Measuring access to high-modulation-rate envelope speech cues in clinically fitted auditory prostheses

DOI:
10.1121/10.0000673

Document Version
Accepted author manuscript

Citation for published version (APA):
Stone, M., Prendergast, G., & Canavan, S. (2020). Measuring access to high-modulation-rate envelope speech cues in clinically fitted auditory prostheses. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/10.0000673

Published in:
Journal of the Acoustical Society of America

The Journal of the Acoustical Society of America

Measuring access to high-modulation-rate envelope speech cues in clinically fitted auditory prostheses

--Manuscript Draft--

Manuscript Number:	JASA-04656R2
Full Title:	Measuring access to high-modulation-rate envelope speech cues in clinically fitted auditory prostheses
Short Title:	High-modulation-rate envelope cues in auditory prostheses
Article Type:	Regular Article
Corresponding Author:	Michael A. Stone Manchester Centre for Audiology and Deafness Manchester, UNITED KINGDOM
First Author:	Michael A. Stone
Order of Authors:	Michael A. Stone
	Garreth Prendergast
	Shanelle Canavan
Section/Category:	Psychological and Physiological Acoustics
Keywords:	dip-listening; glimpsing; modulations; dynamic range compression
Abstract:	The signal processing used to increase intelligibility within the hearing-impaired listener introduces distortions in the modulation patterns of a signal. Trade-offs have to be made between improved audibility and the loss of fidelity. Acoustic hearing impairment can cause reduced access to temporal fine structure (TFS), while cochlear implant processing, used to treat profound hearing impairment, has reduced ability to convey TFS, hence forcing greater reliance on modulation cues. Target speech mixed with a competing talker was split into 12-22 frequency channels. From each channel, separate low-rate (EmodL, < 16 Hz), and high-rate (EmodH, < 300 Hz) versions of the envelope modulation were extracted, which resulted in low or high intelligibility, respectively. The EModL modulations were preserved in channel valleys, and cross-faded to EModH in channel peaks. The cross-faded signal modulated a tone carrier in each channel. The modulated carriers were summed across channels and presented to hearing-aid and cochlear-implant users. Their ability to access high-rate modulation cues, and the dynamic range of this access, was assessed. Clinically fitted hearing aids resulted in 10% lower intelligibility than simulated high-quality aids. Encouragingly, cochlear implantees were able to extract high-rate information over a dynamic range similar to that for the hearing-aid users.

Rebuttal Letter / Helpful/Supporting Material for Reviewer

EditorAndReviewersResponseJan2020.pdf

Measuring access to high-modulation-rate envelope speech cues in clinically fitted auditory prostheses

Michael A. Stone[1], Garreth Prendergast, Shanelle Canavan

Manchester Centre for Audiology and Deafness, School of Health Sciences, University of Manchester, M13 9PL, UK

Corresponding author:

Michael Stone, Manchester Centre for Audiology and Deafness, School of Health Sciences, University of Manchester, M13 9PL, UK

Email : michael.stone@manchester.ac.uk

---

[1] Also at : Manchester Academic Health Science Centre, Manchester University Hospital NHS Foundation Trust, Manchester, M13 9WL, UK

Abstract

The signal processing used to increase intelligibility within the hearing-impaired listener introduces distortions in the modulation patterns of a signal. Trade-offs have to be made between improved audibility and the loss of fidelity. Acoustic hearing impairment can cause reduced access to temporal fine structure (TFS), while cochlear implant processing, used to treat profound hearing impairment, has reduced ability to convey TFS, hence forcing greater reliance on modulation cues.

Target speech mixed with a competing talker was split into 12-22 frequency channels. From each channel, separate low-rate (EmodL, < 16 Hz), and high-rate (EmodH, < 300 Hz) versions of the envelope modulation were extracted, which resulted in low or high intelligibility, respectively. The EModL modulations were preserved in channel valleys, and cross-faded to EModH in channel peaks. The cross-faded signal modulated a tone carrier in each channel. The modulated carriers were summed across channels and presented to hearing-aid and cochlear-implant users. Their ability to access high-rate modulation cues, and the dynamic range of this access, was assessed. Clinically fitted hearing aids resulted in 10% lower intelligibility than simulated high-quality aids. Encouragingly, cochlear implantees were able to extract high-rate information over a dynamic range similar to that for the hearing-aid users.

Keywords: dip-listening, glimpsing, dynamic range compression, modulations, hearing aids, cochlear implants, auditory perception, speech intelligibility

I. INTRODUCTION

The dynamic range of acoustic signals processed by the healthy human hearing system, spanning the range between audibility and discomfort, is around 100 dB, but less below 200 Hz and above 10 kHz. With sensorineural hearing impairment, discomfort thresholds exhibit a wide scatter for the same degree of impairment, but show little increase with increasing impairment until hearing threshold exceeds 60 dB HL (Storey & Dillon, 1998). This reduced dynamic range between audibility and discomfort causes recruitment, (Fowler, 1936; Steinberg & Gardner, 1937), remediated in hearing aids (HA) by the use of multi-channel dynamic range compression (DRC) (Villchur, 1973; White, 1986). For profound hearing losses, direct electrical stimulation of the cochlea can be used to replace acoustic stimulation by using a cochlear implant (CI). Since the dynamic range between threshold of hearing and threshold of discomfort for the electrical signals presented is typically between 5 and 30 dB (Fu & Shannon, 1998; Loizou et al., 2000;), DRC is also essential in CIs (Dillier et al., 1980; Wilson et al., 1988;)

The effectiveness of DRC in the remediation of hearing impairment has resulted in debate (Plomp, 1988; Villchur, 1988) and much experimentation (see reviews in Souza, 2002; Moore, 2008). Although the concept of DRC is old, the flexibility of digital signal processing has seen multiple fields of applicability and configurations proposed for the design of DRC circuits, initially in broadcast audio (McNally, 1984; Stikvoort, 1986; Giannoulis *et al*., 2012), but extending to hearing aids (reviews in Kates, 2005; Dillon, 2012a), and cochlear implants (Stöbich *et al*., 1999; McDermott *et al*., 2002; Boyle *et al*., 2009; Khing & Swanson, 2013). The debate and experimentation centres around the number of frequency bands in which DRC should be performed, the speed with which the DRC should operate, and the balance between the restoration of audibility and the permissible amount of distortion of spectro-temporal modulations, which are important contributors to speech intelligibility and quality (Xu *et al*. 2005; Xu & Zheng 2007; Whitmal, 2007; Souza & Rosen, 2009; Kates, 2010, 2011).

Generally, the restoration-of-audibility-promotes-intelligibility argument proposes the use of fast time constants in multiple channels of DRC (*e.g.* Villchur, 1988) in order to promote audibility of low-level portions of the signal. The low-distortion-promotes-intelligibility argument favours slower time constants, preserving the fidelity of envelope modulations (Plomp, 1988) and in the process, sacrificing short-term audibility. However, this sacrifice appears beneficial in noisy listening situations and where the richness of acoustic cues is impoverished, such as in noise-carrier vocoding (Stone & Moore, 2007). It should also be noted that, apart from DRC, other non-linear signal processing that is used in hearing prostheses, such as adaptive noise reduction and adaptive directional microphones, distorts envelope modulation. Also, like DRC, these are implemented on a multi-channel basis and with varying time constants (Dillon, 2012b). Depending on choice of time constants, these additional signal processing strategies may also be expected to contribute to modulation distortion on perceptually-relevant timescales. Additionally, the sophistication of digital signal processing permits algorithmically-defined modifications of interactions between the component processing blocks of a hearing aid, so that characterising the performance of a system by standard parameters, such as attack and release times, and compression ratios (ANSI, 2003), cannot fully characterise the expected performance in either the short (sub 100-ms) or long-term (several seconds).

Distortion in general, and especially distortion of the signal envelope by signal processing, such as by DRC, is of great interest in the remediation of hearing impairment. Since both hearing loss and age are associated with decreased sensitivity to the temporal fine structure (TFS) of an audio signal (Buss *et al*., 2004; Hopkins & Moore, 2007, 2011; Hopkins *et al*., 2008; Grose & Mamo, 2010), the hypothesis arises that the hearing-impaired (HI) listener is more reliant than a young normal-hearing (NH) listener on the envelope modulations of a signal. CI listeners are almost entirely dependent on the envelope modulations since CI processing either largely, or totally, discards the TFS (Zeng, 2004).

Some recent models for predicting speech intelligibility place emphasis on the important contribution of modulation, specifically considering the signal-to-background ratio (SBR) in the modulation domain (Dubbelboer & Houtgast 2008; Jørgensen *et al*., 2013). There is a subtle difference between the two models: Dubbelboer and Houtgast explicitly model the effect of distortions produced by non-linear signal processing as contributing to the “noise”, so their model requires a reference signal of the clean (unprocessed) target, which is not always practical. The alternative approach, used in the

Jørgensen and Dau family of models (which also require a reference signal, but of the noise alone), only generates an estimate of the audible speech modulations, thereby ignoring intermodulations between target and background produced by non-linear processing. However, the estimated preserved speech modulations will incorporate effects of any non-linear processing, such as the reduced dynamics, and hence altered (long-term) SBR (Naylor & Johannesson, 2009). The successful prediction of results from these models indicate that reduction of the signal envelope power, as well as the addition of distortions and noise, is a major contributor to speech intelligibility, even without consideration of the supporting role of TFS (reviewed in Moore, 2014).

For a wearable hearing prosthesis, there is a need to assess the perceptual consequences, if any, of the device, and assigning the cause of the cost. In doing so, it is possible to identify areas where the technical design of the prosthesis, rather than perceptual limitations of the participant, affect performance. For example, in a speech-intelligibility task, normal-hearing listeners were assessed either wearing hearing aids or unaided (Cubick & Dau, 2016; Cubick et al. 2018). The wearing of high-fidelity hearing aids appeared to show little or no disadvantage in a co-located masker condition, but produced worse performance in a separated masker condition (Cubick et al., 2018), when compared to unaided listening. This pattern of results was attributed to distortion of spatial cues due to the non-ideal location of the microphones (behind-the-ear location and omni-directional pattern). These differences were smaller than those measured when a lower-bandwidth, lower-fidelity hearing aid was used (Cubick & Dau, 2016). Besides the non-linearities often produced by analogue acoustic transducers and their amplifiers, the distortions introduced by signal processing can also be expected to produce modulation- and inter-modulation-distortion components. While these are physical components, their sources of origin, and the relationship between them can also cause perceptual confusion. Stone et al. (2009) reported that the action of fast-acting DRC on a two-talker mixture required greater effort on the part of young, NH, listeners to separate the keywords from the mixture. This they attributed to a loss of independence in the separate modulations patterns of the component talkers: previously independent sound sources had acquired a common component of modulation due to the fast-acting DRC, perceptually making them appear to be less separate.

One way to assess the degree of the perceptual consequences possibly produced by signal processing on modulations is by manipulating the ability of the listener to access them. High-rate envelope modulations (greater than about 15 – 30 Hz) appear to be an important contributor to speech intelligibility, at least in vocoder processing (Dudley, 1936; Whitmal, 2007; Souza & Rosen, 2009). Stone *et al*. (2010) reported measures of manipulating this perceptual access. They band-pass filtered a speech-in-competing-talker signal into either 8 or 15 contiguous channels. Within each channel, they low-pass filtered the full-modulation-bandwidth envelope to produce a restricted-modulation-bandwidth version of the same envelope. The resulting envelope was used to modulate a tone carrier at the centre of the respective channel, before recombining the individual channels. When the full bandwidth envelope signal was used, the resulting intelligibility was high, but it fell markedly when the restricted-bandwidth version was used. Additionally, Stone *et al*. selectively switched in the restricted-bandwidth version as a function of short-term signal level within a channel. In one of their configurations, as the channel signal valleys were progressively filled with restricted-bandwidth information, intelligibility progressively decreased, mapping the ability of the listener to access information in the signal dips. This mapping, relating intelligibility to the relative level of the switch from restricted to full-bandwidth, can be used to define an intensity-importance function (IIF, Boothroyd, 1990), a description of the relative importance of speech information as a function of level in a signal channel. Boothroyd’s IIF was relevant to unmodified speech, i.e. containing both envelope and TFS cues, whereas that for Stone *et al.* was relevant to envelope modulations only. The shape of an IIF for unmodified speech, is described in ANSI (1997) as being rectangular with level, spanning +/-15 dB about the channel mean. Studebaker & Sherbecoe (2002), using band-limited speech reported more rounded functions, tailing away asymetrically to either side from a peak centred near the channel mean. Stone *et al.* (2010) reported that their measured *envelope* IIFs for NH listeners had a rounded shape, similar to those of Studebaker & Sherbecoe (2002). The envelope IIFs measured for HI listeners in Stone *et al.* (2012a) were similar in shape and dynamic range to that they had previously reported for NH listeners (Stone *et al*. 2010), but displaced to a slightly higher level, relative to the channel root-mean-square (rms).

To date, the reported IIFs have been based on group data, assuming a degree of homogeneity in performance across participants. Besides the heterogeneity in audiograms observed between hearing-impaired listeners, supra-threshold performance in psychoacoustic tasks also demonstrates a heterogeneous nature, even for the same audiogram. To account for this heterogeneity, fitting of hearing prostheses usually involves a stage of fine tuning based either on subjective preferences or on objective measures. The experiment to be reported therefore wished to investigate the possible heterogeneity in IIFs in hearing-impaired listeners. Measuring individual access to high-rate modulations may reveal across-participant differences in the ability to benefit from high-rate envelope modulations, providing insight into real-world difficulties, as well as revealing possible differences in the effectiveness of the signal processing used in different hearing prostheses.

The first part of the experiment reported here used the envelope processing technique of Stone *et al*. (2012a) to compare access to high-rate envelope modulations after processing by two types of hearing aids. Stone *et al*. (2012a) only used a linear, simulated high-fidelity hearing aid with their participants, and reported results on a group-average, rather than individual, basis. The experiment reported here extended the previous work on HAs by comparing the performance of the simulated high-fidelity HA to that of a clinically-fitted HA, and additionally measured and reported IIFs on a per-participant, rather than group basis. Additionally, the data allowed the quantification of the loss in performance in access to modulation cues due to the compromises involved in the design of a wearable device (distortions due to non-linear signal processing, as well as possible distortions from analogue stages in the wearable device).

