Auditory acclimatization and hearing aids: Late auditory evoked potentials and speech recognition following unilateral and bilateral amplification

Piers Dawes* and Kevin J. Munro
School of Psychological Sciences, University of Manchester, Manchester M13 9PL, United Kingdom

Sridhar Kalluri and Brent Edwards
Starkey Hearing Research Center, Berkeley, California 94704-1362

(Received 21 August 2013; revised 12 April 2014; accepted 18 April 2014)

The aim of this study was to investigate changes in central auditory processing following unilateral and bilateral hearing aid fitting using a combination of physiological and behavioral measures: late auditory event-related potentials (ERPs) and speech recognition in noise, respectively. The hypothesis was that for fitted ears, the ERP amplitude would increase over time following hearing aid fitting in parallel with improvement in aided speech recognition. The N1 and P2 ERPS were recorded to 500 and 3000 Hz tones presented at 65, 75, and 85 dB sound pressure level to either the left or right ear. New unilateral and new bilateral hearing aid users were tested at the time of first fitting and after 12 weeks hearing aid use. A control group of long-term hearing aid users was tested over the same time frame. No significant changes in the ERP were observed for any group. There was a statistically significant 2% improvement in aided speech recognition over time for all groups, although this was consistent with a general test–retest effect. This study does not support the existence of an acclimatization effect observable in late ERPs following 12 weeks’ hearing aid use.

*Author to whom correspondence should be addressed. Electronic mail: piers.dawes@manchester.ac.uk

PACS number(s): 43.66.Ts, 43.64.Qh, 43.66.Sr, 43.71.Ky [FJG]
Pages: 3560–3569

I. INTRODUCTION

New hearing aid users experience an immediate benefit in speech recognition as audibility of previously inaudible speech cues is restored. Performance may continue to improve over time as the user learns to make optimal use of the auditory input provided by the hearing aid. This further improvement over time—a form of perceptual learning—is referred to as “auditory acclimatization” (Arlinger et al., 1996), or “acclimatization” from here on. Acclimatization effects are of interest because they may need to be taken into account when measuring benefit in research and clinical settings and could suggest avenues for optimizing benefit via improved acclimatization.

Evidence for acclimatization to hearing aids is mixed. Several studies reported evidence for improvements in aided speech perception (Cox and Alexander, 1992; Gatehouse, 1992, 1993; Cox et al., 1996; Horwitz and Turner, 1997; Kuk et al., 2003; Munro and Lutman, 2003; Yund et al., 2006), while others have not (Bentler et al., 1993a,b; Taylor, 1993; Saunders and Cienkowski, 1997; Humes et al., 2002; Humes and Wilson, 2003). These and other studies have been the subject of several reviews (Turner et al., 1996; Palmer et al., 1998; Turner and Bentler, 1998; Munro, 2008). In terms of perceptual evidence for acclimatization, Gatehouse (1989) reported that experienced unilateral hearing aid users showed better speech recognition in the fitted ear for high presentation levels, while in the non-fitted ear recognition was better at low presentation levels. Gatehouse suggested that this may be consistent with selective adjustment to a dynamic range consistent with the range of sound levels experienced at each ear. Gatehouse (1992) then reported improvements in speech recognition in the fitted ear over the first few weeks of hearing aid use in four new unilateral hearing aid users for listening conditions that mimicked the pattern of amplification provided by the hearing aid, but not with an unfamiliar pattern of amplification or in unamplified listening conditions. Munro and Lutman (2003) also found improvements in aided speech recognition over the first three months hearing aid use, particularly for higher intensity speech stimuli over lower intensity stimuli. In one of the largest acclimatization studies to date, Humes et al. (2002) followed 134 hearing aid users tested at one month, six months, and one yr post fitting, with a subgroup of 49 hearing aid users tested at two yr. A further subgroup (n = 9) was tested at three yr (Humes and Wilson, 2003). A variety of speech recognition and questionnaire-based measures of benefit were used. There was no evidence for any systematic improvement in benefit at any time point. According to questionnaire measures, participants reported less hearing aid benefit at six months and one yr compared to the baseline test at one month after fitting. This was not consistent with the hypotheses of acclimatization resulting in increased benefit over time. One limitation in this study was that the initial measures occurred after one month of hearing aid use, and so acclimatization effects may have been underestimated. Some research reviews (e.g., Palmer et al., 1998) concluded that design factors, such as baseline test sessions long after the initial hearing aid fitting, could account for the negative findings of some studies. Other problematic design issues in previous research include the use of participants
with a heterogeneous mix of signal processing or fitting schemes (for example, Saunders and Cienkowski, 1997), lack of a control group (for example, Kuk et al., 2003; Reber and Kompis, 2005), inclusion of participants with previous experience with hearing aids (for example, Bentler et al., 1993a) or participants with relatively mild levels of hearing loss (Palmer et al., 1998). In order to detect acclimatization, it may also be necessary to use assessments that tap frequencies and intensities that hearing aids make newly available (Saunders and Cienkowski, 1997; Dawes et al., 2014). One may be more likely to detect acclimatization effects if participants are tested with stimuli processed by the hearing aid, or with stimuli that are acoustically comparable to the output provided by hearing aids (cf. Gatehouse, 1992).

