detecting envelope following responses to natural vowel stimuli in simulated data and EEG data from normal hearing subjects. Sensitivity was assessed based on the number of detections and the time needed to detect a response for a false-positive rate of 5%. The study also explored the whether a single-channel, multi-frequency HT2 (HT2_3F) and a multichannel, multi-frequency HT2 (HT2_MC) could further improve response detection.

Design: Four repeated words were presented sequentially at 70dB SPL LAeq through ER-2 insert earphones. The stimuli consisted of a prolonged vowel in a /hVd/ structure (where V represents different vowel sounds). Each stimulus was presented over 440 sweeps (220 condensation, 220 rarefaction). EEG data were collected from 12 normal hearing adult participants. After pre-processing and artefact removal, eFFR detection was compared between the algorithms. For the simulation study, simulated EEG signals were generated by adding random noise at multiple signal-to-noise ratios (SNR – 0dB to -60dB) to the auditory stimuli as well as to a single sinusoid at the fluctuating and flattened fundamental frequency (f_0). For each SNR, 1,000 sets of 440 simulated epochs were generated. Performance of the algorithms was assessed based on the number of sets for which a response could be detected at each SNR.

Results: In simulation studies, HT2_3F significantly outperformed the other algorithms when detecting a vowel stimulus in noise. For simulations containing responses only at a single frequency, HT2_3F performs worse compared to other approaches applied in this study as the additional frequencies included do not contain additional information. For recorded EEG data, HT2_MC showed a significantly higher response detection rate compared to MSC and FA-F-Test. Both HT2_MC and HT2_F0 also showed a significant reduction in detection time compared to the FA-F-Test algorithm. Comparisons between different electrode locations confirmed a higher number of detections for electrodes close to Cz compared to more peripheral locations.

Conclusion: The HT2 method is more sensitive than FA-F-Test and MSC in detecting responses to complex stimuli, as it allows detection of multiple frequencies (HT2_F3) and multiple EEG channels (HT2_MC) simultaneously. This effect was shown in simulation studies for HT2_3F and in EEG data for the HT2_MC algorithm. The spread in detection time across subjects is also lower for the HT2 algorithm, with decision on the presence of an eFFR possible within 5 minutes.

Key words: Objective response detection, Envelope Frequency Following Responses, EEG, Auditory Steady-State Responses, Simulation, Natural speech response detection

INTRODUCTION

Improvements in hearing screening methods have greatly reduced the age at which hearing impairment is identified in infants. Hearing impairment is now typically detected when infants are just a few weeks old. At this age, children are not able to respond to behavioral testing methods, which results in reliance on objective methods to test their hearing (Pimperton & Kennedy 2012). The use of Auditory Brainstem Responses (ABR) or Auditory Steady State Responses (ASSR) to objectively estimate hearing thresholds is now well established (Hall 2007). When children are found to have significant hearing loss, early hearing aid fitting is important to prevent auditory deprivation (Kennedy et al. 2005). A challenge then arises in how to best evaluate whether the hearing aids are giving infants access to sound. One approach to test this may be to compare aided and unaided thresholds using periodic short artificial stimuli, such as those used for ASSRs (Picton et al. 1998).

In general, the stimuli used for ABR and ASSR measurement are artificial and mostly use short duration signals or amplitude modulated tones. There are however limitations with the use of artificial stimuli: hearing aids are primarily designed to amplify speech, so noise reduction algorithms in hearing aids may reduce the amplitude of artificial stimuli that are not classified as speech (Easwar, Glista, et al. 2012; Easwar, Purcell, et al. 2012; Jenstad et al. 2012). Moreover, the compression algorithms used in modern hearing aids may respond in unpredictable ways to very short sounds (Jenstad et al. 2012). Furthermore, although some research has suggested that responses to artificial stimuli could indicate difficulties with higher-level language processes (Wible et al. 2004), further research is required to determine if the ability to detect responses to such artificial stimuli implies that the subject will have good access to natural speech as well.

As a result of these issues, there is current interest in using natural speech stimuli to evaluate infant hearing aids objectively. There is considerable literature on the use of cortical responses to speech in adults (Ding & Simon 2013; Kuruvilla-Mathew et al. 2015; O'Sullivan et al. 2015). Some studies have however suggested that detecting cortical responses in infants can be challenging, due to the lack of maturation of the auditory cortex (Billings et al. 2011). An issue for clinical measurement is maintaining infant alertness for sufficient time to obtain a reliable recording (Picton et al. 2000). Auditory brainstem responses have therefore been suggested as an alternative to cortical responses (Krishnan et al. 2004; Aiken & Picton 2008; Choi et al. 2013). Most of these studies have focused on analyzing vowels using single-polarity envelope following responses (EFR), envelope (eFFR) or spectral frequency following responses (sFFR). Some analysis has also been performed on fricatives (Easwar et al. 2015). For the remainder of this paper, a single-polarity EFR is defined as the coherent average to a single-polarity stimulus. The eFFR is defined as the average of the sum of an equal number of responses to the condensed and rarefied stimulus, whereas the sFFR is the average of the difference between responses to condensed and rarefied stimuli (Aiken & Picton 2008).

A range of methods have been proposed to detect the responses in the EEG following speech stimuli. One study focused on speech evoked response analysis based on comparing the average area under the curve of frequency following responses within syllables (signal) to those between syllables (noise), and comparing the SNRs of forward and reversed speech. Results showed that FFRs to speech had a higher SNR than reversed speech when response amplitudes within segments were compared to FFRs between speech segmetns, indicating a higher brainstem activity towards familiar speech properties (Galbraith et al. 2004). Other studies have looked into spectral analysis related to a flattened fundamental frequency (f_0) of voiced speech, allowing the use of a standard Fast Fourier Transform (FFT) for FFR analysis (Krishnan et al. 2004; Russo et al. 2008). However, as natural speech has a fluctuating f_0, recent studies have explored the detection of brainstem responses to vowels based on a fluctuating f_0 trajectory to improve ecological relevance of stimuli. Aiken & Picton (2006) suggested the use of a Fourier

Analyzer (FA) to determine the spectral amplitude of the f_0 trajectory in the EEG response. Results showed that the spectral amplitude of the EEG response to f_0 and f_1 was significantly stronger than those found for neighboring frequencies. By applying an F-test (FA-F-Test) to determine if the peak at f_0 in the FA spectrum is significantly different from neighboring frequencies, it is possible to detect single-polarity EFRs to vowels in sentences (Choi et al. 2013). It appears promising that the FA in conjunction with an F-test can positively contribute towards assessing hearing loss and evaluating hearing aid fitting within clinically feasible test times (Easwar et al. 2015). However, both studies showed that EFRs were not always detected. Response detection appears to be both subject and stimulus dependent. It is not clear if this has an electrophysiological cause (such as variation in response amplitudes or subject myogenic levels), if there can be problems with the FA to follow the f_0 contour when this is highly fluctuating or if the F-test has insufficient sensitivity. The F-test can be used to determine the significance of peaks in the FA amplitude spectrum, yet several studies investigating techniques for detecting brainstem responses have shown that amplitude-based techniques are significantly less sensitive than phase-based or combined (amplitude and phase) approaches (Dobie 1993). In addition, various reports have shown the efficiency of the Hotelling's T^2 (HT2) tests in detecting ASSRs in both spectral (Mijares et al. 2013) and temporal analysis (Van Dun et al. 2015).

In the current work we further consider the benefit of using both phase and amplitude information in detecting the response to speech. To this end, the HT2 and magnitude squared coherence (MSC) were adapted for application in repeated short duration vowel stimuli and compared to the performance of the F-test after applying the Fourier Analyzer (FA-F-Test). More parameters of the EEG response are included in the HT2 algorithm compared to the FA-F-Test and MSC, as it can analyze both phase and amplitude spectral characteristics of multiple selected frequencies simultaneously. The MSC algorithm analyzes both phase and amplitude from a single frequency and would need to be applied multiple times when multiple frequencies are of interest, with the need for compensating for an increased false positive rate after multiple tests. The FA can be used to estimate both phase and amplitude spectra (Aiken & Picton 2006), yet more recent applications of the FA method have focused on the amplitude spectrum of a single frequency (f_0) only (Choi et al. 2013; Easwar et al. 2015). Given the previous work, we therefore propose the use of HT2 for the detection of the eFFR and hypothesize that the HT2, using multiple frequencies, would outperform the MSC and FA-F-Test in eFFR detection. To test this hypothesis, detection rates will be analyzed in a simulation study and on EEG data from normal hearing subjects. To the best of our knowledge, HT2 has not previously been applied to vowel eFFR detection, with no previous mention of HT2 with simultaneous use of multiple frequencies and multiple EEG channels. The overarching aim is to provide more sensitive methods for assessment of hearing in an ecologically relevant way, through speech. In particular we are seeking more powerful objective methods for evaluating hearing aid fitting in patient groups that cannot reliably respond in psychophysiological tests.