The second part of the experiment used the same processing technique with a group of CI users. Compared to HA processing, the envelope distortion produced by CI processing is more complicated. In a CI, at least two stages of DRC are employed, the first acting on the short-term signal in a small number of frequency bands, and the second applying instantaneous DRC to the extracted channel envelope applied at each electrode (Wilson et al., 1988; Fu & Shannon, 1998). Instantaneous DRC can be expected to generate a whole series of distortion components in the modulation frequency domain. However, the ability to present the channel signal directly to frequency-specific regions of the cochlea, bypassing the channel mixing that has to occur prior to acoustic presentation, means that the distortion components may be at least partly cancelled out by the instantaneous non-linearity at the electrode-neural interface, if the correct mapping function is chosen. Early work with CIs showed that, although this mapping function does affect intelligibility, an exact match of the function mapping envelope amplitude to electrode current was not important to produce the highest intelligibility (Fu & Shannon, 1998). Additionally, temporal modulation transfer functions (TMTFs) measured for CI users are similar in shape and absolute sensitivity to those obtained by NH listeners and possibly even better (Shannon, 1992), with some showing a tendency to a relatively improved characteristic in the modulation range of 50-100 Hz. The processing technique of Stone *et al*. (2012a) should therefore be applicable to CI users, albeit with some modifications to accommodate differences in device configurations. Since signal delivery for the CI participants had to be via each users own device, it was not possible to make comparison with a ‘reference’ CI. This second part of the experiment, apart from measuring individual IIFs, permitted the comparison of the performance effected by CI processing against that achieved by another hearing-impaired group, HA users, such as the degree of benefit available from high-rate modulation cues and the dynamic range of individual IIFs in each group.

II. METHODS

A. Participants

Participants were recruited via local UK National Health Service (NHS) audiology clinics (HAs), a volunteer panel (CIs and HAs) and the Richard Ramsden centre for auditory implants at the NHS-run Manchester Royal Infirmary (CIs). The study and access to participants was approved by the NRES Committee North West - Greater Manchester South, approval number 14/NW/1365.

All participants had to be aged over 18 years, fluent speakers of English from birth, and in generally good health for travel to the test site. All participants received a “Participant Information Sheet” at least 24 hours ahead of their first appointment.

1. Selection criteria unique to HA participants

Additional prerequisites for eligibility for this group were having:

1) A sensorineural hearing loss, at least moderate-to-severe in the better (aided) ear, averaged over 1, 2 and 4 kHz, but not exceeding 70 dB HL at any frequency below 8 kHz.
2) A negligible conductive component to the loss (≤10 dB at any frequency, 0.25 to 4 kHz),
3) The absence of an extensive “dead region” in the cochlea of the test ear as assessed using the Threshold-Equalizing Noise test (TEN(HL), Moore *et al*., 2004), but allowing for a ‘fail’ at any single frequency between 0.5 and 4 kHz.
4) No history of middle ear dysfunction and an intact tympanic membrane.
5) Been a daily user of a hearing aid in the test ear for at least 6 months and for at least 5 hours per day.
6) Been fitted with a current generation (primarily NHS) hearing aid (typically < 4 yrs) and were happy with the fit.

HA participants were tested using a single ear. 19 (8 female) HA participants completed the testing. Their details are given in Table I. Figure 1 shows the mean and standard deviation (SD) of the audiograms of the tested ears.

Figure 1 about here

----

2. Selection criteria unique to CI participants

Additional prerequisites for eligibility for this group were:1) that they should be successful users of their device, capable of achieving moderate to high sentence intelligibility (> 70%) at SBRs of +10 dB or less (speech-spectrum weighted noise).

2) that they had been using the device regularly for at least 1 year (> 8 hours/day, 5 days/week) and were happy with the fit.

The requirement for CI participants to be successful users of their device was necessary due to the need to test at SBRs where:

(a) within each channel, the (fluctuating) background would overlap with the dynamic range of the speech-plus-background signal that had previously been shown to be relevant for HI listeners (Stone *et al*., 2012a), typically between about -8 and +8 dB relative to the channel rms, and

(b) their word-intelligibility scores were sufficiently high (> 50%) that the participant was not demotivated by their apparent lack of ability in some (necessarily) low-scoring conditions.

CI participants with unilateral implants were tested through their implant alone. 18 CI participants (8 female) completed the testing. Their details are given in Table II.

B. Stimuli

The target speech consisted of lists of sentences from the IEEE corpus (IEEE, 1969). Either the first 48 lists for the HA participants (24 lists per aid), or the first 24 lists for the CI participants were selected for presentation. They were scored by counting keywords reported correctly. The sentences were spoken by a male speaker of southern British English. Prose passages of 60-sec duration, as well as material from the Adaptive Sentence Lists (ASL, MacLeod & Summerfield, 1990) and IEEE lists 61-72 were used as training material.

The background noise was a single male talker reading a continuous prose passage where pauses for breath had been edited out. Only natural-sounding gaps remained between sentences. All speech material was available at a sampling rate of 22.05 kHz. Although the ASL material used for training only had an 8-kHz bandwidth, all other material had a bandwidth of up to 11 kHz.

C. Processing

The processing was very similar to that used by Stone *et al*. (2012a), the core of which was a 15-channel tone-carrier vocoder which preserved the signal envelope in each channel, while discarding the original TFS within the channel. Here the number of channels, *N*, was constant for group HA (*N*=16), but varied (8 – 22) with participant in group CI, according to their individual device. An *N*-channel filterbank implemented by finite impulse response (FIR) filters was used to separate the signal into channels. In each channel, two versions of the envelope were generated by half-wave rectification and low-pass filtering with a forward and backward pass of a 2-pole Infinite-Impulse-Response filter (Chebyshev Type II design, with stopband ripple set to –36 dB per pass) giving a minimum out-of-band attenuation of -72 dB. This ensured linear phase and good stop-band rejection, but allowed a small amount of "overshoot" in the time-domain response. The different versions were achieved by the use of different corner frequencies. One filter, the low-rate or 'L' filter, had a –3 dB point of 16 Hz. The 10%-90% rise time of its step response was 22 ms. The second filter, the high-rate or 'H' filter, had a –3 dB point which was the lower of either 300 Hz or 0.866 times the channel bandwidth. This latter restriction ensured that a good representation of the original channel envelope was extracted but possibly at the cost of preserving small amounts of the channel carrier (TFS) in low-frequency channels. The channel bandwidth in all processing conditions always exceeded 16 Hz, so that there was no need to modify the corner frequency of any of the L filters.

In each channel a logical 'switching signal' (binary-valued, 0 or 1), was created by comparing the instantaneous value of the L-filtered envelope with an adjustable 'switching threshold'. The switching signal was defined as being 1 if the L-filtered envelope was above the switching threshold and 0 otherwise. The switching signal was then filtered with a 2-pole, minimum overshoot Bessel-derived low-pass filter, whose corner frequency was twice that of the L filter, to give a 10-90% rise time of 11.5 ms, except for low-bandwidth channels, where the corner frequency was scaled so that the rise time was three times the reciprocal of the channel centre frequency, so as to reduce the potential for production of high-level in-channel modulation products.

Stone et al. (2012a) used the switching signal in two ways to generate separate test conditions, which they labelled L/H or H/L. The results from their L/H condition appeared less stable (poorer fits of a model to the data) than those from their H/L condition, so we focus just on this second condition.

Within each channel, a composite envelope signal, *Comp*, was generated by the weighted sum of the H and L signals:

$$Comp = Swf.\mathrm{H} + (1 - Swf).\mathrm{L} \qquad (1)$$

where *Swf* is the filtered switching signal, and H and L are the H-filtered and L-filtered channel envelope signals, respectively. As the switching threshold increased, more of an envelope valley was filled with the L-filtered signal. The filtering of the switching signal acted as a cross-fade between the two logical values so as to remove audible artefacts at the switching points. The composite envelope was used to modulate a sinusoidal carrier at the linear centre frequency of the channel. The FIR filter used to extract the channel signal was then re-applied to ensure no out-of-channel leakage of (modulated) signal energy. The channel signals were then combined for presentation to the participant.

Switching thresholds, expressed as the level relative to the long-term (i.e. sentence-list duration) rms of the channel envelope, were –∞, –13, –7, –2, +2, +7, +13, and +∞ dB. The range, and spacing, were chosen on the basis of the previously measured IIFs for hearing-aided participants (Stone et al. 2012a). These values were assessed in 8 separate conditions, presented in a counterbalanced order using a Latin square. Separate counterbalancing was performed for each participant group. The values of –∞ and +∞ dB meant that the composite envelope comprised either the H-filtered or the L-filtered envelope, respectively. With all-H-filtered envelope, intelligibility should be at its maximum, and with all-L-filtered envelope, it should be at its minimum.

The mean percentage of time that a channel signal was above the switching threshold was 100, 45, 29, 17, 8, 3, 0.5 and 0 %, for thresholds of –∞, –13, –7, –2, +2, +7, +13, and +∞ dB respectively. These percentages are averages across channels 4 to 12 of the 16-channel HA processing, using an SBR of +8 dB, the average across the HA participants when tested using their own aids. These channels span the frequency range 400 to 4000 Hz, (see Table III), where the band importance function of the Speech Intelligibility Index (SII) is maximum (ANSI, 1997).

1. Processing specific to HA participants

The real-ear insertion gain (REIG) of each participant's hearing aid was assessed using the 60-sec duration International Speech Test Signal (ISTS, Holube *et al*., 2010) presented over a loudspeaker. The acoustic signal was recorded in the meatus of the participant by means of an ER-7c (Etymotic Research Inc., Elk Grove, Ill) probe microphone whose tip was placed within 4-6 mm of the tympanic membrane. The output of the microphone pre-amplifier was recorded on a H2 hand-held digital recorder (Zoom Corporation, Tokyo, Japan) for off-line analysis. Two measures were performed, once with no HA present and the meatus open, with a replay level of at least 60 dB(A), and once with the meatus closed by the eartip of the active hearing aid in place, at a replay level of 60 dB(A). The eartips were typically soft dome fittings, providing very little attenuation of the acoustic path directly from the loudspeaker to the meatus. The equal-or-higher replay level in the open-meatus condition was done to ensure that the recorded power spectrum was always above the noise floor of the probe microphone (and recording) system. In practice this lead to an upper frequency limit of 6 kHz. REIGs did not show gain exceeding 0dB for frequencies above 6 kHz in any aid. The higher replay level for the open-meatus condition was compensated for in calculation of the REIG. The REIG was taken as the smoothed difference of the power-spectral densities of the two recordings.

Two versions of the processed training and test speech signals were generated, one with, and one without, the REIG applied to the source speech+interfering speaker signal before the vocoder processing. The two versions were required so that the HA participant could be tested either with their own hearing aid (OWN) or with a simulated, linear hearing aid (SIM) so that they could also be tested open meatus (*i.e.* without their clinical aid), but with the same REIG as for their clinical aid.

The 16-channel vocoder that was used had each channel being 2-ERB$_N$ wide, where ERB$_N$ indicates the width of the normal auditory filter (Glasberg & Moore, 1990). Although Stone et al.

(2012a) used a 15-channel vocoder, due to the higher sampling rate of the speech material used here, the extra channel occupied a frequency region not present in the earlier work. Hence the lowest 15 channels were common in extent between the two experiments. Due to the degree of hearing impairment of the participants, their auditory filter widths were expected to be broader than normal by a factor of around 2 (Moore, 2007) so the modulation sidebands around each tone carrier were unlikely to be resolvable. The edge frequencies and carrier frequencies for the 16 channels are given in Table III.

2. Processing specific to CI participants

The number of channels and channel edges of the vocoder were selected for each participant to match the channel mappings in their clinical fitting. This ensured that the envelope information of each vocoder channel would be presented in the middle of the corresponding analysis channel in the implant. The processing in the devices manufactured by Cochlear only extracted the channel envelope, while some of the devices from MedEl employed processing that attempted to provide information about the channel carrier as well as the envelope. This strategy (FS™) could be applied in up to 4 of the low-frequency channels. For participants with MedEl devices, the corresponding channels were processed by preserving the original channel signal for the peaks and replacing the valleys with a tone carrier modulated by the L-filtered envelope. Although fundamental frequency (f_0) carrier information was still available in the low frequency channels, so permitting the FS processing to operate, its quality was degraded by the progressive removal of this information in the valleys. All other aspects of the experimental processing were the same as for the other CI devices.

D. Equipment and presentation

All signals were processed on a PC and replayed under control of MATLAB™ (Mathworks, Natick, MA) via external sound cards, either an Edirol UA-25 (Roland Corporation, Hamamatsu, Japan) or a Scarlett 2i2 (Focusrite, High Wycombe, UK). The soundcards were connected via a 20-dB attenuator to a Tannoy VXP6 loudspeaker (the version as manufactured by Tannoy Ltd, Coatbridge, Scotland, UK). The loudspeaker was sited at a distance of 1 metre from the participant, who was seated near the middle of a double-walled, sound-treated room with dimensions 4.9 x 3.5 x 2.5 metres (length x width x height). Sound levels were measured at the position of the participant's head, but without the participant present. Due to the use of concentric dual-cone drivers in the VXP6, the loudspeaker behaved like an acoustic point source, and was therefore more robust to errors of relative placement of the participant's head between measures than from two-way loudspeakers with physically separated drivers.

E. Procedure

HA participants attended for three, and CI participants attended for two, test sessions, each of which was intended to last, at most, for 2 hours. In the first session participants had the experiment protocol explained to them before giving written consent to any experimental procedures. Since the level of cognitive ability can be a confounding factor in experiments involving DRC, (Lunner and Sundewall-Thoren, 2007), each participant performed Part A of the Trail Making Task (Reitan, 1958) and two Digit Span Tests (Forward and Backward, Wechsler, 1997). The Trail Making task required a participant to draw connecting lines between number points in ascending order on a sheet of paper while being timed to complete the task. The Digit Span tests measure the longest sequence of digits that can be held in working memory, requiring recall either in order ("Forward") or mentally manipulated into reverse order ("Backward"). One CI participant (C7) failed to complete the Trail Making Task, indicating a mental competence of "pathological" for this test, a diagnosis later confirmed as early vascular dementia. However, he was able to complete the digit span task as well as the intelligibility testing, so apart from analyses requiring a score from the Trail Making Task, his results were retained.

1. Procedure specific to HA participants

For the HA participants, tympanometry, audiometry and the TEN(HL) test (Moore *et al*. 2004) were performed to check for conformance with the selection criteria detailed above.

Any hearing prosthesis in the non-test ear was removed and the meatus plugged with a well-fitted expanding foam earplug. For one participant (H9), with a deeply inserted solid mould, and a mild-to-moderate low-frequency hearing loss, the earplug took the form of the non-test aid inserted but switched off.