Besides attention to design details such as the ones mentioned in the previous paragraph, one of the key recommendations of the mid-1990s Eriksholm workshop on acclimatization (Arlinger et al., 1996) was that electrophysiological measures should be applied to elucidate possible physiological changes as well as the anatomical site of any changes that may accompany perceptual alterations with acclimatization. To date, the few studies that have addressed this recommendation have focused primarily on examining auditory brain stem response (ABR) and cortical auditory event-related potentials (ERPs).

In the case of the ABR, evidence for acclimatization effects is inconsistent. Munro et al. (2007) recorded ABRs in eight experienced unilateral hearing aid users and nine non-hearing aid users using controls. Wave V of the ABR was significantly larger in the fitted ear of the hearing aid users than in the unfitted ears. However, a limitation of this study was the cross-sectional design: the asymmetry may have been due to an idiosyncrasy of the hearing aid user group unrelated to hearing aid use. Philibert et al. (2005) recorded ABRs in five new bilateral hearing aid users at one, three, and six months after commencing hearing aid use. After six months, a significant reduction in the latency of wave V was observed for the right ear only. This was suggested to have been due to an inherent functional asymmetry in the auditory pathway favoring the right ear which had an impact on acclimatization. Participants also perceived tone stimuli as being less loud and showed better intensity discrimination over time, and this seemed consistent with acclimatization to the frequency-dependent gain provided by hearing aids. This study included a small number of participants and no control group. Differences over time may have been due to peculiarities of the small group of participants or retest effects or both. Finally, Dawes et al. (2013) recorded ABRs in eight new unilateral and 10 new bilateral hearing aid users at the time of first hearing aid fitting and again after three months hearing aid use. A control group of long-term hearing aid users was tested over the same timeframe as the new users. No significant change in latency or amplitude of wave V was observed in any group. This did not support plastic changes in the ABR associated with hearing aid use after three months use, although it remains possible that changes may be observable after a longer period of hearing aid use.

In relation to ERPs, two past studies identified asymmetrical ERPs in experienced unilateral hearing aid users, although methodological weaknesses make it difficult to conclude that these asymmetries represent an acclimatization effect. Gatehouse and Robinson (1996) presented a case study of a 69-yr-old long-term unilateral hearing aid user. Difference limens for intensity were obtained, and ERPs recorded for 500 and 2000 Hz tones at 65, 80, and 95 dB sound pressure level (SPL) for both ears. Analysis focused on the amplitude of a composite of the N1 and P2 ERP components. For 500 Hz tones, there was no difference between the ears in the difference limen for intensity or N1-P2 amplitude. For 2000 Hz stimuli however, difference limens were smaller and N1-P2 amplitude higher for the 95 dB SPL stimulus in the fitted ear. Experimental effects appeared specific to the fitted ear and to higher intensity and higher frequency stimuli. This was interpreted as being consistent with an acclimatization effect associated with hearing aid use. It is possible that the ear asymmetry observed may have been an oddity of the single participant in this study rather than being due to an acclimatization effect. However, this case study provided the first suggestion that acclimatization effects related to hearing aid use could be measured with ERPs.

Bertoli et al. (2011) compared ERPs in ten experienced unilateral hearing aid users and ten experienced bilateral users. Ten normally hearing participants were also tested. Following Gatehouse and Robinson’s (1996) findings, higher amplitudes and shorter latency ERPs were hypothesized in fitted ears versus non-fitted ears. ERPs were compared between groups for 500, 1000, and 2000 Hz tones at 55, 70, and 85 dB SPL presented monaurally to the two ears. There were no significant interactions between ear, frequency, and level observed for either latency or amplitude, nor were there any differences between unilateral and bilateral users on metrics of ERP ear asymmetries, as might be expected if there was an acclimatization effect. Nevertheless, effects suggestive of acclimatization were found in a post hoc analysis for the unilateral group only; there was a significant interaction between ear, frequency and level. The finding of increased amplitude P2 responses for the 2000 Hz stimulus in the fitted ear of unilateral users is compatible with Gatehouse and Robinson’s (1996) finding. The tentative explanation for the larger P2 amplitude in the fitted ear of the unilateral group was that this was a reflection of more effortful auditory processing, perhaps as a result of asymmetric amplification. Speech recognition performance did not differ between unilateral and bilateral groups, although the recognition threshold was significantly better in the fitted ear compared to the non-fitted ear of unilateral users. This finding provided some perceptual evidence of an acclimatization effect.