MATERIALS AND METHODS

Participants

This study included 12 normal hearing participants (4 female) aged between 19 and 53 years old (29.0 ± 11.0; mean ± SD). All subjects were recruited locally from the University of Southampton, United Kingdom. Otoscopy was performed to rule out contraindication such as occluding wax, discharge or foreign bodies in the ear. Pure-tone audiometry was performed on all subjects using a GDI-61 audiometer to ensure normal hearing thresholds (< 20 dB HL). It was also confirmed if a conventional ABR could be detected for each participant using 6,000 click stimuli presented at 90 dB peak-to-peak equivalent SPL with a presentation rate of 11.1 Hz. The presence of responses was confirmed using the Fsp method (Elberling & Don 1984). All participants were native English speakers and provided written informed consent to participate. The study protocol was approved by the local ethics committee at the University of Southampton.

Stimuli

Four word stimuli in a hVd format (with V being /a/ pronounced as "had", /ɛ/ pronounced as "hayed", /I/ pronounced as "hid" or /ʊ/ pronounced as "hood", Figure 1) from a single male speaker were taken from a previous study by Hillenbrand et al. (1995), in which participants were asked to speak slowly and prolong the vowel. This was considered beneficial for the current study as it would allow analysis over prolonged time intervals (/a/ 315.4 ms; /ɛ/ 295 ms; /I/ 251.7 ms; /ʊ/ 230.8 ms). Stimuli were presented at 70dB SPL LAeq sequentially via an RME Fireface UC soundcard (Haimhousen, Germany) through ER-2 earphones (Etymotic, IL, USA), with an onset interval of 1 second between onsets of successive words. Using an unweighted intensity-scale (dB SPL Zeq), this resulted in intensities of 71.8 dB SPL Zeq for /a/, 72.3 dB SPL Zeq for /ɛ/, 74 dB SPL Zeq for /I/ and 73.7 dB SPL Zeq for /ʊ/ (see note at end of paper[1]). A total of 440 instances per stimuli were presented (220 rarefaction, 220 condensation), making the total duration of the experiment 29.3 minutes. Breaks were offered according to the subject's convenience. Stimuli were calibrated using a Bruel & Kjaer 2260 Investigator and 4157 occluded ear coupler (Royston, Hertfordshire, UK). Stimulus presentation was controlled using in-house MATLAB scripts (version R2015a, The MathWorks Inc, Natick, MA, USA).

EEG Data Collection

EEG data were collected using a 32-channel ActiveTwo system (BioSemi, Amsterdam, The Netherlands). The electrode locations followed the standard 10-20 setup. Three external electrodes were placed lateral to the eyes and under the chin for artefact detection (eye blinks, swallowing). Throughout the experiment, subjects sat comfortably in a reclining chair and were encouraged to sleep. A control study was performed on 6 subjects in which data were collected when presenting stimuli after the earphones were taken out of the ears and ears were blocked with earplugs. Apart from this, the setup remained the same as when the subject was listening to the stimuli. For all experiments, raw EEG data were collected at 16.384 kHz, and further processed offline using MATLAB. Standard pre-processing of EEG signals was performed similarly for each of the algorithms. After referencing to the average of the mastoid electrodes, EEG data were resampled to 2048 Hz after band-pass filtering between 80 and 1000 Hz using a 7^{th} order Butterworth filter. An artefact rejection threshold was set as the mean+2 x standard deviation (SD) as calculated over the entire EEG signal. In case an epoch reached amplitudes above the threshold, the entire epoch was rejected from the analysis. For each 1-second epoch, EEG signals corresponding to vowel segments including a 10 ms delay due to brainstem processing were extracted (Easwar et al. 2015). Afterwards, each of the algorithms was applied as discussed below.

Detection Algorithms

The following paragraphs will provide a brief theoretical overview of each of the algorithms included in this study, as well as a description on how each algorithm was implemented in the analysis.

F-test on Fourier Analyzer's amplitude spectrum

The Fourier Analyzer can be considered an adapted Fourier Transform, where the reference sinusoids are allowed to vary in frequency within an analysis window. This adaptation is designed to take into account the time-varying fundamental frequency that is typical of normal voiced speech. In cases where the frequency track is known or can be accurately estimated, the FA will outperform the standard Fourier Transform (Aiken & Picton 2006, 2008). To implement the FA, f_0 reference sinusoids were created for each stimulus as suggested previously (Aiken & Picton 2006). Briefly, the vowel stimulus signals were filtered between 50 and 200 Hz using a 7^{th} order Butterworth filter to include only the fundamental frequency

(frequency ranges were determined in Hillenbrand et al. (1995)). The instantaneous phase was then calculated using the four-quadrant inverse tangent on the Hilbert transform of the filtered stimulus, whose gradient (normalized by 2π) gives the instantaneous frequency. This instantaneous frequency was used to create f_0 tracks for each of the vowel (Figure 2). For the four stimuli, the f_0 values ranged from 156-166.4 Hz (had), 161.4-173.9 Hz (hayed), 174-180.2 Hz (hid) and 165.8-169.8 Hz (hood). An f_0-reference sine and cosine were calculated as the sine and cosine of the unwrapped phase, respectively. To compare the energy at f_0 with non-stimulus frequencies, 10 tracks above and 10 tracks below the f_o reference were also created. An F-test with 2 and 40 degrees of freedom was then performed on the ratio of the f_0 power to the mean power of the neighboring frequencies (FA-F-Test). The frequency step between the neighboring tracks was the reciprocal of the duration of the vowel stimulus, corresponding to the standard frequency resolution of an FFT. To determine if the detection rate of the FA-F-Test is frequency-dependent, reference sinusoids and tracks were also derived for the harmonics of f_0 up to $4f_0$. For each stimulus, the spectral power for each track was calculated by multiplying the reference sinusoids with the EEG for a single rarefaction and condensation stimulus presentation (one response pair consists of one rarefaction and one condensation stimulus epoch) and integrating the result after multiplication over the entire vowel length. A response was considered present if the ratio of energy at f_0 over the average energy of the neighboring frequencies was higher than the critical value for an F test (Choi et al. 2013) at a 5% alpha level, which corresponds to a critical value of 3.23. This procedure was repeated, adding one epoch pair (each added pair again consists of one rarefaction and one condensation epoch) to the algorithm over successive iterations. Iterations ran until a response was detected or all epoch pairs were included in the analysis. A subject was considered to have a response when 4 successive iterations produced a significant F-ratio, and a minimum of 8 iterations were included in the analysis to control for false positives, as previously suggested by Choi et al. (2013). Analysis was performed on each of the harmonics of f_0 individually.

Magnitude Squared Coherence

The MSC (γ_{xy}^2) estimates how well a response correlates to its stimulus, in the frequency domain. This association is calculated via a normalized cross-spectral density function, thereby using both amplitude and phase information of the signal (Dobie & Wilson 1989). Mathematically, for N windows and a repeated stimulus,

$$\gamma_{xy}^2(f) = \frac{\left|\sum_{i=1}^{N} y_i(f)\right|^2}{N\sum_{i=1}^{N}|y_i(f)|^2} \tag{1}$$

Where $y_i(f)$, is the Fourier Transform of the i^{th} response. To determine if a response is present, the p-value of the MSC at f_0 can be derived from the F-statistic as follows (Dobie & Wilson 1989):

$$\frac{(N-1)\gamma_{xy}^2(f_0)}{1-\gamma_{xy}^2(f_0)} \sim F_{2,2(N-1)} \tag{2}$$

In this study, the MSC was evaluated for periods with a duration equal to $1/f_0$. To generate these windows, time points of the FA reference sinusoid peaks were identified. These time points were used to window the individual epochs (rectangular window). As the vowel stimuli were of different length for the four different speech stimuli, 23 to 25 f_0 peaks could be detected for the different stimuli, resulting in a total of between 23 and 25 windows for analysis per epoch. The DFT was then applied to each window and the MSC calculated according to equation (1) (Cooley & Tukey 1965). Significance was determined using equation (2). As with the FA-F-Test, this process started by including 1 epoch pair in the analysis and was repeated with increasing numbers of epochs until all EEG pairs were included or a response was detected. The process was also performed on individual harmonics of f_0 up to $4f_0$.

Hotelling's T^2 Test

The HT2 for frequency domain analysis has been used in various forms for the detection of both ABRs and ASSRs (Rodriguez et al. 1986; Valdes-Sosa et al. 1987; Victor & Mast 1991; Valdes et al. 1997; Mijares et al. 2013). In principle, HT2 can be considered a multivariate Student's t-test, in which the difference between the means of Q features and Q hypothesized values is tested for significance. In the current study, the Q features are the real and imaginary parts of the Fourier coefficients ($y_i(f)$) in the relevant frequencies (f_0 and harmonics). Under the null-hypothesis of no response, the expected mean values are all zero. The test statistic is given by (Hotelling 1931):

$$HT2 = N(\bar{q} - \mu_q)S^{-1}(\bar{q} - \mu_q)' \tag{3}$$

where N denotes the number of epochs, $\bar{q}$ the Q-dimensional vector of means of the real and imaginary part of the Fourier transforms, μ_q the vector of hypothesized means, and S^{-1} the inverse of the covariance matrix of the $N \times Q$-dimensional feature matrix. The HT2 algorithm then tests whether the means in vector $\bar{q}$ are significantly different from μ_q. As with the MSC, the HT2 test can be transformed into an F-test according to equation (6) (Valdes et al. 1997):

$$\frac{N-Q}{Q(N-1)}T^2 \sim F_{Q,N-Q} \tag{4}$$

To allow for a direct comparison with the FA-F-Test and MSC algorithm, HT2 analysis was first performed using only one frequency (HT2_F0). The same iteration process was followed as with the MSC: analysis was performed on pairs of EEG responses and additional pairs were averaged in separate iterations until all response were included.