At this point in the first session, the REIG of the test HA was measured, and analysed offline to produce an REIG for the processing required in the SIM condition. If the REIG appeared too low for the degree of loss, the participant was referred back to their dispensing clinic for re-fitting before any further testing was performed. If the re-fitting did not seem appropriate testing was stopped for that participant.

2. Procedure specific to CI participants

For the CI participants, the non-test ear was either (i) left unplugged if the unaided hearing thresholds exceeded 60 dB HL, (ii) de-activated in the case of bilateral implantation (C13) or (iii) the hearing aid function of electro-acoustic stimulation (EAS) was de-activated (C9).

The participant was encouraged to set the processor controls in anticipation, such that conversational speech from the experimenter was at a comfortable level, and to choose their regular clinical program for coping with speech in a moderately noisy environment. Once selected, these settings were checked for comfort during initial training, but not changed throughout the duration of the experiment itself. All signal input to the processor came via the device microphone(s). Table II additionally gives a list of processor features, such as noise reduction, microphone directionality and the ability to select from a range of programs. Only C1 did not have access to a noise-reduction program.

3. Procedure common to both HA and CI participants

For the remainder of the first session, the participants were trained to recognise the vocoded speech, by presentation of material with increasing levels of difficulty, with the giving of feedback to their oral responses. This training lasted about 1 hour. All participants used their own hearing devices during training. Although HA participants could generally recognise the difference between unprocessed and vocoded signals, the CI participants could not.

At the start of the second session, a further 15-20 minutes of training was performed. HA participants were presented with material processed for either the SIM or the OWN HA conditions, in a counterbalanced allocation across their sessions two and three. CI participants were presented with just the vocoded signal, as for their first session. For the final two lists of sentence training material, processed with the H-rate envelope filter, the SBR was adaptively varied between sentences, in steps of 3 dB, in a 1-up, 1-down paradigm. If the participant reported correctly more than three of the five keywords in a sentence, the SBR was decreased. If they reported fewer than three keywords correct, the SBR was increased. If three keywords were reported correctly the SBR was left unchanged. This method therefore tracked the SBR necessary to achieve approximately 60%-correct. A logistic curve was fitted to the data and used to predict the SBR at which the participant would achieve a score of about 70%. The equation of the fitted function was:

$$I = P_{ceil} - \frac{\left(P_{ceil} - P_{floor}\right)}{\left(1 + \exp\left(-\beta\left(SNR - t_{50}\right)\right)\right)} \qquad (2)$$

where I is the intelligibility in percent words correct, P_{ceil} is the asymptotic (high or ceiling) score in silence, P_{floor} is the asymptotic (low or floor) score for chance performance, β is a scaling factor that determines the slope of the transition region, and MP is the value of the switching threshold at the midpoint of the I curve between P_{ceil} and P_{floor}.

This predicted SBR was therefore expected to offer around 70% intelligibility for the all-H-filtered condition, and to decline with increasing removal of H-filtered cues. Results should therefore lie on the steepest portion of the individual's Performance-Intensity (P-I) function and (largely) avoiding floor or ceiling effects. During a short pause in the testing, test signals for the 8 conditions of the test proper were generated at the predicted SBR. The test proper occupied the remainder of the session. The third session (HA participants only) was identical in structure to the second session, except that the processing was adapted for the aid (OWN or SIM) on which the participant had not yet been tested. The results of the logistic regression to set the test SBR were saved for later re-analysis (Sect III.C.1).

The presentation level for HA participants was set to 60 dB(A), around that for speech of "normal" level (62.3 dB SPL, ANSI, 1997). This had several purposes:

(a) to reduce the audibility of the components of speech transmitted via open domes since the domes do not provide much attenuation of external sounds, especially at low frequencies where the attached hearing aids deliver little gain.

(b) to ensure greater reliance on the aided frequency range where the individual channel bandwidths permitted an H-filter bandwidth much greater than the L-filter bandwidth, and therefore greater potential for change in intelligibility.

(c) to ensure that any leakage of sound to the (plugged) non-test ear would be of low level, and make only a small, or nil, contribution to intelligibility.

(d) it was a level sufficiently high that the valleys of a fluctuating signal were either close to, or above, the compression threshold of the HA DRC, ensuring repeated activation of any non-linear signal processing.

Where available in the manufacturer's data sheets, details of the dynamic behaviour of the compressors in the HAs are given in Table I. It should be noted that some of these values are questionable. For instance an attack time of 1 ms appears unlikely for hearing aids employing overlap-add Fourier transform processing where the frame lengths are typically 5-10 ms and the overlap is typically 50% or less. Additionally, for high compression ratios, such fast attack times would produce audible distortion. As previously stated, with many of the signal processing functions interlinked in modern HAs, the actual processing speed is difficult to determine.

For the CI participants, their second session was structured identically to the second session of the HA participants, except that they were tested through their own device with an acoustic presentation level of 65 dB (A) at their unoccupied seat. This was closer to the levels typically used to set up the implant, and, since participants were either ear-plugged or profoundly deaf at low frequencies, was not likely to produce audibility via an acoustic path.

Previous work using HA participants (Stone *et al*., 2012a) showed an average difference in intelligibility of around 36% between maximum and minimum, provided that the SBR was chosen to avoid ceiling and floor effects. Stone *et al*. (2012a) used 2 lists per test condition and relied on inter-list variability being counterbalanced across many participants, a method which worked well with group results. Here we wished to measure the performance of each participant, so we increased the number of lists assessed to 3, as well as assessing 8 rather than 7 switching thresholds. The 8 conditions, with 3 lists per condition, combined with the training and any other administrative procedures, produced a session duration of less than two hours. During all training and testing, the participants were given opportunities for rest and refreshment, preferably during the interval at the end of testing each processing condition.

The processing conditions were presented to the participants within each session in a counterbalanced order by use of a Latin square. For the HA participants, as well as the randomisation of their test session order (SIM then OWN, or vice-versa), a different Latin square was used in each session. Since the presentation order of the lists, as well as the lists used, was fixed within each session, the counterbalancing ensured that, across a square, all groups of three lists were assessed in each of the processing conditions. A fully counterbalanced data set was achieved across either 16 HA participants (two sessions by eight processing orders) or 8 CI participants (eight processing conditions). For participants beyond these minimum numbers, a second Latin square was used, different from the first.

III. RESULTS

The general pattern of results measuring speech intelligibility as a function of switching threshold for each participant was expected to be similar to that for the H/L condition of Stone *et al.* (2012a), the group mean results of which are plotted as filled circles in Figure 2.

Figure 2 about here

These mean data, shown in Fig. 2, were fitted with a logistic function defined by Eqn. (2), but with P_{ceil} and P_{floor} replaced by P_H and P_L, the asymptotic (high) score for an all-H-filter envelope signal, P_L and the asymptotic (low) score for an all-L-filter envelope signal. *SBR* is replaced by *sw*, the switching threshold. β and *MP* are as described for Eqn. (2). For reasons of clarity, we will refer to this variant of Eqn. (2) as Eqn. (2, P_H, P_L, *sw*). For the data of Fig. 2, β and *MP* had values of 0.276 and +0.2, respectively. The total change modelled in this example was 59.3% – 22.7% = 36.6 %, well above the standard deviation likely from assessment of 150 keywords in the triad of lists used to assess each condition (4.1%).

In the results presented below, in order to check for possible bias in mean score from each triad of lists used to assess each condition, the individual data were transformed into rationalised arcsine units (RAU, Studebaker, 1985), and a grand mean was calculated for each triad, across the 16 listeners of each group who were tested in the fully populated Latin Square. The triad means were normalised so that the mean of the individual means was unity. The range of individual triad means was :

(1) for HA participants tested in their first session (IEEE lists 1-24), 0.94 to 1.13,

(2) for HA participants tested in their second session (IEEE lists 25-48), 0.90 to 1.09,

(3) for CI participants tested in their single session (IEEE lists 1-24), 0.89 to 1.12.

The individual scores for each triad were deliberately under-corrected by dividing by the mean for each triad raised to an exponent of 0.75. This method of under correction and the particular exponent value were selected because it minimised the total fitting error across all participants when modelling the P-I functions for each participant, according to Eqn. (2, P_H, P_L, *sw*). All scores were then converted back to percent correct, and all further analyses were performed on these scores.

For the experiments reported here, percent word intelligibility as a function of switching threshold is plotted in separate panels for each participant in Figures 3 (HA data) and 4 (CI data). The HA panels comprise two data sets, one for OWN, and one for SIM aid. The logistic-function fits produced by Eqn. (2, P_H, P_L, *sw*) are plotted as solid lines with different colours (online version, grayscale, print version). The parameters of the fitted functions are shown in Table IV(a) for the HA participants, and Table IV(b) for the CI participants. Table IV also shows the difference score, $(P_H - P_L)$, in percent, to indicate the scale of benefit from the high-rate modulations to both participant groups.

The general pattern of results for each participant in Figs. 3 and 4 follows that expected from Fig. 2 above, with some noteworthy exceptions. In what follows, data were tested for normality before statistical analyses were performed and any instances in which the assumption of normality was violated is explicitly stated.

A. HA intelligibility data

For some participants (H5, H6, H14, H16 and H18), performance with the two aids was remarkably similar, despite some differences in test SBRs. Most participants exhibited the expected 30-40% change in intelligibility between the all-H and all-L conditions $(P_H - P_L)$, for both OWN and SIM aids and the fitted functions fall close to the data. However, H2 (OWN), and H15 (OWN and SIM) showed very small changes (17.7, 13.9 and 21.1 %, respectively). Additionally, the low-P_H score for H15 (OWN) would have been located on a portion of the intelligibility-SBR function that typically has a shallower slope, hence making differences hard to detect. The group mean values of $(P_H - P_L)$ were 34.6 % and 36.4 %, for OWN and SIM aids respectively, which is similar to the previously reported 35.6 % of Stone *et al*. (2012a).

Figure 3 about here

B. CI intelligibilty data

In general the data for the CI participants were much noisier than for the HA participants, with some data difficult to interpret (C1 and C14). C9 and C13 showed an excellent ability to use the high-rate cues, with a near-50% change in intelligibility, in the same range as achieved by the best HA participants. It was more common for the CI participants to exhibit changes of 25 % or less (10/18 participants). The mean value of the benefit measure, $(P_H - P_L)$, was only 25.7 %.

Figure 4 about here

C. Comparisons within the HA data

1. The perceptual 'cost' of a clinically fitted versus a high-quality simulated hearing aid

Due to the bypassing of the miniature microphone and receiver in the SIM aid, as well as a lack of non-linear signal processing, it was expected that the SIM aid would perform at least as well as the OWN aid and usually better. The protocol therefore called for the use of either the same, or lower, SBR as used in the OWN condition. Consequently, 10 out of 19 participants were tested at a lower SBR in the SIM condition, but with only a 1 or 2 dB difference.

In order to compare the data from the two HA systems at an equal SBR, the scores in the SIM condition were adjusted by a per-participant correction factor. The parameters of the logistic regression to the data obtained from the measuring of the adaptive tracks using Eqn. (2), which had been used to set the SBR used for testing, were also used to calculate the mean value of β (β, mean = 0.4182, SD=0.259) in the all-H processing condition. Although β could be expected to vary as a function of SBR, the per-participant adaptive tracks were noisy, and were only available calculated over two sentence lists. Averaging of the individual β over the participants produced a more reliable estimate. Assuming that P_{ceil} can approach 100%, and P_{floor} to 0% (for an open set speech test), Eqn. (2) can be used to map a normalised P-I function using β. The SBR at which the reported value of P_H was achieved in the SIM condition with this normalised P-I function was noted. This SBR was then increased by the difference in SBR between the OWN and SIM condition, and the resulting prediction of intelligibility read off from the normalised P-I function at this new value of SBR. The full set of corrected values are shown in Table IV(a) as the final column, labelled P_{HCorr}. The adjusted values are shown in bold print.

A comparison of P_{HCorr} for the SIM condition against P_H for the OWN condition is shown in Figure 5.

Figure 5 about here

Of the 19 points of the data set, only one point lies below the diagonal line of performance equality. The mean score for the OWN aid was 61.5%, while the mean corrected score for the SIM aid was 71.1%. The mean difference in scores was 9.6%, with an SD of 6.7%. However, the difference data were not normally distributed, so a Wilcoxon signed-rank test was performed, which revealed a significant difference, $z(18) = 8$, $p < 0.0001$. Using the normalised P-I function with the slope defined by β, the mean difference is equivalent to an SBR benefit of 1.0 dB in the SIM compared to the OWN condition.

A comparison of the difference scores, $(P_H – P_L)$, between the SIM and the OWN conditions showed a mean difference of 1.8%, with an SD of 7.7%, which was not significant $t(18) = 1.00$. The 1.0 dB performance benefit of SIM over OWN therefore seems to be due to better access to all modulation rates rather than to just the high-rate cues.

2. The ability to access low-level, high-rate envelope cues compared between the OWN and SIM aids.

Despite the experiment being intended to measure individual performance, the individual data shown in Figs. 3 and 4 were 'noisy', making some results hard to interpret. Since P_H and P_L in Eqn. (2, P_H, P_L, sw) represent asymptotic values that were well sampled in the data, there was a risk that β and *MP* largely co-varied so as to minimise the fitting error. We therefore created a perceptually relevant measure of how far below channel RMS level the participants were able to extract high-rate envelope information, which, at the same time, linked β and *MP*. This 'Valley' measure, V_{10}, was defined as the value of switching threshold at which a 10% decrease in the relative change between P_H and P_L had been reached, *i.e.*, the introduction of low-rate cues in the valleys was just starting to reduce intelligibility. Across participants, the range of $(P_H – P_L)$ varied between 13.9 and 57.2 %. Defining V_{10} as a fixed percentage change in intelligibility below the individual's P_H would translate to operating at a different point on the transition region defined by Eqn. (2, P_H, P_L, sw) for each individual. Choice of a relative measure effectively compares performance at the same relative point on the individual's IIF, independent of its overall magnitude ($P_H - P_L$). Using Eqn. (2, P_H, P_L, sw), normalising by setting P_H and P_L to 100 and 0, respectively, we solve for V_{10}, the switching threshold when I = 90 by using the individual's values of β and MP. This gives:

$$V_{10} = MP - \log_e(9) / \beta \quad \text{(units of dB)} \tag{3}$$

Values for V_{10} are given in Table IV for all participants.

