"Bigger is better: Increasing cortical auditory response amplitude via stimulus spectral complexity"

Fabrice Bardy PhD[1,2]

Bram Van Dun PhD[1,2]

Harvey Dillon PhD[1,2]

[1] HEARing Co-operative Research Centre, Australia

[2] National Acoustic Laboratories, NSW, Australia

Correspondence to:
Fabrice Bardy
Australian Hearing Hub
16 University Avenue
Macquarie University NSW 2109 Australia
P +61 2 94 12 68 14
F +61 2 94 12 67 69
Email: Fabrice.Bardy@nal.gov.au

Conflicts of Interest and Source of Funding

This work was supported in part by the HEARing CRC, established and supported under the Australian Cooperative Research Centres Program, an Australian Government Initiative.

Abstract

Objective: To determine the influence of auditory stimuli spectral characteristics on cortical auditory evoked potentials (CAEPs).

Design: CAEPs were obtained from 15 normal-hearing adults in response to six multi-tone (MT), four pure-tone (PT), and two narrow-band noise (NBN) stimuli. The sounds were presented at +10, +20 and +40 dB above threshold, which were estimated behaviorally beforehand. The root mean square (RMS) amplitude of the CAEP and the detectability of the response were calculated and analyzed.

Results: Amplitudes of the CAEPs to the MT were significantly larger compared to PT for stimuli with frequencies centered around 1, 2 and 4 kHz while no significant difference was found for 0.5 kHz. The objective detection score for the MT was significantly higher compared to the PT. For the 1 and 2 kHz stimuli, the CAEP amplitudes to NBN were not significantly different to those evoked by PT.

Conclusion: The study supports the notion that spectral complexity, and not just bandwidth, has an impact on the CAEP amplitude for stimuli with center frequency above 0.5 kHz. The implication of these findings is that the clinical test time required to estimate thresholds can potentially be decreased by using complex band-limited MT rather than conventional PT stimuli.

Introduction

Objective hearing threshold estimation is convenient for patients who are not able to provide behavioral feedback, such as young children or adults who cannot or will not subjectively cooperate with testing. One way to determine thresholds objectively is through cortical auditory evoked potentials (CAEPs) which reflect the activation at the level of the central auditory system in the supratemporal auditory cortex. Their recording relies on the averaging of synchronous far-field neuronal potentials evoked by auditory stimuli presented multiple times, utilizing non-invasive surface electrodes. For awake adults, the P1-N1-P2 complex generated in the time window 50-200 ms after onset of an acoustical stimulus is the response of interest. CAEPs are appreciated because they can be elicited by highly frequency-specific stimuli (Lightfoot et al. 2006; Lütkenhöner et al. 2007; Ross et al. 1999). In addition, CAEP testing is more preferable than brainstem testing where the subject needs to be asleep. This is often a difficult condition to achieve at ages of 6 months and older.

Several studies have shown that CAEPs are detected at an average level of 10 dB above behavioral threshold when using tone-burst stimuli of varying lengths (Picton 2011).

Considering the practical applicability of CAEPs in a clinical setting, there is much interest in facilitating efficient CAEP detection with the ultimate goal of reducing recording time or increasing the precision of testing during threshold estimation. Previous research has shown the limitations of fast presentation rates (i.e. up to 10 presentations a second) during attempts to decrease testing time (Bardy, Van Dun, Dillon and Cowan 2014). The reason for this failure is that adaptation decreases the CAEP amplitude when using rapid presentation rates. Consequently, this produces averaged responses with a lower signal to noise ratio (SNR) compared to presentation rates of once every 1 or 2 s. These slower rates are therefore recommended clinically (Bardy, Van Dun, Dillon and Cowan 2014).

It has previously been suggested that a variation in the context of the stimulus presentation can be used to improve recording efficiency of CAEP. More specifically, an increase of the CAEP amplitude has been found in response to novel stimuli while varying level, frequency, stimulus onset asynchrony (SOA) and ear of stimulation (Bardy, Van Dun, Dillon and McMahon 2014; Butler 1972; Pantew et al. 1975). However, these benefits are subject to debate (Lightfoot and Kennedy 2006).

Another approach to increase the size of the cortical response lies in the optimization of the stimulus parameters. Several parameters influence the size of the cortical response such as rise-time, duration, bandwidth and spectral content of the auditory stimulus. Studies by Onishi et al. (1968) suggested an optimal rise time of between 10 and 30 ms while Alain et al. (1997) demonstrated an increase of the CAEP amplitude as stimulus durations increased to 70 ms. A combination of EEG and MEG studies in adults have indicated that the amplitude of the cortical response to broadband stimuli is larger when compared to narrow-band stimuli of equal loudness (Mäkelä et al. 1988; Seither-Preisler et al. 2003; Shahin et al. 2005; Tervaniemi et al. 2000). Using EEG, Shahin et al. (2005) demonstrated an increase of the CAEP amplitude to piano tones with three natural upper harmonics, when compared to responses to tone-bursts with only the fundamental frequency. A similar effect was found by Tervaniemi et al. (2000) when investigating mismatch negativity (MMN) using spectrally rich and tone-burst stimuli. Furthermore, a MEG study by Seither-Preisler et al. (2003) found that the amplitude of the cortical N100m component depended significantly on spectral bandwidth. Using complex tone-bursts resulted in a significantly stronger auditory evoked field (AEF) than sinusoidal tone-bursts of equal intensity. Lastly, the largest N100m acoustic change complex was found for the transition from noise to a broadband stimulus when compared to a transition to a pure-tone (Mäkelä et al. 1988).

The two aims of this study were to:

1. Investigate whether complex, multi-tone (MT) stimuli centered around frequencies 0.5, 1, 2 and 4 kHz evoke larger cortical responses than those evoked by sinusoidal pure-tone (PT) at the same centre frequencies.
2. Investigate whether any increase in response amplitude evoked by the complex stimuli is related to their wider bandwidth or to some other factors.

We hypothesized a significant amplitude growth of the CAEPs in response to MT stimuli in comparison to those evoked by PT and narrow-band noise (NBN). The inclusion of NBN stimuli provides the opportunity to investigate whether the growth of the cortical response is driven by the frequency bandwidth or by the arrangement of the frequency components of the auditory stimuli.

Materials and Methods

Subjects

Fifteen normal-hearing test subjects (7 males and 8 females) ranging from 23 to 43 years of age were recruited for this study. None of the participants reported any history of neurological abnormalities. Written consent was obtained from participants and the study was approved and conducted under the ethical supervision of the Australian Hearing Human Research Ethics Committee. Participants received a small monetary compensation for taking part in the study.

Auditory stimuli

Twelve auditory stimuli were generated in Matlab (Mathworks). They comprised four sinusoidal pure-tone (PT) with frequencies 0.5, 1, 2 and 4 kHz, four one-octave multi-tone (MT) stimuli with the same center frequencies, two broadband MT stimuli – the first covering the low frequencies (0.25 to 1 kHz) and the second covering the high frequencies (1.5 to 8 kHz) - and two one-octave narrow bands of noise (NBN) centered around 1 and 2 kHz. All stimuli were 50 ms in duration with 10 ms rise-fall times to minimize spectral splatter.

Multi-tone stimuli

The MT stimuli were constructed by adding together a series of inharmonically related sinusoids. For the one-octave stimuli, the different tonal components were uniformly distributed around the center frequency on a logarithmic frequency scale. The sinusoids all had equal amplitude and a zero phase delay at time = 0 ms. For example, a MT stimulus with a center frequency of 1 kHz contained components with frequencies of 707, 891, 1122 and 1414 Hz with a stimulus bandwidth of one octave (from 707 Hz to 1414 Hz). The spectral characteristics for each stimulus are summarized in Table 1 while stimulus waveform and spectrogram are displayed in Figure 1.

Table 1. Frequency content of multi-tone stimuli.

