# Effects of Additional Low-Pass–Filtered Speech on Listening Effort for Noise-Band–Vocoded Speech in Quiet and in Noise

Carina Pals,[1] Anastasios Sarampalis,[2] Mart van Dijk,[3] and Deniz Başkent[4,5]

**Objectives:** Residual acoustic hearing in electric–acoustic stimulation (EAS) can benefit cochlear implant (CI) users in increased sound quality, speech intelligibility, and improved tolerance to noise. The goal of this study was to investigate whether the low-pass–filtered acoustic speech in simulated EAS can provide the additional benefit of reducing listening effort for the spectrotemporally degraded signal of noise-band–vocoded speech.

**Design:** Listening effort was investigated using a dual-task paradigm as a behavioral measure, and the NASA Task Load indeX as a subjective self-report measure. The primary task of the dual-task paradigm was identification of sentences presented in three experiments at three fixed intelligibility levels: at near-ceiling, 50%, and 79% intelligibility, achieved by manipulating the presence and level of speech-shaped noise in the background. Listening effort for the primary intelligibility task was reflected in the performance on the secondary, visual response time task. Experimental speech processing conditions included monaural or binaural vocoder, with added low-pass–filtered speech (to simulate EAS) or without (to simulate CI).

**Results:** In Experiment 1, in quiet with intelligibility near-ceiling, additional low-pass–filtered speech reduced listening effort compared with binaural vocoder, in line with our expectations, although not compared with monaural vocoder. In Experiments 2 and 3, for speech in noise, added low-pass–filtered speech allowed the desired intelligibility levels to be reached at less favorable speech-to-noise ratios, as expected. It is interesting that this came without the cost of increased listening effort usually associated with poor speech-to-noise ratios; at 50% intelligibility, even a reduction in listening effort on top of the increased tolerance to noise was observed. The NASA Task Load indeX did not capture these differences.

**Conclusions:** The dual-task results provide partial evidence for a potential decrease in listening effort as a result of adding low-frequency acoustic speech to noise-band–vocoded speech. Whether these findings translate to CI users with residual acoustic hearing will need to be addressed in future research because the quality and frequency range of low-frequency acoustic sound available to listeners with hearing loss may differ from our idealized simulations, and additional factors, such as advanced age and varying etiology, may also play a role.

[1]Department of Psychology, University of Utah Asia Campus, Incheon, South Korea, Part of: University of Utah, Salt Lake City, UT, USA; [2]Department of Experimental Psychology, University of Groningen, Groningen, The Netherlands; [3]Department of Work & Social Psychology, Maastricht University, Maastricht, the Netherlands; [4]Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; and [5]University of Groningen, Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, Groningen, The Netherlands.

## INTRODUCTION

A cochlear implant (CI)–mediated speech signal is degraded in acoustic–phonetic details, in both spectral and temporal dimensions, compared with normal hearing. This is due to factors related to the device, the electrode–nerve interface, and the state of the impaired auditory system (for a review, see Başkent et al. 2016). Interpreting a degraded speech signal requires increased top–down cognitive processing (Classon et al. 2013; Gatehouse 1990; Pichora-Fuller et al. 1995; Wingfield 1996). According to the Ease of Language Understanding model, the missing or incomplete segments of the input speech stream cannot be automatically matched to existing phonologic and lexical representations in long-term memory. To fill in the missing information or to infer meaning, a loop of explicit cognitive processing is triggered (Rönnberg 2003; Rönnberg et al. 2013; Rönnberg et al. 2008). This explicit processing increases the cognitive load of speech understanding, referred to as "listening effort." It stands to reason, then, that interpreting the degraded speech heard through a CI may thus be effortful for the listener, and processing strategies or device configurations that improve implant speech signal quality may reduce listening effort for CI users (also see Downs 1982 for a similar argument for hearing impairment and hearing aids). Studies using noise-band vocoders as an acoustic CI simulation suggest that listening effort does indeed increase for the perception of spectrotemporally degraded speech compared with clear speech (Wagner et al. 2016; Wild et al. 2012), and listening effort decreases with increasing spectral resolution (Pals et al. 2013; Winn et al. 2015). The device configuration known as electric–acoustic stimulation (EAS), that is, the combination of a CI with (residual) low-frequency acoustic hearing in either the implanted or the contralateral ear (amplified if necessary), may similarly improve signal quality, potentially reducing listening effort.