A conclusion of an acclimatization effect must be tentative because the cross-sectional design does not permit ruling out that the observed difference between ears in the unilateral group was spurious and unrelated to hearing aid use. For example, stronger ERPs might be found in the fitted ear if there was a tendency to fit the participant’s “better” ear, which may have marginally better sensitivity and therefore produce stronger ERPs. Additionally, one cannot tell if the apparent differences in P2 amplitude and speech recognition in the unilateral group were due to alterations in the fitted ear due to an acclimatization effect, or due to changes in
the non-fitted ear due to a deprivation effect (Silman et al., 1984; Gelfand et al., 1987). One previous study examined speech recognition and ERPs tested longitudinally (McCullagh, 2009). In this unpublished thesis, 10 new hearing aid users and a control group of 10 non-hearing aid users were tested at baseline and after six to eight weeks. There were no significant changes in speech recognition, N1 amplitude, P2 amplitude, or P2 latency. A statistically significantly different change in latency of N1 was reported in the new hearing aid user group compared to the control group. The author concluded that this was evidence of auditory plasticity following amplification. McCullagh’s (2009) finding of a change in N1 latency does not corroborate with Gatehouse and Robinson’s (1996) report of higher amplitude N1-P2 and Bertoli et al. (2011) report of higher amplitude P2 in aided ears. However, McCullagh’s work suggests that alterations in ERPs—but not speech recognition—may be observable after as little as six weeks hearing aid use.

To address the lack of conclusive evidence of an acclimatization effect on electrophysiological measures, in the present study we measured ERPs and speech recognition in new unilateral and bilateral hearing aid users at the time of first fitting and repeated after three months hearing aid use. A control group of experienced users was tested over the same time frame. To date, this is the only controlled longitudinal study of auditory acclimatization to hearing aids that has included both perceptual and electrophysiological measures. The hypotheses were that amplitudes of N1 and P2 ERPs would increase over time for fitted ears in parallel with improvement in aided speech recognition.

II. METHODS

A. Participants

Power calculations for ERP outcome measures suggested that a minimum of six participants would be required to detect a main effect of the size reported by Bertoli et al. (2011) with >80% statistical power using a repeated measures analysis of variance (ANOVA) with an alpha level of 5% (two-tailed) (Faul et al., 2009). Additional participants were recruited and tested to allow for attrition and/or a smaller effect size. Ethics approval was obtained from the appropriate bodies and informed written consent obtained from every participant. Power calculations for the speech recognition measures were based on the magnitude and variance in changes in aided speech recognition over time as reported by Munro and Lutman (2003), where the four alternative auditory feature (FAAF) test was used as in the present study. A sample size of n = 33 across groups would be required to provide 80% power to detect a mean change of 4% (SD 5%) using repeated measures ANOVA with an alpha level of 5% (two-tailed) (Faul et al., 2009). This minimum sample size was met.

Inclusion criteria were symmetrical, mild-to-moderate, sloping high frequency sensorineural hearing loss of at least 40 dB hearing level (HL) at 2–6 kHz, the profile of a typical hearing aid user. Exclusion criteria were (i) fluctuating or recent changes in hearing, (ii) asymmetry in air conduction thresholds greater than 15 dB at two or more frequencies, (iii) abnormal middle ear function assessed using oto-admittance audiometry, and (iv) an air-bone gap greater than 15 dB at any test frequency. For new users, additional inclusion criteria were (i) a history of hearing loss of at least 1 year’s duration, (ii) no prior hearing aid use, and (iii) average daily hearing aid use of at least 6 h per day during the study period (based on data logging information provided by the hearing aid). For the control group of experienced hearing aid users, the additional inclusion criteria were (i) at least 1 yr hearing aid use and (ii) self-reported daily hearing aid use of at least 6 h per day. Finally, participants were excluded if N1 and P2 could not be reliably identified at any point. The reasons for unreliable ERPs were related to poor quality recordings, for example due to movement-related artifact.