Center Frequency	**Frequency (in Hz) of each sinsoidal component**							
0.5 kHz	353	445	561	707				
1 kHz	707	890	1122	1414				
2 kHz	1414	1781	2244	2828				
4k Hz	2828	3563	4489	5656				
Low Freq.	250	315	397	500	630	794	1000	
High Freq.	1500	1889	2381	3000	3779	4762	6000	7559

Fig. 1

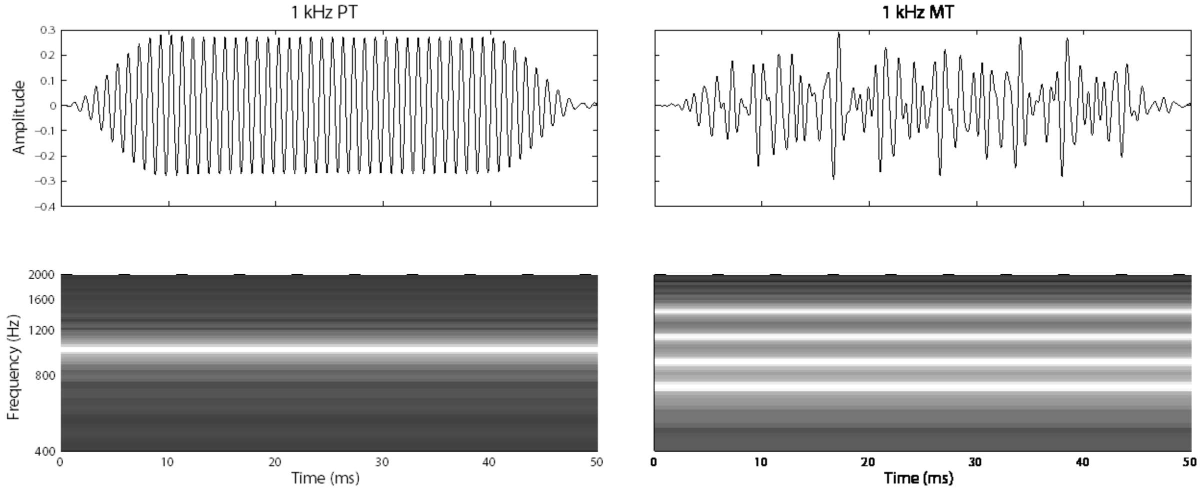


Fig. 1. Stimulus waveform and spectrogram of the 2 kHz pure-tone (PT) (left) and multi-tone (MT) stimuli (right). Stimuli are 50 ms in duration with a 10 ms rise and fall time.

Calibration

All stimuli were acoustically calibrated at 70 dB HL according to the ISO standard 389-2 (ISO 1994) in an HA-2 2-cc coupler, incorporating a 1-inch 4144 microphone, a 1-to-1/2 inch DB0375 adaptor, and a 4230 sound level meter (all Brüel & Kjaer). Continuous stimuli were used for the calibration of tone-bursts, pure-tones and multi-tones.

Behavioral procedure

Automatic threshold estimation

Test parameters of the computerized audiometry implemented using an adaptive staircase were based on Convery et al. (2014). The SOAs, representing the time interval between the onset of two auditory stimuli, were of random duration and ranged from 1000 to 4600 ms. Participants were instructed to respond to the stimuli by pressing a button on a numeric keypad. A response was considered valid if it occurred within a 1.5-second time window commencing from the onset of the stimulus. The test included 3 phases, using a threshold-seeking algorithm. The start level of stimulus presentation was 50 dB SPL. In phase 1, a 10-dB up/down step size was implemented. Phase 1 ended when the first non-response succeeding a positive response to a stimulus presentation was recorded. At this point the staircase "reversed" and intensity was increased by 10-dB prior to the next phase. Phase 2 used a 5-dB up/down step size. A subsequent non-response resulted in an increase in stimulus level in 5-dB increments until a positive response was recorded. After two reversals, a non-response resulted in a 5 dB increase for the next phase. In phase 3, the step size was lowered to 2 dB. Phase 3 ended when four reversals were recorded. A trimmed mean (i.e., removal of the highest and lowest values before averaging the remaining values) of all the presentations in phase 3 was calculated to determine the threshold. This threshold will be referred to as 0 dB SL (sensation level).

Behavioral assessment

Participants underwent a series of audiometric assessments in a sound attenuated booth, to develop local normative data:

(1) Automatic pure-tone air conduction audiometry in both ears using stimuli 500 ms in duration with frequencies 0.25 – 8 kHz. The order of presentation of the pure-tones was

1, 2, 4, 8, 0.5, and 0.25 kHz. Stimuli were presented first in the right ear. Hearing thresholds had to be better than 20 dB HL in both ears to continue the test.

(2) One ear was selected pseudo-randomly such that 7 left and 8 right ears were used in the experiment (N = 15 ears). Automatic air conduction audiometry was conducted using the twelve 50-ms auditory stimuli described in section "Auditory stimuli". The presentation order of the twelve stimuli was randomized.

The thresholds obtained in (1) and (2) allow the difference (in dB) between the 500-ms long and 50-ms short stimuli due to temporal integration (Moore 2012) to be estimated.

All stimuli originated from .wav files stored on a desktop computer and were presented via a RME sound card (Fireface 800). All stimuli were delivered to the test ear through an ER-3A insert earphone (Etymotic Research).

Electrophysiological recording of CAEPs

Sequence generation

Sound sequences used for electrophysiological recording were generated for each participant based on their behavioral thresholds. The twelve stimuli described in section "Auditory stimuli" were presented at three sensations levels (+10, +20, +40 dB SL). Consequently, the total number of conditions in the experiment was 36. Stimulus conditions were randomized such that a full set of 36 stimulus conditions had to be presented before re-iteration. SOAs were jittered uniformly between 1000 and 3000 ms. Each condition was presented 60 times resulting in 2160 trials and a testing time of 72 minutes. MATLAB was utilized to create the sequence file.

Stimulus presentation

The equipment from the behavioral experiment was used in the electrophysiological experiment to present the auditory stimuli. The stimuli were presented monaurally on the selected ear. An earplug was fitted to the opposite ear.

Data acquisition

The electroencephalogram (EEG) was obtained with Neuroscan Synamps2 version 4.3 (Compumedics) by placing four gold-plated electrodes onto the subject's head. Active electrodes were placed at fronto-central midline positions (Cz and FCz). The reference electrode was placed on the mastoid contralateral to the test ear, and the forehead (Fpz) acted as ground electrode (AES 1991). Prior to the placement of electrodes, the subject's skin was prepared using NuPrep EEG abrasive skin prepping gel. Water-soluble electrode paste was used to ensure a good connection between the electrodes and skin to achieve impedances of less than 5 kOhm across all electrode sites. Testing was conducted in an audiometric booth adhering to ANSI standard S.3.1-1999 (R2008). During testing, the subjects were sitting comfortably in a dimmed, sound attenuated booth. The participants watched a muted close-captioned DVD of their choice which effectively captures attention without interfering with auditory processing. Participants were instructed not to pay attention to the stimulus.

Data analysis

Amplitude measurements were analyzed at both FCz and Cz referenced to the mastoid contralateral to the test ear. All EEG channels were amplified with a gain of 2010, digitized at a sampling rate of 1000 Hz, and online bandpass filtered between 0.01 and 30 Hz. All epoched files were exported to MATLAB for off-line processing. The signal processing of the raw EEG files was partly conducted using EEGLAB (Delorme et al. 2003). An epoch of 700 ms (100 ms pre- and 600 ms post-stimulus onset) was used with baseline correction. Artefact and eye-blink were monitored by excluding epochs in excess of ±75 µV. A minimum of 52 accepted epochs was required for each stimulus condition.