Research on the effects of EAS has consistently shown benefits in speech intelligibility, particularly for speech in noise (e.g., Büchner et al. 2009; Dorman & Gifford 2010; Zhang et al. 2010a), as well as improved subjective hearing device benefit (Gstoettner et al. 2008), and improved subjective sound quality (Kiefer et al. 2005; Turner et al. 2005; von Ilberg et al. 1999). The frequency range of residual hearing in CI users is often limited, and the acoustic speech signal alone, without the CI, is not very intelligible (Dorman & Gifford 2010). However, the low-frequency sound does carry additional acoustic speech cues that are not well transmitted through CIs, such as voice pitch, consonant voicing, or lexical boundaries (Başkent et al., Reference

<zdoi; 10.1097/AUD.0000000000000587>

Note 1; Brown & Bacon 2009). Perhaps due to this complementary structure, CI users with residual hearing show significantly improved speech understanding in noise when provided with even as little as 300 Hz low-pass–filtered speech (Büchner et al. 2009; Zhang et al. 2010b), and similar results are observed in normal-hearing listeners with noise-band–vocoded speech in background noise (Dorman et al. 2005; Kong & Carlyon 2007; Qin & Oxenham 2006). The Ease of Language Understanding model would predict that the speech cues available in the low-frequency acoustic signal will improve the match with existing phonetic representations in long-term memory, reducing the need for explicit cognitive processing (Rönnberg 2003; Rönnberg et al. 2013), thus reducing listening effort. In the present study, we therefore, hypothesized that low-frequency acoustic sound in addition to spectrotemporally degraded speech, such as CI mediated speech, can reduce listening effort and free up cognitive resources for concurrent tasks.

**Present Study**

This study systematically investigated how low-pass–filtered speech, provided to complement spectrotemporally degraded, noise-band vocoded speech, affects listening effort for normal-hearing listeners, both in quiet and in noise. The study of listening effort in a clinical context is relatively new, and few studies have addressed factors specific to CI hearing (e.g., Hughes & Galvin 2013; Pals et al. 2013; Steel et al. 2015; Wagner et al. 2016; Winn et al. 2015). Therefore, for a comprehensive investigation, we have included a number of different experimental conditions simulating a wide range of CI-like configurations: noise-band–vocoded speech presented monaurally (simulating monaural CI), noise-band–vocoded speech presented binaurally (simulating bilateral CIs), and noise-band–vocoded speech presented to one ear complemented by low-pass–filtered speech, with cutoff frequencies of either 300 or 600 Hz, presented to the contralateral ear (simulating EAS). A second dimension investigated was the spectral resolution of the noise-band–vocoder signal: each of the four configurations was presented using either six-channel or eight-channel noise-band–vocoded speech.

The specific experimental conditions were chosen based on previous work. Speech understanding of noise-band–vocoded speech has been shown to improve with increasing spectral resolution and result in near-ceiling speech understanding in normal-hearing participants from around six spectral channels onward (Friesen et al. 2001; Pals et al. 2013), while listening effort continues to improve further, at least up to eight spectral channels (Pals et al. 2013) or beyond eight channels, up to 16 or 32 channels (Winn et al. 2015). Similarly, while adding 300 Hz low-pass–filtered speech to spectrotemporally degraded speech significantly improved intelligibility in noise as well as noise tolerance in both CI users (Brown & Bacon 2009) and normal-hearing listeners (Qin & Oxenham 2006), 600 Hz low-pass–filtered speech provided little further improvement in speech intelligibility or noise tolerance (Brown & Bacon 2009; Qin & Oxenham 2006). On the other hand, little is known about the potential benefits of increasing the bandwidth of added low-pass–filtered speech beyond 300 Hz in terms of listening effort. Experimental parameters in this study, therefore, included 300 and 600 Hz low pass–filtered speech, as well as six- and eight-channel noise-band–vocoder stimuli. Prior research has shown lower self-reported listening effort for bilateral CI than for CI combined with a contralateral hearing aid (Noble et al. 2008).