Data from three HA participants (H10, H12 and H18) were removed from this analysis. From Fig. 3 it is seen that these participants had at least one trace that did not reach an asymptotic value of P_H, even by the lowest switching threshold tested. In previous work (Fig. 2), we observed asymptotic performance at about – 15 dB relative to the channel RMS, and the associated V_{10} would be several dB higher. Consequently, in order to remove outliers in V_{10} produced by poor fits to the data, if one of the {SIM, OWN} data pair contained a value below –15 dB, that pair was rejected. The choice of the cut-off of –15 dB was also near where there was a gap in the distribution of the individual data (–21.9 and –18.2 for H12 and H18, respectively).

Figure 6 about here

Figure 6 shows the scatter plot between the measures of V_{10} for both the SIM and the OWN aids. The Pearson correlation coefficient for these data was $r = 0.601$, 14 *df*, ($p < 0.02$ two-tailed). This significant correlation implies that the participants were performing in a similar fashion between conditions. However, there was no significant difference between the V_{10} measures for the OWN and SIM conditions (mean = 1.59, SD = 4.64 dB, $t(15) = 1.38$, NS). We interpret this to mean that the similarity in performance was due to factors related to the participant and, disappointingly, not the change in processing between SIM and OWN conditions in permitting perceptual access to the valleys of the channel envelopes.

3. The possible influence of participant-related factors from the HA data set

Partial correlations were performed between Test SBR and a number of variables, while controlling for the score on the digit span test. The aim of this analysis was to establish the extent to which cognitive factors may mask some interesting relations. The different partial correlations assessed whether Test SBR was related to P_H, access to high-rate envelope cues ($P_H – P_L$), mean low-frequency audiogram (averaged over 250 and 500 Hz), mean high-frequency audiogram (averaged over 2, 3 and 4 kHz), and difference between the low-and high-frequency mean audiogram (a measure of audiogram slope). These revealed a correlation between Test SBR and ($P_H – P_L$) (OWN aid, r = -0.530, 16 *df*, p = 0.024; SIM aid, r = -0.601, 16 *df*, p = 0.008, uncorrected) indicating that participants tested at a high SBR received less benefit from the high-rate envelope cues. We will return to this in the Discussion.

In the same set of partial correlations, a correlation was observed between P_H, and ($P_H – P_L$) (OWN, r = 0.597, 16 *df*, p = 0.009*;* SIM, *n.s.*). This hints that achieving a high difference score was limited by the starting value of P_H, and that, with a low starting value, a possible floor effect was introduced in the all-L condition, despite the goal of adjusting the test SBR so that the all-H condition achieved a score fairly high on the psychometric function. In practice this goal was not always met ; *eg* H15 had a P_H of 28% in the OWN condition (as seen in Table IV(a)). Post-hoc analysis of her data showed erratic performance during the setting of the Test SBR, at least for the OWN condition, which may well have continued during testing as well.

For the SIM aid, a correlation was observed between P_H and mean high-frequency audiogram (r = -0.675, 16 *df*, p = 0.002), indicating a possible lack of audibility of the high-rate modulation components in the valleys of the signal due to the linear processing used. One might expect audibility also to have an effect on the maximum benefit obtained from high-rate modulations, but no correlation between ($P_H – P_L$) and mean high-frequency audiogram was observed.

D. Comparison of HA and CI data

1. The ability to access high-rate envelope cues compared between acoustic and electric hearing prostheses.

Figure 7 shows a histogram of ($P_H – P_L$) values for HA OWN and CI. The choice of bin spacing is arbitrary, but gives a smoothed representation of the data distributions. For both groups there was a wide scatter in the benefit achieved from the high-rate cues; the SDs of the HA and CI data sets were 10.4 and 10.9% respectively. Using a two-tailed, unpaired *t*-test, the mean difference in ($P_H – P_L$) between the HA (OWN) and CI groups was 9.0% (standard deviation, SD, 3.5%), giving $t(35) = 2.55$, $p = 0.015$. CI users do not appear to be as able as the HA users to make use of high-rate modulation information. We will qualify this interpretation later.

Figure 7 about here

----

2. The ability to access low-level high-rate envelope cues compared between HAs and CIs.

The same V_{10} measure was generated from the CI data set using Eqn. (3). Data from participants C2 and C11 were excluded because their V_{10} measure (-28.6 and -18.2 dB, respectively) was less than –15 dB, and likely erratic. Figure 8 shows the histograms of the V_{10} measures for the three hearing prostheses (OWN, SIM and CI). The pairs of mean, (SD) in dB for each device were (i) OWN : -5.9, (4.9) with 17 participants, SIM -6.6, (5.5) with 18 participants, and CI -3.6, (5.7) with 16 participants. Pooling the data for the OWN and SIM conditions due to the non-significant difference reported in III.C.2 above, the difference between the HA and CI conditions gave a value of $t(27) = -1.59$, $p = 0.10$, also non-significant.

Figure 8 about here

E. Relationships between two of the cognitive measures.

For both participant groups, there were strong correlations between scores for the Trail Making task and the Digit Span Backwards tasks, (HA data : $r = -0.686$, 17 *df*, $p < 0.001$; CI data: $r = -0.675$, 16 *df*, $p = 0.003$) but only weaker correlations between the other measures. Participant C7 ('pathological' Trail-making score) was excluded from these correlations. The factor 'Age' did not correlate with any of the cognitive measures in the HA group and was weakly correlated with Digit Span Forward in the CI group (r = -0.491, p = 0.045, 16 *df*).

F. Dependence of Test SBR on Age

In the CI group there was a moderately strong dependence of Test SBR on Age, Pearson r = 0.659, (16 df), p = 0.003. This was primarily due to four of the participants aged in excess of 65 years who were tested at SBRs of +15 dB or higher (C3, C7, C12 and C17). This dependence on Age was not apparent in either of the HA groups, Pearson $r \leq 0.412$, (17 df), $p \geq 0.08$, but a much narrower (and quantised) range of Test SBRs were used with these groups, compared to the CI group. This narrower range may mask a similar effect, but not observable without collecting much more data.

IV. DISCUSSION

The perceptual cost in accessing envelope modulations by using a clinically-fitted non-linear hearing aid compared to a high-quality simulated linear hearing aid was measured as being 1 dB in SBR. This is similar to the disadvantage found for discriminating co-located speech-in-speech masking when comparing binaural linear HAs against unaided listening (Cubick et al. 2018). Although these individual costs, 1dB, appear small, because they come from differing aspects of the acoustic scenario (non-linear and binaural respectively), they have the potential to add up to a more significant disadvantage, especially if there are similar small disadvantages associated with other changes in the acoustic scenario.

Encouragingly, the benefit to intelligibility from high-rate modulation cues to both HA and CI users was similar on some measures, such as the perceptually relevant dynamic range over which these cues were available, but differed on others, such as the gain in intelligibility possible from these cues. In HA users, this gain in intelligibility was very similar to that previously reported by Stone et al. (2012), around 36%, but much less in CI users, around 26%. As with many studies, these interpretation needs to be qualified.

A. Dependence of benefit from high-rate modulations on test SBR : a parallel with observations of masking release

The dependence of ($P_H – P_L$) benefit on Test SBR (section III.C.3 above) has a parallel in the work of Oxenham and Simonson (2009). They investigated the release from masking that occurred when extrinsic modulations were applied to a speech-spectrum-shaped noise. They observed a greater masking release at more negative SBRs. In their work, the masking release disappeared as the SBR approached 0 dB. Further exploration by Bernstein and Grant (2009), and Bernstein and Brungart (2011), demonstrated that this occurred as a by-product of speech statistics, specifically the peaked shape of the speech IIF (*i.e.* the perceptual statistics, and not the statistics of the physical distribution of speech levels, Dunn & White, 1940). At low SBRs only the peaks of speech were not masked by the continuous noise, and these peaks occurred for a very small fraction of time: listeners would only be able to access a small portion of the IIF. Introduction of dips in the masker fluctuations briefly unmasked levels of the speech distribution that had a higher perceptual importance, whereas the peaks of the fluctuating masker only reduced access to parts of the speech signal which had a lower importance. The sum of the importance made more accessible became greater than the sum made less accessible, leading to a benefit in intelligibility. However, at high SBRs, the slope, and hence cumulative importance of the IIF distribution changes. At high SBRs a large proportion of the IIF is available and the introduction of masker fluctuations does not introduce proportionately as much extra cumulative importance in the dips compared to that lost by the peaks of the fluctuating masker. The resulting difference between the gains and losses in cumulative importance is much smaller, hence a smaller benefit from masker fluctuations is observed. A similar explanation would suffice for the dependency of ($P_H – P_L$) benefit being observed here at lower SBRs. However, the fact that this was observed while the SBR was still positive, compared to previously at negative SBRs, is probably related to our use of a speech, rather than a steady masker. The speech masker has a wider distribution of short-term (10-ms or greater) levels than a continuous masker. The peaks of the speech masker extend up to 11 dB above the mean level (1-% exceedance level, the level exceeded by the signal for only 1 % of the time, Moore *et al.*, 2008) and so can reach levels sufficient to interfere with the target speech in its mid-range levels while at positive SBRs.

Using this finding (observed in the HA users), it is therefore possible that the benefit to intelligibility reported here by the CI users from the high-rate envelope cues may have been under-estimated since CI users were, on average, tested at higher SBRs than those used for the HA users. A more nuanced comparison of test SBRs between the two groups is left to Section IV.B below.

B. Comparison of the relative access to high-rate modulations provided by HAs or CIs despite differences in processing and delivery

The means and (SDs) of the Test SBRs for the two groups were 7.9 and (1.4) dB for HA OWN and 12.4 and (4.4) dB for CI, respectively. This 4.5 dB difference was significant, $t(35) = 4.20$, ($p < 0.001$). As previously noted, some of this difference was driven by a small number of the older CI participants tested at SBRs exceeding +15 dB. Excluding these participants altered the CI Test SBR mean and (SD) to 9.6 and (2.1) dB, reducing the difference to 1.7 dB, ($t(28) = 2.45$, $p = 0.02$). Despite the difference in Test SBR, and apart from the difference in amount of benefit of access to high-rate envelope cues (Fig. 7), there was a lack of other differences between other measures of HA and CI performance, such as the across-group scatter in benefit of high-rate cues, and the similarity of the V_{10} measure. This suggests that, on average, the signal processing in CIs is leading to a perceptual performance using envelope modulations that is similar to the delivery of modulation information via an acoustic hearing aid. The perceptual cost of the simplification of modulation information for electrical stimulation does not render too much modulation information inaccessible (as inferred from the loss of about 1.7dB SBR). It should be noted, however, that the CI participants were pre-selected to be at the better end of performance on clinical tests of speech intelligibility (greater than 70% word intelligibility at an SBR of +10 dB for BKB sentences presented in speech-spectrum shaped noise).

C. The overall pattern of the results

For several participants, the fitted functions show a step change in intelligibility over a narrow range of switching threshold, but where the fits look very good (H15 OWN, H15 SIM, C5, C12 and C18). Conversely, there are a few participants whose data also exhibit this step change, but are poorly fitted to the data (H11 SIM, H12 SIM, H15 OWN, H15 SIM, and C1, C3, C10, C14, C15, C17). It is noteworthy that this degree of heterogeneity in response is similar across both groups. One possible explanation is variation in attention or fatigue during the test. Each condition was tested with 30 sentences, and took at least 10 minutes to complete. If attention varied during the test session, performance in some individual conditions may have been unrepresentative of stable performance. A faster rotation of the within-participant test condition by use of smaller blocks of sentences may have provided a better form of counterbalancing.

Although participants exhibited a wide range of abilities to access high-rate modulations in a vocoded signal, on a group basis, the intermediate processing conditions between the all-H and all-L conditions did not produce any extra insight into participant or device behaviour. This may have happened because the processing reported here assesses performance with a wide-band signal (speech) to produce a single IIF. If the abilities of the participant varied across frequency then an average IIF may not reflect frequency-specific IIF performance. It is also possible that the IIF for high-rate modulations varies across frequency (as with audio frequency IIFs, Studebaker & Sherbecoe, 2002). Hence a single figure-of-merit, such as the V_{10} measure, may be insufficient to capture the subtlety of the variation in success of modelling individual results.

The overall pattern of results suggest that refining the accuracy of this processing technique would centre on:

(1) Homogenisation across participants of Test SBR and intelligibility achieved in the P_H condition, such as by the use of audio-visual cues or reduced-size speech corpora (Bernstein & Grant, 2009; Bernstein and Brungart, 2011).

(2) Further interleaving of test conditions to overcome short-term variations in possible fatigue.

(3) Exploration of whether the step change in fitted P-I functions observe in Figs. 3 and 4 is consistent in participants, and to what it could be related, such as an unoptimised match of hearing prosthesis to the participant.

V. CONCLUSIONS

The ability to benefit from high-rate envelope modulations (> 16 Hz) in a two-talker separation task using tone-carrier vocoded processing was explored as a function of depth in the channel envelopes at which the high-rate information was made available. The Signal-to-Background Ratio (SBR) was adjusted for each participant in order to set best performance to about 70% so that the effect of processing was measured on the steepest part of the Performance-Intensity functions.

For HA participants, the 'cost' of a clinically fitted non-linear hearing aid over a simulated linear aid with the same insertion gain in accessing these higher-rate modulations was estimated as a 1.0 dB loss of SBR (for a competing talker background). The dynamic range of modulations made accessible by the processing did not seem to differ between the clinical and the simulated aids. There was no evidence of a distortion of high-rate cues by the OWN aid over and above the generally poorer performance across all rates of modulation.

The finding of a negative correlation between the Test SBR and the degree of benefit obtained from the high-rate modulations by the HA participants appears to be another example where the statistics of speech perception (as measured by modulation-based intensity importance functions, IIFs), and not just hearing ability, influence results. This is similar to the findings of Bernstein and Grant (2009) who explained their results in terms of full-audio IIFs. The similarities between the two unrelated experiments re-emphasises the need to control for Test SBR when comparing results across participants.

CI participants were much less able than HA participants to make use of high-rate envelope cues. Apart from one star performer, C15, CI participants were generally tested at higher SBRs. The demonstration among HA participants, that Test SBR influences the degree of benefit obtained, suggest that even the lower degree of benefit from the high-rate cues among the CI participants could partly be due to speech statistics rather than the underlying deficits caused by the more severe hearing losses that are a pre-requisite for cochlear implantation.

A further note of optimism for the CI population was that the dynamic range over which they could access high-rate modulation cues was very similar to that of the HA population. Despite the highly synthetic signal delivered via a CI, and uncertainties about matching stimulation to the non-linearities at the neural interface, it is encouraging to see that CI processing is able to enable a functionally similar access to speech cues.