The effect of adding harmonics on response detection was analyzed by including the $2f_0$ and $3f_0$ bins (HT2_3F) in the HT2 algorithm. As the average bin width for the stimuli was 3.49±0.40 Hz, it was possible to select a single bin for each of the harmonics and still capture most of the power in the harmonics, in spite of the time-varying fundamental frequency (see

Figure 1) without losing signal power as fundamental frequency moves beyond the bin-width. This combination of frequencies showed the highest detection rate out of all possible combinations of harmonics up to $4f_0$. Lastly, the possibility of further increasing detection rate (and decreasing detection time) by merging several EEG channels in the HT2_3F analysis (multichannel HT2 – HT2_MC) was explored, by augmenting the Q features with those from the additional channels. In the following, results of all algorithms will be compared with a multichannel HT2 including 5 channels at locations Cz, Pz, Fz, C3 and C4, as this combination showed the highest detection rate after performing the test using 1, 3, 5, 9, 18 and 32 electrodes.

Simulation Data Analysis

To test the potential benefits of adding multiple frequencies for detection of brain responses, a simulation study was set up. A simulation allows the algorithms to be assessed in well-controlled conditions and identify sources that contribute to the relative performance of the methods. The simulation aims to emulate the time-varying frequency characteristics of the vowel stimulus. Therefore, signals made up of the vowel part of the /had/ stimulus were generated, as well as its time-varying f_0 reference sinusoid extracted for the FA algorithm were used. Copies of these references were then combined with white Gaussian noise at signal-to-noise ratios (SNR) varying between 0dB and -60dB SNR when measured over the full stimulus spectrum. For each test, the number of epochs generated was made equal to that obtained in the EEG recordings (440). Response detection was then performed using the FA-F-Test, MSC, HT2_F0 and HT2_3F algorithm as discussed above. The simulation was repeated 1,000 times for each SNR. Results are reported as a percentage of detection at the estimated SNR in each frequency bin. The SNRs were also estimated for $2f_0$ and $3f_0$ as these frequencies were included in the algorithm. As the SNRs at these frequencies were consistently 0.5 dB and 1.9 dB lower compared to the SNR at f_0 (since the stimulus amplitude is lower at higher frequencies and the noise is white), results were not reported separately, but correspond to those at the fundamental frequency, shifted by 0.5 and 1.9 dB respectively.

A final simulation was performed on a constant sinusoid (i.e. a flat f_0) at 159 Hz. This frequency was chosen as it is within the range of the fundamental frequencies of the vowel stimuli. One additional simulation was performed on 1,000 iterations of pure white Gaussian noise (SNR $= -\infty$) to check the false-positive rate for each of the tests.

Statistics

Data were tested for normality using the Shapiro-Wilk test. For both simulation and EEG data, a Cochran's Q-test was used to determine if the ratio of no detection vs detection was significantly different over the algorithms, combining results from all four stimuli and across all subjects. Between-method differences were analyzed using a McNemar test. For the EEG data, analysis was performed by first including all four stimuli, followed by an analysis on individual stimuli to assess if different words led to similar results or not. Furthermore, a Friedman test was used to determine if the detection time was significantly different over the algorithms. For this task, only stimuli for which each of the algorithms was able to detect an eFFR were included (28 out of 48 tests), to avoid biasing results by including some stimuli with some algorithms but not with others. A Wilcoxon signed-rank test was used to compare differences between pairs of algorithms. In order to compare results with previous work, single-channel response analysis was first performed on the vertex electrode (Cz) only. Then analysis was extended to include all 32 channels in order to compare differences in detection rate and time for different EEG locations. Statistical significance was determined at an alpha level of 0.05. A Bonferroni correction was applied in case of repeated measures. All statistical tests were performed using SPSS Statistics 22 (IBM, Armonk, NY, USA). Graphs were created using GraphPad Prism 7.0 (La Jolla, CA, USA).

RESULTS

Simulations

Results from the simulation studies are shown in Figure 3. When detecting a vowel stimulus in noise (Figure 3A), the HT2_3F algorithm appears to significantly outperform the other algorithms for SNRs between -3.9 dB and -9.95 dB (Cochran's Q test, $p<0.001$, all SNRs based on an estimated SNR in a 10 Hz frequency band around f_0). The FA-F-Test and HT2_F0 achieved an almost equal detection rate, and both performed significantly better than the MSC algorithm in this range ($p < 0.03$ in all cases). For the HT2_3F algorithm, 100% detection can be achieved above an SNR of about -4.95 dB. The other algorithms appear to drop below a 100% detection rate from an SNR equal to -1.43 dB. In comparison, when only the constant-frequency sinusoid at f_0 was present (Figure 3B), the MSC and HT2_F0 algorithm significantly outperform the FA-F-Test and HT2_3F algorithm between SNRs of -12.9 and -15.09 dB ($p<0.001$). In the same SNR range, the FA-F-Test significantly outperforms HT2_3F ($p<0.001$), the latter being handicapped by including two frequency bins without a response present. Detection rates of 100% appear to be achievable for SNRs down to -12.5 dB for all algorithms. When the f_0 frequency was flattened to 159 Hz, all algorithms had a 100% detection rate down to an SNR of -10.7 dB, and the FA-F-Test, MSC and HT2_F0 significantly outperformed ($p<0.001$) the HT2_3F algorithm between -11.6 and -13.7 dB (Figure 3C). In summary, the results show that in the presence of harmonics, the HT2_3F algorithm outperforms the others. When there are no harmonics in the stimulus-response, HT2_3F performs worse than the alternatives, as might be expected. The frequency adaptive behavior of the FA applied in the FA-F-Test algorithm however provided no advantage compared to the fixed-frequency detectors of HT2_F0 and MSC. This clearly demonstrates the potential benefit of the HT2 methods, but so far only in simulations. Results from recorded signals are presented in the next section.

EEG Data Analysis

Figure 4 shows the detection rates for single-frequency algorithms for different harmonics of f_0 on the EEG data. The detection rates for the optimized HT2_3F (using multiple frequencies for a single channel) and HT2_MC (using multiple frequencies for multiple channels) algorithm are also indicated. For each of the single-frequency algorithms, detection rate at f_0 is significantly higher (Cochran's Q test, $p<0.01$) compared to the other harmonics, which do not perform significantly different from one another. Detection rates at f_0 are always above 68%, whereas detection rates for other harmonics are below 50%. Further analysis was therefore only focused on detection rates at f_0 for the single-frequency algorithms.

Detection rates for the different statistical methods are shown in Figure 5. When comparing the detection rates combining all stimuli (4 stimuli for 12 participants = 48 tests, Figure 5A), the HT2_MC algorithm has a significantly higher detection rate than the FA-F-Test and MSC algorithms (Cochran's Q test, $p<0.05$). Although not shown in the figure, for the no stimulation (control) condition, detection rates stayed below the 5% chance level used in the current work for all algorithms. Detection rates for individual stimuli did not significantly differ, but the HT2 algorithms consistently showed (visually) equal or higher detection rates than the other algorithms (Figure 5B). Between individual subjects, it could be observed that there was high variability in detection and time needed for detection of an eFFR. The detection rates using the HT2_3F algorithm for two subjects shown in Figure 5C provide an example of this inter-individual variation. Subject 1 appeared to need a much lower number of repetitions for stimuli /hayed/ and /hid/ than subject 2. On the other hand, an eFFR to /hood/ could be detected for subject 2, which was not the case for subject 1. The opposite occurred for stimulus /had/. Three subjects were considered poor responders, as for these subjects a response could only be observed for a maximum of 2 stimuli for each of the algorithms.

Besides number of detections, a comparison was also performed on the time needed for each of the algorithms to detect a response. Figure 6A shows the spread of detection time for each algorithm over all stimuli. To facilitate interpretation, results are only shown for stimuli for which each of the algorithms was able to detect a response (28 in total). Datasets were removed from this part of this study due to three subjects with poor responses over all algorithms (removing 12 of a total of 48 comparisons). Other removed datasets (8) were due to the FA-F-Test and/or MSC algorithm not detecting a response. Inclusion of these datasets would have led to inequitable comparisons. Median detection times over all stimuli were 74s (50.5s-135s; interquartile range) for FA-F-Test, 82.5s (40s-122.75s) for MSC, 70s (42.5s-108.5s) for HT2_F0, 72s (50s-121s) for HT2_3F and 63s (49.25s-85.75s) for HT2_MC. A Friedman test with Bonferroni correction showed significant differences in detection time between HT2_MC and FA-F-Test (<0.001) and HT2_F0 and FA-F-Test (<0.05). It may be noted that it is particularly the upper quartile which is reduced with the HT2 methods. For individual stimuli, detection time for HT2_MC was significantly shorter than for FA-F-Test ($p<0.01$) for the /hayed/ stimulus (Figure 6B). The HT2_F0 also detected responses significantly quicker than the FA-F-Test for this stimulus ($p<0.05$).

