The role of envelope periodicity in the perception of masked speech with simulated and real cochlear implants

Kurt Steinmetzger[a)] **and Stuart Rosen**

Speech, Hearing and Phonetic Sciences, University College London, Chandler House, 2 Wakefield Street, London WC1N 1PF, United Kingdom

[a)] Author to whom correspondence should be addressed. Electronic mail: kurt.steinmetzger.12@ucl.ac.uk

Current address: *Section of Biomagnetism, Department of Neurology, Heidelberg University Hospital, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany*

Running title: Envelope periodicity in cochlear implants

Keywords: Cochlear implants, speech, noise, envelope periodicity

ABSTRACT

In normal hearing, complex tones with pitch-related periodic envelope modulations are far less effective maskers of speech than aperiodic noise. Here, it is shown that this *masker-periodicity benefit* is diminished in noise-vocoder simulations of cochlear implants (CIs) and further reduced with real CIs. Nevertheless, both listener groups still benefitted significantly from masker periodicity, despite the lack of salient spectral pitch cues. The main reason for the smaller effect observed in CI users is thought to be an even stronger channel interaction than in the CI simulations, which smears out the random envelope modulations that are characteristic for aperiodic sounds. In contrast, neither interferers that were amplitude-modulated at a rate of 10 Hz nor maskers with envelopes specifically designed to reveal the target speech enabled a masking release in CI users. Hence, even at the high signal-to-noise ratios at which they were tested, CI users can still exploit pitch cues transmitted by the temporal envelope of a non-speech masker, whereas slow amplitude modulations of the masker envelope are no longer helpful.

I. INTRODUCTION

A crucial limitation when listening through a cochlear implant (CI) is the restricted access to pitch information, which impairs the abilities to perceive prosodic cues and to segregate competing auditory signals such as speech embedded in background noise (Oxenham, 2008; Rosen, 1992). Compared to normal acoustic hearing, the spectral resolution offered by a CI is markedly lower and the electric pulse trains emitted by the device also lack the temporal fine structure if the original signals (e.g., Macherey and Carlyon, 2014; Moore, 2008; Wilson and Dorman, 2008). CI users therefore must rely on the periodicity of the temporal envelope when attempting to extract the pitch of a sound, rather than the much more salient spectral pitch cues. This reliance on temporal voice pitch cues at the rate of the fundamental frequency ($F0$) has, for example, repeatedly been demonstrated when CI users had to identify the gender of a talker and serves to explain the lower performance compared to normal-hearing listeners in this task (Fu et al., 2005; Fuller et al., 2014; Gaudrain and Başkent, 2018; Meister et al., 2016). Similarly, CI users can to some extent discriminate between questions and statements, based on temporal $F0$ cues (Chatterjee and Peng, 2008; Green et al., 2005; Meister et al., 2009). There is, however, conflicting evidence regarding whether CI users can also exploit temporal $F0$ cues when attempting to understand speech in the presence of a masker. Stickney and colleagues (Stickney et al., 2007; Stickney et al., 2004) have reported no effect of increasing the $F0$ difference between two competing talkers or varying the gender of the talkers, respectively. On the other hand, Cullington and Zeng (2008) found that a female voice is a less effective masker of a male talker. More generally, studies employing a variety of tasks with speech and non-speech materials (Deeks and Carlyon, 2004; Gaudrain et al., 2008; Kreft et al., 2013) have shown that temporal periodicity cues appear not to be sufficient to induce stream segregation in CI users and simulated CIs.

Yet, none of the studies mentioned so far measured speech intelligibility in CI users and CI simulations with non-speech maskers specifically designed to vary regarding the presence or absence of $F0$ cues, which would enable a more direct investigation of the role of temporal periodicity. The current study seeks to do so by re-using materials introduced in Steinmetzger and Rosen (2015), where it was investigated whether periodicity cues in both target speech and masker affect the ability of normal-hearing listeners to understand spoken sentences. Specifically, periodic maskers based on harmonic complex tones with dynamically varying $F0$ contours derived from real speech were contrasted with aperiodic speech-shaped noise maskers. Listeners were found to substantially benefit from masker periodicity, while manipulating the periodicity of the target speech using different vocoders had little effect. Factors that are thought to explain this *masker-periodicity benefit* (MPB) in normal hearing include the use of the masker pitch to segregate (e.g., Oxenham, 2008) and possibly subtract it from the signal mixture (i.e., harmonic cancellation; de Cheveigné et al., 1995; de Cheveigné et al., 1997); the glimpsing of sections of the target speech in between the resolved masker harmonics (Deroche et al., 2014a, 2014b; Leclère et al., 2017); and the absence of random envelope modulations in periodic sounds (i.e., modulation masking; Stone et al., 2011; Stone et al., 2012) that could interfere with the low-frequency envelope modulations of the target speech which are critical for speech intelligibility (Drullman et al., 1994; Elliott and Theunissen, 2009). However, the exact contribution of each of these factors remains to be specified.

Due to the limited access to spectral information with CIs, neither harmonic cancellation nor spectral glimpsing are hypothesised to play a role in the current study. Additionally, as suggested by Oxenham and Kreft (2014), channel interaction effects appear to smear out random envelope modulations when listening through a CI, which would further reduce the acoustic contrast between the periodic and aperiodic maskers. Hence, the remaining part of the MPB observed in CI users can likely be attributed to the weak pitch percept caused by the *F*0-related envelope modulations of the periodic maskers. Compared to normal acoustic hearing, these *F*0-related modulations may even be stronger when listening through a CI, as the current spread along the electrode array should emphasise the temporal regularity of the pulse trains presented to the individual electrodes (Geurts and Wouters, 2001).

Additionally, the current study further investigated the ability to benefit from slow amplitude modulations of the masker in simulated and real CIs. The motivation for this was, firstly, to assess whether the *fluctuating-masker benefit* (FMB) is affected by the periodicity of target speech and masker, and secondly, to estimate the size of the FMB relative to the MPB. For normal-hearing listeners, the MPB has been found to be markedly larger than the FMB obtained from sinusoidal 10-Hz modulations of the masker envelope at a modulation depth of 100% (~8.5 vs. ~4 dB, respectively; cf. Figs. 5 & 6 in Steinmetzger and Rosen, 2015). However, CI simulation studies have usually found hardly any benefit from masker envelope fluctuations (Cullington and Zeng, 2008; P. B. Nelson and Jin, 2004; Qin and Oxenham, 2003), while CI users often even show a small decline in performance (Fu and Nogaki, 2005; P. B. Nelson et al., 2003; Stickney et al., 2004). The absence of an FMB in CI users has also been attributed to the reduced spectral resolution (Fu et al., 1998) and the limited access to *F*0 information (Stickney et al., 2007; Stickney et al., 2004), as well as increased forward masking (D. A. Nelson and Donaldson, 2001). At least in part, however, it can also be explained by the elevated speech reception thresholds (SRTs) compared to normal-hearing listeners (Bernstein and Grant, 2009), as the FMB is generally larger at lower signal-to-noise ratios (SNRs; Freyman et al., 2012).

Importantly, in all previously mentioned studies concerned with the benefit obtained from slow masker fluctuations, target and masker envelope varied independently of each other. Kwon and colleagues (2012), in contrast, introduced maskers that are intended to maximise (+MR) or minimise (-MR) the *masking release* by altering the temporal overlap with the target speech, without changing the overall level of the masker. In their study, the masker envelopes were adjusted in inverse proportion to the target sentence envelope (+MR) or proportionally to it (-MR). In other words, the +MR maskers have most of their energy at times when the speech level is low, and vice versa. The current study included +MR maskers in addition to the steady and 10-Hz modulated maskers used in Steinmetzger and Rosen (2015), with the intention to parametrically increase opportunities to glimpse sections of the target speech (steady < 10-Hz modulated < +MR). The reasoning behind this was that if glimpsing is possible at all with a CI, then it should be observed with the +MR maskers. However, contrary to what would be expected in the near-absence of energetic masking and modulation masking caused by random envelope fluctuations, only the few CI users in Kwon et al. (2012) whose intelligibility rates in quiet were at least 90% showed a substantial masking release when tested with the +MR maskers. The authors concluded that it may be particularly difficult to identify the segmental boundaries between speech and noise when listening through a CI. The present study aimed to test whether this finding can be replicated and if the results also depend on the presence of periodicity cues in target speech and masker.

