Cortical encoding of speech acoustics: Effects of noise and amplification

Abin Kuruvilla-Mathew, Suzanne C Purdy & David Welch


To link to this article: http://dx.doi.org/10.3109/14992027.2015.1055838

Published online: 23 Jul 2015.

Article views: 185

View related articles

View Crossmark data
Cortical hearing loss leads to rapid and wide ranging changes in the auditory cortex which are yet to be understood completely (Lomber & Eggermont, 2006). These can be investigated using event-related potentials, particularly the obligatory cortical auditory-evoked potentials (CAEPs), which reflect the audibility and physical properties of a stimulus (Hyde, 1997; Martin et al., 2008). Historically CAEPs have been recorded using brief stimuli such as clicks and tone bursts. More recently studies have used synthetic and natural speech sounds, including natural vowels, consonants, consonant-vowel (CV) syllables, and synthetic speech stimuli (Sharma et al., 2000; Tremblay et al., 2003; Agung et al., 2006; Korczak & Stapells, 2005, 2010). Auditory cortex neurons are more sensitive to transient changes at the onset of a stimulus than the presence of an ongoing stimulus (Philips & Hall, 2002). CAEPs are primarily onset responses but are also produced by the offset of a stimulus (Pratt et al., 2008). Hence, CAEPs evoked by consonant-vowel (CV) syllables have overlapping onset responses evoked by the consonant and change responses evoked by the consonant-vowel transition and offset (Ostroff et al., 1998; Sharma et al., 2000; Martin et al., 2008; Digeres et al., 2009).

Previous studies using speech stimuli have demonstrated the ability of CAEPs to show encoding of speech features (Tremblay et al., 2003; Purdy et al., 2006; Korczak et al., 2010; Doellinger et al., 2011). For example, Digeres et al. (2009) found significant differences in CAEP waveforms in response to the spectro-temporally different consonant-vowel (CV) syllables, /da/ and /ta/. Group waveform differences in CAEPs to different speech stimuli are also evident in individual subject waveforms recorded in quiet for unaided and aided conditions (Tremblay et al., 2003, 2006).

It has been proposed that onset CAEPs reflect stimulus level relative to the level of background noise (SNR) rather than the absolute stimulus level (Billings et al., 2009; Baltzell & Billings, 2013). In general, as the SNR becomes unfavourable, the morphology of the CAEPs becomes poorer. Changes in tone and speech evoked CAEP morphology occurs for different masking stimuli, including white noise and speech (Whiting et al., 1998; Kaplan-Neeman et al., 2006;
Cortical encoding of speech acoustics

Billings et al, 2011). Obligatory CAEPs are sensitive to the acoustic characteristics of speech signals in noise (Kaplan-Neeman et al, 2006). Kaplan-Neeman compared CAEPs evoked by /da/ and /ga/; N1 latencies for both speech stimuli increased with the addition of background noise. N1 latencies were longer for /ga/ than for /da/ in quiet and in noise. The authors proposed that this was because of the long /ga/ burst duration. Acoustic cues specific to the speech stimuli were encoded even in the presence of background noise. Studies such as this could help understanding of difficulties discriminating specific acoustic cues in the presence of noise in people with hearing loss.

More recently CAEPs have been used as a clinical tool for validation of hearing-aid fittings in children (Dillon, 2005; Purdy et al, 2005). This may be useful in infants and difficult to test populations where behavioral information regarding hearing-aid benefit for speech perception is limited and unreliable (Carter et al, 2010). However, several studies have highlighted uncertainty in the clinical use of aided CAEPs and there is variability in the methodology and results across individual studies and participants (Korczak et al, 2005; Tremblay et al, 2006; Billings et al, 2007; Marynewich et al, 2012; Billings et al, 2011; Munro et al, 2011; Easwar et al, 2012b). Recent studies (Munro et al, 2011; Billings et al, 2012) investigating the clinical utility of aided CAEPs suggest that it is valid to use CAEPs as an indicator of speech stimulus detection, but not discrimination.

Hearing aids may use several nonlinear processing algorithms, such as compression and noise suppression, and may produce unwanted nonlinear distortion (Kim & Loizou, 2011). Prior studies indicate that SNR and stimulus onset modification (e.g. altered rise time) are two stimulus characteristics that have significant effects on aided CAEPs (Billings et al, 2007, 2009, 2013; Jenstad et al, 2012). Hearing-audio output also varies with different speech contexts (phonemes in running speech vs. isolation), which may influence interpretation of aided CAEPs (Easwar et al, 2012a). Only a few studies have measured hearing-audio output in the ear and correlated this to CAEP findings (Tremblay et al, 2006; Billings et al, 2007, 2009). Natural speech tokens are preferred to non-speech sounds for unaided and aided CAEP testing if the goal is to better understand speech processing in people with hearing loss (Aiken & Picton, 2008). Aided CAEPs could reflect changes in the short-term envelope of the stimulus triggered by changes in acoustic spectro-temporal features, level, and SNR due to the hearing aid. To our knowledge no study has examined the effects of speech stimulus onset characteristics on aided CAEPs.