Lastly, differences in detection rate and detection time were compared for individual electrode locations using the FA-F-Test, MSC, HT2_F0 and HT2_3F algorithm (Figure 7). Regarding detection rates, the FA-F-Test algorithm had a significantly lower detection rate than the HT2_3F algorithm ($p<0.01$) due to a significantly lower detection rate following the /had/ stimulus ($p<0.05$). When comparing electrode locations, it could be observed that detection rates around the vertex were higher than for more peripheral locations. These differences occasionally reached significance. Detection times followed a similar pattern to the detection rate.

DISCUSSION

In the evaluation of hearing aid fittings, especially for infants, a possible approach is to record responses to speech sounds in the EEG signal (Picton et al. 1998; Van Dun et al. 2012). EEG systems have good clinical applicability for testing infants (John et al. 2004), they can be made portable and can potentially even be integrated into hearing aid devices using in-the-ear EEG systems (Looney et al. 2012) which may allow extended measurement time. One reason to use natural speech over artificial stimuli, such as those used in ABR or ASSR, is that current hearing aids are programmed to detect and process speech stimuli in a different way to non-speech stimuli. A second reason to use natural speech stimuli is that it has face validity: in the real world, we generally want to listen to natural speech and not clicks or pips. This work compared statistical methods for the detection of responses to naturally produced vowels. These methods differ in the number of signal features they include when analyzing the response. The FA, as discussed in recent papers (Choi et al. 2013; Easwar et al. 2015), estimates the amplitude spectrum determined from non-stationary frequency tracks. Detection can be based on comparing the amplitude of a frequency of interest (i.e. f_0) to neighboring frequencies (Choi et al. 2013), or the amplitude of f_0 and its harmonics (Aiken & Picton 2008). The MSC uses both amplitude and phase features at f_0 to determine if a response is present based on a Fast Fourier Transform (FFT) (Dobie & Wilson 1989). The HT2 test detects responses by including phase and amplitude features (given by the real and imaginary) from an FFT from predetermined harmonics (Valdes et al. 1997). In this study, the HT2 tests were performed using just f_0 (HT2_F0) as well as using the first 3 harmonics (HT2_3F) as this showed the highest detection rate out of all possible combinations of harmonics between f_0 and $4f_0$. Finally, the HT2 test can also very readily combine responses from multiple channels, as implemented here in the HT2_MC test.

Having proposed the use of these methods for the detection of speech evoked responses, the main aim of this work was to determine which (if any) of these techniques has a higher sensitivity in detecting a response. Responses were detected objectively using simulation studies as well as EEG data collected from normal hearing participants.

Results from simulation studies showed that the HT2_3F algorithm significantly outperformed the other algorithms in detecting a response based on a vowel stimulus in noise. As the simulated response contains the vowel's full frequency content, HT2_3F can detect a response at any of the harmonics included in the algorithm. It therefore uses more information from the response which increases its ability to detect responses in noise compared to other frequency-domain algorithms which only use a single frequency to perform statistical analysis. The MSC and HT2_F0 are quite similar in concept, but while the MSC assumes that the real and imaginary part of the Fourier coefficients are uncorrelated and have equal variance, the HT2_F0 (and indeed HT2_3F and HT2_MC) use covariance estimates (S^{-1} in equation (3)) to account for deviations from this assumption. The HT2 is thus also well suited for combining different frequencies with different amplitudes (as in HT2_3F) or the signals from multiple electrodes (HT2_MC), as it has normalization as part of the calculation. Including multiple frequencies or multiple EEG channels in for example the MSC is not such a simple extension of the basic algorithm, as variances differ between frequencies, and signals from adjacent channels are correlated (De Sá et al. 2004).

When only f_0 was present, MSC and HT2_F0 performed equally well for both a time-varying and flattened f_0, while the FA-F-Test and HT2_3F algorithm performed significantly worse. When a response is mainly determined by a single frequency (as is the case in the f_0 simulation), the HT2_3F is expected to perform worse, as it includes additional frequencies which contain no information on the response but only noise (at $2 \cdot f_0$ and $3 \cdot f_0$) into the calculations. The similarity in performance of HT2_F0 and MSC is expected, given the similarity in theoretical formulation as discussed in the previous paragraph. The likely reason for HT2_F0 and MSC outperforming FA-F-Test in detecting the f_0 response is the inclusion of phase in the algorithm, rather than only spectral power. This is in agreement with previous studies on ABR detection to click stimuli, which have shown that the inclusion of an increased number of EEG features to determine if a response is present increases detection rates in simulations (Dobie & Wilson 1993; Cebulla et al. 2006), although in our case the effect of including phase seems smaller than previously suggested. Phase characteristics can be estimated from the FA algorithm (Aiken & Picton 2006), which could potentially raise its performance to the level of the of the MSC or HT2_F0 algorithm. The simulations clearly illustrate the potential benefit of using multiple frequencies, as well as limitations of different algorithms. The simulations can show these under very tightly controlled conditions, without the confounding effects of within and between individual differences, and residual noise or artefacts. This facilitates understanding of inherent limitations of the algorithms and their performance on specific recorded datasets.

Simulation studies as used in this paper might however not directly infer how the objective tests perform on real EEG data, as they do not approximate eFFR detection in skewed EEG noise (Özdamar & Delgado 1996). To determine if the simulation results could be reproduced on EEG data, detection rates and times were analyzed on 12 participants with normal hearing thresholds. As the vertex is known to be the preferred location for ABR detection (Jewett & Williston 1971), and some previous studies detected EFR responses to vowels at the vertex only (Krishnan et al. 2004; Choi et al. 2013; Easwar et al. 2015), it was decided to first compare the different test methods at this single location. All participants showed a response to at least one of the stimuli using at least one of the methods. From Figure 4, it can be observed that a response could be detected in at least 68% of measurements, depending on the detection method used. This detection rate is similar to the rates observed in a previous study on participants with normal hearing thresholds, which reported detection rates between 60 and

100% for FA-F-Test at 70 dBA SPL (Choi et al. 2013). It should be noted that the calibration procedures are not entirely compatible between their work and the present one, and it is possible that in their work the words (or vowels) themselves were presented at higher intensity levels than in our study, which is known to lead to stronger responses.

For the single-frequency algorithms, it could be observed that a significantly higher detection rate was achieved when analyzing responses to f_0 compared to its harmonics. Previous studies on vowels also showed a higher response strength to f_0 when a coherent average of alternating stimulus polarities was presented (Aiken & Picton 2008). Recent studies on eFFRs to vowels have also mostly focused on f_0 (Choi et al. 2013; Easwar et al. 2015), whereas higher harmonics have been used for assessing response to the temporal fine structure of speech or spectral FFRs (Krishnan, 2016). These results might also be relevant to the HT2_3F algorithm not significantly outperforming the single-frequency algorithms on EEG data (Figure 4), and goes some way to explain why including additional spectral characteristics at harmonics add only relatively small benefit in detecting eFFRs. Considering the encouraging results from the simulation, the results from EEG signals suggest that the time-varying properties of the stimulus are not the main limitation of the HT2_3F algorithm in this data, but rather the relative strength of responses at different harmonics. Including multiple channels in the HT2 algorithm did however show a significant increase in detection rate, further indicating that performing analysis on multiple harmonics and locations could improve response detection. The lack of a significant difference between single-channel algorithms could be due to a low sample size (48 tests). Some studies measuring ASSRs to tone stimuli did show an increase in detection rate when using phase features or both amplitude and phase features for response detection compared to only using amplitude features (Picton et al. 1987; Cebulla et al. 2001; Picton et al. 2001; Cebulla et al. 2006). One previous study detecting responses to vowels did however show a higher detection rate for the FA-F-Test when comparing to a single-frequency circular

HT2 test (Aiken & Picton 2006), which differs from the traditional HT2 test in that it assumes variables to be tested are independent quantities with a Gaussian distribution of equal variance (Victor & Mast 1991). Finally, the inter-individual variation in eFFR detection rate (Figure 5B), which was also found by Choi et al (2013) should be noted, together with the inconsistent results from different stimuli. This indicates that for clinical purposes, eFFR detection will potentially benefit from using not just a single word stimulus, but rather a combination of different stimuli.

The median detection time appeared similar for the different algorithms. A Friedman test with post-hoc analysis did however show that the time elapsed until detection for the HT2_F0 and HT2_MC algorithm was significantly shorter than for the FA-F-Test algorithm, even in a small sample of 12 subjects. HT2 algorithms also seemed to have a reduced spread in detection time (and in particular a reduction in the upper quartile), indicating that it might be a more convenient test to use in a clinical environment, where longer-duration tests are more difficult to achieve. Based on the responses detected in the current sample, it was possible to detect an eFFR within 5 minutes of stimulus presentation in at least 68% of subject independent of the HT2 algorithms. This is similar to the number of sweeps indicated for a detectable response using the FA-F-Test algorithm by Choi et al. (2013) – though as indicated above, the results are not necessarily comparable. It also further indicates the potential for improved eFFR detection when analyzing spectral amplitude and phase characteristics of the response simultaneously.