A total of 42 new and 17 experienced hearing aid users were recruited from local audiology clinics. At the commencement of the study, new hearing aid users were allocated to bilateral or unilateral fitting, alternating between the two as users were recruited to the study. Additional unilateral users were recruited towards the end of data collection to compensate for the many who did not comply with the requirement to use hearing aid(s) for the minimum 6 h per day. By the completion of the study, 25 new users were fit unilaterally and 17 bilaterally. No participant expressed a strong preference for unilateral or bilateral aiding, and all participants were able to switch to unilateral or bilateral aiding at the completion of the study. 12 new unilateral users and three new bilateral users were excluded from the study due to not wearing their hearing aid(s) the required 6 h per day. An additional two new unilateral users and one new bilateral user were excluded due to unreliable ERPs. Thus 11 new unilateral (six male) and 13 bilateral (eight male) users remained. A single unilateral user did not complete speech in noise testing due to illness, and therefore was excluded from this analysis only. Of the experienced users, four were excluded from the ERP analysis due to unreliable ERPs. Of the 13 remaining experienced users (seven male), six were unilateral and seven were bilateral users.

The mean age of each group was 70 (SD = 12), 69 (SD = 8), and 72 (SD = 6) years for the new unilateral, new bilateral and experienced user groups, respectively [F(2,34) = 0.5, p = 0.58]. Mean hearing loss across 2 to 6 kHz for each group was 50 dB HL (SD = 7), 50 dB HL (SD = 8), and 55 dB HL (SD = 14) for the new unilateral, new bilateral, and experienced user groups, respectively. There was no statistically significant difference in hearing level between groups [F(2) = 0.8, p = 0.45]. Hearing thresholds were retested at the completion of the study. Mean change in threshold across groups was 0.7 dB (SD = 3.1). Overall, 93% of hearing thresholds changed by <5 dB, and 99% changed by <10 dB. This is in line with reported test–retest differences for audiometry (Robinson, 1991).

B. Amplification, fitting, and electroacoustic measures

Starkey radius behind-the-ear (BTE; n = 4) or destiny completely-in-the-canal (CIC; n = 20) hearing aids were fitted to new users. Both CIC and BTE aids had similar levels of technology; an eight-channel compressor with gain adjustable in 12 bands, with active noise management. Compression
operated with an attack time of 20 ms and release time of 2000 ms (per the ANSI S3.22 standard). Adaptive listening programs were switched off for the duration of the study. BTE users had standard shell earmolds with a 0.8 mm vent. Hearing aids were fit according to National Acoustic Laboratories Non-Linear 1 prescription targets (NAL-NL1; Dillon et al., 1998). The match to target was verified with real ear insertion gain measurement, carried out using a Siemens Unity system according to recommended procedures (British Society of Audiology and the British Academy of Audiology, 2008). All new hearing aid users were given at least one day’s trial use of the hearing aid(s), and gain was adjusted if the participant felt that he or she could not tolerate the prescribed level of gain for the duration of the study. No further adjustments in gain were made during the study period, and participants were not able to alter hearing aid gain themselves. Real ear insertion gain was measured following any adjustment based on participant preference. Mean prescription targets and user gain for a 65 dB input signal level are shown for new and experienced users (Fig. 1). Mean target gain between 1 and 4 kHz (where hearing impairment is greatest) ranged between 10 and 18 dB. Mean aided Speech Intelligibility Index3 (SII; ANSI, 1997) across all new users for sentences in quiet at 65 dB SPL was 0.78 versus 0.50 unaided; hearing aids significantly improved the audibility of conversational-level speech. The maximum output of the hearing aids was determined by reference to the loudness discomfort level obtained during the initial fit and adjusted if the participant reported any undue discomfort. New users were encouraged to allow time to adjust to the hearing aid(s), and to use their hearing aid at least 6 h per day. Experienced users continued to use their hearing aids as normal, with no alteration to hearing aid gain made over the duration of the study. Experienced users’ hearing aids were Danalogic 6070 BTEs (n = 7), Oticon Spirit3 BTEs (n = 2), Oticon Synchro CIC (n = 2), Oticon Spirit 2 BTE (n = 1), and Siemens Reflex 1012 (n = 1), all non-linear hearing aids. Real ear insertion gain was re-measured at the end of the study in order to check the stability of amplification over time. The mean difference in REIG was −0.17 (SD = 1.7) across groups, with a range of −4 to +4 dB across frequencies, and no change in gain was statistically significant based on repeated measures ANOVA. The variation in gain is consistent with short-term test–retest variability of real ear insertion gain measures (Hawkins et al., 1991). As a check on the consistency of hearing aid use, the mean number of hours per day of hearing aid use was recorded via the hearing aid data logging facility, which was calculated and reported via the manufacturer’s hearing aid fitting program. After excluding those users who failed to wear their hearing aids for the required 6 h per day, average daily hearing aid use in the new user groups was 10 h (SD = 2.3).