Response amplitude

Using the grand averages of the epoched waveforms, the "signal + noise" amplitude was expressed as the root mean square (rms) value within a window of 250 ms beginning 30 ms after stimulus onset. Due to the non-homogeneity of the variance across stimuli conditions and the dependence of the standard deviation on the mean response amplitude a log transform was applied on the amplitude data prior to statistical analysis, to stabilize the variance across conditions (Zacharias et al. 2011).

Measure of response detection

The Hotelling's T^2 statistic was used to provide an objective measure of CAEP response presence. Before applying the detection method, each recorded epoch was reduced to 9 averaged voltage levels, covering the range from 51 to 347 ms, with each bin being 33 ms wide. The bin width and number of bins were chosen based on earlier data (Golding et al. 2009). Response detection was based on the p-value obtained from a one-sample Hotelling's T^2 test on the bin-averaged data. The one-sample Hotelling's T^2 test is the multivariate extension of the ordinary one-sample t-test; instead of testing a null hypothesis that a scalar true mean equals a specified value, the Hotelling's T^2 test takes vector data and tests a null hypothesis that the true mean vector equals the zero vector. For every testing condition, the p-value was calculated after the collection of 9 epochs and subsequently, every additional two epochs. As the average SOA was 2 s, the p-value versus testing time could be presented for every subject. The p-values were afterwards converted into z-scores (assuming a normative z-distribution) and a measure of response detection was calculated by cumulative summation of the z-score values. As two conditions (MT versus PT stimuli) were compared using the same sequential statistical testing, and no detection sensitivity was evaluated, no multiple testing adjustments needed to be performed.

Statistical analysis

Repeated measures ANOVAs

For statistical analysis, a three-way repeated measures analysis of variance (ANOVA) was performed on the log-transformed rms amplitudes and the measures of response detection. Greenhouse–Geisser corrections for sphericity were applied, as indicated by the cited ε value. Post-hoc comparisons were calculated using Tukey's test. Statistical analyses were conducted using Statistica 7.1 (StatSoft, Inc.) and R (R Development Core Team 2013), with the additional packages car (Fox et al. 2012), reshape (Wickham 2011), nlme (Pinheiro et al. 2013), and multcomp (Hothorn et al. 2013).

Results

Behavioral thresholds

Table 2 presents the behavioral mean thresholds and standard deviations (in dB SPL) across 15 subjects for six 500-ms audiometric pure-tones (250 - 8000 Hz) and twelve 50-ms auditory stimuli. The mean threshold differences (in dB) across all subjects between 500-ms pure-tones and 50-ms tone-bursts for the frequencies 500, 1000, 2000 and 4000 Hz are shown in Table 2. The mean reaction time over all stimulus condition was 0.56 s. (SD = 0.21). As expected the 50-ms tone-bursts had elevated thresholds when compared to 500-ms pure-tone thresholds. The mean behavioral threshold differences for the four tested frequencies ranged between 5 and 9 dB. The average threshold differences between 50-ms tone-bursts and 50-ms multi-tone stimuli ranged between 0 and 9 dB. These results can be used as corrections to account for the difference between the behavioral hearing thresholds estimated using 500-ms pure-tones and 50-ms tone-bursts.

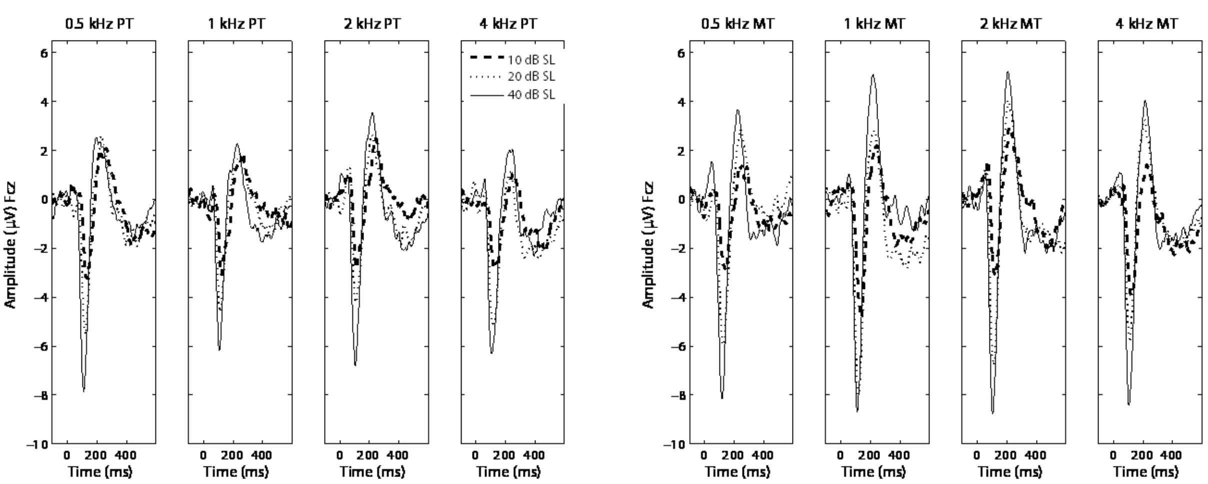
Table. 2

Stimuli	Threshold (dB SPL)	RET 0 dB HL (dB SPL)	PT- PT$_{500}$ (dB)	MT – PT$_{500}$ (dB)	MT - PT (dB)
PT$_{500}$ 0.25 kHz	18.7 ± 5.9	14			
PT$_{500}$ 0.5 kHz	7.4 ± 5.1	5.5			
PT$_{500}$ 1 kHz	0.3 ± 5.6	0			
PT$_{500}$ 2 kHz	9.7 ± 4.2	3			
PT$_{500}$ 4 kHz	4.8 ± 6.8	5.5			
PT$_{500}$ 8 kHz	-1.0 ± 8.2	0			
PT 0.5 kHz	15.8 ± 6.0		8.4 ± 4.4		
PT 1 kHz	9.3 ± 4.0		9.0 ± 4.4		
PT 2 kHz	14.9 ± 3.5		5.1 ± 2.2		
PT 4 kHz	9.9 ± 6.2		5.1 ± 3.8		
MT 0.5 kHz	17.9 ± 4.3			10.5 ± 4.0	2.1 ± 4.5
MT 1 kHz	12.1 ± 3.2			11.8 ± 5.0	2.8 ± 3.7
MT 2 kHz	15.0 ± 3.9			5.3 ± 4.0	0.1 ± 3.3
MT 4 kHz	18.5 ± 5.1			13.7 ± 8.2	8.6 ± 5.7
LF MT	14.8 ± 3.1				
HF MT	16.1 ± 4.5				
NBN 1 kHz	12.4 ± 3.0				
NBN 2 kHz	16.5 ± 3.2				

Table. 2. Behavioral mean thresholds and standard deviations across 15 subjects for six 500-ms pure-tones (PT$_{500}$) (0.25 – 8 kHz), and twelve 50-ms auditory stimuli used for the recording of CAEPs. The twelve stimuli consisted of four PT with frequencies 0.5, 1, 2 and 4 kHz, four band-limited (one-octave) multi-tone (MT) stimuli with the same center frequencies, two broadband MT stimuli covering the low (LF MT: 0.25 to 1 kHz) and high frequencies (HF MT: 1.5 to 8 kHz) and two one-octave narrow bands of noise (NBN) centered around 1 and 2 kHz. The reference equivalent threshold (RET, i.e. 0 dB HL) according to ISO Organization (1994) is provided. The mean threshold difference and standard deviation between 50 and 500 ms PT, between MT and PT$_{500}$ and between MT and 50 ms PT are provided.