We, therefore, chose to include a bilateral CI condition as an extra control condition: to distinguish between effects of contralateral low-frequency acoustic speech in addition to vocoded speech on listening effort and effects of binaural hearing, that is, binaural vocoder, compared with monaural hearing. Benefits of EAS in intelligibility and noise tolerance have been previously documented (Büchner et al. 2009; Kong & Carlyon 2007; Zhang et al. 2010b), and we are, therefore, specifically interested in additional effects of low-frequency acoustic speech on listening effort. In this study, the auditory stimuli for the different experimental conditions were, therefore, presented at equal levels of intelligibility, so that changes in listening effort can be observed independently of intelligibility.

Listening effort was quantified using a dual-task paradigm that combines a speech intelligibility task with a secondary visual response time (RT) task. If low-pass–filtered speech in addition to noise-band–vocoded speech reduces listening effort and, therefore, frees up cognitive resources for the secondary task, this should result in shorter RTs on the secondary task (Kahneman 1973). A recent review suggested that, although the specific dual-task designs used differ from study to study, in general, the dual-task paradigm is a successful method for quantifying listening effort (Gagné et al. 2017). Previous research using a dual-task paradigm similar to the one used in this study has shown that changes in signal quality, such as increased spectral resolution (Pals et al. 2013) or noise reduction (Sarampalis et al. 2009), can result in decreased listening effort even when no change in intelligibility is observed. As in our previous study (Pals et al. 2013), we included the NASA Task Load indeX (NASA-TLX; Hart & Staveland 1988) as a subjective self-report measure for listening effort. If a self-report measure could capture the same effects as an objective measure of listening effort, this would be a powerful tool quantifying listening effort in diverse settings. Studies using both objective and subjective measures of listening effort, however, often find different patterns of results for the two types of measures (Feuerstein 1992; Fraser et al. 2010; Gosselin & Gagné 2011b; Pals et al. 2013; Zekveld et al. 2010).

On the basis of the observations from previous research summarized earlier, we propose the following specific hypotheses: (1) for near-ceiling speech intelligibility, higher spectral resolution, as manipulated by number of vocoder channels, will result in faster RTs on the secondary task of the dual-task paradigm; (2a) the presence of low-frequency acoustic speech will result in faster dual-task RTs; (2b) if the improvement in listening effort (dual-task RTs) is indeed due to the low-frequency acoustic sound and not an effect of binaural hearing, then the vocoded speech combined with contralaterally presented low-frequency acoustic speech should result in faster dual-task RTs than binaurally presented vocoded speech; (3) increasing the low-frequency acoustic signal from 300 to 600 Hz low-pass–filtered speech will result in faster dual-task RTs; (4) we expect to see differences in subjective listening effort, (i.e., NASA-TLX scores) between different intelligibility levels, however, not within intelligibility levels. We test these hypotheses in three experiments, in which speech intelligibility was fixed at three different levels: Experiment 1 for speech in quiet at near-perfect intelligibility (similar to Pals et al. 2013), and Experiments 2 and 3 for noise-masked speech at 50% and 79% intelligibility, respectively, to investigate effects on listening effort at different parts of the psychometric function.

# EXPERIMENT 1: SPEECH IN QUIET AT NEAR-CEILING INTELLIGIBILITY

## Motivation

In Experiment 1, we examined how the addition of low-frequency acoustic speech affects listening effort for the understanding of noise-vocoded speech without background noise and with intelligibility near-ceiling. When intelligibility is near-ceiling, there is little room for further improvement in intelligibility; however, we hypothesized that the additional low-frequency acoustic speech will still serve to reduce listening effort independently of changes in intelligibility.