Poor performance in noise is one of the frequent complaints reported by hearing-audio users (Kochkin, 2007). It may be possible to use aided-speech-evoked CAEPs to better understand suprathreshold speech perception difficulties of people with sensorineural hearing loss using hearing aids. The present study investigated the effects of natural speech stimuli on CAEP components across a range of conditions, unaided and aided in quiet and in noise, in adults with normal hearing. When combined with acoustic analysis of the speech stimuli, CAEPs elicited by a number of naturally spoken speech sound contrasts may help to better understand neural processing of place, manner, and voicing cues (Tremblay et al, 2003; Digeser et al, 2009; Korczak et al, 2010). The selection of speech contrasts in the current study was based on several factors, but was primarily based on speech features (place, manner, voicing). In addition, speech stimuli were selected that encompassed commonly observed speech errors (e.g. Warner-Czyz et al, 2010) and common consonant confusion errors in background noise of people with hearing loss (Doyle & Edgerton, 1981; Phatak et al, 2009; Woods et al, 2010). Studying these effects first in adults with normal hearing wearing a hearing aid allows us to separate the effects of amplification from hearing loss, since hearing loss is accompanied by frequency and temporal deficits and reduced auditionability (Moore, 1996).

The current study investigated speech stimuli and background noise dependent changes in CAEP components in unaided and aided conditions and determined amplification effects on CAEPs in young adults with normal hearing. Specifically this study investigated effects of: (1) noise (multi-talker babble) and aiding across speech stimuli (2) speech contrasts, (3) amplification (unaided versus aided) on speech CAEPs, and (4) speech stimulus onset characteristics. In-the-ear probe-microphone measurements were used to determine effects of hearing aid on SNR, since previous research by Billings and colleagues (Billings et al, 2007, 2009) highlighted the impact of SNR on aided CAEPs.

**Materials and Methods**

Testing was conducted in two sessions of three hours each. Unaided and aided CAEPs were recorded in adults with normal hearing using seven speech stimuli (Figure 1). Participants were monaurally fitted in the right ear when testing the aid condition. For all conditions, the left ear was plugged using a foam ear plug. Conditions (aided/ unaided, quiet/noise) were counterbalanced and stimulus presentation order was randomized across participants and test sessions. In-the-canal signal acoustic measurements were done with and without the hearing aid to determine stimulus levels, amplification effects (unaided versus aided) on the SNR, and to compare stimulus acoustic characteristics to CAEP findings.

**Participants**

Ten young adults with normal hearing were recruited (seven females, three males) aged 19 to 35 years (M 24 years, SD 4.6). All were right handed, English speakers, with pure-tone audiometric thresholds of 20 dB HL or better at 250 to 8000 Hz and normal Type A tympanograms. CAEPs were recorded in quiet and in noise, with and without a hearing aid in place. For all conditions, the left ear was plugged using a foam ear plug. All participants gave informed consent before testing.

**Stimuli**

Stimuli consisted of naturally produced speech syllables (/di, ti, gi, mi, pi, si, fi/) and multi-talker babble as the masker. Speech stimuli and contrasts were selected based on differences in: (1) single speech features such as place (/di-gi/ , /fi-si/ , /ti-pi/) and voicing (/di-ti/), (2) multiple speech features including place, manner, and voicing (/mi-pi/ , /mi-ti/ , /gi-ti/), and (3) consonants dominant in low (< 3000 Hz) (/di, gi, mi/) versus high frequency (> 3000 Hz) (/pi, ti, fi, si/)
Speech syllables were recorded using a native New Zealand female speaker in a soundproof room via an AKG HC 577 L omnidirectional headset microphone placed 3 cm from the lips of the speaker, attached to an M-Audio MobilePre, using Adobe Audition version CS6 sound editing software, with a sampling rate of 44100 Hz and 16 bit quantization rate. Total duration of each syllable was 246 ms after editing; before editing each syllable was approximately 400–500 ms in duration. Segments were removed during the steady-state part of the vowel, starting and ending at zero crossings to prevent audible clicks. Stimulus onsets were not changed in an effort to minimize the effect on CAEP waveforms, however the reduction in stimulus duration does reduce ecological validity of the stimuli. Stimuli were individually root mean square (RMS) normalized using Adobe Audition and presented via an Impact 50 Turbosound loudspeaker at 0-degrees azimuth at 1 m distance at 65 dB SPL (overall RMS). Spectrograms and time-domain waveforms for the seven stimuli derived using Praat version 5.3.53 are shown in Figure 1.

The competing noise signal was eight-talker babble presented via a DELL laptop. The babble noise comes from an anechoic recording of four males and four females reading out loud from different materials at the same time (NAL CD Speech and Noise for Hearing Aid Evaluation; Keidser et al, 2002). The spectra of the speech stimuli and multi-talker babble, showing the substantial differences in the region of dominant spectral energy across stimuli are shown in the Figure 2. Spectra were derived for the unaided and aided stimuli using in-the-canal measurements and Adobe C S6 software. For each stimulus, babble noise was continuously presented for the entire duration of the noise condition, beginning a few seconds before the speech stimuli, at 55 dB SPL (+10 dB SNR).