Grand average CAEP waveforms

Fig. 2 shows the mean CAEP waveforms, averaged across all fifteen subjects, in response to tone-bursts, one-octave-band multi-tone stimuli, broadband multi-tone stimuli and one-octave-band noise, all 50 ms long and presented at +10, +20 and +40 dB SL. Clear CAEPs characterized by the P1-N1-P2 complex are identifiable by visual inspection for all conditions.


a) Pure-tone

b) Multi-tone

c) Multi-tone & Narrow band noise

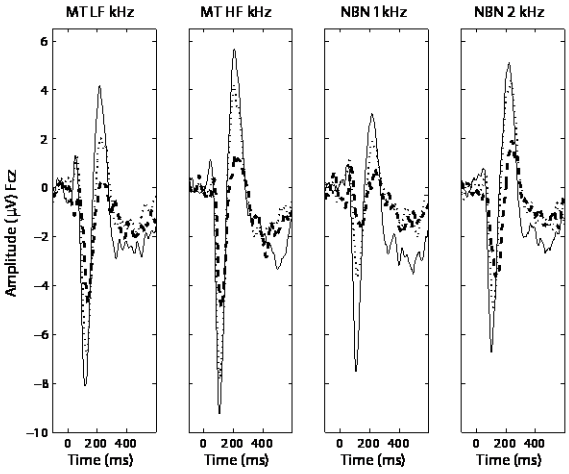


Fig. 2. Grand average CAEP waveforms (n=15) generated by a) four pure-tone (PT), b) four one-octave-band multi-tone (MT), c) two broadband MT and two one-octave narrowband noise (NBN) auditory stimuli. Responses are presented for three presentation levels, +10 dB (thick dashed line), +20 dB (thin dashed line), +40 dB (thin solid line).

CAEP amplitudes

PT versus MT stimuli

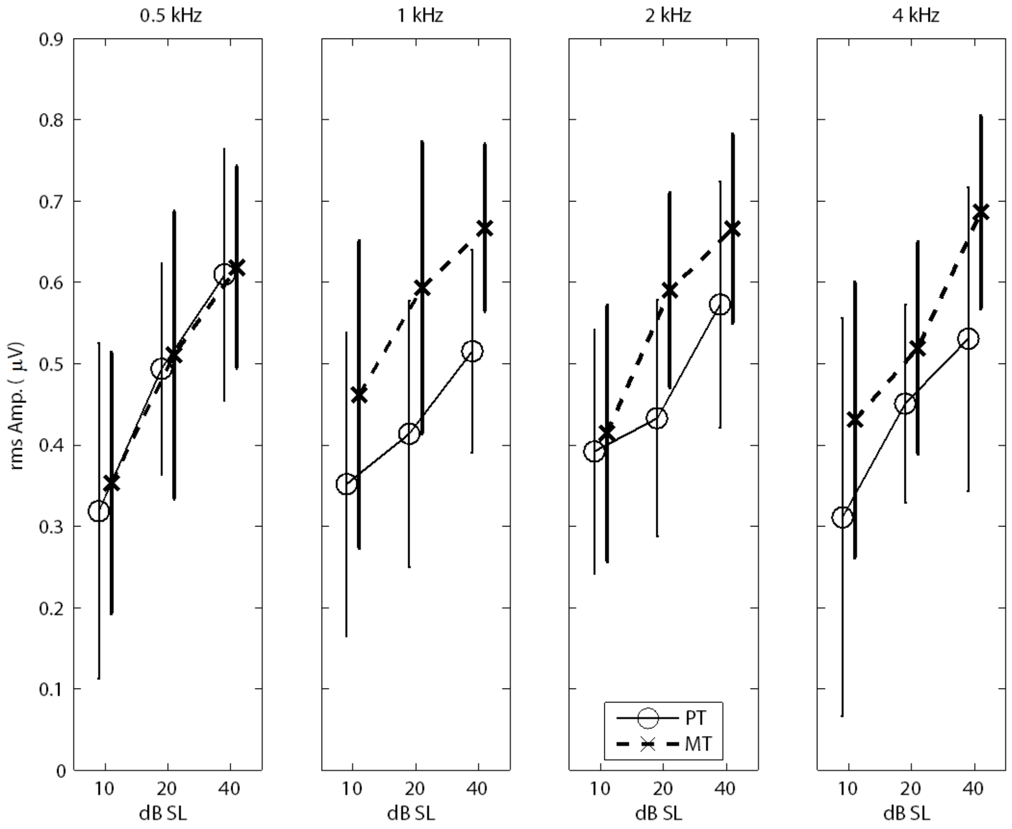


Fig. 3 summarizes the CAEP rms amplitudes in the time window 30-280 ms after stimulus onset as a function of stimulus (PT, one-octave MT stimuli), frequency (500, 1000, 2000 and 4000 Hz) and sensation level (10, 20, and 40 dB SL), while collapsed over EEG channels (Cz and FCz). A 2 x 2 x 4 x 3 repeated-measures ANOVA with EEG channel, stimulus, frequency and sensation level was performed on the rms amplitude data.

Effects of stimuli (PT versus MT) and frequency (0.5, 1, 2 and 4 kHz)

The repeated-measures ANOVA revealed a main effect of stimulus (F(1,14) = 67.36; p = 0.000001; ε = 1). The MT stimuli elicited significantly higher response amplitudes than PT. Moreover, an interaction effect was found between stimulus and frequency (F(3,42) = 4.56; p = 0.01; ε = 0.01). Tukey pairwise comparisons showed no significant difference between PT and MT stimuli for 0.5 kHz (p = 0.99) while significant differences were found for the other frequencies (i.e. 1, 2 and 4 kHz) ($p < 0.05$). Table 3 shows the rms amplitude ratio between MT and PT stimuli and the time reduction (in %) to achieve the same SNR for MTs as PTs. Time reduction is calculated based on the MT/PT ratio, assuming that the residual noise in the averaged waveform decreases with the square root of the number of epochs. When collapsing the data across the three frequencies 1, 2 and 4 kHz and all levels, an average rms amplitude ratio of 1.32 (95 % confidence interval 1.25 – 1.37) was found for MT stimuli when compared to PT, which corresponds to a potential 46 % average time reduction.

Frequency (kHz)	Level (dB SL)	Rms amplitude ratio MT/PT	Estimated time reduction (%)	p-value
0.5	10	1.08 (0.88; 1.33)	14.7	0.97
	20	1.04 (0.85; 1.27)	7.4	1.00
	40	1.02 (0.83; 1.25)	4.1	1.00
1	10	1.29 (1.05; 1.58)	39.8	0.004
	20	1.51 (1.23; 1.85)	56.3	<0.0001
	40	1.42 (1.16; 1.74)	50.2	<0.0001
2	10	1.05 (0.86; 1.29)	9.8	1.00
	20	1.44 (1.17; 1.76)	51.5	<0.0001
	40	1.24 (1.01; 1.52)	34.8	0.03
4	10	1.32 (1.07; 1.62)	42.4	0.001
	20	1.17 (0.96; 1.44)	27.1	0.28
	40	1.43 (1.17;1.76)	51.2	<0.0001

Table. 3. Mean and 95% confidence intervals of the rms amplitude ratio MT/PT at 10, 20 and 40 dB SL for stimuli with center frequencies at 0.5, 1, 2 and 4 kHz. An estimation of the time reduction using MTs when compared to PTs to reach a similar SNR is provided. The last column shows a p-value calculated using a mixed-effects model. It displays whether the difference between MT and PT stimuli is significant.

Effect of sensation level (10, 20 and 40 dB SL)

A main effect of sensation level was found (F(2,28) = 122.66; $p < 0.000001$; $\varepsilon = 0.86$) with higher intensities eliciting larger CAEP amplitudes.

Effect of channel (FCz versus Cz)

A main effect of channel was found (F(1,14) = 17.74; $p = 0.0008$; $\varepsilon = 1$) with a 11% rise of rms amplitudes obtained from channel FCz-mastoid than from Cz-mastoid (95% confidence interval 7-14%).