## Methods

**Participants** • Twenty normal-hearing, native Dutch-speaking, young adults (age range, 18 to 21 years; mean, 19 years; 5 female, 15 male) participated in this experiment. Participants were recruited via posters at university facilities and were screened for normal-hearing thresholds of 20 dB HL or better at audiometric frequencies between 250 and 6000 Hz, measured in both ears. Dyslexia or other language or learning disabilities were exclusion criteria in this and subsequent experiments.

We provided written information about the experiment to all participants, explained the procedure in person during the laboratory visit, and gave the opportunity to ask questions before signing the informed consent form. Participants received a financial reimbursement of €8 per hr, plus traveling expenses, for their time and effort. The local ethics committee approved the procedures for this and the subsequent experiments.

**Speech Task and Stimuli** • The primary intelligibility task was to listen to processed Dutch sentences presented in quiet and to repeat each sentence as accurately as possible. The sentence onsets were 8 sec apart. The average duration of sentences was about 1.8 sec, leaving about 6.2 sec available for the verbal response. The verbal responses were recorded for offline scoring by a native Dutch speaker. Speech intelligibility was scored based on the percentage of full sentences repeated entirely correctly.

The sentences used for the primary intelligibility task were taken from the Vrije Universiteit (VU) corpus (Versfeld et al. 2000), which consists of conversational, meaningful, and unambiguous Dutch sentences, rich in semantic context, each eight to nine syllables long. The corpus is organized into 78 unique lists of 13 sentences, half recorded with a female speaker and half with a male speaker. The lists are balanced such that the phoneme distribution of each list approximates the mean phoneme distribution of the full corpus, and each sentence is of approximately equal intelligibility in noise (Versfeld et al. 2000). In this experiment, we used the 39 lists spoken by the female speaker, the last six of these lists were used for training and a random selection of the remaining lists was used in each experiment, such that each sentence was presented no more than once to each participant.

In Experiment 1, three different device configurations (monaural CI, bilateral CIs, and monaural CI + contralateral low-frequency acoustic hearing) were approximated and compared in a total of eight different experimental conditions. Both six-channel and eight-channel noise-band–vocoded speech were used to create two versions of four different listening modes: monaural vocoded speech, binaural vocoded speech, and monaural vocoded speech with contralaterally presented low-pass–filtered speech at 300 or 600 Hz. See Table 1 for an overview of all the experimental conditions.

The noise-band vocoder used was implemented in MAT-LAB as follows (Dudley 1939; Shannon et al. 1995). The original audio recordings of the sentences were filtered into six or eight spectral bands (analysis bands) between 80 and 6000 Hz using sixth-order Butterworth band-pass filters with cutoff frequencies that simulate frequency bands of equal cochlear distance (Greenwood 1990). The carrier bands (synthesis bands) were generated with white noise band-pass filtered using the same filters. The carrier bands were then modulated using the envelopes of the analysis bands, extracted with half-wave rectification and third-order low-pass Butterworth filter with −3 dB cutoff frequency of 160 Hz. The modulated carrier noise bands were postfiltered, again using the same band-pass filters, and combined to form the final noise-band–vocoder CI-simulated speech signal.

The low-frequency acoustic speech was obtained by low-pass filtering at 300 and 600 Hz, values similar to earlier EAS simulation studies (Başkent 2012; Qin & Oxenham 2006; Zhang et al. 2010b), using sixth-order Butterworth low-pass filters (Qin & Oxenham 2006). Because sixth-order Butterworth filters have a 36 dB per octave roll-off, and the low-frequency sound is paired with noise-band–vocoded speech in the conditions of interest in this study, we believe that what low-frequency sound would still be audible in the higher frequencies in quiet will be masked by the noise-band–vocoded speech and therefore rendered useless. See Başkent and Chatterjee (2010) for spectra of stimuli including low-pass–filtered speech with an 18 dB per octave roll-off combined with noise-band–vocoded speech. Even with the 18 dB per octave roll-off, the overlap appears minimal. The