II. COCHLEAR IMPLANT SIMULATIONS

A. Short introduction and rationale

Normal-hearing listeners were presented with three types of target speech (aperiodic, mixed, or periodic), each of which was combined with two types of maskers (aperiodic or periodic) that had three different kinds of envelopes (steady, 10-Hz modulated, or +MR). The periodic maskers had speech-like dynamically varying F0 contours. For each of these 18 conditions, SRTs at the 50%-correct level were measured. CI processing was simulated by noise-vocoding the mixture of target speech and masker with 8 channels and an envelope low-pass filter cut-off of 400 Hz. Due to the noise carrier used in the vocoder, random envelope modulations were added to any input signal, irrespective of whether it was initially periodic or aperiodic. To evaluate the modulations contained in the final stimulus materials, modulation spectrograms were computed using the front end of the mr-sEPSM speech intelligibility model (Jørgensen et al., 2013).

B. Methods

1. Participants

Eleven normal-hearing listeners (6 females, 5 males) were tested. Their ages ranged from 18–21 yrs, with a mean of 19.5. All participants were native speakers of British English and had audiometric thresholds of less than 20 dB hearing level (HL) at octave frequencies between 125 and 8000 Hz.

2. Stimuli

The target speech materials used in this experiment were recordings of the Basic English Lexicon sentences (BEL; Calandruccio and Smiljanic, 2012), spoken by an adult male Southern British English talker that were normalised to a common root-mean-square (RMS) level. The talker had speaking rate of 4.2 syllables/s (Praat script ‘Syllable Nuclei’; De Jong and Wempe, 2009), the median F0 frequency of the recordings was 110.1 Hz, and the first and third quartiles ranged from 103.0 to 120.1 Hz (Praat script ‘ProsodyPro’ version 5.7.7; Xu, 2013). The original sentences were slightly modified for appropriate British vocabulary. The BEL sentence corpus consists of 20 lists with 25 sentences each and the individual sentences contain 4 keywords. The sentences are characterised by a simple syntactic structure, high semantic predictability, and the use of basic

English vocabulary that would be expected to be known by non-native speakers (e.g., *'The <u>annoying</u> <u>student</u> <u>asks</u> too many <u>questions</u>.')*.

The masker materials were the same as in Steinmetzger and Rosen (2015): Harmonic complex maskers were based on *F*0 contours extracted from recordings in the EUROM database of English speech in which different speakers read five- to six-sentence passages (Chan et al., 1995). Sixteen different male talkers with Southern British English accents, and a similar speaking rate and voice quality to that of the target talker were chosen. The median F0 frequency of these 16 passages was 122.9 Hz and the first and third quartiles ranged from 107.0 to 144.1 Hz. Noise maskers were based on a 23.8-second passage of white noise.

3. Signal processing

Three target speech conditions with different degrees of source periodicity were synthesised prior to the experiment using TANDEM-STRAIGHT (Kawahara et al., 2008) implemented in MATLAB (MathWorks, Natick, MA). TANDEM-STRAIGHT is a vocoder that, unlike a classic channel vocoder, does not filter the input speech into distinct frequency bands, but separates the periodic and aperiodic components of the source from the spectral filter. In contrast to typical channel vocoder applications, this software was employed to manipulate the periodicity of the speech signals without compromising their intelligibility.

By default, TANDEM-STRAIGHT produces natural-sounding speech with a mixed source excitation, but the source estimation procedure can be adapted to produce fully aperiodic or fully periodic speech as well. Aperiodic speech was synthesised by keeping the default settings of TANDEM-STRAIGHT but setting the *F*0 to 0 Hz throughout. To synthesise speech with a natural mix of periodicity and aperiodicity, the default settings were kept, but the values of the sigmoid parameter in the source estimation routine were fixed to 1 and -40, to minimise the level of the aperiodic component in voiced speech segments. This avoids higher harmonics being noisier than lower ones, as is the case in natural speech, and hence emphasises the contrast of voiced and unvoiced speech. The same technique was used to produce fully periodic speech, but here interpolated *F*0 contours were used as input for the source extraction routine. These interpolated *F*0 contours were obtained by first extracting the original *F*0 contours. Secondly, the original *F*0 contours were interpolated through unvoiced sections and periods of silence, using a piecewise cubic Hermite interpolation in logarithmic frequency. The start and end points of each contour were anchored to the median frequency of the sentence.

The same interpolation procedure was used to obtain the *F*0 contours for the harmonic complex maskers. The waveforms for these maskers were synthesised on a period-by-period basis using the Liljencrants-Fant model (Fant et al., 1985), which closely approximates a typical adult male glottal pulse [see Green and Rosen (2013) for details]. Both the harmonic complexes and the noise maskers were matched in spectrum to the long-term average of speech (LTASS), using a fast Fourier transform-based (FFT) finite impulse response filter (FFT size 512, Greenwood-spaced 1-octave smoothing, filter order 1024).

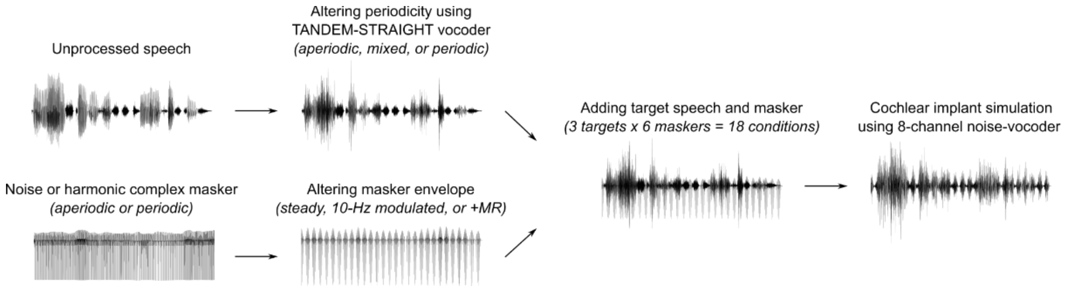

Figure 1. Cochlear implant simulations: signal processing scheme. The periodicity of the target speech was altered using the TANDEM-STRAIGHT vocoder. The aperiodic and periodic maskers were both processed to have three different types of envelopes. Target speech and masker were then added together at a given signal-to-noise ratio and additionally noise-vocoded to simulate cochlear implant signal processing.

Masker envelopes were either steady, sinusoidally amplitude-modulated at a rate of 10 Hz with a modulation depth of 100%, or inversely proportional to the target sentence envelope, adjusted in 50-ms steps (+MR; Kwon et al., 2012). As in the paper by Kwon and colleagues (2012), the level of the +MR masker was restricted to vary between -50 to -10 dB below full scale, to generate a noise floor and avoid clipping, respectively. Silent portions before and after the stimulus sentences have been removed to avoid adding significant amounts of masker energy at these locations, and to prevent potential forward masking effects[1]. For the additional portions of the masker inserted before and after the stimulus sentences, the resulting inverse envelopes were then simply extended at the levels where they started and stopped.

The onset of all maskers was 600 ms before that of the target sentence and they continued for another 100 ms after the end of the target sentence. An onset and offset ramp of 100 ms was applied to the mixture of target and masker. The masker level was kept constant and the speech level was adjusted to achieve a specific SNR.

To simulate CI processing, the signal mixture was additionally noise-vocoded before each trial, using a channel vocoder implemented in MATLAB. The mixture of target sentence and masker was first band-pass filtered into eight bands (sixth-order Butterworth). The filter spacing was based on equal basilar membrane distance (Greenwood, 1990) across a frequency range of 70 Hz–4 kHz. The output of each filter was full-wave rectified and low-pass filtered at 400 Hz (fourth-order Butterworth) to extract the amplitude envelope. The high cut-off value was chosen to ensure that temporal periodicity cues were preserved. The envelope from each band was then multiplied with a white noise carrier and the resulting signals were again band-pass filtered using the same

---

[1] *The interpretation of the results of Kwon and colleagues in the +MR condition is complicated by the fact that their stimuli appear to include substantial periods of silence before and after the target sentences (see their Fig. 2).*

filters as in the first stage of the process. Finally, before summing the individual bands together, the output of each band was adjusted to the same RMS level as found in the original recording.

A schematic depiction of the complete signal processing pipeline is shown in Fig. 1 and examples of the stimuli after CI simulation processing are shown in Fig. 2.