ACKNOWLEDGEMENTS

This work was funded by the Manchester University Hospitals NHS Foundation Trust, Trust Charitable Funds (Manchester University Hospital NHS Foundation Trust Charity), and facilitated by the Manchester Biomedical Research Centre and the NIHR Greater Manchester:

Clinical Research Network. The authors extend thanks to Danny Kearney at Manchester Royal Infirmary, and Greg Nassar at Trafford General Hospital, for help in HA participant recruitment. Especial thanks are extended to Deborah Mawman of the Richard Ramsden Centre for Auditory Implants, Manchester Royal Infirmary, for recruitment of the CI participants.

ANSI (1997). ANSI S3.5-1997, “Methods for the calculation of the Speech Intelligibility index” (American National Standards Institute, New York).

ANSI (2003). ANSI S3.22-2003, “Specification of hearing aid characteristics” (American National Standards Institute, New York).

Bernstein, J. G., and Brungart, D. S. (2011). “Effects of spectral smearing and temporal fine-structure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio,” J. Acoust. Soc. Am. 130, 473–488.

Bernstein, J. G., and Grant, K. W. (2009). “Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners,” J. Acoust. Soc. Am. 125, 3358–3372.

Boothroyd, A. (1990) “Articulation index: Importance function in the intensity domain,” J. Acoust. Soc. Am. Suppl. 1 88 S31.

Boyle P.J., Büchner A., Stone M.A., Lenarz T., and Moore B.C.J. (2009) “Comparison of dual-time-constant and fast-acting automatic gain control (AGC) systems in cochlear implants”. Int J Audiol 48, 211-221.

Buss, E., Hall, J. W., 3rd, and Grose, J. H. (2004). “Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss,” Ear Hear. 25, 242-250.

Cubick, J., and Dau, T. (2016). “Validation of a virtual sound environment system for testing hearing aids,” Acta Acust. united Acust. 102, 547–557.

Cubick J., Buchholz J.M., Best V., Lavandier M., and Dau T. (2018) “Listening through hearing aids affects spatial perception and speech intelligibility in normal-hearing listeners”. J. Acoust. Soc. Amer. 144, 2896-2905. DOI 10.1121/1.5078582

Dillier N., Spillmann T., Fisch U. P., Leifer L. J. (1980) “Encoding and decoding of auditory signals in relation to human speech and its application to human cochlear implants. Audiology, 19, 146-163. DOI: 10.3109/00206098009072657

Dillon, H., and Storey L. (1998) “The National Acoustic Laboratories' Procedure for selecting the Saturation Sound Pressure Level of hearing aids: theoretical derivation”. Ear & Hearing 19, 255-266

Dillon H. (2012a) “Compresssion systems in hearing aids”. Ch. 6 in Hearing Aids, 2nd ed., Thieme, Stuttgart.

Dillon H. (2012b) “Directional microphones and arrays” and “Advanced signal processing schemes” Chs. 7 and 8 in Hearing Aids, 2nd ed., Thieme, Stuttgart.

Dubbelboer, F., and Houtgast, T. (2008). “The concept of signal-to-noise ratio in the modulation domain and speech intelligibility,” J. Acoust. Soc.Am. 124, 3937–3946.

Dunn, H. K., and White, S. D. (1940). “Statistical measurements on conversational speech,” J. Acoust. Soc. Am. 11, 278–288.

Ewert S., and Dau T. (2000) “Characterizing frequency selectivity for envelope fluctuations”. J. Acoust. Soc. Am. 108, 1181-1196.

Fowler, E. P. (1936). “A method for the early detection of oto-sclerosis”. Archives of Otolaryngology, 24, 731-741.

Fu Q.-J., and Shannon R.V. (1998) “Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners”. J. Acoust. Soc. Am. 104, 2570-2577.

Giannoulis D., Massberg M., and Reiss J.D. (2012) “Digital Dynamic Range Compressor Design- A Tutorial and Analysis”. J Audio Eng Soc 60, 399-408.

Glasberg, B. R., and Moore, B. C. J. (1990). “Derivation of auditory filter shapes from notched-noise data." Hear. Res. 47, 103-138.

Grose J.H. and Mamo S.K. (2010) “Processing of temporal fine structure as a function of age” Ear. Hear 31:755-760.

Holube, I., Fredelake, S., Vlaming, M., and Kollmeier, B. (2010). “Development and analysis of an International Speech Test Signal (ISTS)”. Int Jour Audiol, 49, 891-903.

Hopkins, K., and Moore, B.C.J. (2007). “Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am. 122, 1055-1068.

Hopkins, K., and Moore, B.C.J. (2009). “The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise,” J. Acoust. Soc. Am. 125, 442-446.

Hopkins, K., Moore, B.C.J., and Stone, M.A. (2008). “Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech,” J. Acoust. Soc. Am. 123, 1140-1153.

IEEE (1969). “IEEE recommended practice for speech quality measurements.” IEEE Trans. Audio Electroacoust. 17, 225–246.

Jørgensen, S., and Dau, T. (2011). “Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing”. J. Acoust. Soc. Am. 130, 1475–1487.

Jørgensen, S., Ewert, S. D., and Dau, T. (2013). “A multi-resolution envelope-power based model for speech intelligibility”. J. Acoust. Soc. Am. 134, 436–446.

Kates J.M. (2005) “Principles of Digital Dynamic-Range Compression”. Trends Amplif 9, 45-76.

Kates J.M. (2010) “Understanding compression: Modeling the effects of dynamic-range compression in hearing aids”. Int J Audiol 49, 395-409

Kates J.M. (2011) “Spectro-temporal envelope changes caused by temporal fine structure modification”. J Acoust Soc Am, 129, 3981–3990.

Khing P.P., Swanson B.A., and Ambikairajah E. (2013) "The Effect of Automatic Gain Control structure and release time on cochlear implant speech intelligibility". PLoS ONE 8(11): e82263. doi:10.1371/journal.pone.0082263

Loizou P.C., Dorman M., Poroy O., Spahr T. (2000) Speech recognition by normal-hearing and cochlear-implant listeners asa funciton of intensity resolution. J.Acoust. Soc. Amer. 108, 2377-2387.

Lunner, T., and Sundewall-Thoren, E. (2007). "Interactions between cognition, compression, and listening conditions: effects on speech-in-noise performance in a two-channel hearing aid," J. Am. Acad. Audiol. 18, 604-617.

MacLeod, A., and Summerfield, Q. (1990). "A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use". Br. J. Audiol. 24, 29-43.

McNally G.W. (1984) "Dynamic range control of digital audio signals". J Aud Eng Soc 32, 316-327.

McDermott H.J., Henshall K.R., and McKay C.M. (2002) "Benefits of syllabic input compression for users of cochlea implants". J Am Acad Audiol. 13, 14-24.

Miller, G. A., and Licklider, J. (1950). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. 22, 167–173.

Moore, B.C.J. (2007). "Cochlear Hearing Loss: Physiological, Psychological and Technical Issues", 2nd Ed. (Wiley, Chichester), pp. 332.

Moore B.C.J. (2008) "The choice of compression speed in hearing aids: theoretical and practical considerations and the role of individual differences". Trends Amplif 12, 103-112.

Moore B.C.J. (2014) "Auditory processing of temporal fine structure". World Scientific. DOI: 10.1142/9789814579667_0004

Moore, B. C. J., Glasberg, B. R., and Stone, M.A. (2004) "New version of the TEN test with calibrations in dB HL," Ear Hear. 25, 478–487.

Naylor G., and Johannesson R.B. (2009) "Long-Term Signal-to-Noise Ratio at the input and output of amplitude-compression Systems". J Am Acad Audiol 20:161–171. DOI: 10.3766/jaaa.20.3.2

Oxenham, A. J., and Simonson, A. M. (2009). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference." J. Acoust. Soc. Am. 125, 457-468.

Plomp, R. (1988). "The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function". J. Acoust. Soc. Amer., 83, 2322-2327.

Reitan R.M. (1958). "Validity of the Trail Making test as an indicator of organic brain damage". Percept Motor Skills 8, 271-276.

Shannon R.V. (1992) "Temporal modulation transfer functions in patients with cochlear implants". J. Acoust. Soc. Am. 91, 2156-2164.

Souza P.E. (2002) “Effects of compression on speech acoustics, intelligibility and sound quality”. Trends Amplif 6, 131-165.

Souza, P., and Rosen, S. (2009). “Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech,” J. Acoust. Soc. Am. 126, 792–805.

Steinberg, J.C., Gardner, M.B. (1937). “The Dependence of Hearing Impairment on Sound Intensity,” J. Acoust. Soc. Am. 9, 11-23.

Stikvoort E.F. (1986) “Digital Dynamic Range Compressor for Audio”. J. Aud. Eng. Soc. 34, 3-9.

Stöbich, B., Zierhofer, C.M., and Hochmair, E.S. (1999) “Influence of Automatic Gain Control Parameter Settings on Speech Understanding of Cochlear Implant Users Employing the Continuous Interleaved Sampling Strategy”. Ear Hear. 20, 104-116

Stone M.A., and Moore B.C.J. (2008). “Effects of spectro-temporal modulation changes produced by multi-channel compression on intelligibility in a competing-speech task”. J. Acoust. Soc. Am. 123, 1063-1076.

Stone M.A., Füllgrabe C., Moore B.C.J. (2009). “High-rate envelope information in many channels provides resistance to reduction of speech intelligibility produced by multi-channel fast-acting compression”. J. Acoust. Soc. Am. 126, 2155-2158.

Stone, M.A., Anton, K., and Moore, B.C.J. (2012a). “Use of high-rate envelope speech cues and their perceptually relevant dynamic range for the hearing impaired,” J. Acoust. Soc. Am. 132, 1141–1151.

Stone, M.A., Füllgrabe, C., and Moore, B.C.J. (2012b). “Notionally steady background noise acts primarily as a modulation masker of speech,” J. Acoust. Soc. Am. 132, 317-326.

Stone, M.A., Moore, B.C.J., Alcántara, J. ., and Glasberg, B.R. (1999). “Comparison of different forms of compression using wearable digital hearing aids”. J Acoust Soc Am, 106, 3603-3619.

Studebaker, G.A., and Sherbecoe, R.L. (2002). “Intensity-importance functions for bandlimited monosyllabic words,” J. Acoust. Soc. Am. 111, 1422–1436.

Villchur E. (1973) “Signal processing to improve speech intelligibility in perceptive deafness”. J. Acoust. Soc. Am. 53, 1646-1657.

Villchur E. (1988) “Comments on ‘‘The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function’’ [J. Acoust. Soc. Am. 8 3, 2322-2327 (1988)] J. Acoust. Soc. Am. 86, 425-427.

Warren, R.M., Bashford J.A. & Lenz P.W. (2005) “ Intelligibilities of 1-octave rectangular bands spanning the speech spectrum when heard separately and paired. J. Acoust. Soc. Am. 118, 3261-3266. DOI: 10.1121/1.2047228

Wechsler, D. (1997). “Wechsler Adult Intelligence Scale”. Third Edition. Oxford: Pearson Assessment.

White M.W. (1986) Compression systems for hearing aids and cochlear implants. J. Rehab. Res. Devel. 23:25-39.

Whitmal, N.A., Poissant, S.F., Freyman, R.L., and Helfer, K.S. (2007). “Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience,” J. Acoust. Soc. Am. 122, 2376–2388.

Xu, L., Thompson, C.S., Pfingst, B.E. (2005). “Relative contributions of spectral and temporal cues for phoneme recognition” J. Acoust. Soc. Am. 117, 3255–3267.

Xu, L., and Zheng, Y. (2007). “Spectral and temporal cues for phoneme recognition in noise, J. Acoust. Soc. Am. 122, 1758–1764.

Zeng, F.-G. (2004). Trends in Cochlear Implants. Trends in Amplification, 8(1), 1–34. https://doi.org/10.1177/108471380400800102

TABLE I. Demographic data for the participants using hearing aids, including cognitive test results (Digit Span and Trail Making, see section II.C for precise details). (Abbreviations: RsndDan = Resound Danalogic, Prog = Progressive, NG = not given in data sheet, NIHL = Noise Induced Hearing Loss, * denotes that other compression systems occur in series, so other attack and release times may be in play).

Subject Identifier	Gender	Age at Testing (Years)	Diagnosis	HA Make	HA Model	Years of Hearing Impairment	Compression: number of channels (*N*), attack & release times (ms)	Years of Hearing Aid Use	*(L=Left, R=Right)*Ear Tested	Pure Tone Average, dBHL (1,2 & 4 kHz)	Test SBR dB 'OWN'	Test SBR dB 'SIM'	Digit Span (Forward)	Digit Span (Backward)	Trail Making (secs)
H1	F	64.4	Ototoxicity	Oticon	Spirit Zest	2	*N*=NG NG,NG	2	L	43	5	5	14	9	16
H2	M	73.4	Presbyacusis	Oticon	Spirit Zest	Prog	*N*=NG NG,NG	5	R	53	8	8	10	6	30
H3	M	69.1	Presbyacusis	RsndDan	i-FIT 71	> 3	*N*=17 12*,30*	>3	L	37	8	8	12	8	22
H4	F	59.8	Presbyacusis	RsndDan	i-FIT 71	>2	*N*=17 12*,30*	1.5	L	37	8	6	11	7	18
H5	M	70.2	Unknown	Oticon	Spirit Zest	10	*N*=NG NG,NG	9	R	50	6	6	9	5	30
H6	M	71.3	NIHL/Ototoxic	Oticon	Spirit Synergy	8	*N*=NG NG,NG	6	L	33	7	5	14	9	16
H7	F	67.2	Unknown	RsndDan	i-FIT 71	>10	*N*=17 12*,30*	9	R	32	7	7	8	4	35
H8	M	72.8	NIHL	RsndDan	i-FIT 71	>10	N=17 12*,30*	8	L	38	7	5	13	11	19
H9	M	69.6	NIHL	Oticon	Spirit Zest	7	*N*=NG NG,NG	7	L	53	8	8	10	3	50
H10	M	73.7	Menieres	Oticon	Spirit Zest	> 10	*N*=NG NG,NG	>10	L	50	10	9	9	8	36
H11	M	74.2	Unknown	Oticon	Spirit Zest	> 20	*N*=NG NG,NG	20	R	47	8	7	16	8	21
H12	M	75.5	Presbyacusis	RsndDan	i-FIT 71	10	*N*=17 12*,30*	5	R	52	7	5	11	6	30
H13	F	74.4	Unknown	RsndDan	i-FIT 71	>10	*N*=17 12*,30*	>10	L	62	10	10	13	6	30

H14	F	57.6	Familial	RsndDan	i-FIT 71	>20	*N*=17 12*,30*	12	R	43	7	5	8	2	48
H15	M	72.7	NIHL	Phonak	Nathos Micro	>20	*N*=16 1, 50	14	L	58	10	10	9	6	27
H16	F	68.4	Presbyacusis	RsndDan	i-FIT 71	>10	*N*=17 NG,NG	11	L	50	7	5	6	5	34
H18	F	75.5	Trauma / Presbyacusis	RsndDan	i-FIT 71	>10	*N*=17 12*,30*	>6	R	40	8	6	9	11	20
H19	F	71.8	Presbyacusis	Phonak	Nathos Micro	>10	*N*=16 1, 50	2	R	58	9	9	8	6	21
H25	M	72.5	Presbyacusis	RsndDan	Linx	>10	*N*=17 NG,NG	2	R	58	10.3	9.6	11	8	47

TABLE II. Demographic data for the participants using cochlear implants, including cognitive test results (Digit Span and Trail Making, see section II.C for precise details). Abbreviations: NR = noise reduction, Dir = directional microphone, Multi-prog = multiple programs available, NK = not known, CSOM = Chronic suppurative otitis media.