Lastly, analysis over all electrode locations shows a significantly higher detection rate for the HT2_3F algorithm compared to the FA-F-Test. Comparison of different electrode locations shows that detection rates around Cz are higher (and detection times significantly lower) than for more peripheral electrode locations. The observation of better detection at or around Cz has been confirmed previously (Plourde 2006), and is believed to be caused by the tangential activity of the brainstem source being stronger than the radial activity (Herdman et al. 2002). Early studies on detection of click ABR responses also showed robust detection at the Cz location (Jewett & Williston 1971). In cases where EEG data are to be used for diagnosis of hearing loss and/or optimizing hearing aids when limited electrodes are available, it is recommended to position these electrodes at or close to the vertex.

STUDY LIMITATIONS

This study focused on detection of responses in participants with normal hearing levels, whereas the aim, ultimately, is to use speech evoked responses to evaluate hearing aids in hearing impaired subjects. A study comparing the different methods in a cohort of mild to moderate hearing loss subjects to determine the performance of the different detection methods in more clinically relevant situations is currently underway. Besides this, other methods have been suggested for the detection of brainstem responses, such as the modified Rayleigh test (Moore 1980) and the q-sample test (Cebulla et al. 2006). Future studies could perform more extensive comparisons by including additional methods in the analysis. While the simulation study provided a means of testing relative benefits and limitations of the different algorithms by performing tests in well-controlled conditions, it does not cover the range of challenges encountered in clinical tests. Apart from neglecting individual variability in responses and the different responses to different vowels, it also assumes white stationary noise. Non-stationary noise is a major challenge whose effects were not simulated.

CONCLUSION

This paper proposes the HT2 algorithm applied in the frequency-domain as a means for objectively detecting envelope frequency following responses to natural vowels. It compared the performance of HT2 against the FA-F-Test and MSC algorithm, which use different features of the EEG signal to determine if a response is present. Differences in detection rates at different measurement locations were also analyzed. Simulation studies showed that the HT2 algorithm can achieve a higher detection rate than the FA-F-Test for single-frequency responses. When more complex responses were simulated, the ability of HT2 to analyze multiple frequencies of interest improved detection rate. Detection rates can be improved yet further if the HT2 is extended to analyze variables of interest at multiple locations of the scalp (multi-channel HT2). For detection of eFFRs, we propose that these additional electrodes should be positioned around Cz (e.g. Pz, Fz, C3 or C4) to be beneficial. Lastly, HT2 algorithms appeared to be most consistent in the time needed to detect a response, allowing detection within 5 minutes, which is important for potential clinical applications.

ACKNOWLEDGEMENTS

The experiments were designed by F.J.V., S.L.B. and D.M.S.; F.J.V. and M.C. performed the experiments and analyzed the data; F.J.V. wrote the paper; and S.L.B. M.C. and D.M.S. provided critical revision. The authors would like to thank Louise Goodwin for her technical support in the experimental setup. This research project is funded by the Engineering and Physical Sciences Research Council (EPSRC, grant No. EP/M026728/1). All data supporting this study are openly available from the University of Southampton repository at http://doi.org/10.5258/SOTON/D0005.

NOTES

[1]At the recent International Evoked Response Audiometry Study Group Biennial Symposium (IERASG2017, Krakow, Poland), it was argued that stimulus levels should be reported using flat weighting (dB Z) as low frequencies of formants can be affected by A-weighting. As previous papers in the field have used dB A, we report both to indicate the need for conversion in future studies, but allow comparison of the results of this paper with previous work

REFERENCES

Aiken, S. J., Picton, T. W. (2006). Envelope following responses to natural vowels. *Audiol Neurootol, 11*, 213-232.

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

Billings, C. J., Tremblay, K. L., Miller, C. W. (2011). Aided cortical auditory evoked potentials in response to changes in hearing aid gain. *Int J Audiol, 50*, 459-467.

Cebulla, M., Stürzebecher, E., Elberling, C. (2006). Objective detection of auditory steady-state responses: Comparison of one-sample and q-sample tests. *J Am Acad Audiol, 17*, 93-103.

Cebulla, M., Stürzebecher, E., Wernecke, K.-D. (2001). Objective Detection of the Amplitude Modulation Following Response (AMFR). *Audiology, 40*, 245-252.

Choi, J. M., Purcell, D. W., Coyne, J.-A. M., et al. (2013). Envelope following responses elicted by English sentences. *Ear Hear, 34*, 637-650.

Cooley, J. W., Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Math Comput, 19*, 297-301.

De Sá, A. M., Felix, L. B., Infantosi, A. F. C. (2004). A matrix-based algorithm for estimating multiple coherence of a periodic signal and its application to the multichannel EEG during sensory stimulation. *IEEE Trans Biomed Eng, 51*, 1140-1146.

Ding, N., Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J Neurosci, 33*, 5728-5735.

Dobie, R. A. (1993). Objective response detection. *Ear Hear, 14,* 31-35.

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. (1993). Objective response detection in the frequency domain. *Electroen Clin Neuro, 88*, 516-524.

Easwar, V., Glista, D., Purcell, D. W., et al. (2012). Hearing aid processing changes tone burst onset: effect on cortical auditory evoked potentials in individuals with normal audiometric thresholds. *Am J Audiol, 21*, 82-90.

Easwar, V., Purcell, D. W., Aiken, S. J., et al. (2015). Evaluation of Speech-Evoked Envelope Following Responses as an Objective Aided Outcome Measure: Effect of Stimulus Level, Bandwidth, and Amplification in Adults With Hearing Loss. *Ear Hear, 36*, 635-652.

Easwar, V., Purcell, D. W., Scollie, S. D. (2012). Electroacoustic comparison of hearing aid output of phonemes in running speech versus isolation: Implications for aided cortical auditory evoked potentials testing. *Intl J Otolaryng, 2012*.

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

Galbraith, G. C., Amaya, E. M., Diaz de Rivera, J. M., et al. (2004). Brain stem evoked response to forward and reversed speech in humans. *NeuroReport, 15*, 2057-2060.

Hall, J. W. (2007). *New handbook of auditory evoked responses*. (1 ed.). London: Pearson.

Herdman, A. T., Lins, O., Van Roon, P., et al. (2002). Intracerebral sources of human auditory steady-state responses. *Brain topogr, 15,* 69-86.

Hillenbrand, J., Getty, L. A., Clark, M. J., et al. (1995). Acoustic characteristics of American English vowels. *J Acoust Soc Am, 97,* 3099-3111.

Hotelling, H. (1931). The generalization of Student’s ratio. *Ann of Math Stat, 2*, 360-378.

Jenstad, L. M., Marynewich, S., Stapells, D. R. (2012). Slow cortical potentials and amplification—Part II: Acoustic measures. *Int J Otolaryngol, 2012*.

Jewett, D. L., Williston, J. S. (1971). Auditory-Evoked Far Fields Averaged from the Scalp of Humans. *Brain, 94*, 681-696.

John, M. S., Brown, D. K., Muir, P. J., et al. (2004). Recording auditory steady-state responses in young infants. *Ear Hear, 25*, 539-553.

Kennedy, C. R., McCann, D., Campbell, M. J., et al. (2005). Universal newborn screening for permanent childhood hearing impairment: an 8-year follow-up of a controlled trial. *Lancet, 366*, 660-662.

Krishnan, A., Xu, Y., Gandour, J. T., et al. (2004). Human frequency-following response: representation of pitch contours in Chinese tones. *Hearing Res, 189*, 1-12.

Kuruvilla-Mathew, A., Purdy, S. C., Welch, D. (2015). Cortical encoding of speech acoustics: Effects of noise and amplification. *Int J Audiol, 54*, 852-864.

Looney, D., Kidmose, P., Park, C. S., et al. (2012). The in-the-ear recording concept: User-centered and wearable brain monitoring. *IEEE Pulse, 3*, 32-42.

Mijares, E., Pérez Abalo, M. C., Herrera, D., et al. (2013). Comparing statistics for objective detection of transient and steady-state evoked responses in newborns. *Int J Audiol, 52*, 44-49.

Moore, B. R. (1980). A modification of the Rayleigh test for vector data. *Biometrika, 67*, 175-180.

O'Sullivan, J. A., Power, A. J., Mesgarani, N., et al. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb Cortex, 25*, 1697-1706.

Oostendorp, T. F., van Oosterom, A., Huiskamp, G. (1989). Interpolation on a triangulated 3D surface. *J Comput Phys, 80*, 331-343.

Oostenveld, R., Fries, P., Maris, E., et al. (2010). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci, 2011*.

Özdamar, Ö., Delgado, R. E. (1996). Measurement of signal and noise characteristics in ongoing auditory brainstem response averaging. *Ann Biomed Eng, 24*, 702-715.

Picton, T. W., Bentin, S., Berg, P., et al. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology, 37*, 127-152.

Picton, T. W., Dimitrijevic, A., John, M. S., et al. (2001). The use of phase in the detection of auditory steady-state responses. *Clin Neurophysiol, 112*, 1698-1711.

Picton, T. W., Durieux-Smith, A., Champagne, S. C., et al. (1998). Objective evaluation of aided thresholds using auditory steady-state responses. *J Am Acad Audiol, 9*, 315-331.