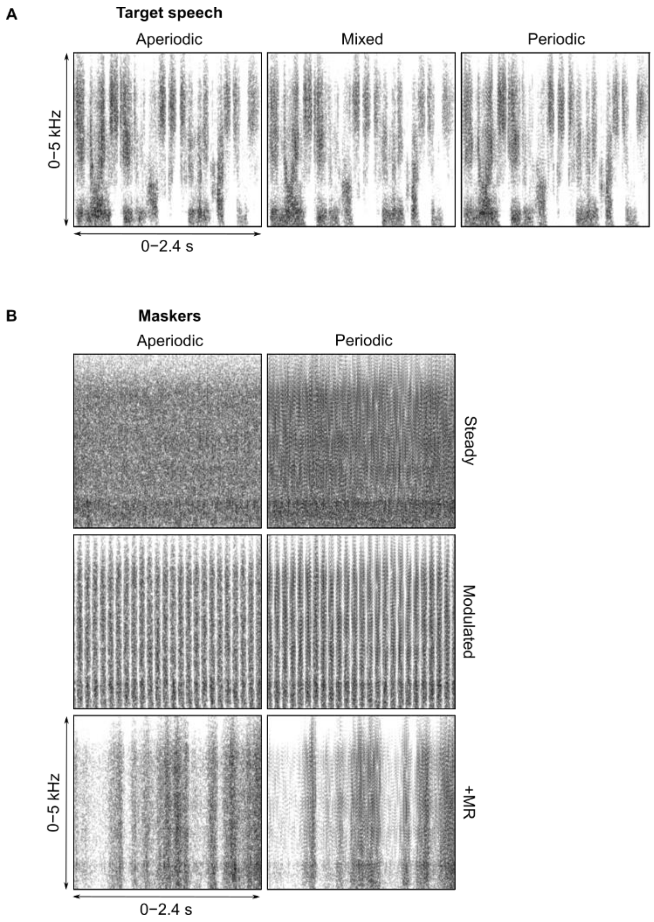

Figure 2. Cochlear implant simulations: stimuli. Panel A shows narrow-band spectrograms of one example sentence ('The annoying student asks too many questions.'), processed to have an aperiodic, mixed, or periodic source excitation. Panel B shows narrow-band spectrograms of examples of the six different maskers. Masker sources were either aperiodic or periodic and masker envelopes were either steady, 10-Hz modulated, or the inverse of the target speech (+MR). The +MR masker example is tailored to the example sentence shown above. All stimuli are shown after cochlear implant simulation processing. See Fig. 5 for an alternative depiction of the stimulus materials (modulation spectrograms) in which the subtle differences between the target speech conditions are more apparent.

4. Procedure

Participants were presented with 1 BEL sentence list in each of the 18 experimental conditions (3 target speech conditions x 6 maskers). Only the first 20 sentences of each list were used to reduce the testing time required. The SRT for every processing condition was determined by tracking the SNR necessary to repeat 50% of the keywords correctly, using a 1-up/1-down adaptive procedure. The initial SNR was set to +10 dB and adjusted up or down by 11 dB before the first reversal, 7 dB before the second reversal, and 3 dB after that. If fewer than half of the keywords in the first trial were incorrect, the SNR was set to +24 dB and the procedure started over again. The SRT was calculated by taking the mean of the largest even number of reversals with a 3-dB step size.

The verbal responses were logged by the experimenter before the next sentence was played. A so-called loose keyword scoring technique was applied, in which the roots of the four keywords had to be correctly identified. No feedback was given following the responses. The presentation and logging of the responses was carried out using locally developed MATLAB software. The order of the 18 conditions was fully randomised using a Latin Square design and the order of the BEL lists was randomised as well. For each trial of the experiment, a random portion of the respective masker was picked and presented along with the target sentence, except for the tailored +MR maskers. For the periodic maskers, the order of the talkers was also randomised, ensuring that all 16 of them were picked before any of them was repeated.

Before being tested, the participants were familiarised with the materials by listening to 4 example sentences of each of the 3 target speech conditions in quiet and 1 example sentence of each of the 18 speech-in-noise conditions at an SNR of +10 dB. As in the main experiment, no feedback was given following the responses. The first BEL list was reserved for the familiarisation procedure and not used in the main experiment. The total duration of the experiment, including hearing screening and familiarisation procedure, was about 45 mins and the participants could take breaks whenever they wished to.

The experiment took place in a double-walled sound-attenuating booth. The stimuli were converted with 24-bit resolution at a sampling rate of 22.05 kHz using an RME Babyface soundcard and presented diotically over Sennheiser HD650 headphones. The level of the signal mixture was set to about 70 dB SPL over a frequency range of 70 Hz–4 kHz, as measured on an artificial ear (Brüel & Kjær, Type 4153).

C. Results and discussion

The SRTs obtained in each of the 18 processing conditions are shown in Fig. 3. The data were analysed by fitting a general linear mixed-effects regression model in a top-down manner, with *p*-values based on the Satterthwaite approximation of the degrees of freedom. Neither the main effect of target periodicity [$F(2,168.97) = 0.48$, $p = 0.62$] nor any of the fixed-effects interactions ($p \geq 0.54$) were significant. The final model thus only included the highly significant fixed effects of masker periodicity [$F(1,184.00) = 148.27$, $p < 0.001$] and masker envelope [$F(2,184.00) = 19.28$, $p < 0.001$], and participants as random effect.

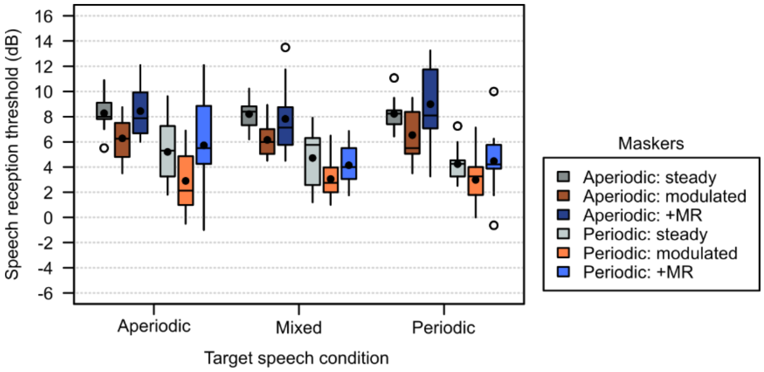

Figure 3. (Colour online) Cochlear implant simulations: speech reception thresholds. Values on the y-axis indicate the signal-to-noise ratios required to correctly repeat 50% of the keywords. The black horizontal lines in the boxplots indicate the median and the black dots the mean. The boxes range from the first to the third quartile, the whisker length is up to 1.5 times the interquartile range, and the black circles represent outliers.

The same data were re-plotted as MPBs in Fig. 4A, i.e., the SRTs of the periodic maskers were subtracted from their aperiodic counterparts, where positive values indicate that listeners benefitted from masker periodicity. In Fig. 4B, the same data are again re-plotted as FMBs, i.e., the SRTs of the modulated and +MR maskers subtracted from those of the steady maskers. Here, positive values indicate that listeners were, on average, able to benefit from 10-Hz or +MR masker envelope fluctuations. MPBs were generally larger than the FMBs and a Bonferroni-corrected post-hoc *t*-test confirmed that the SRTs for aperiodic maskers were significantly higher than for periodic ones [estimated mean difference = 3.5 dB, $t(184) = 12.18$, $p < 0.001$]. Bonferroni-corrected post-hoc *t*-tests of the SRTs also showed that there was a significant FMB for the 10-Hz modulated maskers [estimated mean difference = 1.8 dB, $t(184) = 5.19$, $p < 0.001$], but not the +MR maskers [estimated mean difference = -0.1 dB, $t(184) = -0.35$, $p = 1$].

In summary, as for the normal-hearing listeners in Steinmetzger & Rosen (2015), the amount of target periodicity had little effect on the SRTs and the MPB was larger than the FMB, even with less salient pitch cues compared to normal hearing. In addition, although they hardly overlapped with the target sentences, the +MR maskers led to similar SRTs as the steady interferers.