Subject Idenifier	Gender	Age at Testing (Years)	Diagnosis	Implant Make	Implant Model	Number of Active Channels	(NR/Dir/Multi-prog)Processor features	(pre implantation)Years of Deafness	Years of Implant Use	Test SBR (dB)	Digit Span (Forward)	Digit Span (Backward)	Trail Making (secs)
C1	M	49.9	Meningitis	Cochlear	Freedom	18	N/Y/N	5	24.9	10	14	8	23
C2	M	50.6	Progressive	Medel	Opus 2	12	Y/N/Y	Progressive over 20 years	1.4	12	9	8	18
C3	M	70.6	Progressive	Medel	Opus 2	12	Y/N/Y	5	1.5	18	9	6	21
C4	M	45.9	Progressive	Medel	Opus 2	12	Y/N/Y	9	1.0	8	8	7	20
C5	F	55.8	Progressive	Cochlear	CP810	20	Y/Y/Y	20	4.2	9	9	7	17
C6	F	44.6	Progressive (Idiopathic)	Cochlear	CP910	22	Y/Y/Y	11	13.1	12	11	8	21
C7	M	70.9	Progressive (Idiopathic)	Cochlear	CP910	20	Y/Y/Y	NK	NK	18	9	3	> 300
C8	M	65.4	CSOM	Medel	Opus 2	12	Y/N/N	2	1.8	13	8	8	30
C9	F	22.1	Progressive	Medel	EAS Duet Opus 2	9	Y/Y/Y	6	9.7	10	12	4	37
C10	F	43.3	Meningitis	Cochlear	CP810	20	Y/Y/Y	0.25	17.1	9	15	11	16
C11	M	43.6	Head Injury	Cochlear	CP810	21	Y/Y/Y	3	9.8	13	9	6	18
C12	F	67.2	Congenital Rubella	Medel	Opus 2	9	Y/N/N	Progressive over many years	15.1	20	8	4	30
C13	M	65.8	Menieres	Medel	Tempo+/Opus 2 (Bilateral - tested Tempo+ only)	8	Y/N/Y	3	20.0	10	12	10	19.5
C14	M	82.0	Idiopathic (Sudden)	Cochlear	Freedom	21	Y/Y/Y	4	19.0	19	8	5	25
C15	F	51.6	Progressive (Familial)	Cochlear	CP810	20	Y/Y/Y	Progressive over many	10.4	5	12	8	16

								years					
C16	F	63.6	Progressive	Medel	Rondo	10	Y/N/Y	7	17.6	10	12	6	22
C17	F	76.6	Familial Progressive	Cochlear	CP910	18		Progressive over many years	13.0	18	7	4	29
C18	M	55.8	Head Injury	Cochlear	CP910	20	Y/Y/Y	18	16.5	10	12	9	18

TABLE III. Channel edge, tone-carrier and bandwidth (BW) frequencies, all in Hz, for the 16-channel vocoder used in the HA processing experiments.

Edge :	100	179	276	397	548	734	964	1250	1604	2043	2588	3263	4099	5136	6422	8015	9990
Carrier :	139	228	337	473	641	849	1107	1427	1824	2316	2925	3681	4618	5779	7218	9002	
BW:	79	97	121	151	184	230	286	354	439	545	675	836	1137	1286	1593	1875	

TABLE IV. Parameters from curve fittings to the data of Figs. 3 and 4, using Eqn. (2, P_H, P_L, *sw*). "Id" denotes "Subject Identifier", as used in Tables I and II. Values of P_H and P_L, and their differences, are expressed in percent. For the HA SIM group, P_{HCorr} is the predicted P_H as if tested at the same SBR as HA OWN. Values to which corrections have been made are shown in bold. See text for details of this correction. V_{10} indicates the 'valley measure', in dB, described in section III.C.2.

(a) Hearing aid (HA) participants.

	OWN						SIM						
Id	P_H	P_L	β	MP	$P_H - P_L$	V_{10}	P_H	P_L	β	MP	$P_H - P_L$	P_{HCorr}	V_{10}
H1	56.1	13.6	0.216	0.2	42.4	-10.0	65.9	15.5	0.130	4.3	50.5	65.9	-12.6
H2	40.4	22.7	1.777	-1.6	17.7	-2.8	49.8	24.8	0.232	-3.9	25.0	49.8	-13.4
H3	72.4	44.1	0.142	5.8	28.2	-9.7	83.2	55.8	0.197	1.6	27.3	83.2	-9.6
H4	74.2	37.3	1.386	5.2	37.0	3.6	69.5	26.0	0.304	3.2	43.6	**84.0**	-4.0
H5	43.8	3.3	0.506	2.8	40.5	-1.5	48.7	5.1	0.543	5.2	43.6	48.7	1.1
H6	76.4	45.4	0.201	-3.1	31.1	-14.1	71.2	27.4	0.172	-1.5	43.7	**85.1**	-14.3
H7	50.3	20.4	0.229	2.6	29.9	-7.1	67.8	40.2	0.324	1.2	27.5	67.8	-5.6
H8	71.8	14.6	0.270	4.6	57.2	-3.5	63.1	19.5	0.325	2.5	43.6	79.**8**	-4.3
H9	53.9	26.1	0.168	4.1	27.8	-9.0	66.0	25.5	0.204	2.2	40.5	66.0	-8.5
H10	65.9	34.3	0.184	5.1	31.5	-6.8	64.1	28.8	0.132	1.5	35.3	**73.1**	-15.2
H11	55.7	26.7	0.175	-0.9	29.1	-13.5	69.4	54.8	6.215	-7.0	14.6	**77.5**	-7.4
H12	64.9	25.0	0.140	-6.2	39.9	-21.9	61.5	26.6	7.700	-2.0	34.9	**78.7**	-2.3
H13	47.9	19.6	0.340	-0.2	28.3	-6.7	56.6	23.5	0.203	1.7	33.1	56.6	-9.1
H14	76.9	41.6	0.348	4.8	35.3	-1.5	73.0	34.1	0.340	-0.1	38.9	86.2	-6.6
H15	28.0	14.1	21.52	0.2	13.9	0.1	47.1	26.0	7.521	5.9	21.1	47.1	5.6
H16	73.2	26.6	0.284	5.1	46.6	-2.7	68.1	14.4	0.236	2.9	53.7	**83.2**	-6.5
H18	78.2	29.4	0.224	-0.7	48.8	-10.6	74.7	26.7	0.166	1.7	48.0	**87.2**	-11.5
H19	77.6	37.8	0.165	-4.9	39.8	-18.2	67.4	35.1	7.387	1.8	32.3	67.4	1.5
H25	60.5	28.2	0.509	0.3	32.3	-4.1	56.3	22.8	0.144	3.4	33.4	**63.3**	-11.9

Means and (SD) of ($P_H - P_L$) : OWN 34.9 (10.4), SIM 36.4, (10.3) percent.

TABLE IV (cont).

(b) Cochlear implant (CI) participants

Id	P_H	P_L	β	MP	$P_H - P_L$	V_{10}
C1	46.9	30.7	7.288	-2.0	16.2	-2.3
C2	61.0	33.4	0.084	-2.4	27.6	-28.6
C3	80.3	56.6	7.054	2.1	23.8	1.8
C4	58.4	29.9	0.219	1.7	28.5	-8.4
C5	36.8	16.4	6.063	2.2	20.4	1.8
C6	78.6	46.0	0.169	-1.7	32.6	-14.7
C7	61.4	46.5	0.569	0.7	14.9	-3.2
C8	60.2	36.1	0.889	0.7	24.1	-1.8
C9	54.5	5.8	0.418	1.1	48.7	-4.2
C10	44.4	32.1	18.48	10.0	12.4	7.7
C11	59.8	34.5	0.147	-3.3	25.3	-18.2
C12	56.7	21.2	6.902	1.8	35.4	1.5
C13	48.1	0.0	0.241	3.0	48.1	-6.1
C14	52.8	39.9	5.622	-6.7	12.8	-7.1
C15	54.5	32.8	6.553	1.9	21.6	1.6
C16	58.2	23.3	0.346	-1.2	34.9	-7.5
C17	63.7	46.4	0.842	-9.1	17.3	11.7
C18	62.5	44.8	15.87	-4.8	17.7	-4.9

Mean and (SD) of $P_H - P_L$: 25.7 (10.9) percent.

Figure captions

FIG. 1. Mean audiogram data for the tested ears of the HA group. Errorbars show ±1 standard deviation (SD).

FIG. 2. Data from condition H/L of Stone et al. (2012a), plotting average word intelligibility as a function of switching threshold for a group of 12 subjects with a simulated HA, all tested at an SBR of +5 dB.

FIG. 3. Word intelligibility plotted as a function of switching threshold for individual subjects using HAs and fits to the data. For all panels, the ordinate spans a range of 80%, but with varying offsets. Each panel is labelled with the participant identifier as well as the test SBRs in dB as a subscript pair (OWN, SIM). (Color online)

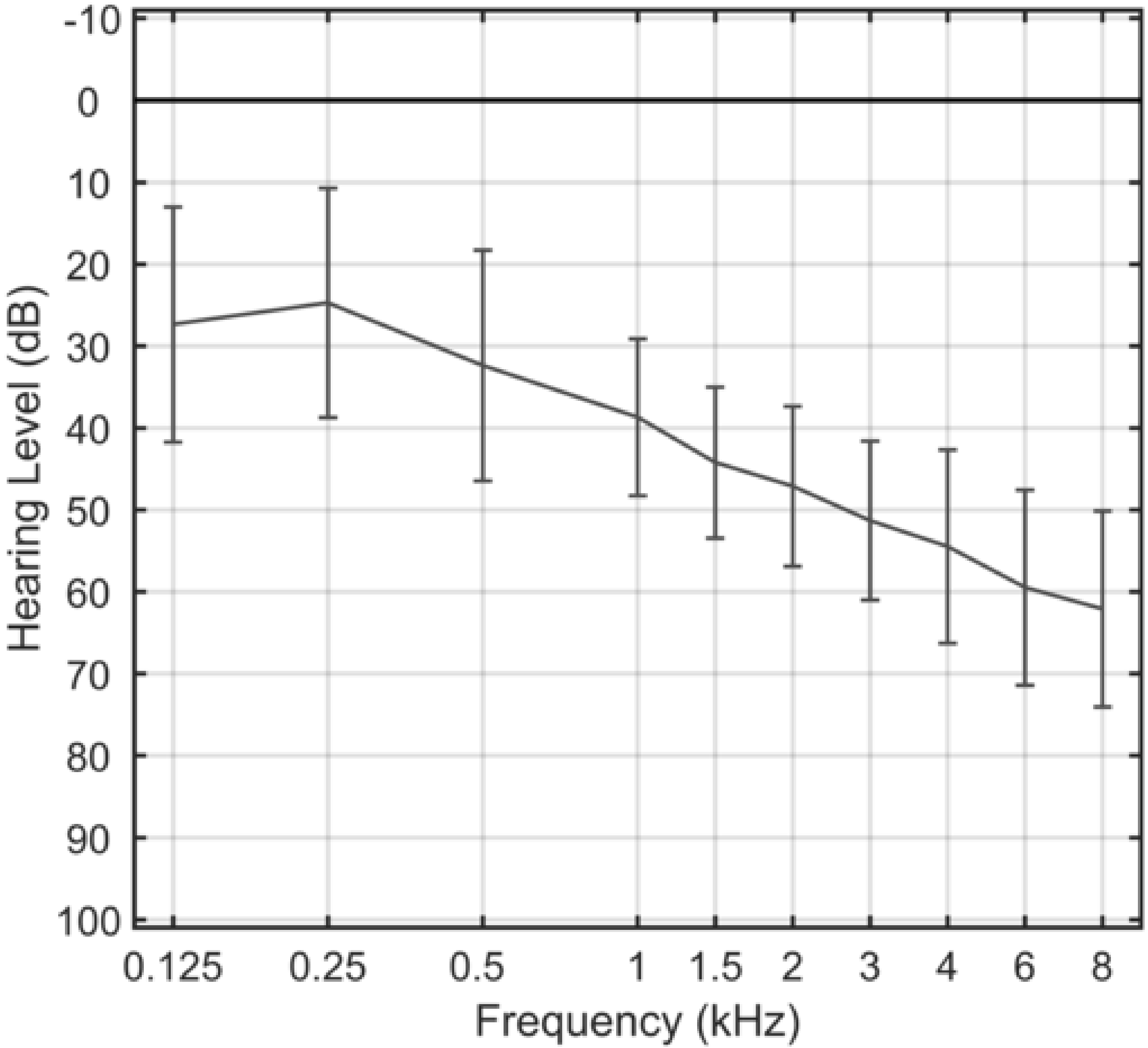
FIG 4. Individual CI participant data plotted as word intelligibility as a function of switching threshold, as well as regression fits to these sets. For all panels the ordinate spans a range of 65%, but with varying offsets. Each panel is labelled with the participant identifier as well as the test SBR in dB, as a subscript.