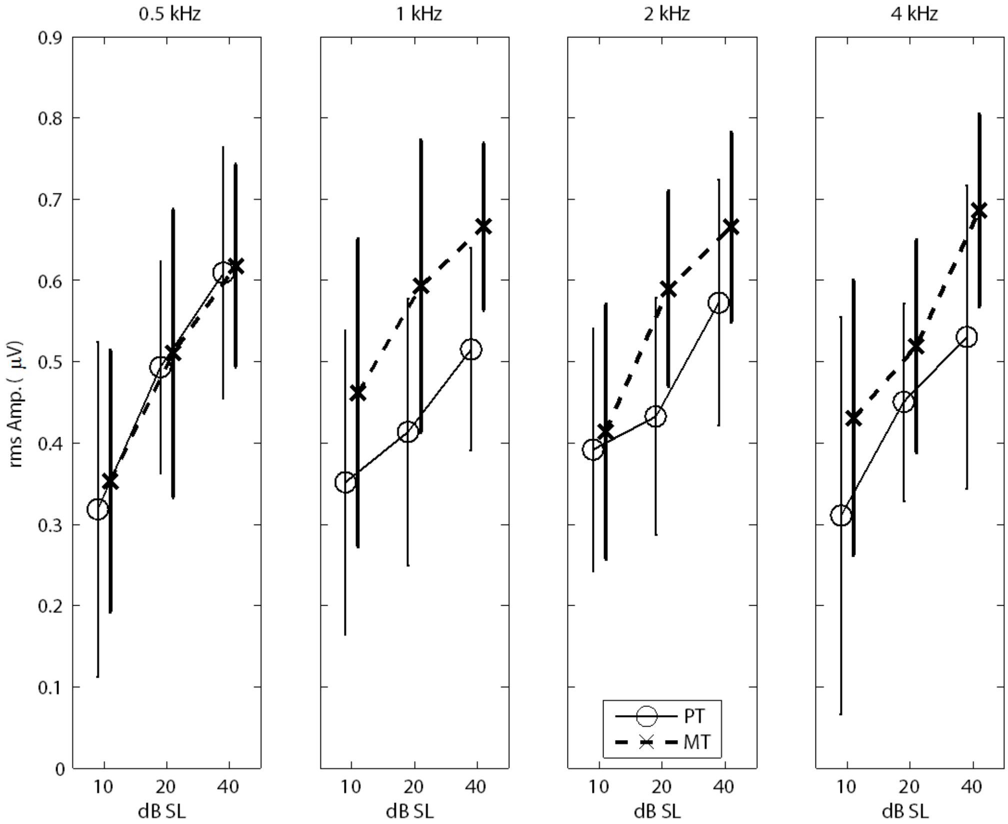


Fig. 3. CAEP rms amplitudes, collapsed over electrode positions FCz and Cz, for 0.5, 1, 2 and 4 kHz MT and PT stimuli for three sensation levels +10 dB, +20 dB, +40 dB SL. Vertical lines represent standard deviations between participants.

NBN versus PT and MT stimuli at frequencies 1 and 2 kHz

Rms amplitudes of the CAEP elicited by NBN were compared to responses of both MT and PT stimuli in a 2 x 3 x 2 x 3 repeated-measures ANOVA with channel, stimulus, frequency and level. Fig. 4 shows rms CAEP amplitudes as a function of stimulus (i.e. PT, one-octave MT stimulus, and one-octave NBN), for the two frequencies (1000 and 2000 Hz) and the stimulus level (+10, +20, and +40 dB SL). Of interest, a main stimulus effect was found ($F(2,28) = 26.23$; $p = 0.000003$; $\varepsilon = 0.92$). Tukey pairwise comparisons revealed no significant difference between PT and NBN ($p = 0.37$) but a significant difference between MT stimuli and both NBN and PT ($p < 0.001$). A significant interaction between stimulus and frequency was present ($F(3,42) = 4.55$; $p = 0.01$; $\varepsilon = 1$). That is, the effect of stimulus is larger at 1 kHz than at 2 kHz.

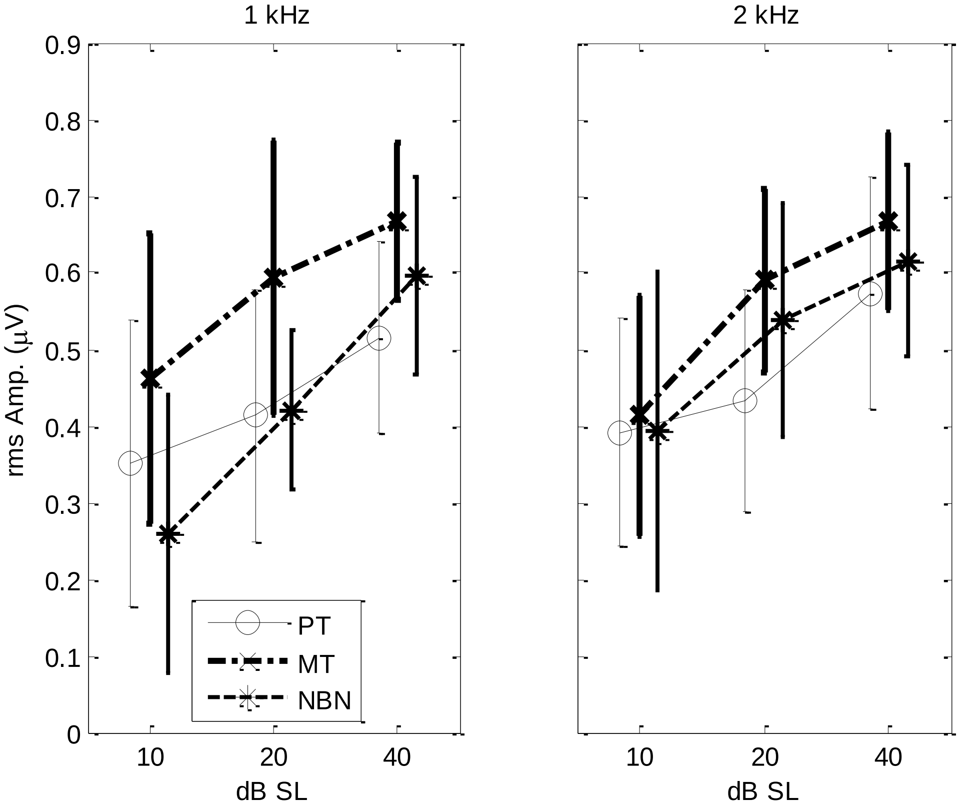


Fig. 4. CAEP rms amplitudes, collapsed over electrode positions FCz and Cz, for 1 and 2 kHz PT, NBN and MT stimuli for three sensation levels +10 dB, +20 dB, +40 dB SL. Vertical lines represent standard deviations between participants.

One-octave (0.5 and 2 kHz) versus broadband (LF and HF) MT stimuli

Fig. 5 shows rms CAEP amplitudes for one-octave and broadband MT stimuli. Although the mean rms amplitude for the broadband MT stimuli was larger in every condition, a one-way repeated-measures ANOVA did not show a significant difference between the two stimuli (F(1,14) = 2.65; p = 0.12; ε = 1).