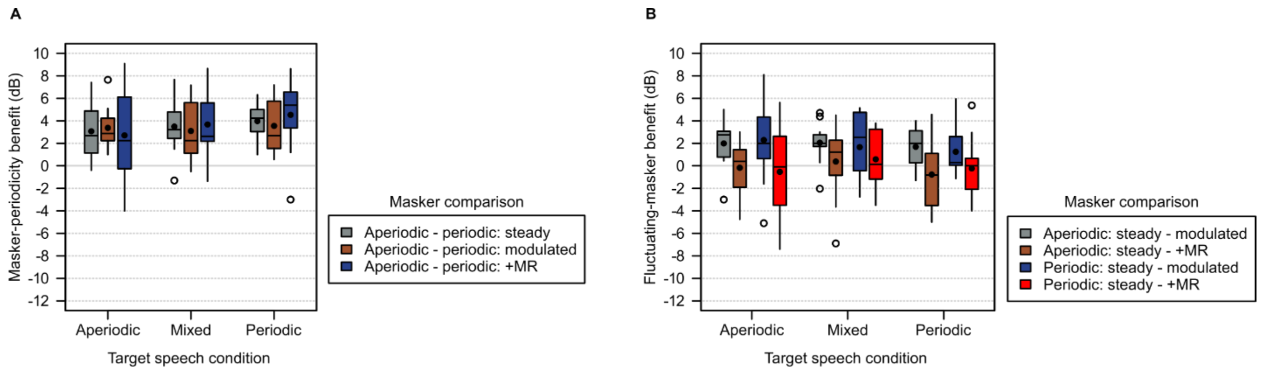

Figure 4. (Colour online) Cochlear implant simulations: masker-periodicity benefits (Panel A) and fluctuating-masker benefits (Panel B). Masker-periodicity benefits were obtained by subtracting the SRTs obtained with the periodic maskers from those obtained with the aperiodic version of the same masker. Fluctuating-masker benefits were obtained by subtracting the SRTs obtained with the 10-Hz modulated or +MR maskers from those obtained with the steady masker versions. In both panels, positive numbers on the y-axis indicate a benefit, i.e., improved performance.

To further examine the hypothesis that the better performance with periodic maskers is due to a combination of *F*0-related envelope modulations and less pronounced random envelope modulations, the front end of the mr-sEPSM speech intelligibility model (Jørgensen et al., 2013) was used to compute modulation spectrograms of the stimulus materials. These spectrograms depict the modulation power for each combination of auditory and modulation filter, after CI simulation processing and averaged across all individual files in each stimulus condition, allowing for a detailed evaluation of the differences between conditions. Firstly, this analysis revealed that there is little difference between the modulations of the three target speech conditions (Fig. 5A), in line with the behavioural results and the spectrograms shown in Fig. 2A. All three conditions have a diffuse modulation pattern, with the most energy in the lower modulation filters (2–8 Hz) crucial for speech intelligibility. The only feature that varies between the three conditions are, as expected, the *F*0-related temporal modulations in the higher modulation filters (64–256 Hz), which show a small parametric increase along with the degree of source periodicity. The masker modulation spectrograms (Fig. 5B), on the other hand, differ markedly at these high modulation rates. In auditory filters with centre frequencies higher than about 1250 Hz, all three periodic maskers show a prominent $F0$-related peak that distinguishes them from their aperiodic counterparts. Importantly, when subtracting the modulation spectrograms of the periodic maskers from that of the aperiodic ones (Fig. 5C), it also becomes apparent that the aperiodic maskers have stronger random modulations in the lower auditory filters. This difference is most pronounced when comparing the steady aperiodic and periodic interferers at modulation rates below about 64 Hz, where no other modulations are superimposed on these random fluctuations. Hence, the linear but time-varying process of amplitude-modulating a noise carrier with an envelope that also contains random modulations resulted in a signal with more pronounced random modulations, compared to when the carrier was periodic. The aperiodic maskers thus have stronger random modulations than the periodic maskers before as well as after the materials were noise-vocoded.

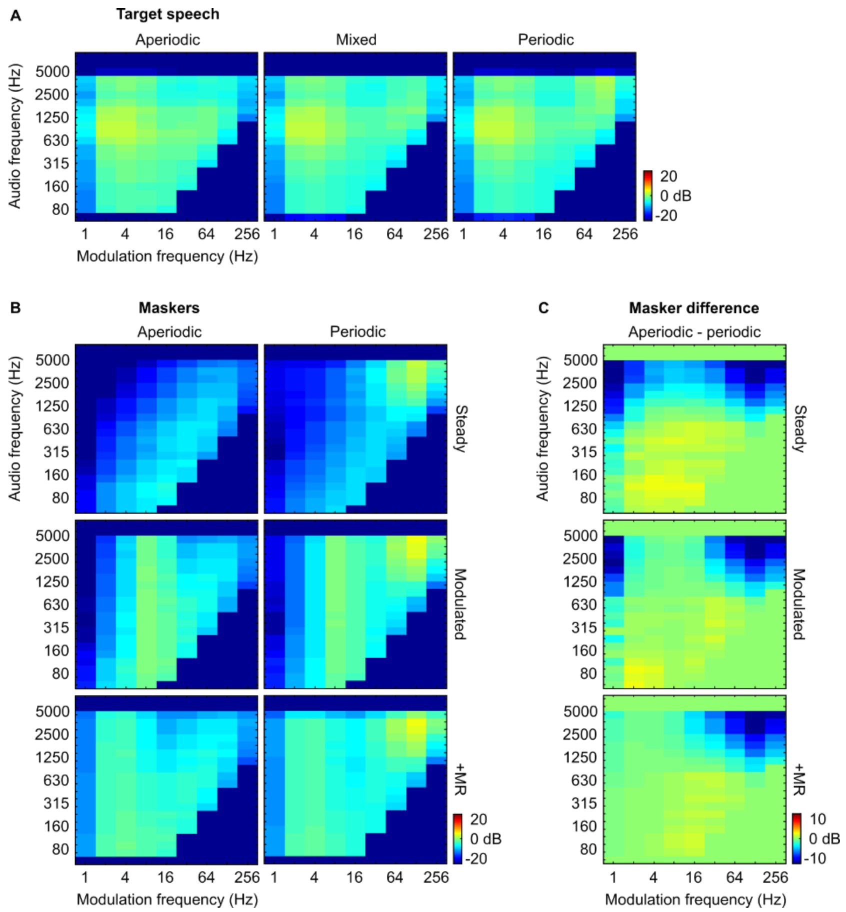

Figure 5. (Colour online) Cochlear implant simulations: stimulus modulation spectrograms. Panel A shows the average envelope modulation power of the three target speech conditions, Panel B that of the six maskers. The modulation power was computed for each combination of auditory (y-axes) and modulation filter (x-axes) using the front end of the mr-sEPSM speech intelligibility model. In Panel C, the modulation power of the periodic maskers was subtracted from that of the aperiodic ones to facilitate their comparison.

While the reduced FMBs obtained with maskers modulated at a rate of 10 Hz agree with the results of previous CI simulation studies (Cullington and Zeng, 2008; P. B. Nelson and Jin, 2004; Qin and Oxenham, 2003), it is a surprising finding that performance with the steady and +MR maskers was almost identical. In the study of Kwon et al. (2012), a masking release with the +MR maskers required the CI users to have intelligibility rates of at least 90% in quiet. Although not explicitly tested, similar performance levels can be assumed in the current experiment. For comparison, even with the much more difficult IEEE sentences, the normal-hearing listeners in

Steinmetzger and Rosen (2015; cf. Fig. 2) perceived almost 90% of the keywords correctly when tested with 8-channel noise-vocoded speech. As the +MR maskers hardly overlap with the target speech, CI simulation processing thus appears to make it particularly difficult to distinguish target speech and masker. This may in large part be because spectral and pitch cues that aid stream segregation are mostly unavailable with simulated CIs. However, it has also been shown that CI users and listeners in CI simulations have problems fusing auditory information across temporal gaps, even in the absence of a masker (P. B. Nelson and Jin, 2004). In that study, participants were presented with sentences interrupted by periods of silence and recognition performance was severely impaired across all gap frequencies, which ranged from 1 to 32 Hz. Similar results have been obtained by Ardoint et al. (2014), who tested normal-hearing listeners and found that 5-Hz interruptions affect the intelligibility of vocoded speech much more than that of unprocessed speech. Importantly, their study has also shown that this seems to be due to the lower intelligibility of uninterrupted vocoded speech *per se*, rather than acoustic properties such as its spectral resolution or the availability of pitch cues.

Additionally, in contrast to the sinusoidal amplitude modulations of the 10-Hz modulated maskers, the amplitudes of the +MR maskers fluctuate in a non-deterministic manner. More specifically, listeners were confronted with an inverted copy of the target speech envelope, which therefore also contains speech-like modulations (cf. Fig. 5). With simulated CIs, this type of slow-rate modulation masking that makes it difficult to tell target speech and masker apart appears to be particularly detrimental.

III. COCHLEAR IMPLANT USERS

A. Short introduction and rationale

The design of the current experiment is identical to the preceding one, apart from two modifications: Firstly, to make the experiment less demanding for the participants and because no effect of target periodicity was observed with simulated CIs, periodic target speech was omitted. The remaining two types of target speech (with aperiodic or mixed sources) were each combined with the same six maskers as before (aperiodic or periodic sources; steady, 10-Hz modulated, or +MR envelopes), resulting in twelve speech-in-noise conditions.