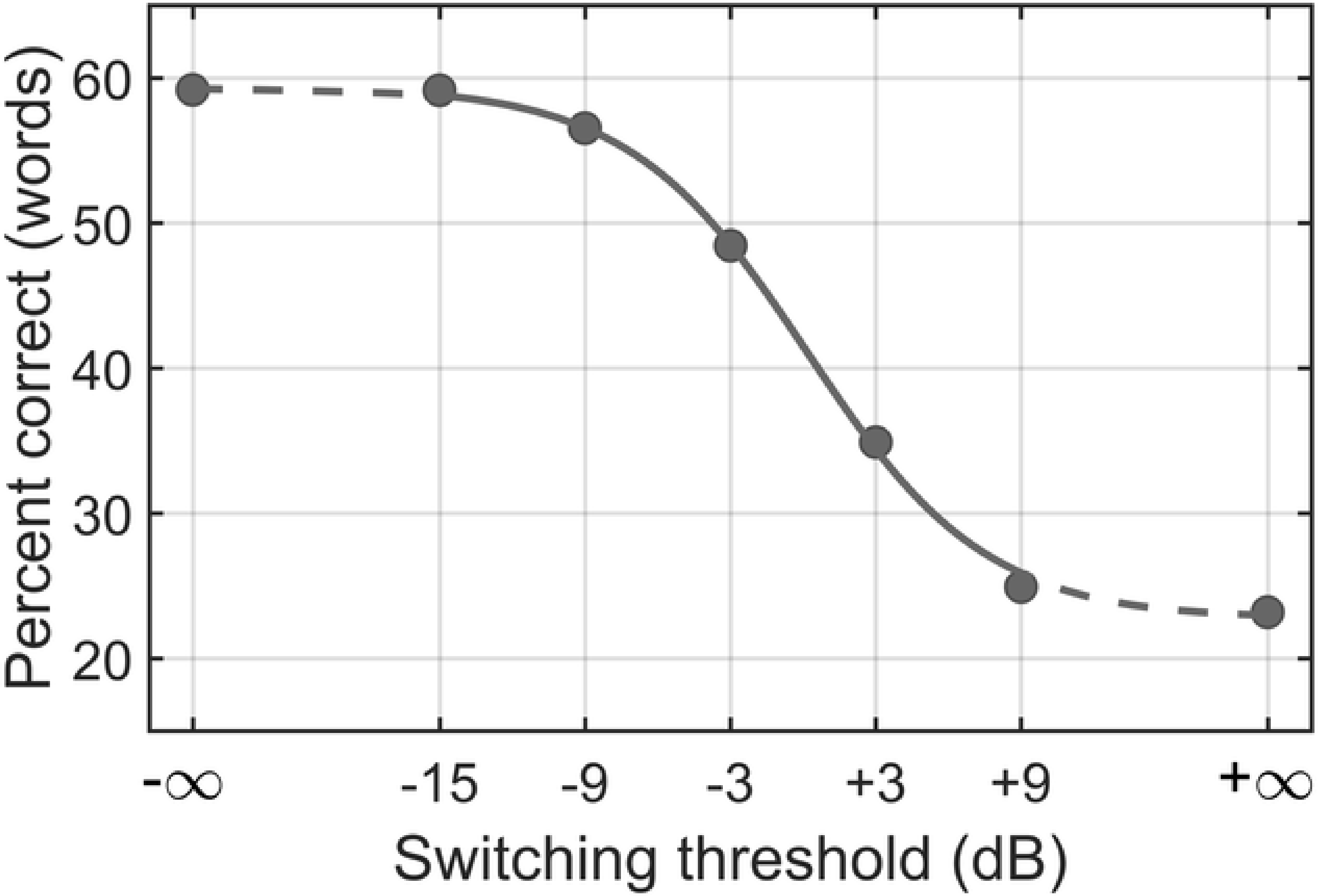
FIG. 5. Scatterplot of SBR-corrected P_H (i.e. P_{HCorr}) for the HA SIM condition against P_H for the HA OWN condition. All measures are in percent correct. The diagonal line indicates where the points would lie if intelligibility were equal for the two conditions. (Color online)

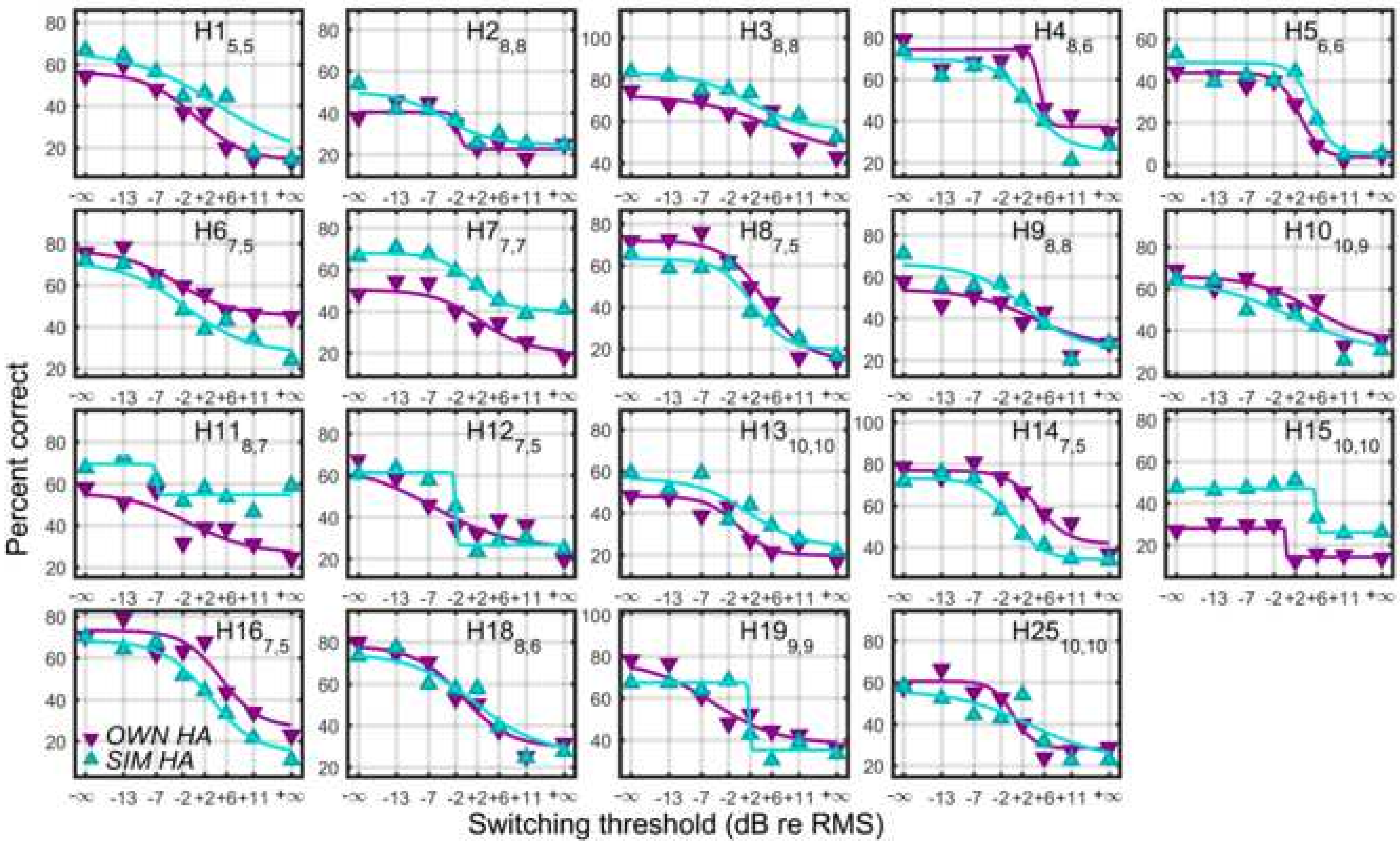
FIG. 6. Comparison of the Valley measures, V_{10}, in the SIM condition and the OWN condition, excluding pairs where V_{10} was less than -15 dB. (Color online)

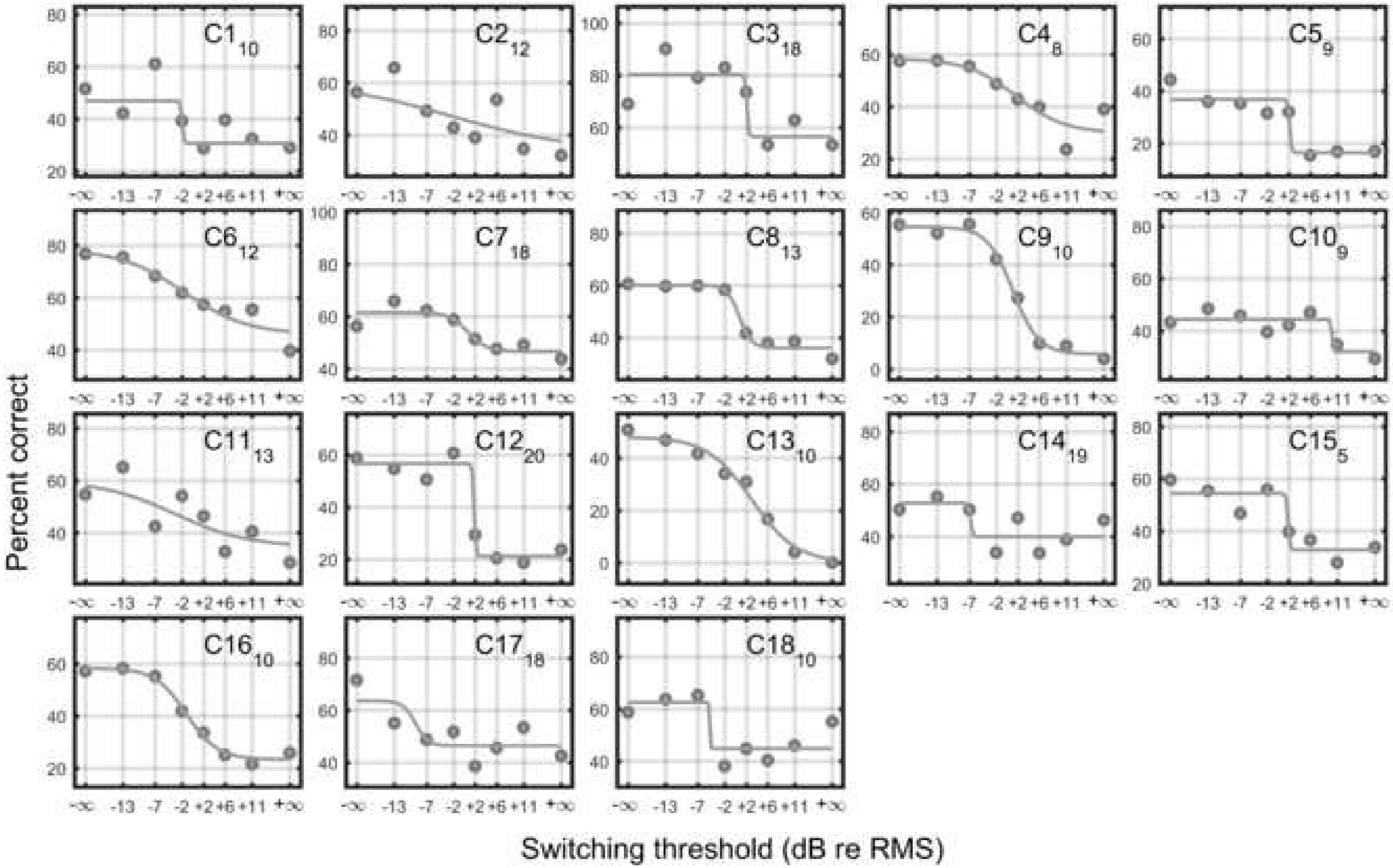
FIG. 7. Histograms of ($P_H - P_L$), the benefit in percentage intelligibility gained from the high-rate cues for the CI group (light/yellow bars) and the HA OWN group (dark/blue bars). (Color online)

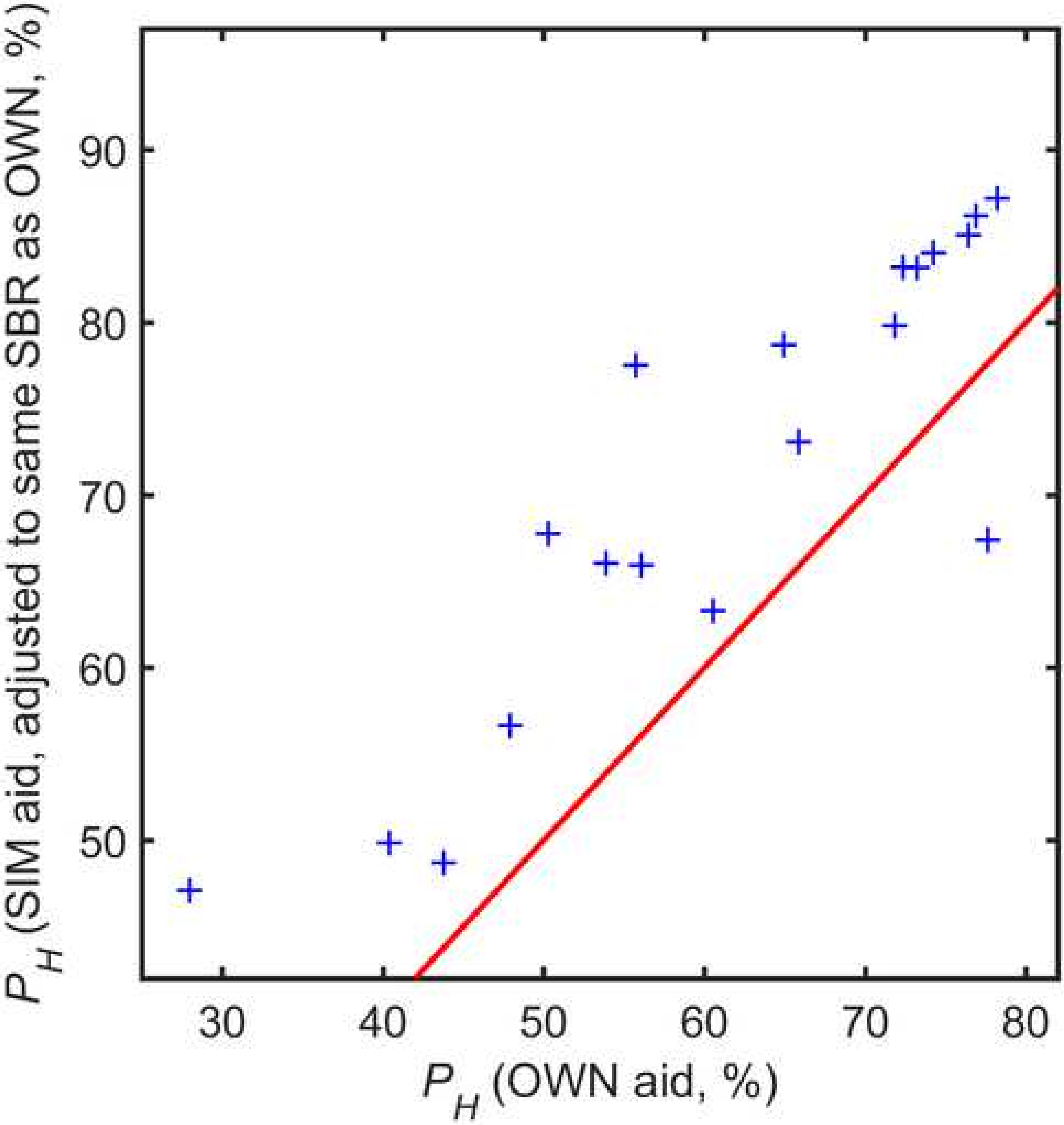
FIG. 8. Histogram of the Valley measure, V_{10}, across all three aiding conditions, but excluding individual data where the valley measure was less than –15 dB. (Color online)