Picton, T. W., Vajsar, J., Rodriguez, R., et al. (1987). Reliability estimates for steady-state evoked potentials. *Electroen Clin Neuro, 68*, 119-131.

Pimperton, H., Kennedy, C. R. (2012). The impact of early identification of permanent childhood hearing impairment on speech and language outcomes. *Arch Dis Child, 97*, 648-653.

Plourde, G. (2006). Auditory evoked potentials. *Best Prac Res Clin Anaesthesiol, 20*, 129-139.

Rodriguez, R., Picton, T. W., Linden, D., et al. (1986). Human auditory steady state responses: effects of intensity and frequency. *Ear Hear, 7*, 300-313.

Russo, N. M., Skoe, E., Trommer, B., et al. (2008). Deficient brainstem encoding of pitch in children with Autism Spectrum Disorders. *Clin Neurophysiol, 119*, 1720-1731.

Valdes-Sosa, M. J., Bobes, M. A., Perez-Abalo, M. C., et al. (1987). Comparison of auditory-evoked potential detection methods using signal detection theory. *Audiology, 26*, 166-178.

Valdes, J. L., Perez-Abalo, M. C., Martin, V., et al. (1997). Comparison of statistical indicators for the automatic detection of 80 Hz auditory steady state responses. *Ear Hear, 18*, 420-429.

Van Dun, B., Carter, L., Dillon, H. (2012). Sensitivity of cortical auditory evoked potential detection for hearing-impaired infants in response to short speech sounds. *Audiol Res, 2*, e13.

Van Dun, B., Dillon, H., Seeto, M. (2015). Estimating hearing thresholds in hearing-impaired adults through objective detection of cortical auditory evoked potentials. *J Am Acad Audiol*, *26*, 370-383.

Victor, J. D., Mast, J. (1991). A new statistic for steady-state evoked potentials. *Electroen Clin Neuro, 78*, 378-388.

Wible, B., Nicol, T., Kraus, N. (2004). Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol Psychol, 67*, 299-317.

FIGURE CAPTIONS

Figure 1: Stimuli used for this study (top images) along with their spectrograms (bottom) indicating f_0 and its harmonics. A threshold between 0 and 200 dB (with respect to an arbitrary reference) was applied to the spectrograms to highlight the harmonics.

Figure 2: Tracks of the fundamental frequencies for each of the vowels.

Figure 3: Average detection rates for simulated responses (1000 simulations, 440 epochs per stimulus) based on a vowel stimulus input (A), a time-varying f_0 reference sinusoid input (B), and a flattened f_0 (C, 159 Hz) sinusoid mixed with white Gaussian noise at different signal-to-noise ratios (SNR). Note that in (B) the results for MSC and HT2_F0 almost coincide. SNRs are reported as SNRs in the bin of the DFT around the fundamental of the stimulus. The SNRs at $2f_0$ and $3f_0$ were about 0.5 dB and 1.9 dB lower, respectively.

Figure 4: Comparison of detection rates for individual harmonics of f0 using the FA-F-Test, MSC and HT2_F0 algorithm in EEG data. Detection rates for HT2_3F (red bar) and HT2_MC including channels Cz, Pz, Fz, C3 and C4 (blue bar) are also indicated. For all single-frequency algorithms, a significantly higher number of detections was achieved with f_0 compared to its harmonics (2F0, 3F0, 4F0).

Figure 5: A, Detection rates for the different statistical methods with real EEG data. A detection rate of 100% corresponds to a detection for 4 (stimulus types)*12 (subjects) = 48 tests. Significant differences in detection rates could be found between the FA-F-Test and multichannel HT2_3F (HT2_MC), as well as the MSC and multichannel HT2_3F algorithm ($p<0.05$, Bonferroni correction). B, Detection rates per stimulus for the different detection methods. No significant differences could be found, although the HT2 algorithms performed better or equally well compared to the FA-F-Test and MSC algorithm (12 tests per stimulus). C, Detection time for two subjects for each stimulus using HT2_3F. A response is detected at

a significance level of 5% (dotted line). Response detection, as well as detection time, clearly varies for different stimuli and different participants. Asterisks indicate statistical significance ($p<0.05$).

Figure 6: A, Detection time over all stimuli for individual algorithms. The HT2 and HT2_MC algorithms require significantly shorter times to detect a response compared to the FA-F-Test algorithm. B, Detection time for individual stimuli. Significant differences could be found between the FA-F-Test and HT2 algorithm and the FA-F-Test and HT2_MC algorithm for the /hayed/ stimulus. Asterisk indicate statistical significance at alpha = 0.05 (*), alpha = 0.01(**) and alpha = 0.001 (***).

Figure 7: Percentage of detection over all patients for different electrode locations. Data are presented for each method and each word. Colour interpolation was performed using the surface Laplacian algorithm (Oostendorp et al. 1989). Darker colours means a lower detection rate. Contour lines indicate changes of 10% in detection rate. The head model was obtained from the FieldTrip toolbox (Oostenveld et al. 2010).