C. Event-related potential measurement

The paradigm is similar to that of Gatehouse and Robinson (1996) and Bertoli et al. (2011), informed by Thornton’s (2007) recommendations for ERP design. Stimuli were 500 and 3000 Hz pure tones presented at three

---

![Figure 1](image-url)  
**FIG. 1.** Mean real ear user gain and prescribed gain targets for 65 dB input signal level. Error bars show standard deviation.
levels of intensity: 65, 75, and 85 dB SPL. Tones were 300 ms in duration including 20 ms rise/fall times. Stimulus generation and presentation was performed with a Neuroscan STIM system. Stimuli were presented monaurally to both ears via insert earphones (Etymotic EAR 3 A) with an inter-stimulus interval that varied randomly between 1100 and 1400 ms. In order to avoid mismatch effects, the order of stimuli was pseudo-randomly determined, with the constraint that identical stimuli could not occur sequentially. There were a total of 12 conditions (two frequencies by three intensity levels by two ears). Stimuli were presented in three blocks of 504 stimuli each. Each stimulus was played 42 times in each block, with a different random order of stimuli for each block. Thus each stimulus was played 126 times across the three blocks. The duration of each block was approximately 11 min. Participants were instructed to ignore the test stimuli.

Participants were seated in a comfortable chair in a sound treated booth and watched a subtitled DVD of their choice with the sound off. Silver/silver chloride recording electrodes were applied to the scalp at the vertex (Cz) according to the International 10/20 system with a ground electrode on the high forehead and a nose-tip reference electrode. Eye movements were monitored with electrodes above and below the right eye. Impedances were maintained below 5 kΩ. EEG data acquisition was performed with a Neuroscan SCAN evoked potential system. EEG data were recorded continuously with a recording bandwidth of 0.1–100 Hz with a 250 Hz sampling rate for off-line averaging and analysis.

Off-line analysis consisted of +/−50 μV artifact rejection, baseline correction, and digital filtering from 1 to 30 Hz with a slope of 24 dB/octave and re-windowed from −200 to 600 ms relative to stimulus onset. Three runs were combined to provide a maximum of 126 sweeps for each condition. In addition to a grand mean waveform for each condition, mean waveforms were generated for each run separately.

A grand mean waveform was calculated based on ERPs from all participants across conditions. Mean amplitude (referred to as “amplitude” from here on) was taken as an area-under-the-curve measure for the 40 ms range centered on grand mean peak latencies for N1 and P2, in accordance with Luck (2005). Peak latency was automatically identified in each ERP as the latency of the lowest and highest amplitude within the ranges 70–150 ms and 180–260 ms for N1 and P2, respectively. Wave forms were visually inspected for each ERP sub-average by the first author (PD), and the automatically detected peak latency accepted only if peaks could be reliably identified in each sub-average. Amplitudes and latencies for P1 are not reported, as P1 could not be reliably identified across conditions for all participants.

D. Speech in noise recognition

The FAAF test (Foster and Haggard, 1979, 1987) is a closed-set word identification task which has been used in previous acclimatization studies (Gatehouse, 1992, 1993; Munro and Lutman, 2003). The FAAF test contains 20 sets of four binary and minimally paired words based on auditory/phonetic distinctions, giving an 80-item test. Target words differ either in the initial or final consonant within a set. Identification of the target largely depends on high-frequency auditory information, ensuring that the FAAF is sensitive to frequencies newly available by amplification. The participant hears the phrase “Can you hear the word X clearly?” spoken by an English-accented male speaker. The participant must then select the target word “X” from among the four rhyming words within the set (for example, mail, bail, daile or nail) shown on the computer monitor. Responses are scored as correct or incorrect, giving an overall percentage correct.

Targets were presented at two levels; 65 and 75 dB SPL, which correspond roughly to “conversational” and “raised” speech levels for male speakers in realistic listening environments (Pearsons et al., 1977) and to stimulus levels used in previous acclimatization studies (Munro and Lutman, 2003). Targets were presented against a background of steady noise with the same long-term spectrum as the target and from the same loudspeaker at zero degrees azimuth. The overall SPL of the speech and noise were calibrated at the reference point, the approximate center of the participant’s head with the participant removed. Speech recognition testing was carried out in a sound-treated audiometric booth. Participants were seated in the center of the booth with the loudspeaker positioned directly in front of the participant at ear level at a distance of 1.5 m from the reference point. A computer monitor was positioned below the level of the loudspeaker in front of the participant to display response options. Computerized stimulus presentation was via a 16-bit sound card at a sample rate of 20 kHz with a Tannoy 607 loudspeaker and Denon amplifier (PMA-250III).