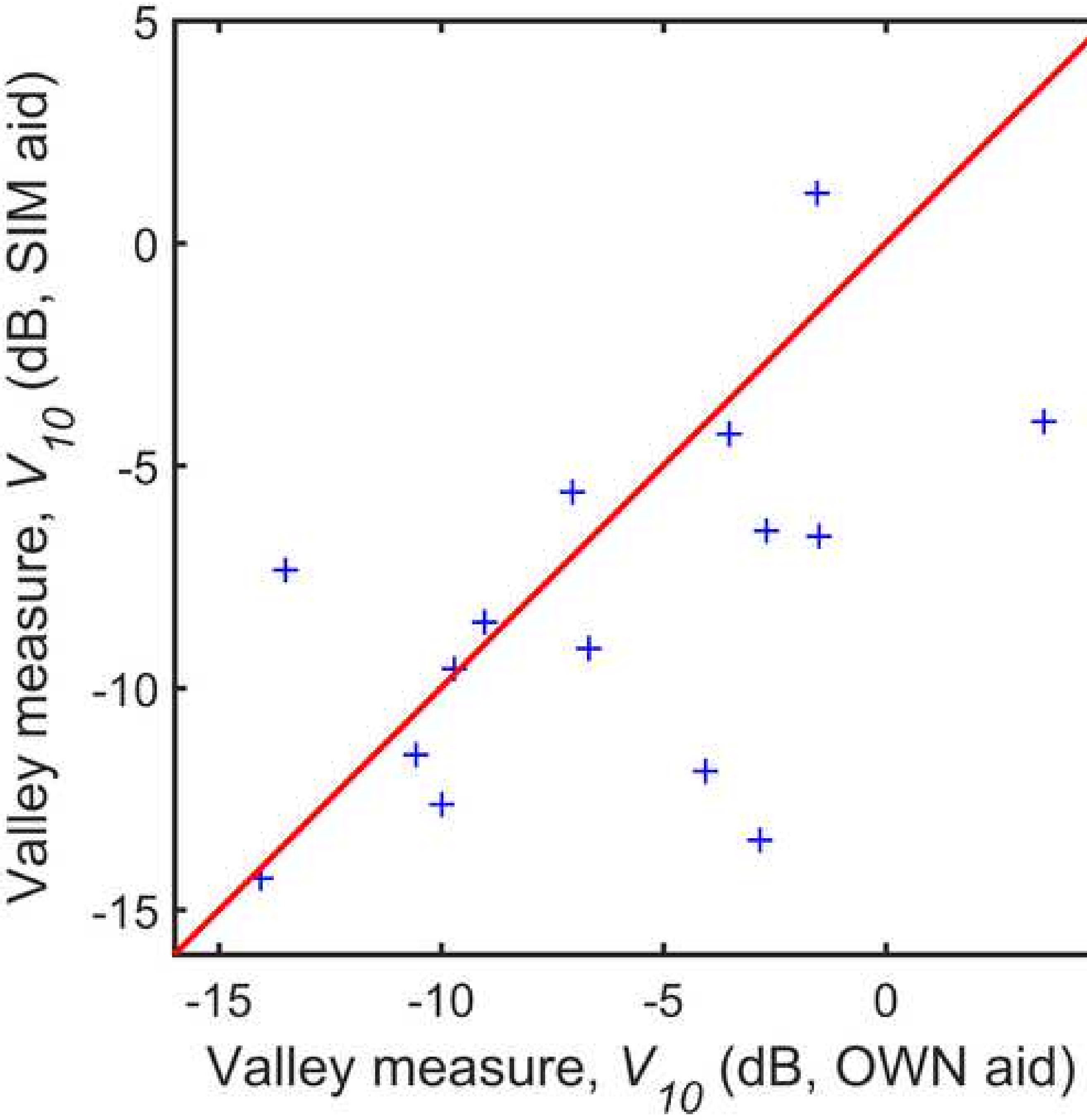

-10
0
10
20
30
40
50
60
70
80
90
100
Hearing Level (dB)
0.125
0.25
0.5
1
1.5
2
3
4
6
8
Frequency (kHz)

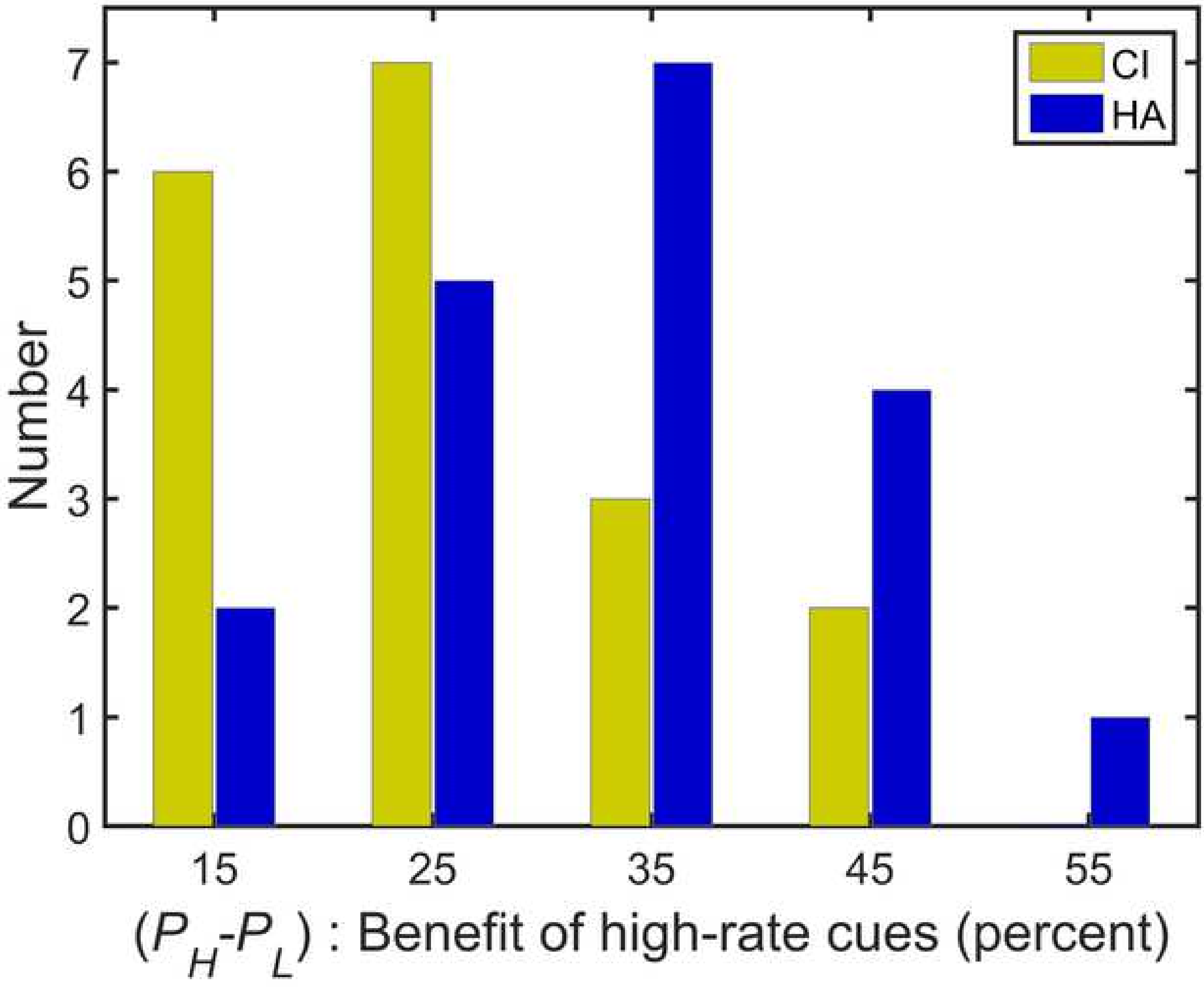

Percent correct (words)
60
50
40
30
20
-∞
-15
-9
-3
+3
+9
+∞
Switching threshold (dB)

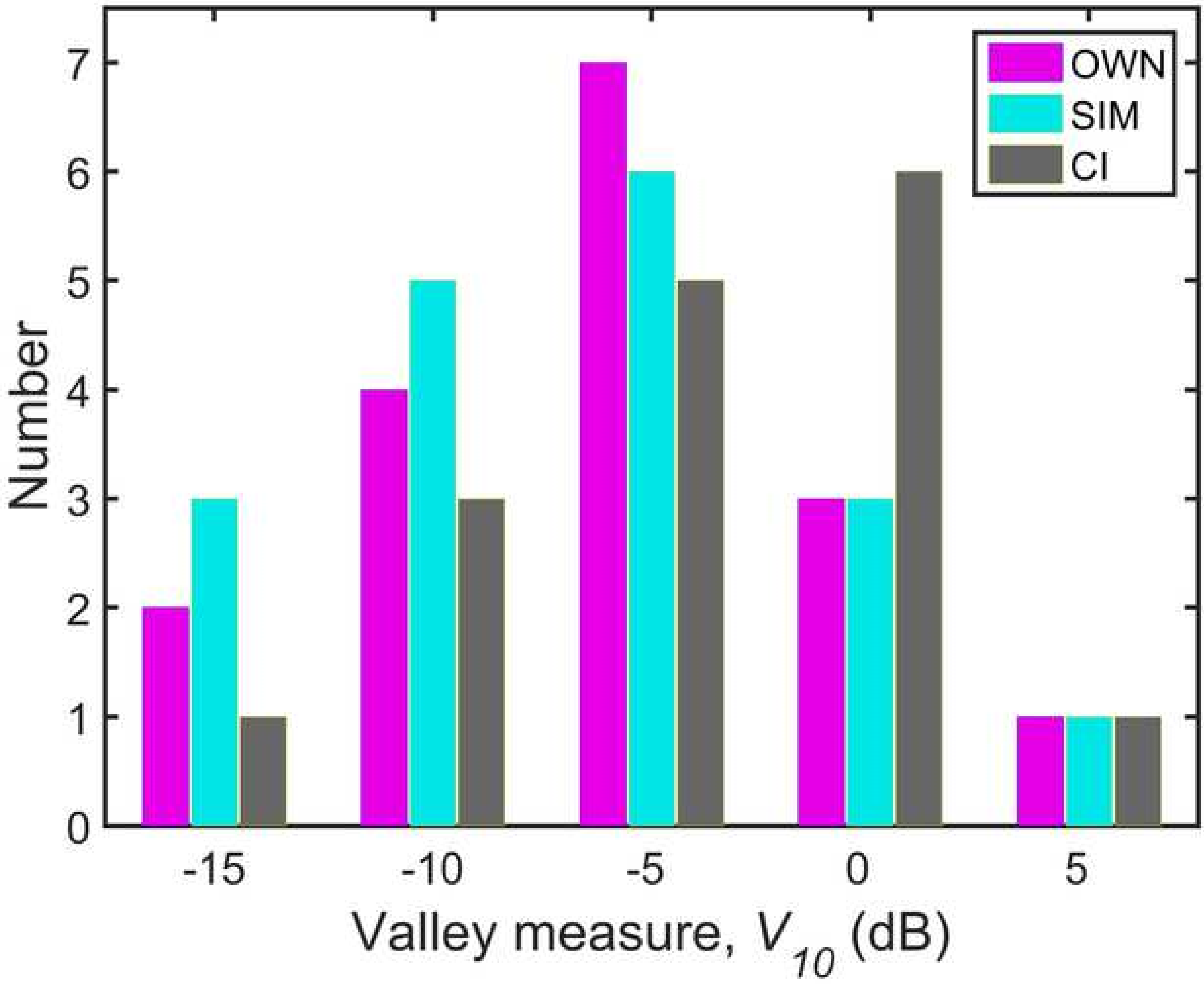

H1 5,5
H2 8,8
H3 8,8
H4 8,6
H5 6,6
H6 7,5
H7 7,7
H8 7,5
H9 8,8
H10 10,9
H11 8,7
H12 7,5
H13 10,10
H14 7,5
H15 10,10
H16 7,5
H18 8,6
H19 9,9
H25 10,10
OWN HA
SIM HA
Percent correct
-∞ -13 -7 -2 +2 +6 +11 +∞
Switching threshold (dB re RMS)


Percent correct
Switching threshold (dB re RMS)
-∞ -13 -7 -2 +2 +6 +11 +∞
C1 10
C2 12
C3 18
C4 8
C5 9
C6 12
C7 18
C8 13
C9 10
C10 9
C11 13
C12 20
C13 10
C14 19
C15 5
C16 10
C17 18
C18 10
0
20
40
60
80
100


P_H (SIM aid, adjusted to same SBR as OWN, %)
90
80
70
60
50
30
40
50
60
70
80
P_H (OWN aid, %)


5
0
-5
-10
-15
Valley measure, V_{10} (dB, SIM aid)
-15
-10
-5
0
5
Valley measure, V_{10} (dB, OWN aid)


CI
HA
Number
0
1
2
3
4
5
6
7
15
25
35
45
55
(P_H-P_L) : Benefit of high-rate cues (percent)


OWN
SIM
CI
Number
0
1
2
3
4
5
6
7
-15
-10
-5
0
5
Valley measure, V_{10} (dB)