Figure 1

Click here to download Figure Figure_1_HR.tif

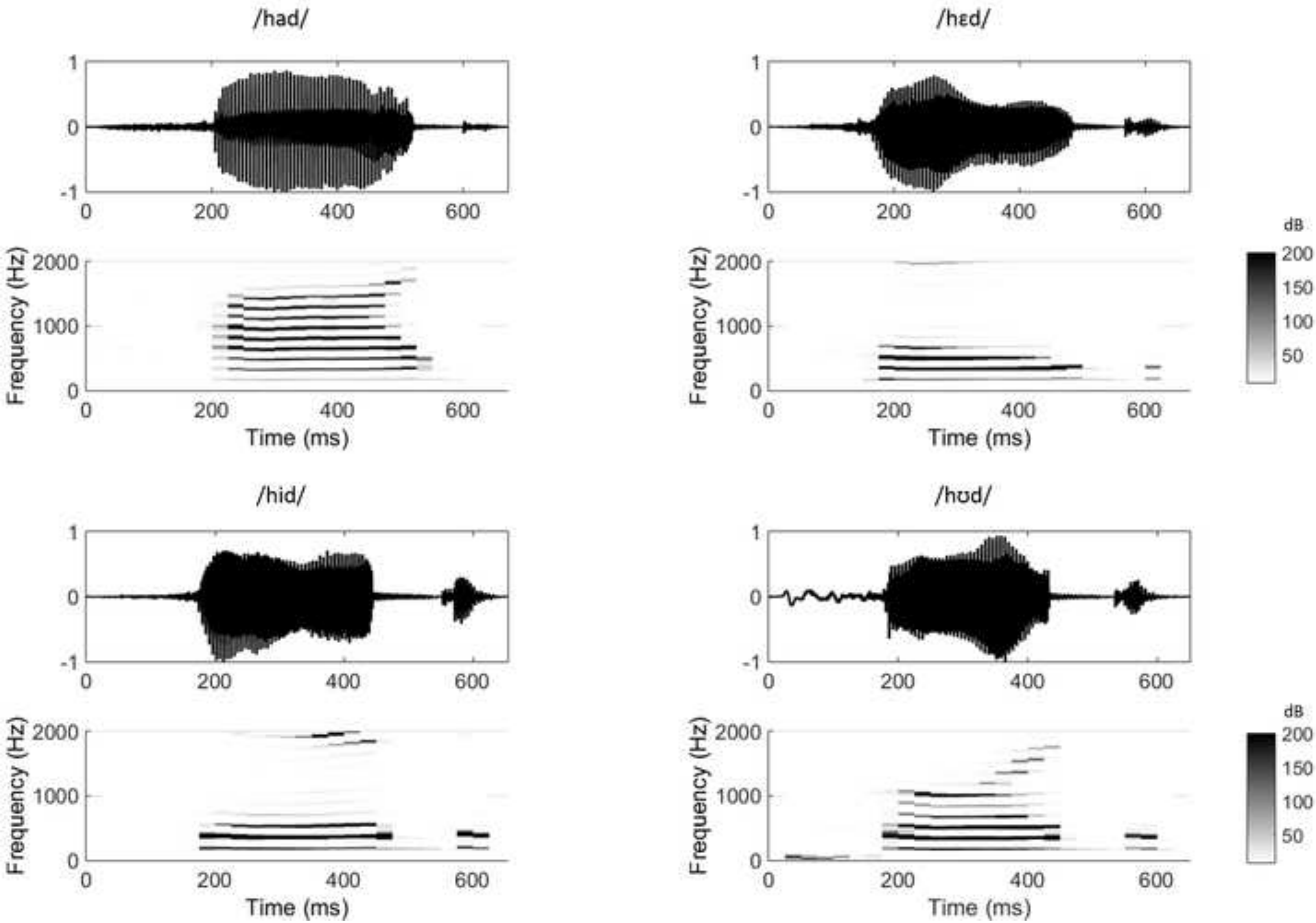

Figure 2

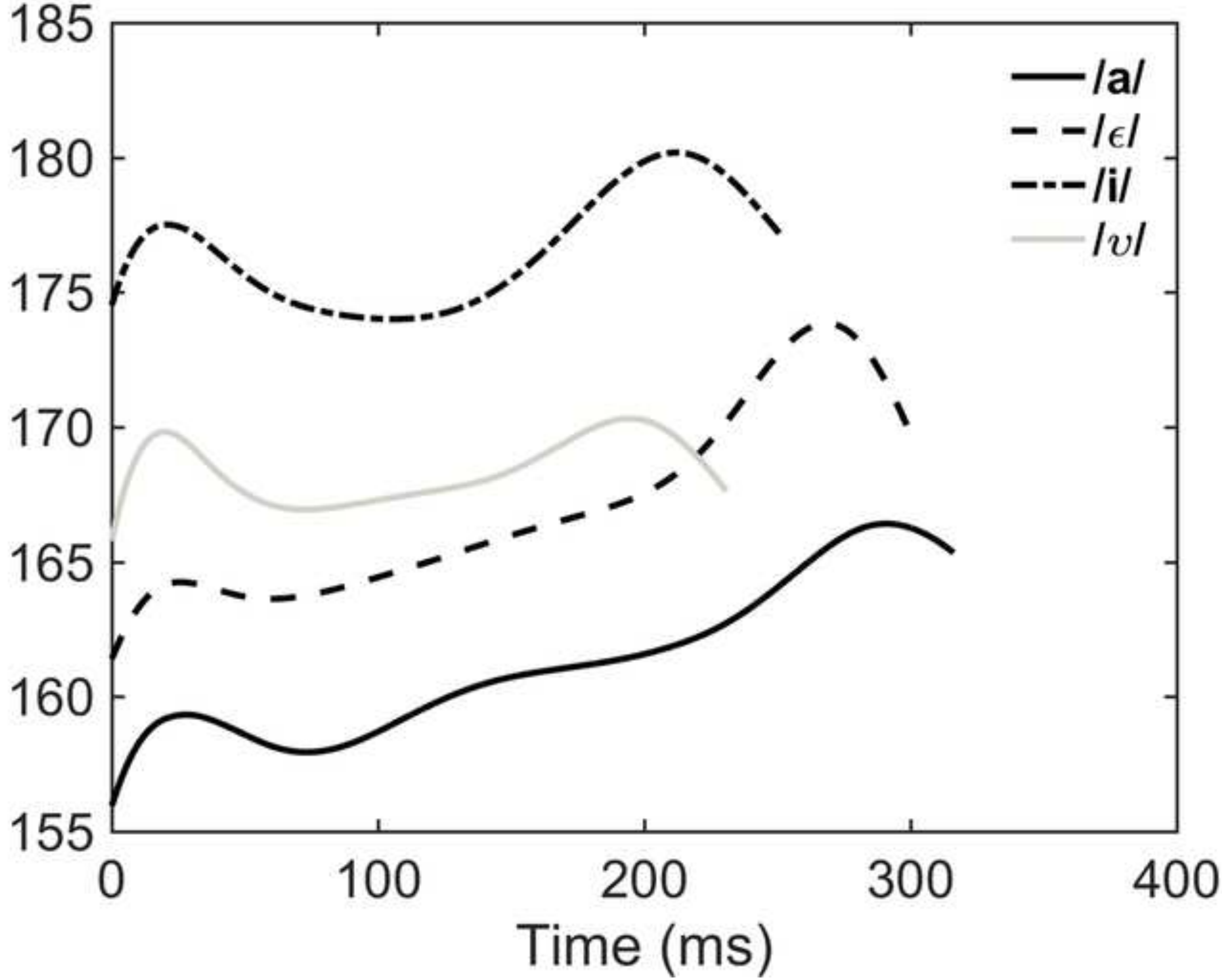

Figure 3

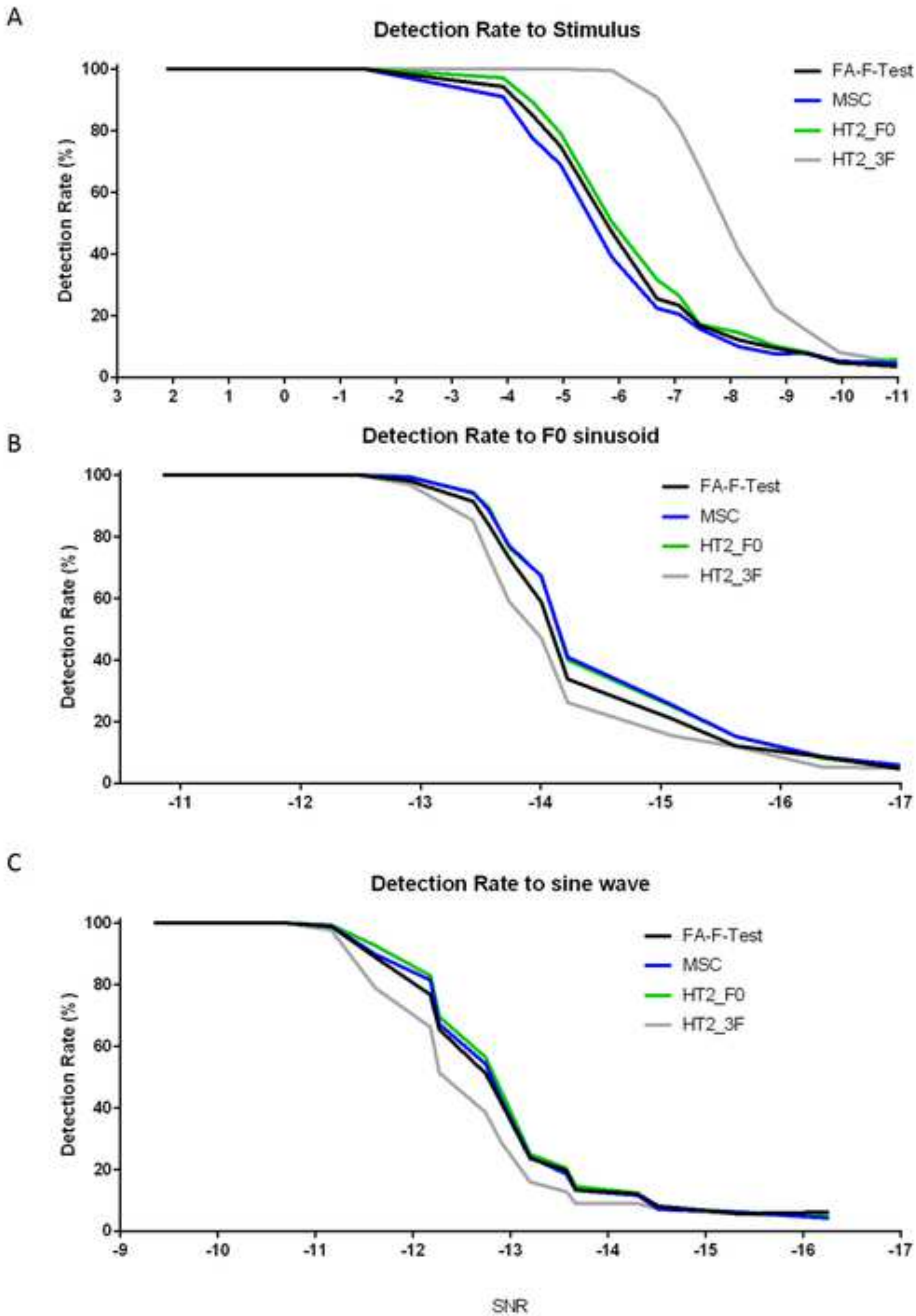
A
Detection Rate to Stimulus
FA-F-Test
MSC
HT2_F0
HT2_3F
Detection Rate (%)
B
Detection Rate to F0 sinusoid
C
Detection Rate to sine wave
SNR