Secondly, to account for the typically large variability between CI users, SRTs were determined at an individual performance level. As in Kwon et al. (2012), half the percentage of keywords that the participant correctly perceived in quiet listening conditions was tracked adaptively. This approach required that each participant was first tested with the two target speech conditions in quiet, resulting in a total of 14 experimental conditions.

B. Methods

1. Participants

Eight CI users that were post-lingually deafened in both ears were tested. Their mean age was 67.9 yrs. The participants were required to be native speakers of British English and to have used their devices for at least two years at the time of testing. Detailed information regarding the participants is provided in Table 1.

Participant	Age	Sex	Age at onset of deafness	Years of implant use	Aetiology of deafness	Implant fitting	Implant type (Processing strategy)
1	70	M	45	2	Sensorineural	Right	CI522 (ACE)
2	69	F	53	3	Ménière’s	Right	CI422 (ACE)
3	82	F	70	3	Unknown	Right	CI422 (ACE)
4	65	F	38	9	Unknown	Left	HiRes 90K (HiRes Optima)
5	60	F	25	2	Unknown	Left	CI512 (ACE)
6	49	F	23	2	Sensorineural	Right	HiRes 90K Adv. (HiRes Optima)
7	75	F	35	3/3	Hereditary	Both	CI422 (ACE) & CI422 (ACE)
8	73	F	50	13/11	Ménière’s	Both	CI24R (ACE) & CI24RE (ACE)

Table 1. Cochlear implant users: participant information.

2. Stimuli and signal processing

Materials and signal processing were the same as in the preceding experiment, but the current one did not include periodic target speech and the signal mixture was not additionally noise-vocoded to simulate CI signal processing. Approximations of the electrical stimulation received by the CI users for each target speech condition and masker are shown in Fig. 6. These example electrodograms were computed with the Nucleus Matlab Toolbox (Version 4.31, Cochlear Limited Australia; Fuller et al., 2014), using the ACE strategy with a default frequency map and 12 maxima. In addition to showing the $F0$-related envelope modulations of the periodic stimuli at the individual electrodes, these plots also demonstrate that activation was much more scattered across electrodes for the aperiodic maskers[2].

---

[2] It should be noted that the ACE strategy was only used in six participants and that the conclusions drawn from depictions of the HighRes Optima strategy used in the remaining two participants might differ slightly.

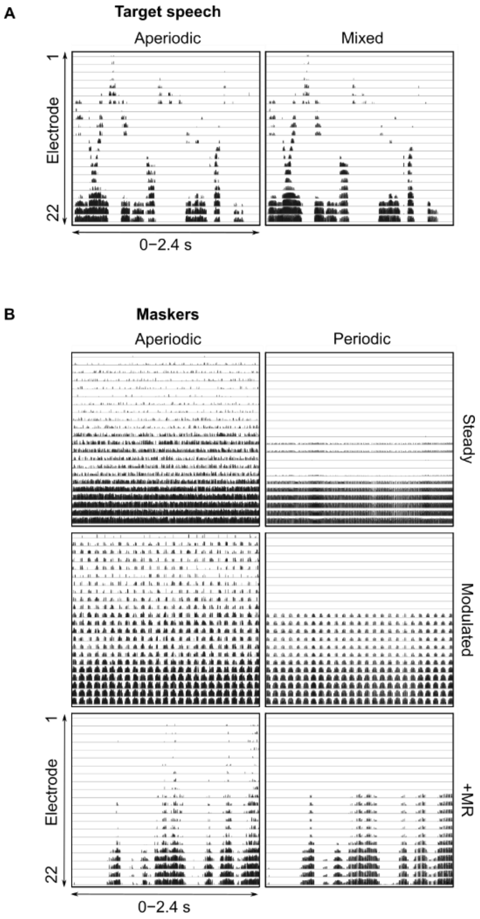

Figure 6. Cochlear implant users: stimuli. Example electrodograms showing approximations of the electrical stimulation patterns received by listeners using the ACE strategy. Panel A shows an example sentence of the two target speech conditions and Panel B shows examples of the six different maskers. The examples are the same as in the CI simulation experiment (cf. Fig. 2).

3. Procedure

The experimental procedure was largely the same as for the CI simulation experiment and details that remained unchanged are omitted here. Participants were presented with 1 complete BEL sentence list in each of the 14 conditions (2 conditions in quiet & 12 speech-in-noise conditions). SRTs for each of the speech-in-noise conditions were determined by tracking the SNR necessary to correctly repeat 50% of the keywords that the respective participant achieved in quiet listening conditions with the same target speech condition (Kwon et al., 2012). This approach was implemented by applying the weighted up-down rule (Kaernbach, 1991). Hence, for less than 100% correct keywords in quiet, the SNR was adjusted with step sizes upwards (S_{up}) that were smaller than steps downwards (S_{down}), as determined by the following formula:

$$S_{up} = S_{down} * \frac{Percentage\ to\ track}{100 - Percentage\ to\ track}\,. \qquad (1)$$

Before being tested, the participants were familiarised with the materials by listening to 5 example sentences of the 2 target speech conditions in quiet and one example sentence of each of the twelve speech-in-noise conditions at an SNR of +10 dB. The first BEL list was again reserved for the familiarisation procedure and not used in the main experiment. The total duration of the experiment, including the familiarisation procedure, was about 45 mins and participants could take breaks whenever they wished to.

The stimuli were converted with 24-bit resolution and a sampling rate of 22.05 kHz using an RME Babyface soundcard and presented over a Genelec 8030A speaker. The speaker was placed directly in front of the listener, approximately 1.5 m away and level with the participant's ears. The level of the signal mixture was set to about 69 dB SPL over a frequency range of 60 Hz–10 kHz, as measured with a sound level meter (Brüel & Kjær, Type 2231).

C. Results and discussion

1. Speech intelligibility in quiet

The data of the first experiment, where the CI users were presented with the two different target speech conditions in quiet, are shown in Fig. 7 and were analysed using a generalised linear mixed-effects logistic regression model. The model included target periodicity as fixed effect and participants as random effect. On average, the participants correctly perceived 94.6% of the keywords in the aperiodic condition and 95.4% in the mixed condition. A Wald χ^2-test indicated no significant performance difference between the two conditions [$\chi^2(1) = 0.51$, $p = 0.48$].

These results demonstrate, firstly, that a group of very high-performing CI users participated in the study. In combination with the relatively easy BEL sentence materials, this led to a ceiling effect in both experimental conditions. While this restricts the ability to conclude that there is indeed no intelligibility difference between speech with aperiodic and mixed sources in CI users, this result is in line with previous findings. Even when vocoded with few channels, so that performance was far below ceiling level, there was little difference between these two processing conditions for listeners with normal hearing (cf. Fig. 2 in Steinmetzger & Rosen, 2015).

Moreover, the primary aim of the present experiment was to assess the condition-specific performance of each individual listener, which was required as a starting point for the ensuing speech-in-noise experiment. Due to the unexpectedly high intelligibility rates in quiet, however, the individually adjusted SRT levels hardly differ from the 50%-level tracked in the CI simulations, which simplifies comparison with the CI simulation experiments.

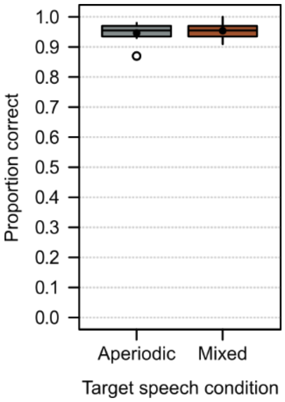

Figure 7. (Colour online) Cochlear implant users: speech intelligibility in quiet. Proportion of correctly perceived keywords in the two target speech conditions.

2. Speech intelligibility in noise

The SRTs obtained during the speech-in-noise experiment are shown in Fig. 8 and were analysed by fitting a general linear mixed-effects regression model in a top-down manner, with *p*-values based on the Satterthwaite approximation of the degrees of freedom. The final model included the significant fixed effect of masker periodicity [$F(1,81.11) = 10.64$, $p < 0.01$] as well as the non-significant and marginally non-significant fixed effects of target periodicity [$F(1,76.78) = 2.52$, $p = 0.12$] and masker envelope [$F(2,81.46) = 2.86$, $p = 0.063$], as the interaction of the latter two factors was highly significant [$F(2,81.55) = 8.64$, $p < 0.001$]. Participants and sentence lists were both included as random effects.