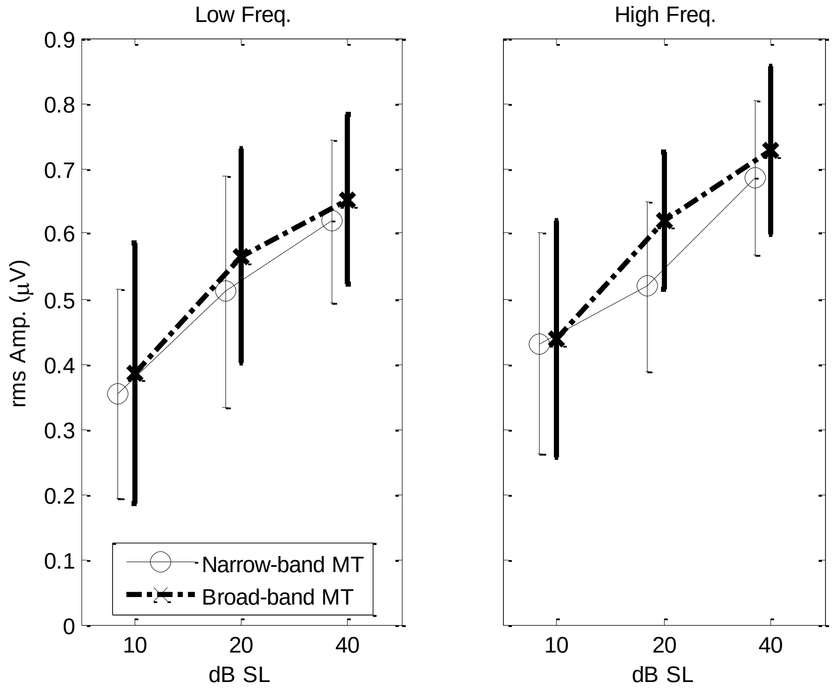


Fig. 5. CAEP rms amplitudes, collapsed over electrode positions FCz and Cz, for 0.5 and 2 kHz one-octave MT and broadband (LF and HF) MT stimuli at three presentation levels +10 dB, +20 dB, +40 dB SL. Vertical lines represent standard deviations between participants.

Objective detection scores of the CAEP

Fig. 6 shows an example of the representation of z-score traces for the tone-burst (left) and multi-tone stimulus (right) for the fifteen subjects. A more negative z-score represents a smaller p value, and therefore a higher response detection. Hence, in this example, the response to the multi-tone stimulus is more likely to be objectively detected than to the tone-burst. It is valuable to have this measure as the Hotelling's T^2 is clinically used for the detection of cortical responses. Mean cumulative z-scores are displayed in Fig. 7 for all stimuli, frequencies and sensation levels. Once again, more negative cumulative z-scores translate in higher detections of the responses. A 2 x 2 x 4 x 3 repeated-measures ANOVA with EEG channel, stimulus, frequency and sensation level was performed to assess their effects on the z-score data. It revealed a main effect of stimulus (F(1,14) = 41.22; p = 0.00001; ε = 1). Significantly more negative mean cumulative z-scores for the MT stimuli were observed when compared to z-scores from PT. A main effect of level was observed as well (F(2,28) = 100.70; p < 0.00001; ε = 0.79) with higher sensation levels showing significantly more negative z-scores. No main effect of channel was observed, indicating no advantage for a specific channel (i.e. FCz-M versus Cz-M) (F(1,14) = 0.60; p =0.45; ε = 1.00). This is in contrast with the main channel effect for CAEP amplitudes, which indicated significantly larger amplitudes at FCz. This is likely caused by increased noise at this electrode position.

A significant interaction between stimulus and level (F(2,28) = 4.70; p = 0.02; ε = 1) was observed. Tukey pairwise comparisons indicated no difference in z-scores between PT and MT stimuli for +10 dB SL (p = 0.85) while significant differences were found for the other levels (i.e. +20 and +40 dB SL) (p < 0.001). A significant interaction was present between stimulus and frequency (F(3,42) = 13.23; p < 0.00001; ε = 0.82). Similarly to the CAEP amplitudes, Tukey pairwise comparisons revealed a significant effect of the stimulus for the frequencies 1, 2 and 4 kHz (p < 0.001) but not for 0.5 kHz (p = 0.22).

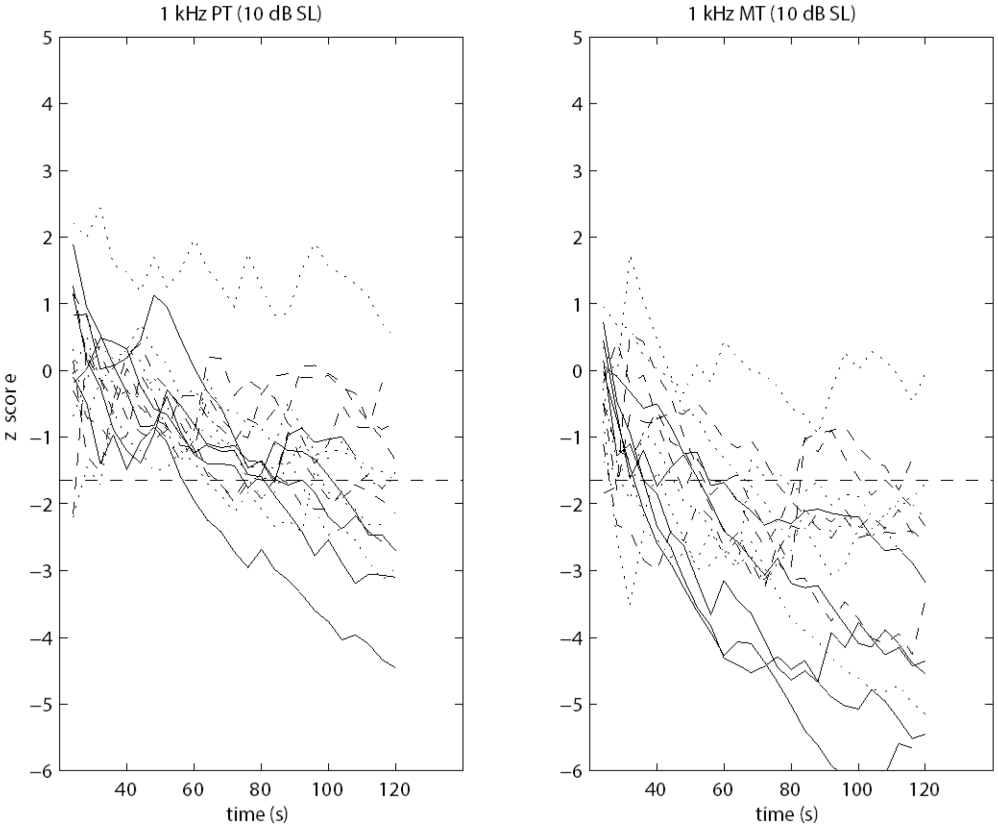


Fig. 6. Representation of the objective detection measure of 15 subjects (i.e. one trace per subject), for PT (left panel) and MT (right panel) stimuli for a 1kHz centre-frequency at +10 dB SL. The results of the Hotelling's T2 which are calculated every 4 s are converted into z-score and presented across time. For z-score lower than –1.64 (i.e. p< 0.05), the CAEP is considered to be present, assuming that the z score is compared to this threshold value at just one pre-specified point within each trace.

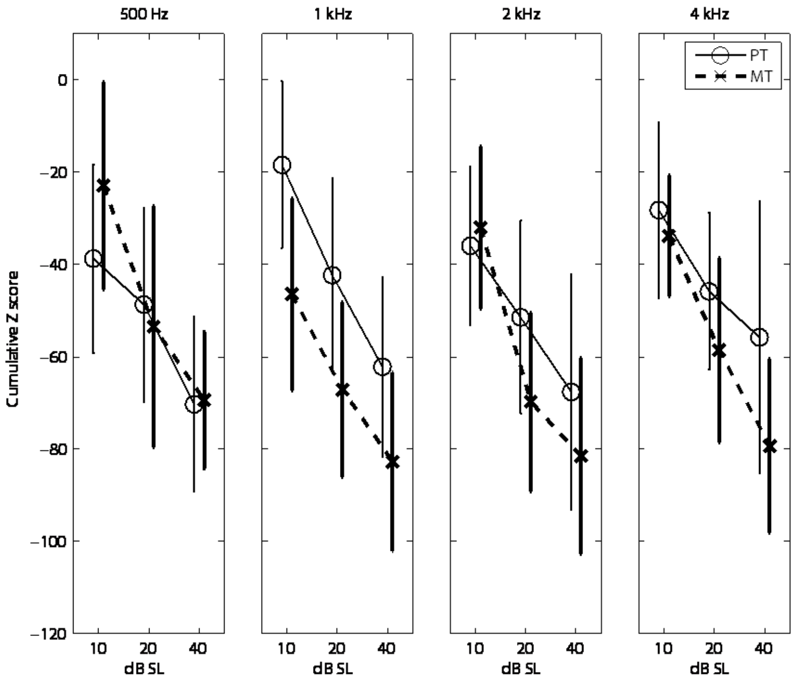


Fig. 7. Cumulative CAEP detection z-scores, collapsed over electrode positions FCz and Cz, for 0.5, 1, 2 and 4 kHz MT (thick dashed line) and PT stimuli (thin solid line) for three sensation levels, +10 dB, +20 dB, +40 dB SL. Vertical lines represent standard deviations between participants.