For aided cortical recordings a digital nonlinear Oticon Alta Pro behind-the-ear (BTE 13) hearing aid was coupled to a plastic closed temporary tip that occluded the right ear canal for all participants. Alta Pro uses a 10-channel sound processor. The processing delay time of the hearing aid is 5–6 ms (Schum & Beck, 2006), with little variation across frequencies. The frequency range of this hearing aid extends from 100 to 7700 Hz; equivalent input noise is 18 dB SPL in a 2cc coupler. The hearing aid was programmed using Oticon’s voice aligned compression (VAC) fitting rationale (Flynn, 2004), assuming a N4 audiogram, which has thresholds ranging between 55 and 80 dB HL and a pure-tone average across 500, 1000, and 2000 Hz of 55 dB HL (Bisgaard et al, 2010). The VAC fitting algorithm provides curvilinear compression with both low-level compression, as well as increased linearity for high-level signals. The hearing aid was set to omnidirectional mode with all other auto-listening support features (noise reduction, feedback cancellation) disabled. In the VAC algorithm the hearing aid is programmed to provide more gain for soft-moderate speech and in the moderate to loud range, the compression ratio is unchanging and low even in the presence of background noise. Manufacturer specific ‘first fits’ to a particular audiogram are significantly different and thus hearing output levels can be highly variable (Keidser et al, 2003).

In-the-canal acoustic measurements
The level of the hearing aid processed speech stimuli should affect amplitudes and latencies of CAEP components. Therefore, stimulus levels were measured with and without the hearing aid in place.
Output levels of the hearing-aid transduced stimuli for all conditions were measured in the ear canal of participants using an Etymotic Research Inc ER-7C probe microphone and preamplifier set to +20 gain (see Figure 2) connected to a DELL laptop. Stimuli were recorded and analysed using Adobe C S6 software. Measurements were performed by placing a probe tube 28 mm past the intertragal notch. To maintain correct positioning throughout testing, the probe tube was taped to the earlobe. For these measurements CAEP stimuli were delivered via the equipment used for evoked potential recordings. Figure 2 indicates that in the unaided condition, energy was most intense in frequencies below 500 Hz, presumably showing contributions from the vowel. The onset consonants were generally less intense than the vowel (by about 10 dB) as observed in running speech (Easwar et al, 2012a). In the aided condition, the high frequency emphasis onset consonants were more intense than the lower frequency sounds (see Figure 2), as expected as the hearing aid was programmed to fit a sloping N4 audiogram configuration.

Mean output levels of the speech stimuli in the presence of background noise were slightly more intense (by 6 dB on average) than for the quiet condition, presumably due to the VAC algorithm, which provides a more linear response when background noise is present. The peak output level of speech in the low and high frequency range respectively was 69 and 58 dB SPL in the unaided condition and 72 and 82 dB SPL in the aided condition. Participants reported that the aided speech stimuli were 'loud but comfortable'.

SNRs for the unaided and aided conditions (in quiet) were measured for nine of the 10 participants using the in-the-canal recordings (Adobe C S6 software). SNRs for the unaided and aided conditions were obtained from one repetition of the stimulus presentation. Signal level was computed based on the average RMS levels of each speech stimulus. The level of the noise floor was measured from the interstimulus interval. The noise floor of the hearing aid was 40 to 58 dB SPL between 129 Hz and 7019 Hz, with predominant spectral energy extending from about 990 to 7019 Hz. As gain was greater at higher frequencies, the noise floor of the hearing aid would also be greater at higher frequencies. SNRs were calculated individually for the initial 50-ms and the entire 246-ms of the speech stimuli. The 50-ms SNRs were computed because of the important influence of stimulus onset on CAEPs.

**Electrophysiology**

The Neuroscan SCAN (version 4.3) was used for recording CAEPs. CAEPs were obtained using four EEG channels with 10-mm silver-silver-chloride disc electrodes placed at Cz and Fz, referenced to M1 and M2. Data presented here are for the contralateral (M1) reference electrode (sound was presented frontally but the left ear was plugged). The ground electrode was located on the forehead and eye blink activity was monitored using electrodes placed above and below the right eye. Electrode impedances were under 3 kΩ. The Neuroscan STIM system was used to present the speech stimuli. Two different randomized sequences of the seven stimuli were presented, with two blocks of 150 sweeps for each stimulus and each condition. Interstimulus interval was 920 ms. EEG was amplified with a gain of 1000 and sampled at the rate of 1000 Hz. EEG data were pre-processed using Neuroscan’s built-in functions. Recordings with eye blink artifacts were corrected using a regression procedure, the ocular artifact rejection function in Neuroscan software. First, the vertical electro-oculogram (VEOG) channel was scanned for the maximum eye movement potential. EOG deviations of more than 10% from the maximum were used as indicators of blinks. A minimum of 20 blinks was required to estimate an average blink. The procedure discarded artifacts starting < 400 ms before a previous artifact, to avoid double detection. From the average VEOG ocular artifact, transmission coefficients were computed for each EEG channel by estimating the covariance of the averaged potentials of the VEOG channel with the EEG channels. The contribution of the average blink from the VEOG channel was then subtracted from all other EEG channels on a point-by-point basis. EEG epochs with pre-stimulus to 600 ms post-stimulus time windows were extracted and baseline corrected before averaging. The artifact rejection threshold was set to ± 50 μV.
Testing was performed in a double-walled sound attenuating booth. Sounds were calibrated using a Bruel and Kjaer 2215 sound level meter measured at 1 m distance from the loudspeaker located in the participant’s midline. Short breaks were given between testing conditions. Participants were seated comfortably on a reclining chair while watching a close captioned DVD of their choice (Lavoie et al, 2008). The study was approved by the University of Auckland Human Participants Ethics Committee.