At each test session, participants first completed a familiarization run of 40 words with easy levels of target and noise (65 dB and 55 dB SPL, respectively). At the first test session (baseline), the signal to noise ratio (SNR) for 50% correct for each participant and each listening condition was determined using the adaptive version of the FAAF so that (i) testing would occur at an equally challenging level for each participant and (ii) ceiling and floor effects would be avoided. Participants were then tested at this SNR with the 80-item version of the FAAF at baseline and week 12. In the present study, aided recognition for the fitted ear of unilateral users and the left ear of bilateral users is reported. The non-test ear was plugged and muffed, and a different random order of testing was employed for each test session.

E. Procedure

Otoscopy, pure tone audiometry, oto-admittance measures, real ear insertion gain measures, and the speech in noise task were completed at baseline. ERPs for new users were recorded either before or within seven days of hearing aid fitting; the range was 22 days before fit to seven days after fitting (M = 2 days before fitting, SD = 6). All procedures were repeated after 12 weeks. A 12 week retest interval was selected following previous studies that demonstrated acclimatization effects within this timescale (for example, Munro and Lutman, 2003). The average time between baseline and week 12 was 95 days (SD = 16, range 71 to 110 days). The
interval between test sessions varied due to the availability of participants for testing.

**F. Statistical analyses**

For new unilateral participants, the ear of testing was classified according to fitted/non-fitted ear. For the remaining groups, ear was classified as left/right. Changes in average N1 and P2 latency and amplitude between baseline and week 12 were tested with repeated measures ANOVA, with time (baseline/week 12), level (65/75/85 dB SPL), frequency (500/3000 Hz), and ear (left or fitted/right or non-fitted) as within-participants factors for each group. For the new unilateral group, the effects of time and interaction of time and ear (fitted/non-fitted) are of interest. For the bilateral group, the effect of time is of primary interest. Bonferroni’s *post hoc* measures were used to explore any significant main effects or interactions. Changes in speech recognition accuracy were also tested with repeated measures ANOVA, with time (baseline/week 12), level (65/75 dB SPL) as within-participants factors and participant group (new unilateral/new bilateral/experienced user) as a between-participants factor.

**III. RESULTS**

ERP wave forms averaged across intensities are shown for each group in Figs. 2–4. Average ERP wave forms for each stimulus intensity and frequency are shown in Fig. 5.

**A. Evoked response potentials**

For the new unilateral group (Fig. 2) there was a significant effect of level on N1 latency [F(2,9) = 14.9, *p* < 0.01], but no significant main effect or time or time–ear interaction. For P2 latency, there were no significant main effects or interactions. There were significant effects of level on N1 and P2 amplitude [F(2,9) = 10.2, *p* < 0.01; F(2,9) = 6.2, *p* = 0.02, respectively]. There was no main effect of time or time–ear interaction for N1 or P2 amplitude.

For the new bilateral group (Fig. 3), there was a significant effect of level on N1 and P2 latency [F(2,11) = 16.0, *p* < 0.01; F(2,11) = 6.5, *p* < 0.01], but no significant effect for time or time–ear interaction for N1 or P2 latency. There were significant effects of ear on N1 and P2 amplitude [F(2,12) = 37.9, *p* < 0.001; F(2,12) = 24.7, *p* < 0.001,
respectively], and also a significant effect of level \( F(2,11) = 17.9, p < 0.001; F(2,11) = 11.9, p < 0.001 \) on N1 and P2 amplitude. There was no significant effect of time or interactions with time for N1 or P2 amplitude.

For the experienced user group (Fig. 4), there was a significant effect of level on N1 latency \( F(2,11) = 15.1, p < 0.01 \) but not P2 latency. There was no significant effect of time or interactions with time on N1 or P2 latency. There were significant effects of ear \( F(1,12) = 8.7, p = 0.01; F(1,12) = 6.4, p = 0.03 \) and level \( F(2,11) = 17.2, p < 0.001; F(2,11) = 4.5, p = 0.04 \) for N1 and P2 amplitude, respectively. There was no significant effect of time or interactions with time for N1 or P2 amplitude.

For new users, there was no significant correlation between change in N1/P2 amplitude or latency and age, average level of hearing loss or amount of hearing aid use (Pearson’s \( r \) between 0.01 and 0.35; all \( p > 0.05 \)). Across groups, correlations between baseline and week 12 amplitudes ranged between 0.50 and 0.63 (Pearson’s \( r \); all \( p < 0.05 \)). Correlations between baseline and week 12 latencies ranged between 0.41 and 0.75 (Pearson’s \( r \); all \( p < 0.05 \)) across conditions.