Discussion

In the present study, we designed narrowband multi-tone (MT) stimuli centered around 0.5, 1, 2 and 4 kHz. We compared the cortical auditory evoked potentials (CAEPs) they elicited with responses to sinusoidal pure-tone (PT), one-octave, narrow-band noise (NBN), and one-octave, multi-tone complexes. In total, electrophysiological responses were recorded for 12 different stimuli at 3 sensation levels (+10, +20 and +40 dB SL) for which clear P1-N1-P2 waveforms could be discerned. In a group of subjects with normal hearing it was found that the amplitude of the CAEP was influenced by the spectral composition of the auditory stimuli, with all auditory stimuli matched in sensation level. First, the effect of stimulus will be discussed by comparing the CAEP amplitudes of both the MT and the NBN stimulus with those of the PT. The latter group served as the reference. Second, we reflect on the physiological reasons underlying the change of the cortical response characteristics. Finally, the potential benefit of using MT stimuli for threshold estimation in a clinical setup is considered.

PT versus MT stimuli

Responses elicited by MT stimuli with frequencies centered around 1, 2 and 4 kHz showed a significantly larger rms CAEP amplitude than responses to PT stimuli. These results show that, not surprisingly, the neural response to a pure tone is different from that to a complex tone centered on the same frequency. The effect of spectral complexity found in this study is broadly in agreement with previous research presented in the introduction (Mäkelä et al. 1988; Seither-Preisler et al. 2003; Shahin et al. 2005; Tervaniemi et al. 2000). Of interest, for the 0.5 kHz frequency, no significant amplitude difference was found between the cortical responses.

PT versus NBN stimuli

The inclusion of two NBN stimuli at 1 and 2 kHz in the experimental design allowed investigation as to whether the growth of the cortical response was driven by the frequency bandwidth or by the arrangement of the frequency components i.e. spectral fine structure. The main effects in Fig. 4 showed no significant amplitude differences for CAEPs elicited by NBN and PT stimuli. Conversely, significantly larger amplitudes were observed for MT stimuli when compared to both PT and NBN. This suggests that the spectral fine structure of the sound, rather than its bandwidth, is principally affecting the cortical response. This observation is reinforced by the results in Fig. 5, which showed no significant main differences between one-octave and multi-octave MT stimuli. A limitation of the present study is that the small sample size could be the factor explaining the lack of any significant difference between the two types of MT stimuli. However, the observed small effect size makes any differences clinically unimportant.

These findings are partially inconsistent with (Hirata et al. 1999) who reported smaller N100m responses for NBN when compared to piano or PT stimuli. However, in the Hirata et al. (1999) study the frequencies of the stimuli were not matched, which may be a reason for the difference in findings. In addition, there was a significant interaction effect between frequency and stimulus in Fig. 4 for which there is no immediately obvious explanation. Further studies will be required to investigate the effect of frequency for different stimulus types.

Possible functional reasons

There are at least four possible reasons why complex stimuli may elicit larger responses than pure tones.

First, the tonotopic arrangement of the auditory system, including the primary auditory cortex (Howard III et al. 1996) means that stimuli with wider bandwidths may evoke cortical activity in a more widespread group of neurons immediately surrounding those that respond best to pure tones at the centre frequency. If the total number of neurons increases, so too may the magnitude of the cortical responses. This would be analogous to the way that, for sounds at moderate input levels, loudness increases with bandwidth when total intensity is held constant.

Second, rather than a larger number of neuronal firings, the MT stimulus could somehow cause the same neurons to fire more synchronously with each other, which by itself would increase the magnitude of the cortical response on the scalp.

Third, the MT stimulus may excite neuron firing in cortical regions remote from those excited by a pure tone. Functional magnetic resonance imaging (fMRI) studies found that the complexity of the auditory stimulus has an effect on the area of activation in the auditory cortex (Strainer et al. 1997; Wessinger et al. 2001). Wessinger et al. (2001) indicated that whereas sinusoidal stimuli elicited activity principally in the core region of the auditory cortex, narrow-band noise stimuli elicited activity in the surrounding belt regions. Strainer et al. (1997) showed that complex stimuli, such as speech, activate association areas, while pure-tones primarily activate areas in the lateral and medial temporal gyrus. An additional study by Norman-Haignere et al. (2013) showed an increase of activity by stimuli containing ‘resolved harmonics frequency components’ in regions localized to the anterior half of the auditory cortex. These studies support the idea that the MT stimuli may cause activity in more widespread regions of the auditory cortex, which could potentially lead to larger cortical responses.

Fourth, a complex spectrum where frequency regions of high intensity alternate with regions of low intensity (i.e. a line spectrum, whether harmonically or inharmonically related) may give rise to complex excitatory and especially inhibitory stimulation between adjacent tonotopic regions within the cortex. Such interactions may occur to a much lesser degree with stimuli that have a more diffuse spectrum, even when the two stimuli extend over the same total bandwidth.

Of these four possibilities, the fourth and possibly the third are the most consistent with the data in this experiment. The first reason (more locally extensive activity as a result of increased bandwidth) cannot be responsible. This follows because of the lack of difference between the response to PT and NBN (Fig. 3), the significant difference between the response to MT stimuli and NBN of the same bandwidth (Fig. 4), and the lack of difference in the response to narrow band and wide-band MT (Fig. 5). Increased bandwidth therefore seems *not* to be the feature of the stimulus that causes a larger response with complex stimuli, so we can reject the idea that the increased amplitude comes just from locally enlarging the response region of auditory cortex in a manner tonotopically related to stimulus bandwidth.

Although we certainly cannot rule out the second reason, we cannot identify any temporal feature in the MT stimulus that seems capable of inducing greater synchronicity of firing. Because the components of the MT stimulus are inharmonically related, the phase relationship between each pair of components within the set is constantly changing. The only temporal aspect they have in common is their onset and offset, and it is difficult to see why the same set of neurons would respond more synchronously to the onset of the MT stimulus than they do to the onset of the pure tone stimulus.

The third explanation, more remotely extensive neuron firing for the MT stimulus, seems possible. If so, again it is certainly not just the increased bandwidth of the MT stimuli that induces the more widespread remote activity, as the amplitude increase did not occur for the NBN stimuli.

Potential benefit of using MT stimuli in a clinical setup

CAEPs are increasingly used in clinical applications for both hearing aid evaluation (Van Dun et al. 2012) and hearing loss diagnostics (Lightfoot and Kennedy 2006). As a result, reduction of measurement time is of great interest. An advantage of stimuli that elicit larger CAEP responses is a reduction of the number of averages required to extract the response from background noise, resulting in a shorter test duration (see table 3). The use of frequency-specific MT stimuli may therefore be of clinical use in assessing hearing thresholds objectively. A disadvantage of the MT stimuli is that in the case of steeply sloping audiograms, the wider bandwidth of the MT stimuli will likely lead to some under-estimation of the threshold at the centre frequency (Walker et al. 1984).