Data analysis
Grand average CAEP waveforms were created for each participant by averaging two blocks of 150 runs for each condition. CAEP peak amplitudes and latencies were identified for each subject by two independent observers. Waves P1, N1, and P2 were analysed at Cz, as this electrode site gave the largest response waveforms. Repeated measures analyses of variance were performed separately on latencies and amplitudes of each component (P1, N1, and P2). Analyses included the 2 × 2 × 7 factors of noise condition (quiet and noise), aiding condition (unaided and aided), and speech stimuli. Post-hoc analyses of repeated measures ANOVA interaction effects were performed using paired t tests. A Bonferroni correction was used to adjust the alpha level of .05 to correct for the number of post-hoc paired comparisons. The amplitude of P1 was defined as the largest positive deflection occurring between 50–100 ms after stimulus onset. The amplitude of N1 was identified as the largest negative deflection between 80–160 ms after stimulus onset. P2 amplitude was defined as the largest peak occurring between 170 and 270 ms. Peak latency was measured at the centre of the peak. When the waveform contained a double peak of equal amplitude or a peak with a plateau, latency was measured at the midpoint of the peak. Using these criteria it was possible to pick peaks for all participants and all speech stimuli. Peaks appeared to be present but were difficult to distinguish from the noise floor for responses to /si/ in babble for four individuals. For these four measurements the residual noise level present in the averaged response within the latency region of interest was used as the estimate of peak amplitude and the latency was picked at the point of peak amplitude within the latency region of interest. To evaluate variations in SNR for the quiet condition a repeated measures analysis of variance was performed for the 2 × 2 × 7 factors of amplification (unaided versus aided SNRs), stimulus time window (50 ms and 246 ms), and speech stimuli.

Results
CAEP latencies and amplitudes
CAEPs with characteristic morphology were elicited for all participants across all conditions. The repeated measures ANOVA (Table 1) showed significant main and interaction effects on N1 and P2 latencies and amplitudes for the three independent variables (noise, speech stimuli, and aiding). Effects on P1 latencies and amplitudes were only seen for noise and speech stimuli.

Effects of noise across speech stimuli
A general trend evident in Figure 3 (a) & (b) was an increase in latency and decrease in amplitude of all CAEP components in noise for the unaided and aided conditions across all speech stimuli. A significant main effect of noise was seen for latencies of P1, N1, and P2 (P1: F[1,9] = 55.61, p < .001; N1: F[1,9] = 235.13, p < .001; and P2: F[1,9] = 249.13, p < .001). Significant main effects of noise