**TABLE 1. Summary of the experimental conditions for Experiment 1**

| No. | Listening Mode | Spectral Resolution | Left Ear (Level) | Label | Right Ear (Level) | Label |
|-----|----------------|---------------------|------------------|-------|-------------------|-------|
| 1 | Monaural vocoder | 6-channel vocoder | | | 6-channel vocoder (65 dBA) | Voc6 |
| 2 | | 8-channel vocoder | | | 8-channel vocoder (65 dBA) | Voc8 |
| 3 | Binaural vocoder | 6-channel vocoder | 6-channel vocoder (60 dBA) | Voc6 | 6-channel vocoder (60 dBA) | Voc6 |
| 4 | | 8-channel vocoder | 8-channel Vocoder (60 dBA) | Voc8 | 8-channel vocoder (60 dBA) | Voc 8 |
| 5 | Monaural vocoder + 300 Hz LPF | 6-channel vocoder | 300 Hz LPF (60 dBA) | LPF300 | 6-channel vocoder (60 dBA) | Voc6 |
| 6 | speech | 8-channel vocoder | 300 Hz LPF (60 dBA) | LPF300 | 8-channel vocoder (60 dBA) | Voc8 |
| 7 | Monaural vocoder + 600 Hz LPF | 6-channel vocoder | 600 Hz LPF (60 dBA) | LPF600 | 6- channel vocoder (60 dBA) | Voc6 |
| 8 | speech | 8-channel vocoder | 600 Hz LPF (60 dBA) | LPF600 | 8-channel vocoder (60 dBA) | Voc8 |

*Conditions are divided into factors "listening mode" and "spectral resolution," listed in the first two columns. The middle two columns show the stimuli presented to the left ear, including presentation levels, and their labels. The last two columns similarly show the stimuli presented to the right ear and their labels.*
*LPF, low-pass filtered.*

roll-off for our stimuli is twice as steep and will be masked by the noise-band–vocoded speech quite soon beyond the −3 dB cutoff frequency (Table 1).

The vocoder signal was always presented to the right ear. In the binaural conditions, the vocoder signal was presented to both ears. In the EAS conditions, the low-pass−filtered speech was presented to the left ear in addition to the vocoder signal in the right ear. In the monaural vocoder conditions, no sound was presented to the left ear, and the stimulus in the right ear was presented at 65 dBA. In the remaining conditions, a signal was presented to each ear, which can result in an increase in perceived loudness corresponding to an increase of about 5 dB for stimuli presented over headphones (Epstein & Florentine 2009). Loud or amplified speech can be perceived as more intelligible (Neel 2009) and can potentially affect listening effort as well. Therefore, in these binaural conditions, the signal was presented at 60 dBA to each ear. The presentation level of the stimuli was calibrated using the KEMAR head (G.R.A.S., Holte, Denmark) and the SVANTEK 979 sound level meter (Svantek, Warsaw, Poland), and the speech-shaped noise was provided with the VU corpus, which matches the long-term speech spectrum of the sentences spoken by the female speaker (Versfeld et al. 2000).

**Visual Task and Stimuli** • The secondary task in the dual-task paradigm was a visual rhyme judgment task. This task involved indicating as quickly as possible whether a pair of monosyllabic Dutch words presented one above the other on a monitor in front of the participant rhymed or not. The accuracy of responses and the RTs were recorded by the experimental software. The RT was defined as the interval from visual stimulus onset to the key-press by the participant. The participant was instructed to look at a fixation cross in the middle of the screen. At the onset of each trial, a randomly chosen pair of words would appear on the screen, one above the other. The chance of a rhyming word pair being selected was set to 50%. The words would stay on the screen until either the participant had pressed the response key or the time-out duration of 2.7 sec was reached, the latter of which would be logged as a "miss." After completion of a trial, the fixation cross would reappear for a random duration between 0.5 and 2.0 sec before the next word pair would appear. The timing of the presentation of the visual rhyme words was not coupled to the timing of the auditory stimulus; therefore, a secondary task trial could start at any time during or between auditory stimuli for the primary task.