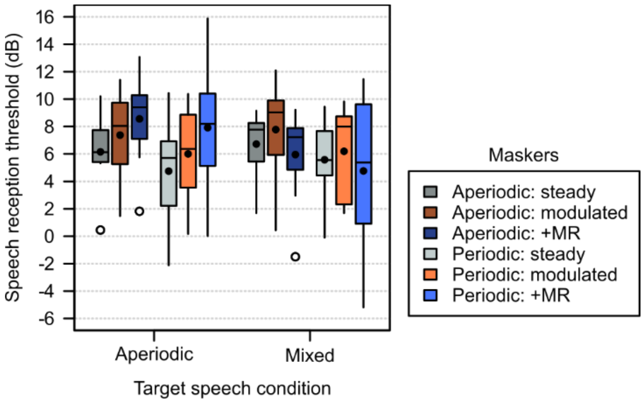

Figure 8. (Colour online) Cochlear implant users: speech reception thresholds. Values on the y-axis indicate the signal-to-noise ratios required to correctly perceive 50% of the keywords the listeners achieved in quiet. To aid comparison, the same scaling as for the results of the CI simulation experiment was used (cf. Fig. 3).

In Fig. 9A, the SRT data are again re-plotted as MPBs. Although the size of the effect was reduced in comparison to the CI simulation experiment reported above, a post-hoc *t*-test revealed that MPBs were significant, regardless of masker envelope and target periodicity [estimated mean difference = 1.2 dB, $t(81.11) = 3.26$, $p < 0.01$]. Lastly, the SRTs were re-plotted as FMBs (Fig. 9B). In contrast to the results obtained in the CI simulations, CI users performed slightly worse with the 10-Hz modulated maskers, compared to the steady ones. However, a Bonferroni-corrected post-hoc *t*-test showed that this trend did not reach significance [estimated mean difference = -0.9 dB, $t(81.9) = -1.87$, $p = 0.195$]. FMB (Bernstein and Grant, 2009; Freyman et al., 2012) as well as the MPB (Steinmetzger and Rosen, 2015) have been shown to depend on the SNR at which a test is carried out. In both cases, lower SNRs have been found to enable larger benefits. However, this cannot explain the difference between the CI simulation and CI experiments, as the SRTs in steady noise were relatively similar (~8 and ~6 dB, respectively).

Crucially, another Bonferroni-corrected post-hoc *t*-test confirmed that SRTs were significantly lower for the +MR maskers when the target speech had a mixed source excitation rather than an aperiodic one [estimated mean difference = -2.8 dB, $t(80.9) = -4.29$, $p < 0.001$], in agreement with the significant interaction of target periodicity and masker envelope. However, even with the mixed target speech condition, no masking release was observed with the +MR maskers. Hence, the results obtained with these maskers again do not agree with those reported in Kwon et al. (2012), even though all our participants apart from one achieved scores of at least 90% in quiet.

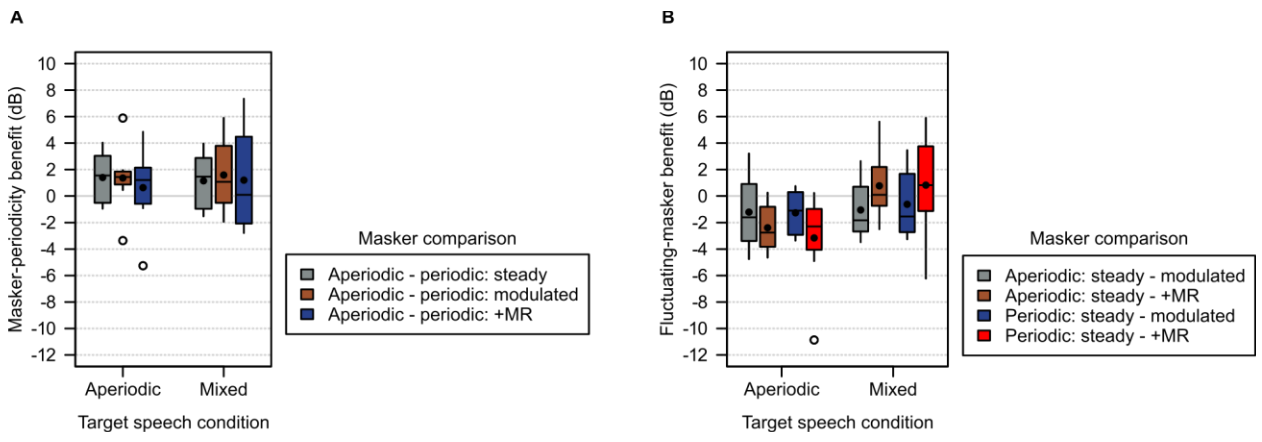

Figure 9. (Colour online) Cochlear implant users: masker-periodicity benefits (Panel A) and fluctuating-masker benefits (Panel B). To aid comparison, the same scaling as for the results of the CI simulation experiment was used (cf. Fig. 4).

In summary, as for normal hearing and simulated CIs, the presence of periodicity cues in the target speech did not affect performance. The MPB, on the other hand, was further reduced compared to the CI simulations, but CI users still significantly benefitted from masker periodicity. In contrast to the results obtained with simulated CIs, no FMB was observed with the 10-Hz modulated maskers, but a trend for deteriorated performance. Additionally, SRTs for the +MR and steady maskers were similar, as in the CI simulations, but only if the target speech had a mixed source excitation. With aperiodic target speech, on the other hand, performance was markedly worse.

IV. GENERAL DISCUSSION

A. Possible age effects

A factor that requires consideration when interpreting the current results is the large age difference between the normal-hearing listeners in the CI simulation experiment and the CI users (mean ages of ~20 and ~68 yrs, respectively). Older normal-hearing listeners without substantial hearing impairment generally have greater difficulties to understand speech in the presence of a masker than younger listeners (Füllgrabe et al., 2015; Pichora-Fuller and Souza, 2003), which has been explained by a combination of impaired auditory temporal processing and cognitive declines. However, the differences between groups are usually more pronounced with competing speech or multi-talker babble than non-speech maskers such as steady or modulated noise (Başkent et al., 2014; Schoof and Rosen, 2014), which may be due to the higher cognitive demands imposed by speech maskers. In addition, studies using vocoded stimuli have reported that the ability to use temporal envelope cues may be impaired for older listeners in CI-like listening conditions (Arehart et al., 2014; Souza and Boike, 2006), although it could also be argued that they perform worse than younger adults because they find it more difficult to adapt to the unusual sound of the vocoded materials. Nevertheless, these two studies suggest that the MPB observed in CI users might have been somewhat larger if the listeners would have been younger.

In summary, it is assumed that possible age effects in the current study should be more pronounced with the speech-like +MR maskers, for which the pattern of results indeed differed markedly across groups (discussed further in Sec. IV.D below). For the steady and 10-Hz modulated maskers, in contrast, age effects are expected to be less critical if they exist at all.

These considerations also suggest future studies which could attempt to compare age-matched participant groups or the performance of younger and older CI users. Additionally, the maskers used in the current study could be substituted for periodic and aperiodic speech maskers, to investigate to what extent informational masking effects alter the results observed in the present experiments, and how strongly the performance with speech maskers is affected by the age of the participants.

B. Masker-periodicity benefit

For normal-hearing listeners tested with simulated CIs, the MPB was markedly larger than for the CI users (3.5 vs. 1.2 dB). This raises the question whether the detrimental effects of current spread have been accurately simulated with an 8-channel noise-vocoder. As suggested by Oxenham and Kreft (2014), one crucial effect of current spread may be that random envelope modulations are smeared out when listening through a CI. They attempted to demonstrate this by using a vocoder CI simulation algorithm with a relatively high number of analysis channels (16), in which the individual channel envelopes were subsequently determined by the weighted average of the surrounding channels, to account for current spread. Their results showed that this algorithm indeed reduced the modulation power of the stimuli considerably and led to very similar performance rates of normal-hearing listeners and CI users, when attempting to understand speech in the presence of steady noise. This approach stands in contrast to commonly used vocoder simulations, such as the one used in the present study, where effects of current spread are emulated by using fewer channels in the initial analysis (4–8; e.g., Friesen et al., 2001; Fu and Nogaki, 2005; Whitmal III et al., 2007). However, these two simulation approaches – spectral smearing through envelope summation or via a filter bank – have not been compared explicitly to date and it hence remains to be seen if they differ substantially. Presumably, the MPB in the CI simulation experiment could also have been reduced to the level of the CI users by simply using filters with shallower slopes than the sixth-order Butterworth filters.