<p>| Table 1. ANOVA Results. Repeated measures analyses of variance (ANOVA) results for data collected at electrode Cz. Results for latency and amplitude across components P1, N1, and P2 are included. |</p>
<table>
<thead>
<tr>
<th>Noise</th>
<th>Stimulus</th>
<th>Aiding</th>
<th>F-stat (df)</th>
<th>P-value</th>
<th>F-stat (df)</th>
<th>P-value</th>
<th>F-stat (df)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latency</td>
<td>P1</td>
<td>55.6 (1,9)</td>
<td>.001</td>
<td>8.5 (6,54)</td>
<td>.001</td>
<td>23.1 (6,54)</td>
<td>.001</td>
<td>10.5 (1,9)</td>
</tr>
<tr>
<td>N1</td>
<td>225.1 (6,54)</td>
<td>.001</td>
<td>10.4 (1,9)</td>
<td>.001</td>
<td>42.9 (6,54)</td>
<td>.001</td>
<td>10.5 (1,9)</td>
<td>.001</td>
</tr>
<tr>
<td>P2</td>
<td>295.3 (6,54)</td>
<td>.001</td>
<td>10.4 (1,9)</td>
<td>.001</td>
<td>42.9 (6,54)</td>
<td>.001</td>
<td>10.5 (1,9)</td>
<td>.001</td>
</tr>
<tr>
<td>Amplitude</td>
<td>P1</td>
<td>7.3 (1,9)</td>
<td>.024</td>
<td>11.5 (6,54)</td>
<td>.001</td>
<td>5.3 (6,54)</td>
<td>.001</td>
<td>4.0 (1,9)</td>
</tr>
<tr>
<td>N1</td>
<td>3.6 (1,9)</td>
<td>.089</td>
<td>11.5 (6,54)</td>
<td>.001</td>
<td>5.3 (6,54)</td>
<td>.001</td>
<td>4.0 (1,9)</td>
<td>.001</td>
</tr>
<tr>
<td>P2</td>
<td>3.6 (1,9)</td>
<td>.089</td>
<td>11.5 (6,54)</td>
<td>.001</td>
<td>5.3 (6,54)</td>
<td>.001</td>
<td>4.0 (1,9)</td>
<td>.001</td>
</tr>
</tbody>
</table>
on amplitudes was seen only for P1 (P1: $F[1,9] = 7.36, p = .024$), and there was also a trend for smaller P2 amplitudes in noise (Table 1). There were significant Noise × Stimuli interaction effects for all latencies (P1: $F[6,54] = 7.94, p < .001$; N1: $F[6,54] = 29.24, p < .001$; and P2: $F[6,54] = 29.72, p < .001$). Post-hoc paired t-tests to determine the influence of noise for different speech stimuli (averaged across aiding conditions) showed no significant latency differences ($p \geq .007$) in quiet versus noise for P1 for two speech sounds: (/i/: $t[9] = -.006, p = .995$); /si/: $t[9] = .67, p = .519$) and for P2 for three speech sounds (/i/: $t[9] = -.65, p = .595$); /j/: $t[9] = 1.11, p = .296$); /si/: $t[9] = .34, p = .742$). All other speech stimuli had significantly longer latencies in noise than in quiet ($p \leq .001$). A three-way interaction between Aiding × Noise × Speech stimuli was seen for N1 latency ($F[6,54] = 2.85, p = .016$) and hence post-hoc paired t-tests were conducted to determine the effect of noise on N1 latencies for each stimulus separately for aided and unaided conditions. Two speech stimuli failed to show significant latency differences for quiet versus noise in the unaided condition: /si/: $t[9] = -1.78, p = .108$); /j/: $t[9] = -2.01, p = .075$). Three speech stimuli did not show a difference in the aided condition: /i/: $t[9] = -2.06, p = .013$); /si/: $t[9] = .96, p = .362$); /j/: $t[9] = -2.62, p = .028$). In general, the noise did not affect latencies for high-frequency emphasis speech stimuli.

There were significant two-way interactions for Noise × Stimuli for N1 and P2 amplitudes (N1: $F[6,54] = 4.84, p < .001$; and P2: $F[6,54] = 5.52, p < .001$). As was observed for latencies, the effect of noise on N1 amplitude was inconsistent across stimuli. Post-hoc comparisons of N1 amplitudes in quiet versus noise showed no significant effects of noise, $p > .007$. Although a general trend of reduced P2 amplitude in noise was seen across all speech stimuli, post-hoc comparisons also showed no significant effect of noise on P2 amplitude ($p > .007$). A Noise × Aiding interaction was seen only for N1 amplitude ($F[1,9] = 8.3, p = .018$); N1 amplitude was slightly increased by noise for the aided condition, by 12% on average, and reduced on average by noise in the unaided condition by 15%.

**Effect of speech contrasts**

A significant main effect of speech stimulus was seen for latencies of P1, N1, and P2 (P1: $F[6,54] = 8.53, p < .001$; N1: $F[6,54] = 23.16, p < .001$; and P2: $F[6,54] = 42.94, p < .001$). Significant main effects of stimulus on amplitudes were seen for N1 and P2 (N1: $F[6,54] = 11.50, p < .001$; P2: $F[6,54] = 5.30, p < .001$). CAEP latencies and amplitudes were compared for speech contrasts in Table 2. Results of paired t-tests for unaided speech contrasts along

**Figure 3.** (a, upper) Unaided grand mean CAEP waveforms (N = 10) recorded at Cz for each speech stimulus in quiet and in noise. The waveform in black is the grand average of all stimuli for each condition (quiet and noise). (b, lower) Aided grand mean waveforms (N = 10) recorded at Cz for each speech stimulus in quiet and in noise. The waveform in black is the grand average of all stimuli for each condition (quiet and noise).
Effects of amplification (aided versus unaided)

Although a main effect of hearing-aid amplification was seen only for P2 latency ($F_{[1,9]} = 10.56, p = .010$), the statistical trend was observed for P2 amplitude (Table 1). In general, compared to the unaided condition, P2 was later and larger for the aided condition. Two-way interactions between Aiding × Stimuli were found for N1 ($F_{[6,54]} = 2.17, p = .019$) and P2 latencies ($F_{[6,54]} = 3.49, p = .007$), however post-hoc comparisons (averaged across noise condition) showed that latencies, after correcting for hearing-aid delay, were not significantly different between unaided and aided conditions across speech stimuli, $p \leq .007$.