**B. Speech in noise recognition**

Aided speech recognition improved over time; there was a significant main effect of time \( F(1,33) = 6.77, p = 0.01 \), but no significant interactions between time, level, or group (Fig. 6). Correlations between baseline and week 12 recognition performance ranged between 0.5 and 0.8 (Pearson’s \( r \); all \( p < 0.001 \)) across conditions. There were no significant correlations between change in speech recognition and change in N1/P2 latency or amplitude, age, average

**FIG. 4.** Average ERP wave form for the left and right ear of experienced hearing aid users at baseline and following 12 weeks hearing aid use.

**FIG. 5.** Average ERP wave form for the left ear of participants at baseline for each stimulus intensity level and frequency.

**FIG. 6.** Speech recognition accuracy. Error bars show standard deviation.
level of hearing loss, or amount of hearing aid use (Pearson’s $r$ between 0.03 and 0.44; all $p > 0.05$). Figure 7 shows example correlations between change in ERP amplitude and speech recognition.

IV. DISCUSSION

We did not observe any significant changes in N1 or P2 over time consistent with a hearing-aid related acclimatization effect. The finding of no change in cortical ERPs is consistent with the finding of no change in the ABR, which was observed in a separate study that included some of the participants in the present study (Dawes et al., 2013). Likewise, although there was a significant improvement in aided speech recognition over time in the new user groups, a similar improvement was evident in the control group of experienced hearing aid users. This generalized improvement is more consistent with a practice effect than with an acclimatization effect specific to newly fitted ears.

In the present study, hearing aid gain and use were monitored, and a control group was included. The study was statistically well-powered and included an informative prospective longitudinal design. Design factors may therefore account for the lack of acclimatization effects on electrophysiological measures observed in the present study compared to previous ones, which relied on cross-sectional testing of experienced hearing aid users, lacked a control group or both (Philibert et al., 2005; Munro et al., 2007; Bertoli et al., 2011).

Alternatively, other potentially significant differences between the present study and previous ones are (i) the duration of hearing aid use and (ii) the use of non-linear hearing aids. Previous studies that showed an asymmetry of ERP (Gatehouse and Robinson, 1996; Bertoli et al., 2011) utilized hearing aid users with several years of experience with hearing aids. Perhaps the effect of hearing aid use on ERPs only becomes apparent after long-term use. A definitive answer to this question would require longitudinal studies with a retest interval longer than 12 weeks, possibly up to several years post-fitting. The majority of previous studies that reported acclimatization effects based on perceptual measures utilized linear hearing aids [for example, Munro and Lutman (2003)]. Acclimatization effects may be more apparent with linear amplification (Dawes et al., 2014), or with signal processing such as frequency lowering that causes dramatic change to the acoustic environment. This explanation does not account for the discrepancy between previous ERP studies and the present one, however, as the participants in Bertoli and colleagues’ (2011) study also used non-linear hearing aids.

A final possibility is that small acclimatization effects may occur, although these may have been obscured by between-participant variability in outcome measures. Variability of outcome is a noted feature of previous acclimatization research (Turner et al., 1996). Individual differences, such as the level of hearing loss and the amount of hearing aid use are suggested as accounting for some of this variability (Turner et al., 1996; Tyler and Summerfield, 1996; Palmer et al., 1998). In the present study, there were no significant correlations between change in ERP indices and change in aided speech recognition, and no correlation between ERP indices and potential determinants of acclimatization including age, level of hearing loss, or the amount of hearing aid use. Therefore, the variability in outcome measures may have been due to either (i) unknown determinants of acclimatization that were not controlled in this study or (ii) inherent variability of ERP indices and speech recognition scores. Correlations between baseline and week 12 ERP indices and speech recognition scores ranged between 0.4 and 0.8 (i.e., medium- to large-sized correlations) across

![N1 Diagram](image)

$r^2 = -0.04$, ns

![P2 Diagram](image)

$r^2 = 0.03$, ns

Notes: ERPs are for the aided ear for new unilateral users or the left ear for new bilateral users for the 3 kHz stimulus at 85 dB SPL condition. The $y$ axis is a measure of area amplitude. Units of measurement are $\mu V \cdot ms$ (microvolts x milliseconds). Positive values indicate an increase in area amplitude. Change in speech recognition is for the 75 dB SPL condition. Positive values indicate an improvement in aided speech recognition. The solid line shows the regression line of best fit.