Figure 4

Click here to download Figure Figure _4_HR.tif

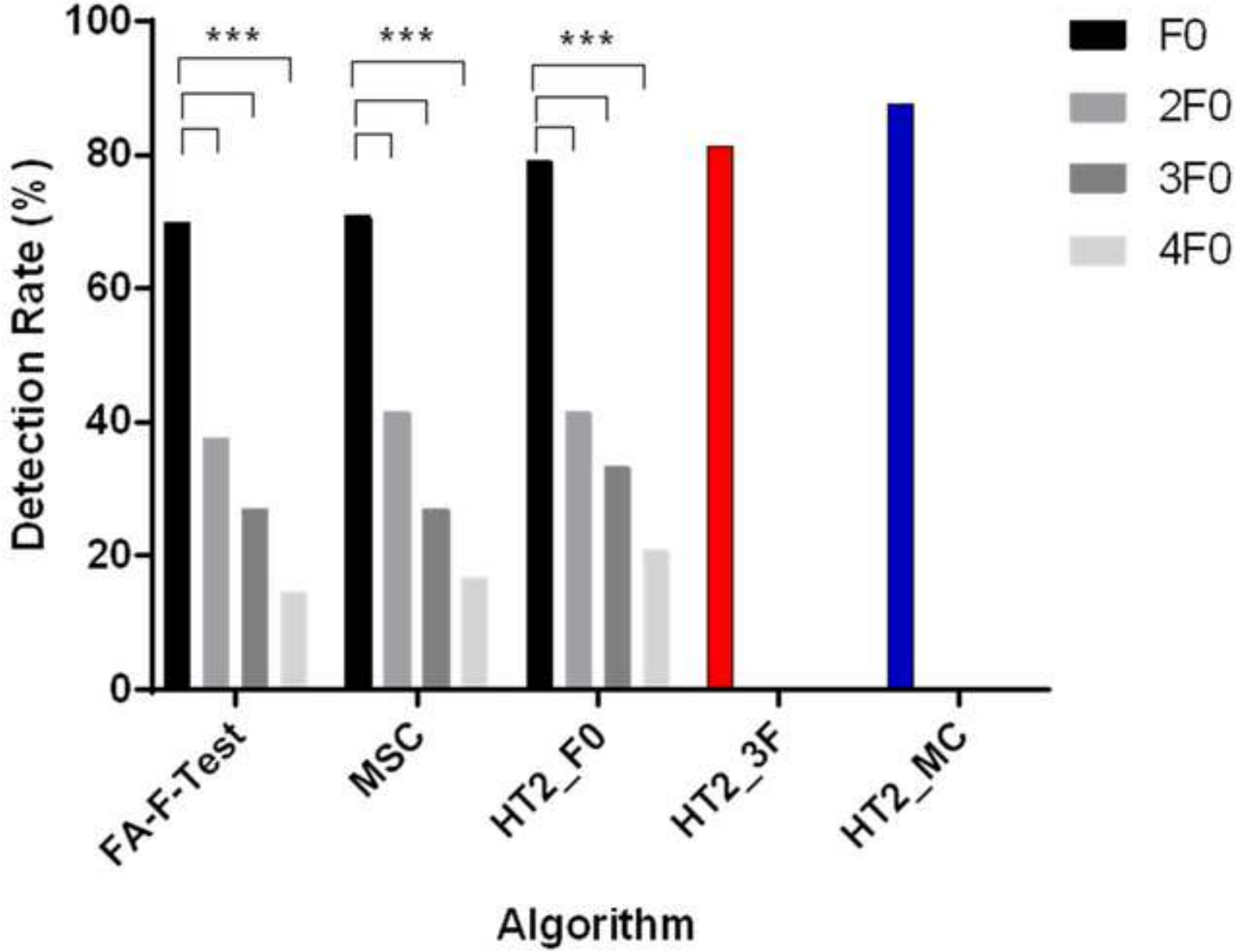

Figure 5

Click here to download Figure Figure_5_HR.tif

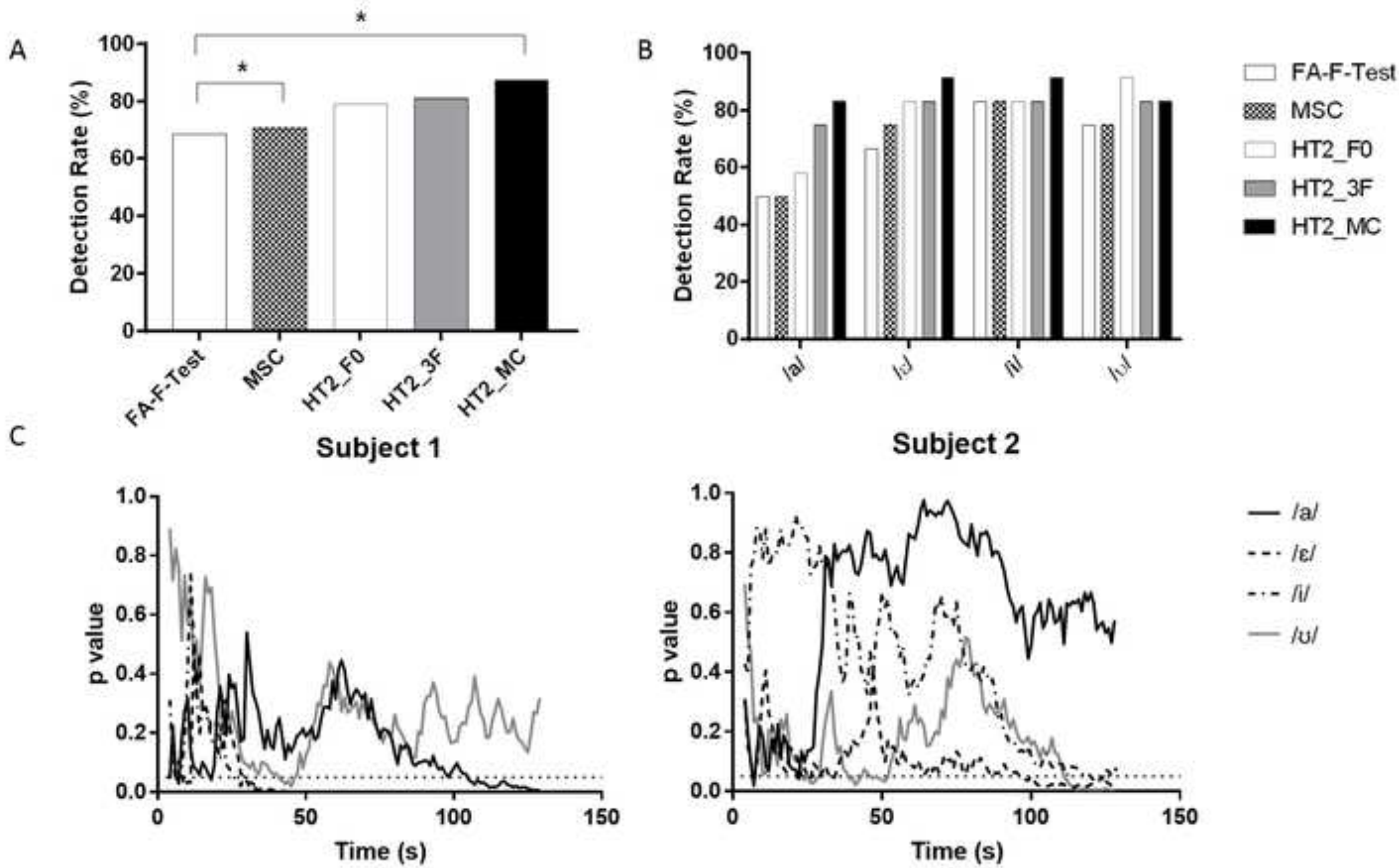

Figure 6

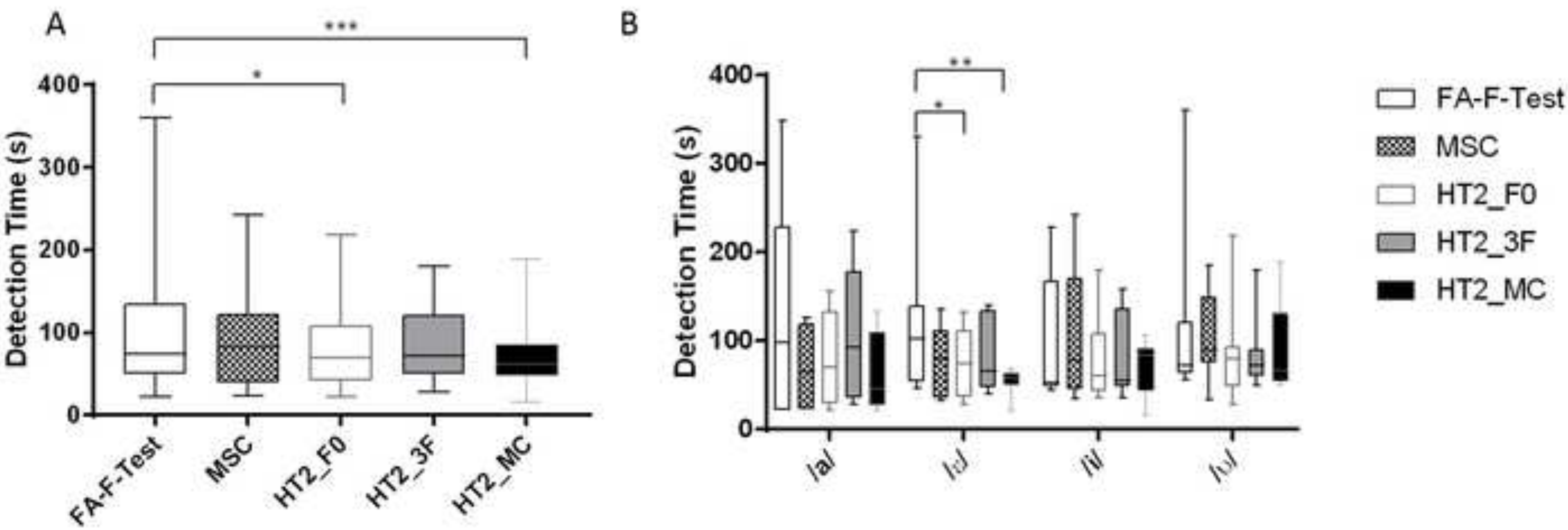
A
***
*
400
300
200
100
0
Detection Time (s)
FA-F-Test
MSC
HT2_F0
HT2_3F
HT2_MC
B
**
*
400
300
200
100
0
Detection Time (s)
/a/
/i/
FA-F-Test
MSC
HT2_F0
HT2_3F
HT2_MC

Figure 7