Corrections due to temporal integration

Auditory stimuli used for CAEP recording are generally shorter than those used for behavioral assessment, due to optimal stimulus lengths for CAEP recording being up to 70 ms (Alain et al. 1997). As stimulus duration lengthens, the perceived loudness of a sound increases and detection threshold lowers (Moore 2012). In this case, it is important to apply corrections to compensate for the higher thresholds found when using short duration stimuli. The results from the behavioral aspect of this study allowed determination these corrections, which account for the difference between hearing thresholds for long and short stimuli due to temporal integration. These values were provided in Table 2. The mean behavioral threshold differences between 50-ms short and 500-ms long stimuli ranged from 5 to 9 dB. It is important to account for these differences in order to determine behavioral hearing thresholds and optimize subsequent hearing aid fitting.

Future work

The present work compared CAEP amplitudes to MT, PT and NBN stimuli, and has been conducted on adults with normal-hearing. It is important to extend this work to subjects with hearing impairment and the newborn population since utilizing the MT stimuli may provide a more efficient approach in objective hearing threshold estimation and/or hearing aid fitting evaluation. Moreover, further studies will need to investigate the generators' location and orientation differences in the human auditory cortex between complex sounds and pure-tones. They could explain the difference of amplitude response observed in this study. This can potentially be achieved using multi-electrode EEG or MEG recording or the functional magnetic resonance imaging (fMRI) technique, which offer a better spatial resolution.

Acknowledgements

This work was supported in part by the HEARing CRC, established and supported under the Australian Cooperative Research Centres Program, an Australian Government Initiative and by support from the Australian Government Department of Health. F.B. designed and performed experiments, analyzed data and wrote the paper; B.V.D. and H.D. wrote the paper and provided critical revision. The authors gratefully thank Mark Seeto for his support with statistical analysis as well as Robert Cowan and Lizzie Whittle for comments on the manuscript.

References

AES. (1991). Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology, 8*, 200-202.

Alain, C., L Woods, D., Covarrubias, D. (1997). Activation of duration-sensitive auditory cortical fields in humans. *Electroencephalography and Clinical Neurophysiology, 104*, 531-539.

Bardy, F., Van Dun, B., Dillon, H., et al. (2014). Least-squares (LS) deconvolution of a series of overlapping cortical auditory evoked potentials: a simulation and experimental study. *Journal of Neural Engineering, 11*, 046016.

Bardy, F., Van Dun, B., Dillon, H., et al. (2014). Deconvolution of overlapping cortical auditory evoked potentials (CAEPs) recorded using short stimulus onset-asynchrony (SOA) ranges. *Clinical Neurophysiology, 125*, 814-826.

Butler, R. A. (1972). Frequency specificity of the auditory evoked response to simultaneously and successively presented stimuli. *Electroencephalography and Clinical Neurophysiology, 33*, 277-282.

Convery, E., Keidser, G., Seeto, M., et al. (2014). Identification of Conductive Hearing Loss Using Air Conduction Tests Alone: Reliability and Validity of an Automatic Test Battery. *Ear and hearing, 35*, e1-e8.

Delorme, A., Makeig, S. (2003). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*, 9-21.

Fox, J., Weisberg, S. (2012). car: Companion to applied regression (Version 2.0-19). Available from http://cran.r-project.org/package=car.

Golding, M., Dillon, H., Seymour, J., et al. (2009). The detection of adult cortical auditory evoked potentials (CAEPs) using an automated statistic and visual detection. *International journal of audiology, 48*, 833-842.

Hirata, Y., Kuriki, S., Pantev, C. (1999). Musicians with absolute pitch show distinct neural activities in the auditory cortex. *Neuroreport, 10*, 999-1002.

Hothorn, T., Bretz, F., Westfall, P. (2013). multcomp: Simultaneous inference in general parametric models (Version 1.3-1). Available from http://cran.r-project.org/package=multcomp.

Howard III, M. A., Volkov, I. O., Abbas, P. J., et al. (1996). A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain research, 724*, 260-264.

Institute, A. N. S. (2008). Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms (ANSI S3.1-1999, R2008). In A. N. S. Institute (Ed.), *ANSI S3.1-1999, R2008*. New York, NY.

Lightfoot, G., Kennedy, V. (2006). Cortical Electric Response Audiometry Hearing Threshold Estimation: Accuracy, Speed, and the Effects of Stimulus Presentation Features. *Ear and Hearing, 27*, 443-456.

Lütkenhöner, B., Klein, J.-S. (2007). Auditory evoked field at threshold. *Hearing research, 228*, 188-200.

Mäkelä, J., Hari, R., Leinonen, L. (1988). Magnetic responses of the human auditory cortex to noise/square wave transitions. *Electroencephalography and clinical neurophysiology, 69*, 423-430.

Moore, B. C. (2012). *An introduction to the psychology of hearing*. (6th ed.): Emerald.

Norman-Haignere, S., Kanwisher, N., McDermott, J. H. (2013). Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *The Journal of Neuroscience, 33*, 19451-19469.

Onishi, S., Davis, S. (1968). Effects of duration and rise time of tone bursts on evoked V potentials. *Journal of the Acoustical Society of America, 44*, 582-590.

Organization, I. S. (1994). Acoustics: Reference zero for the calibration of audiometric equipment. Part 2: Reference equivalent threshold sound pressure levels for pure tones and insert earphones. In: International Organization for Standardization Geneva.

Pantew, C., Kevanishvili, Z. S., Galle, E., et al. (1975). ERA-stimulation with free-programable frequency and intensity Sequence—A method to reduce the examination time. *European Archives of Oto-Rhino-Laryngology, 211*, 43-49.

Picton, T. (2011). Late auditory evoked potentials: Changing the things which are. In *Human auditory evoked potentials*. Brisbane, Australia: Plural Pub.

Pinheiro, J., Bates, D., DebRoy, S., et al. (2013). nlme: Linear and nonlinear mixed effects models (Version 3.1-113). Available from http://cran.r-project.org/package=nlme.

R Development Core Team. (2013). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2012. In: ISBN 3-900051-07-0. Available from: http://www. R-project. org.

Ross, B., Lütkenhöner, B., Pantev, C., et al. (1999). Frequency-specific threshold determination with the CERAgram method: basic principle and retrospective evaluation of data. *Audiology and Neuro-Otology, 4*, 12-27.

Seither-Preisler, A., Krumbholz, K., Lütkenhöner, B. (2003). Sensitivity of the neuromagnetic N100m deflection to spectral bandwidth: a function of the auditory periphery? *Audiology and Neurotology, 8*, 322-337.

Shahin, A., Roberts, L. E., Pantev, C., et al. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport, 16*, 1781-1785.

Strainer, J. C., Ulmer, J. L., Yetkin, F. Z., et al. (1997). Functional MR of the primary auditory cortex: an analysis of pure tone activation and tone discrimination. *American journal of neuroradiology, 18*, 601-610.

Tervaniemi, M., Schröger, E., Saher, M., et al. (2000). Effects of spectral complexity and sound duration on automatic complex-sound pitch processing in humans–a mismatch negativity study. *Neuroscience letters, 290*, 66-70.

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. *Audiology Research, 2*, e13, 65-76.

Walker, G., Dillon, H., Byme, D. (1984). Sound field audiometry: Recommended stimuli and procedures. *Ear and hearing, 5*, 13-21.

Wessinger, C., VanMeter, J., Tian, B., et al. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience, 13*, 1-7.

Wickham, H. (2011). reshape: Flexibly reshape data (Version 0.8.4). Available from http://cran.r-project.org/package=reshape.

Zacharias, N., Sielużycki, C., Kordecki, W., et al. (2011). The M100 component of evoked magnetic fields differs by scaling factors: Implications for signal averaging. *Psychophysiology, 48*, 1069-1082.