FIG. 7. Correlations between change in N1 and P2 amplitude and change in speech recognition.
conditions. These statistics suggest that over the relatively long 12 week retest interval in the present study, ERP and speech recognition indices were reasonably reliable. The present study was adequately powered to detect effects of the size reported by previous studies, and which would be large enough to be of clinical significance. Small acclimatization effects may have been undetectable, however. For future acclimatization research, it may be necessary to establish the reliability of measures for older, hearing impaired adults over the relatively long time periods needed for acclimatization studies.

There was a significant effect of higher intensity stimuli resulting in shorter latency, higher amplitude N1 and P2 ERPs. This has been well-documented previously (Naätänen and Picton, 1987; Crowley and Colrain, 2004; Luck, 2005; Hyde, 2009). Such effects were not a focus of interest in the present study, and shall not be discussed further. One unexpected finding was of an ear asymmetry, with shorter latency and higher amplitude responses in the left ear over the right ear. Comparison of left and right ear ERPs for individual participants confirmed that this was a consistent pattern. Note that in the analysis reported above, the new unilateral group was classified according to ear of fitting rather than left or right, and this classification is responsible for the lack of a significant effect for ear for this group. Analyzing according to left/right ear, there was a significant effect of ear for the new unilateral group, similar to the new bilateral and experienced user groups. Hearing loss was symmetrical on average, with any extreme asymmetries of hearing loss excluded. The ear asymmetry in ERPs was apparent at baseline in the new user groups, and so was unrelated to hearing aid use. Headphones were calibrated, subject to listening checks during data collection and re-checking after data collection; they provided equal levels of stimulation. Asymmetries of cortical auditory ERPs have previously been reported. First, Hine and Debner (2007) showed greater hemispheric asymmetry in activity when tones or noise were presented to the left compared to the right ear. Ear differences have also been reported by Valtonen and colleagues (2003). Possible reasons for this include preferential processing of monaurally presented tones and noise by the right hemisphere (Hine and Debener, 2007), or different anatomy of the supratemporal planes in each hemisphere, which may be modified with aging so that the orientation of the source becomes more favorable in the right hemisphere when activation is recorded at Cz (Jyrki Mäkelä, personal communication, 12 July 2012; Andrew R. Dykstra, personal communication, 13 July 2012), or due to age-related demyelination and loss of information transfer across corpus callosum (Jerger et al., 1995) (although this last observation was in relation to the P3, not N1/P2). It should also be noted that a nose-tip reference electrode was used in the present study. Use of different reference electrode locations (such as mastoid or earlobe) may result in a different pattern of findings with respect to the cortical asymmetries observed here. The reason for the ear asymmetry remains speculative, and additional data are required to elucidate this question. However, the primary measure of interest in the present study was hearing aid use-related change over time based on repeated measures within-participant. The ear difference would not be expected to affect these repeated measures.

V. CONCLUSIONS

There were no acclimatization effects in terms of alterations in ERPs or aided speech recognition in noise in first-time hearing aid users after the first 12 weeks hearing aid use. Test–retest variability may have obscured small acclimatization effects, although if such effects do exist they are not likely to be large enough to be clinically significant. More robust acclimatization effects following hearing aid amplification may be observable with longer duration of hearing aid use or with signal processing strategies that provide a more dramatic alteration of auditory input.

ACKNOWLEDGMENTS

The authors thank Aneea Greval and Andrea Curran for hearing aid fitting, and Keith Wilbraham for technical support. This research was funded by the Starkey Hearing Research Center, and part-funded and facilitated by the NIHR Biomedical Research Center.

1Bertoli et al. (2011) reported a significant difference in P2 peak amplitude between the aided and unaided ear of experienced unilateral hearing aid users. The F statistic was 6.1, and the sample size was 10, giving an effect size of d = 1.5 (Cohen’s d). Power calculations were based on the hypothesis of a similar size effect for N1 and P2 amplitude in the present study as observed for P2 amplitude in Bertoli et al. (2011) study.

2A mixture of BTE and CIC devices was used because CIC devices became available to participants shortly after the commencement of the study. It was found that the option of CIC hearing aids increased participation rates, and this is the reason for the greater number of CIC over BTE users.

3SII is calculated as the weighted sum of band audibility according to the relative importance of each frequency band to the understanding of speech and ranges between 0 and 1. SII calculation assumes that the speech area covers a dynamic range of 30 dB. The SII was determined based on overall levels of 65 SPL for sentences in quiet. Aided values were obtained by adding the mean user insertion gain to speech levels.

4The booth had a reverberation time across 0.125–8 kHz of 0.1 s, which is slightly more favorable than a typical domestic living room (around 0.3 s; Burgess and Utley, 1984).


Dawes et al.: Auditory acclimatization and hearing aids