Effects of amplification on SNR

The effects of amplification (unaided vs. aided) on the in-the-canal measurements of SNR for each stimulus were investigated using repeated measures ANOVA (Figure 4). The main effect of higher SNRs in the aided compared to the unaided condition was significant ($F_{[1,8]} = 78.13, p < .001$), averaged across speech stimuli and time window. Apart from some minor variations, unaided SNRs were similar (within 2 dB) across stimuli when computed across the entire speech stimulus (246 ms), as expected because the stimuli were amplitude-normalized. There was a significant two-way interaction between Aiding × Speech stimuli ($F_{[6,48]} = 106.17, p < .001$); the effect of amplification SNR varied across stimuli. In general, aided SNRs were better than unaided and reflected the frequency response of the hearing aid, with higher SNRs for high-frequency emphasis speech stimuli. The hearing aid was programmed for a sloping moderate-severe hearing loss, with more gain at high frequencies. The drop in SNR for /si/ may be due to the spectral peak for /si/ at > 8000 Hz being outside the hearing aid’s effective amplification range (100–7700 Hz). Three high frequency emphasis speech sounds (/pi, ti, fi/) showed very high SNR values, particularly for the aided condition.

SNR calculations based on comparison of pre-stimulus noise floor to the stimulus level could overestimate SNR as this calculation does not consider the effect of input gain on the frequency spectrum of the hearing-aid noise floor. However, a comparison of real ear measures of output (Figure 2) to noise floor values also indicates better aided SNRs for the high-frequency emphasis speech stimuli.

Effects of speech contrast were similar for aided and unaided conditions (Table 3), however, a few contrasts had significant latency and amplitude differences in the aided, but not the unaided condition. Enhanced CAEP differences between speech contrasts for the aided condition were mainly seen for latency measures.
Cortical encoding of speech acoustics

A short burst duration; for this stimulus SNR values did not differ between onset and the total stimulus. Time window did not influence SNR for /mi/ for the unaided condition, but the onset SNR was reduced in the aided condition because the hearing aid changed the onset amplitude (Figure 5). The onset SNR is increased for /ti/ for the aided condition due the effect of the hearing-aid gain on the onset burst (Figure 5).

Discussion

Summary of key findings

Our results demonstrate that speech stimuli and background noise can affect CAEP components for both unaided and aided conditions. The addition of background noise resulted in increased N1 and P2 latencies and decreased P2 amplitudes for most speech stimuli. Furthermore, CAEPs showed neural encoding of different speech stimuli and acoustic features. There was a strong association between stimulus onset characteristics and CAEPs. Acoustic measures of hearing-aid output (SNR and real-ear output level) were quantified. Overall, aiding resulted in later and larger P2 responses.

Effects of noise across speech stimuli

We investigated the neural representation of unaided and aided CV syllables in quiet and in noise (10 dB SNR). Noise and aiding are known to affect CAEPs (Whiting et al., 1998; Korczak et al., 2010) but previous studies have not investigated a wide range of natural speech stimuli. Both amplitudes and latencies were affected by background noise although, overall, noise had more effect on latencies. N1 and P2 latencies increased and P2 amplitudes decreased in noise. Increased CAEP latencies in noise is a consistent pattern seen across studies (Kaplan-Neeman et al., 2006; McCullagh et al., 2012; Billings et al., 2013). This can be attributed to disruption of synchronized neural discharges to stimulus onset. Although significant increases for N1 and P2 latencies and a trend for reduction of P2 amplitudes were evident, N1 amplitude was not consistently affected by background noise. This could be because cortical neurons adaptively adjust their thresholds relative to the background noise (Philips, 1990). Therefore, a proportional decrease in N1 amplitude with background noise level may be evident only once the masker level is above a certain threshold for effecting a reduction in N1 amplitude. This is consistent with Whiting et al.’s finding that N1 amplitude significantly reduced once SNR reduced to < 5 dB. N1 amplitude increase in noise for the aided condition may be the result of slightly intense hearing-aid output when background noise was present (by 6 dB on average).

Effects of multi-talker babble on CAEP morphology varied across stimuli, which may reflect the spectral relationship between the noise and speech stimuli. Latencies for the three high-frequency emphasis sounds (/ti, si, fi/) were not affected by noise, presumably because the multi-talker babble had little energy at high frequencies (Figure 2), and hence had minimal impact on the onset envelope. Gordon-Salant (1985) reported similar spectrum-dependent energetic speech babble masking of CV syllables based on behavioural measures of speech

Table 3. Statistically significant CAEP latency (L) & amplitude (A) measures for post hoc comparisons of speech contrasts where the main effect of stimuli was significant.

<table>
<thead>
<tr>
<th>Speech contrast</th>
<th>Quiet</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>/di/ vs /ti/</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>/di/ vs /gi/</td>
<td>#</td>
<td>*</td>
</tr>
<tr>
<td>/ti/ vs /mi/</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>/ti/ vs /gi/</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>/si/ vs /fi/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* = p ≤ .007 unaided. # = p ≤ .007 aided.

Figure 4. Unaided (left column) and aided (right column) acoustic SNRs across speech stimuli for the initial 50 ms and the entire 246 ms of the stimuli. Measurements were made with and without the hearing aid in place and the speech stimuli presented at 65 dB SPL. The hearing aid was programmed assuming an N4 audiogram. Aided SNRs were better than unaided, with higher SNRs for high-frequency emphasis speech stimuli.
perception. This frequency-specific masking effect was less apparent for amplitudes. Although P2 amplitudes consistently reduced in noise this did not reach statistical significance. N1 amplitude also did not show a consistent reduction in amplitude across stimuli in noise. The +10 dB SNR used in the present study was not low enough for the noise to substantially decrease CAEP amplitudes consistently across all speech stimuli (Whiting et al., 1998).