In general, studies that have investigated the ability of CI users to detect amplitude modulations via direct stimulation of individual electrodes have found a good modulation sensitivity (Fu, 2002; Shannon, 1992), suggesting that the reduced MPB is indeed due to the interaction of the stimulated electrodes and not the inability to perceive random modulations *per se*. Similarly, CI users have been shown to discriminate *F*0-related envelope modulations equally well as normal-hearing listeners (Kreft et al., 2013). While the ability to perceive temporal modulations declines sharply at frequencies above about 150 Hz (Green et al., 2004), the median *F*0 of the concatenated sentences (~110 Hz) and periodic masker materials (~123 Hz) used in the current study lies well below this upper limit. Hence, it can be assumed that these cues were available to the CI users, as well as with simulated CIs. The pitch cues conveyed by the temporal envelopes of the periodic maskers are thus assumed to be the reason for the MPB observed in CI users.

The stimulus electrodograms in Fig. 6 might suggest that an alternative explanation for the MPB observed in CI users is that electrical activity for the aperiodic maskers is simply more scattered across electrodes, thereby making them more effective maskers. However, although this scattering is much less pronounced for the aperiodic +MR masker, the size of the MPB was similar for all three types of masker envelopes, confirming that *F*0-related temporal modulations are the crucial factor.

It is also worth noting that the listeners in the CI simulation experiment showed a greater MPB than the CI users despite the use of a noise-excited vocoder simulation. The inherent random modulations of a noise carrier are known to make it more difficult to detect a target modulation (Dau et al., 1997) and in line with this, CI simulations using tone-vocoders (Whitmal III et al., 2007) and pulse-spreading harmonic complexes (Mesnildrey et al., 2016) have reported better speech perception in the presence of a masker. Accordingly, using these types of carriers would likely resulted in an even larger MPB. Nevertheless, the present study has demonstrated that, when using a noise-vocoder CI simulation, the random modulations of the noise carrier and the random modulations contained in the signal envelope to some extent add up (cf. Fig. 5C), preserving the difference between the modulation spectra of the original aperiodic and periodic maskers.

Compared to the normal-hearing listeners in Steinmetzger and Rosen (2015), the total size of the MPB was markedly reduced in the current CI simulation and CI experiments (~8.5 to 3.5/1.2 dB; cf. Fig. 6 in Steinmetzger and Rosen, 2015). However, when the higher SRTs in steady noise that were measured in the current study are considered and the results are compared at a similar SNR level (+7 dB), the MPB in the previous study amounts to about 4.5 dB only (This value was extracted from the estimated psychometric functions; cf. lower row of Fig. 8 in Steinmetzger and Rosen, 2015). This further supports the notion that the absence of random modulation in the periodic maskers is the crucial factor explaining the MPB, at least at positive SNR levels. Even in normal hearing, pitch-related effects such as streaming appear to be far less important.

C. Fluctuating-masker benefit with 10-Hz modulated maskers

In line with earlier findings (e.g., Cullington and Zeng, 2008; Fu and Nogaki, 2005; Stickney et al., 2004), the masking release obtained from slow-rate modulations of the masker was limited with simulated CIs (1.8 dB) and even turned negative in CI users (-0.9 dB). As for the MPB, the difference between listener groups can be explained by the apparent inability of the CI users to perceive random envelope modulations, resulting from the interaction of the CI electrodes (Oxenham and Kreft, 2014). While the superimposed 10-Hz modulations led to a release from the modulation masking caused by these random fluctuations in the CI simulation experiment, the same does not apply to the CI users. As can be seen in the modulation spectrograms in Fig. 5, the sinusoidal 10-Hz masker modulations coincide with the slow envelope modulations of the target speech and hence pose an additional source of modulation masking, resulting in slightly higher SRTs in the CI experiment. Similarly, Fu and Nogaki (2005) found that performance in gated noise with simulated CIs became more similar to that of CI users when the degree of spectral smearing in the noise-vocoder simulation was increased. Akin to the simulation algorithm used by Oxenham and Kreft (2014), where the weighted mean of the surrounding channels determined the individual channel envelopes, using filters with very shallow roll-offs resulted in an effective flattening of the channel envelopes.

Compared to the data from Steinmetzger and Rosen (2015), the total size of the FMB was also markedly reduced in the current CI simulation and CI experiments (~4 to 1.8/-0.9 dB; cf. Fig. 5 in Steinmetzger and Rosen, 2015). In contrast, a comparison at the same SNR of +7 dB here revealed a strongly negative FMB of about -4 dB in normal-hearing listeners. As their performance was already close to ceiling level at this high SNR when the maskers were steady, this suggests that the detrimental effect of the additional modulation masking caused by the 10-Hz fluctuations of the maskers was particularly strong.

D. Interaction of +MR maskers and target periodicity in CI users

The performance of the CI users with the +MR maskers worsened markedly (by 2.8 dB SRT) if the target speech had an aperiodic rather than a mixed source excitation, while there was no such effect with simulated CIs. Even taking into account the earlier results obtained in normal hearing (Steinmetzger and Rosen, 2015), this constitutes the most distinct effect associated with periodicity cues in the target speech. As they are the only acoustic feature distinguishing the two target speech conditions, this effect clearly demonstrates that the CI users are sensitive to $F0$-related envelope modulations.

Firstly, due to the speech-like envelopes of the +MR maskers, $F0$ cues in the target speech might be particularly helpful when attempting to distinguish it from this type of masker. Moreover, if the degree of spectral smearing was indeed underestimated by the 8-channel noise-vocoder CI simulation, the greater current spread in real CIs may have emphasised these $F0$ cues (Geurts and Wouters, 2001). This might be one reason for the large performance difference with the two target speech conditions for CI users.

Secondly, and perhaps more importantly, it has been shown (Bhargava et al., 2016) that similar intelligibility levels of interrupted speech with simulated and actual CIs require the age as well as the performance with uninterrupted speech to be matched across groups, possibly because age-related declines affect the ability of older listeners to integrate the individual speech segments. As the +MR maskers act to interrupt the target speech too, the poor performance of the CI users in the absence of $F0$ cues in the target speech may thus be caused by the age difference between listener groups in the present study. However, the more general finding that the +MR maskers did not enable any masking release still holds, irrespective of this possible age effect.

V. SUMMARY AND CONCLUSIONS

The present study has shown that CI users can exploit temporal pitch cues conveyed by the envelope of a periodic non-speech masker when attempting to segregate target speech from interferer, whereas no similar effect with respect to periodicity cues in the target speech was observed. Compared to previous results obtained with normal-hearing listeners, the overall size of this *masker-periodicity benefit* (MPB) was smaller with simulated CIs (~8.5 to 3.5 dB) and further reduced with real CIs (1.2 dB). However, when compared at the higher signal-to-noise ratios (SNRs) measured in the current study, the MPB for normal-hearing listeners amounts to about 4.5 dB only and the differences are less pronounced.

In contrast, the CI users neither showed a benefit when the maskers were amplitude-modulated at a rate of 10 Hz nor when the masker envelopes were tailored to reveal the target sentence, which was intended to promote a masking release. Moreover, the listeners in the corresponding CI simulation experiment similarly did not perform better with the latter type of interferer, although they did show a fluctuating-masker benefit (FMB) of 1.8 dB with the 10-Hz modulated maskers.

In summary, these results demonstrate that CI users can exploit the temporal pitch cues conveyed by a masker when attempting to understand speech in noise, while they fail to benefit from slow-rate masker envelope modulations. Despite being much older than the listeners in the CI simulations, the smaller MPBs and FMBs in CI users can best be explained by the inability of present CI devices to transmit random envelope modulations. Firstly, this effect reduces the contrast between aperiodic and periodic sounds, and secondly, it diminishes the release from modulation masking that is the main reason for the FMB. Consequently, the noise-vocoder CI simulation algorithm used in the current study likely underestimated the current spread in real CIs.

ACKNOWLEDGEMENTS

We thank Tim Green for providing his recordings of the BEL sentences and for helping with the recruitment of the CI users, Etienne Gaudrain for computing the electrodograms, and Gaston Hilkhuysen for helpful comments. This project has been funded with support from the European Commission under Contract FP7-PEOPLE-2011-290000, the Medical Research Council of the UK (Grant Number G1001255), and the Dietmar Hopp Stiftung (Grant Number 2301 1239).

Ardoint, M., Green, T., and Rosen, S. (**2014**). "The intelligibility of interrupted speech depends upon its uninterrupted intelligibility," J. Acoust. Soc. Am. **136**, EL275–EL280.

Arehart, K. H., Croghan, N. B., and Muralimanohar, R. K. (**2014**). "Effects of age on melody and timbre perception in simulations of electro-acoustic and cochlear-implant hearing," Ear Hear. **35**, 195–202.

Başkent, D., van Engelshoven, S., and Galvin III, J. J. (**2014**). "Susceptibility to interference by music and speech maskers in middle-aged adults," J. Acoust. Soc. Am. **135**, EL147–EL153.