Effects of speech contrasts on CAEPs

While CAEPs may be a suitable approach for studying speech processing in people with hearing loss, age-related hearing and cognitive deficits, and auditory processing disorder (Billings et al., 2013; Sharma et al., 2014; Wilson, 2013), the ability of CAEPs to show encoding of different speech features in people with hearing loss wearing a hearing aid is not yet established. The different natural speech stimuli investigated in the current study evoked distinct neural response patterns based on group data. These distinct patterns were also reliably recorded in individuals (see Figure 6). Feature differences between speech contrasts, CAEP differences, and findings from previous studies are summarized in Table 2. Overall results indicate that CAEPs were sensitive to the various speech features investigated here. There were significant latency and/or amplitude differences for each contrast, with the exception of /ʃi/-/si/.
In general, CAEP responses to stop consonants (/di, ti, pi, gi/) were earlier and larger compared to steady consonants (/mi, si, fi/) (Figure 7). A possible explanation for the stop consonant effect on CAEPs might be the sharp onset and higher burst energy of the stop consonants compared to the steady/non-stop consonants (Gage et al., 1998; Young, 2008). Similar findings were reported by Golding et al. (2006) for the unvoiced stop consonant, /ti/, which produced greater CAEP amplitude and shorter latency than a voiced nasal consonant, /mi/. Low-frequency stop consonants /di/ and /gi/ produced larger amplitude responses compared to high-frequency emphasis stops, /ti/ and /pi/. This is consistent with previous studies showing that high-frequency sounds produce smaller CAEPs than low-frequency sounds of the same intensity (Picton et al., 1978). Antinoro et al. (1969) also observed a decrease in peak-to-peak amplitude with increase in tone-burst frequency. Also, aperiodic speech sounds have been shown to evoke smaller N1m responses than periodic sounds (Yrttiaho et al., 2008). CAEPs to /si/ and /fi/ were generally smaller compared to other speech stimuli in the current study. This difference in CAEP morphology could be attributed to the high frequency content of the onset consonant or the slower rise time and lack of burst energy at the onsets of these stimuli. Responses to fricative-vowel stimuli are usually dominated by prominent later peaks associated with the transition from the consonant to the vowel (Tremblay et al., 2003).

A previous study by Sharma et al. (1999) proposed that N1 morphology reflected changes in VOT. A later paper, however, concluded that N1 morphology is not the cortical correlate of VOT differences between voiced and unvoiced speech stimuli (Sharma et al., 2000). An interesting finding in the current study was the difference in morphology between CAEPs evoked by voiced and voiceless consonants (Figure 8). P2 and the following negativity were broader and significantly later for voiced consonants. This distinction could be attributed to the continuity in the spectral and temporal characteristics of voiced CVs compared to voiceless CVs. Voiceless consonants have a short VOT and an onset burst with spectral energy similar to the following vowel's formant frequencies. In contrast voiceless consonants have longer VOT and little energy in the first formant (F1) frequency region before the onset of the vowel (Figure 1). The pause in the F1 frequency region for voiceless consonants could be contributing to the negativity/positivity in the 200–400 ms region seen for voiceless but not voiced CVs. For example, in Figure 8, latencies of the negative peaks between 200 and 300 ms for the voiceless speech stimuli are arranged in time as one would expect given the increasing delays of voicing onset (/pi < ti < fi < si/). This pattern was also reported by Ostroff et al. (1998) for /sei/, and by Tremblay et al. (2006) for /si/ and /fi/. Digerer et al. (2009) also found that short duration CV syllables evoked CAEP components that interfered with each other. The overlap in CAEPs to successive portions of speech was noted by Aiken et al. (2008) for responses to sentences. The current study suggests that CAEPs to CV stimuli are composite responses to short-term spectro-temporal characteristics such as rise time, primarily influencing the onset response at N1, and a P2/acoustic change complex dependent on the consonant vowel transition (Ostroff et al., 1998; Martin et al., 2008).

Effects of amplification on onset CAEPs

N1's sensitivity to the stimulus envelope/rise time at stimulus onset is well documented (Onishi & Davis, 1968; Easwar et al., 2012c). Sensitivity of CAEPs to stimulus onset characteristics is especially important when considering hearing-aid complex automatic signal processing, hearing-aid processing delay, and expansion and compression characteristics where the time constants vary, but can be as short as a few milliseconds.

Jenstad et al. (2012) measured the effect of hearing-aid processing on the onset time of their 1000-Hz stimulus and linked this to aided CAEP amplitudes. Depending on the hearing-aid gain and frequency response the prominence of different frequency regions in the stimulus will change with amplification. In the present study the hearing-aid gain characteristics typically changed the stimulus envelope at the onset, introducing short-term amplitude contrasts between /mi/ and /ti/, for example (Figure 5). The contrast /mi/ vs. /ti/ produced a 23-ms greater N1 latency difference in the aided condition compared to the unaided condition. N1 latencies for /mi/ were further increased in the aided condition, which may be due to the effect of the hearing aid altering stimulus-specific rise times or the effect of frequency-specific gain characteristics. The latter seems more likely as the gain varied between the two speech stimuli (Figure 5) and the hearing aid caused a greater change for /mi/ than /ti/. It is difficult to separate these two factors, however.
Previous research has demonstrated that interpretation of cortical neural processing of amplified speech signals can be problematic because of known confounding factors such as SNR, hearing-aid processing delay, and onset modifications induced by the type of hearing aid (Billings et al., 2012; Marynewich et al., 2012). In the current study, aiding was associated with significantly later (and a trend for larger) P2 responses. Ear canal acoustic measurements showed that aided stimulus SNR was higher than unaided for most of the stimuli (Figure 4). Larger P2 amplitudes could be linked to the higher SNR and effects of higher output levels in the aided condition. Previous studies have reported an increase in P2 and N1-P2 amplitude with 20-dB of gain compared to unaided (Billings et al., 2007; Marynewich et al., 2012). N1 amplitude has been shown to asymptote at intensities ~70 dB, which may account for the hearing aid not affecting N1 amplitude in the present study (Adler & Adler, 1989).

A few speech contrasts produced significant latency and amplitude differences for the aided, but not the unaided condition (Table 3), which may reflect the hearing aid altering envelope rise time due to the effects of nonlinear gain and higher noise floor. Because of complex effects of amplification on speech stimuli a better understanding of the effects of hearing-aid processing on CAEPs may be gained using hearing aid transduced signals controlling for factors such as signal processing delay, SNR, stimulus levels, and aided spectrum (Billings et al., 2012; Easwar et al., 2012b), in individuals with and without hearing loss.

Effects of stimulus parameters on N1 versus P2

Although N1 and P2 components co-vary they can be differentiated using experimental conditions such as those used in the present study. Stimulus conditions (noise and aiding) affected N1 only when the stimulus envelope was altered. For example, in the aided condition N1 was affected only for the speech stimuli that had changed envelope shape as a result of amplification e.g. /mi/ vs. /ti/. This is consistent with the view that N1 is a transient response evoked as a response to envelope change (Onishi et al., 1968). A relatively high SNR (+10 dB) was used in the current study for the noise condition, which could explain why N1 amplitude did not reduce consistently in noise. Latencies increased only when the noise interacted with the stimulus envelope for stimuli with spectral profiles similar to the background noise (Martin et al., 1999).

P2 is sensitive to attention and stimulus parameters such as intensity and pitch (Crowley et al., 2004). A recent study associated P2 with processing of altered phrase boundaries in music, especially in the context of harmonics (Istók et al., 2013), suggesting important effects of pitch and harmonics on P2. Amplification from a hearing aid introduces modulations or additional harmonics across the entire bandwidth of the signal, which changes the timbre of the aided output signal (Chasin & Russo, 2004). A listener with normal hearing can distinguish these differences in timbre, which may account for morphological changes in the P2 component in the aided compared to the unaided condition.

Although N1 and P2 originate from different sources and are functionally different responses (Ross & Tremblay, 2009), the distances between their sources are small and it is likely that P2 sources in planum temporale overlap with N1 sources with a centre of activity near Heschl’s gyrus (Crowley et al., 2004). Planum temporale has been described as the ‘computational engine’ for the segregation and spectro-temporal matching of complex sounds (Griffiths & Warren 2002, p. 348), responsible for pitch processing and melodic perception (Zatorre et al., 1998; Keenan et al., 2001). Activity in Heschl’s gyrus and planum temporale is enhanced by stimulus pitch differences (Schadwinkel & Gutschalk, 2010). Differential effects of noise and aiding on N1 and P2 suggest differential effects of processing of stimulus pitch, timbre, and envelope on these components.

Summary and Conclusions

Stimulus characteristics and hearing-aid model and settings are important factors to consider when investigating aided CAEP responses. Overall the current study showed that CAEPs evoked using natural stimuli are sensitive to adverse effects of background noise, onset characteristics of the stimuli, and spectro-temporal differences between speech stimuli. N1 and P2 components varied differently across noise and aiding conditions. Future studies differentiating N1 and P2 components in relation to pitch, timbre, and envelope cues will be useful. Nonlinear hearing aids with adaptive features may process speech differently in babble than they do in stationary noise and hence it would be useful to determine hearing-aid effects on speech-CAEPs for different types (and levels) of background noise. Establishing that CAEPs can reflect differences in speech acoustic features is a first step, further studies linking this to hearing-aid signal processing strategies and central auditory processing of speech stimuli are needed to further explore the potential use of CAEPs for evaluating hearing aids.

Acknowledgements

This research was supported by the Oticon Foundation, Denmark and New Zealand.

Declaration of interests: No conflict of interest and none declared.

References


Cortical encoding of speech acoustics

863