Bernstein, J. G. W., 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.

Bhargava, P., Gaudrain, E., and Başkent, D. (**2016**). "The intelligibility of interrupted speech: Cochlear implant users and normal hearing listeners," J. Assoc. Res. Otolaryngol. **17**, 475–491.

Calandruccio, L., and Smiljanic, R. (**2012**). "New sentence recognition materials developed using a basic non-native English lexicon," J. Speech. Lang. Hear. Res. **55**, 1342–1355.

Chan, D., Fourcin, A., Gibbon, D., Granström, B., Huckvale, M., Kokkinas, G., Kvale, L., Lamel, L., Lindberg, L., and Moreno, A. (**1995**). "EUROM – A spoken language resource for the EU," *Proceedings of Eurospeech*, pp. 867–880.

Chatterjee, M., and Peng, S.-C. (**2008**). "Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition," Hearing Res. **235**, 143–156.

Cullington, H. E., and Zeng, F.-G. (**2008**). "Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects," J. Acoust. Soc. Am. **123**, 450–461.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.

de Cheveigné, A., McAdams, S., Laroche, J., and Rosenberg, M. (**1995**). "Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement," J. Acoust. Soc. Am. **97**, 3736–3748.

de Cheveigné, A., McAdams, S., and Marin, C. M. (**1997**). "Concurrent vowel identification. II. Effects of phase, harmonicity, and task," J. Acoust. Soc. Am. **101**, 2848–2856.

De Jong, N. H., and Wempe, T. (**2009**). "Praat script to detect syllable nuclei and measure speech rate automatically," Behav. Res. Methods **41**, 385–390.

Deeks, J. M., and Carlyon, R. P. (**2004**). "Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation," J. Acoust. Soc. Am. **115**, 1736–1746.

Deroche, M. L., Culling, J. F., Chatterjee, M., and Limb, C. J. (**2014a**). "Roles of the target and masker fundamental frequencies in voice segregation," J. Acoust. Soc. Am. **136**, 1225–1236.

Deroche, M. L., Culling, J. F., Chatterjee, M., and Limb, C. J. (**2014b**). "Speech recognition against harmonic and inharmonic complexes: Spectral dips and periodicity," J. Acoust. Soc. Am. **135**, 2873–2884.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.

Elliott, T. M., and Theunissen, F. E. (**2009**). "The modulation transfer function for speech intelligibility," PLoS Comput. Biol. **5**, e1000302.

Fant, G., Liljencrants, J., and Lin, Q.-G. (**1985**). "A four-parameter model of glottal flow," STL-QPSR **4**, 1–13.

Freyman, R. L., Griffin, A. M., and Oxenham, A. J. (**2012**). "Intelligibility of whispered speech in stationary and modulated noise maskers," J. Acoust. Soc. Am. **132**, 2514–2523.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q.-J. (**2002**). "Temporal processing and speech recognition in cochlear implant users," Neuroreport **13**, 1635–1639.

Fu, Q.-J., Chinchilla, S., Nogaki, G., and Galvin III, J. J. (**2005**). "Voice gender identification by cochlear implant users: The role of spectral and temporal resolution," J. Acoust. Soc. Am. **118**, 1711–1718.

Fu, Q.-J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Fu, Q.-J., Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.

Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q.-J., Free, R. H., and Başkent, D. (**2014**). "Gender categorization is abnormal in cochlear implant users," J. Assoc. Res. Otolaryngol. **15**, 1037–1048.

Füllgrabe, C., Moore, B. C., and Stone, M. A. (**2015**). "Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition," Front. Aging Neurosci. **6**, 347.

Gaudrain, E., and Başkent, D. (**2018**). "Discrimination of voice pitch and vocal-tract length in cochlear implant users," Ear Hear. **39**, 226–237.

Gaudrain, E., Grimault, N., Healy, E. W., and Béra, J.-C. (**2008**). "Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation," J. Acoust. Soc. Am. **124**, 3076–3087.

Geurts, L., and Wouters, J. (**2001**). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," J. Acoust. Soc. Am. **109**, 713–726.

Green, T., Faulkner, A., and Rosen, S. (**2004**). "Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants," J. Acoust. Soc. Am. **116**, 2298–2310.

Green, T., Faulkner, A., Rosen, S., and Macherey, O. (**2005**). "Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification," J. Acoust. Soc. Am. **118**, 375–385.

Green, T., and Rosen, S. (**2013**). "Phase effects on the masking of speech by harmonic complexes: Variations with level," J. Acoust. Soc. Am. **134**, 2876–2883.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species – 29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

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.

Kaernbach, C. (**1991**). "Simple adaptive testing with the weighted up-down method," P&P **49**, 227–229.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., and Banno, H. (**2008**). "TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation," *Proceedings of in Proceedings of International Conference on Acoustics, Speech and Signal Processing*, pp. 3933–3936.

Kreft, H. A., Nelson, D. A., and Oxenham, A. J. (**2013**). "Modulation frequency discrimination with modulated and unmodulated interference in normal hearing and in cochlear-implant users," J. Assoc. Res. Otolaryngol. **14**, 591–601.

Kwon, B. J., Perry, T. T., Wilhelm, C. L., and Healy, E. W. (**2012**). "Sentence recognition in noise promoting or suppressing masking release by normal-hearing and cochlear-implant listeners," J. Acoust. Soc. Am. **131**, 3111–3119.

Leclère, T., Lavandier, M., and Deroche, M. L. (**2017**). "The intelligibility of speech in a harmonic masker varying in fundamental frequency contour, broadband temporal envelope, and spatial location," Hearing Res. **350**, 1–10.

Macherey, O., and Carlyon, R. P. (**2014**). "Cochlear implants," Curr. Biol. **24**, R878–R884.

Meister, H., Fürsen, K., Streicher, B., Lang-Roth, R., and Walger, M. (**2016**). "The use of voice cues for speaker gender recognition in cochlear implant recipients," J. Speech. Lang. Hear. Res. **59**, 546–556.

Meister, H., Landwehr, M., Pyschny, V., Walger, M., and Wedel, H. v. (**2009**). "The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients," Int. J. Audiol. **48**, 38–48.

Mesnildrey, Q., Hilkhuysen, G., and Macherey, O. (**2016**). "Pulse-spreading harmonic complex as an alternative carrier for vocoder simulations of cochlear implants," J. Acoust. Soc. Am. **139**, 986–991.

Moore, B. C. (**2008**). "The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people," J. Assoc. Res. Otolaryngol. **9**, 399–406.

Nelson, D. A., and Donaldson, G. S. (**2001**). "Psychophysical recovery from single-pulse forward masking in electric hearing," J. Acoust. Soc. Am. **109**, 2921–2933.

Nelson, P. B., and Jin, S.-H. (**2004**). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2286–2294.

Nelson, P. B., Jin, S.-H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.

Oxenham, A. J. (**2008**). "Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants," Trends Amplif. **12**, 316–331.

Oxenham, A. J., and Kreft, H. A. (**2014**). "Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing," Trends Hear. **18**, 2331216514553783.

Pichora-Fuller, M. K., and Souza, P. E. (**2003**). "Effects of aging on auditory processing of speech," Int. J. Audiol. **42**, 11–16.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Rosen, S. (**1992**). "Temporal information in speech: acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. Lond. B Biol. Sci. **336**, 367–373.

Schoof, T., and Rosen, S. (**2014**). "The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners," Front. Aging Neurosci. **6**, 307.

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

Souza, P. E., and Boike, K. T. (**2006**). "Combining temporal-envelope cues across channels: Effects of age and hearing loss," J. Speech. Lang. Hear. Res. **49**, 138–149.

Steinmetzger, K., and Rosen, S. (**2015**). "The role of periodicity in perceiving speech in quiet and in background noise," J. Acoust. Soc. Am. **138**, 3586–3599.

Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F.-G. (**2007**). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentencesa)," J. Acoust. Soc. Am. **122**, 1069–1078.

Stickney, G. S., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Stone, M. A., Füllgrabe, C., Mackinnon, R. C., and Moore, B. C. (**2011**). "The importance for speech intelligibility of random fluctuations in “steady” background noise," J. Acoust. Soc. Am. **130**, 2874–2881.

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

Whitmal III, 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.

Wilson, B. S., and Dorman, M. F. (**2008**). "Cochlear implants: a remarkable past and a brilliant future," Hearing Res. **242**, 3–21.

Xu, Y. (**2013**). "ProsodyPro – A tool for large-scale systematic prosody analysis," *Proceedings of in Tools and Resources for the Analysis of Speech Prosody*, pp. 7–10.