The stimuli used for this task were the same monosyllabic, meaningful Dutch words used by Pals et al. (2013). For each of the five Dutch vowels (a, e, i, u, o), Pals et al. created lists of monosyllabic rhyme words with several word endings [e.g., (stok, vlok, wrok) or (golf, kolf, wolf)]. They excluded words that could be pronounced in more than one way, as well as the 25% least frequently occurring words, according to the CELEX lexical database of Dutch (Baayen et al. 1995). Due to the nature of the Dutch language, it was not possible to control for orthographic similarity. For each trial, two words were simultaneously displayed one above another, centered on a computer monitor in large, black capital letters on a white background, each letter approximately 7-mm wide and 9-mm high, with 12-mm vertical whitespace between the words.

**Equipment** • Participants were seated in a soundproof booth, approximately 50 cm from a wall-mounted computer screen. The experiment interface was programmed in MATLAB using the Psychophysics Toolbox Version 3 and run on an Apple Mac Pro computer. This program coordinated the presentation of the speech stimuli for the primary task and the visual stimuli for the secondary task. A PalmTrack 24-bit digital audio recorder (Alesis, L.P., Cumberland, RI) was used to record the verbal responses on the primary listening task. The digital audio stimuli were routed via the AudioFire 4 external soundcard (Echo Digital Audio Corporation, Santa Barbara, CA) to the Lavry digital-to-analog converter and on to the open-back HD600 headphones (Sennheiser Electronic GmbH & Co. KG, Wedemark, Germany).

**Procedure** • Before each new task, the experimenter explained the procedure in detail to ensure that the participant understood the task. The participants were first given 3 min to practice the rhyme judgment task alone, during which the experimenter monitored their performance to see whether they understood the task and provided additional instructions if this proved necessary. This was followed by a 20-min intelligibility training session (based on Benard & Başkent 2013), in which participants familiarized themselves with the different processing conditions of the speech stimuli. The intelligibility training session consisted of six blocks of 13 sentences each, one block each for six of the eight processing conditions (the two monaural CI and the four EAS conditions), which were presented in random order. The participant's task was to repeat the sentences as best they could. After each response, the participants received both visual and auditory feedback. First, the sentence was displayed as text on the monitor, and then the audio recording was played back twice, once unprocessed and once processed. The sentences used during training were not used again in the rest of the experiment.

The data collection phase of the experiment consisted of 16 blocks: both a single-task and a dual-task block for each of the eight experimental conditions. The single tasks consisted of 13 sentences and served to obtain a measure of intelligibility for each of the experimental conditions. The dual task combined the intelligibility task and visual rhyme task, and for each dual task, two sets of 13 sentences each were used. This ensured that during each dual task, a sufficient number of secondary task trials could be presented and thus a sufficient number of RTs could be recorded. Approximately, three secondary task trials were presented for each sentence in the primary intelligibility task, and on average, 80 RTs were recorded per participant per dual-task block. The presentation order of the conditions was randomized using the MATLAB random permutation function seeded to the system clock.

After completing each test with one of the processing conditions, either single or dual task, the participants were instructed to fill out a multidimensional subjective workload rating scale, the NASA-TLX.

The procedure for Experiment 1, including audiometric tests and training, lasted approximately 2 hr.

## Analysis

Each of the 20 participants completed 2 × 4 dual tasks; each task comprised 26 sentences and approximately 80 rhyme judgment RT trials, resulting in an estimated 1600 data points per condition for the RT measure of listening effort. The presentation of the rhyme judgment task depended, in part, on the individual participants' response speed. Wrong answers were excluded from the dataset (approximately 4 to 5% of the data points for each of the experiments) because these could result

from accidental button presses and thus could have introduced unrealistically short RTs. Therefore, the exact number of data points per participant per condition varied. A data set such as this, with numerous and unequal number of data points per participant per condition, would violate the independence assumption of analysis of variance. We have therefore chosen to use linear mixed-effects (LMEs) models to analyze these RT data. LME models offer the opportunity to include random effects for each participant and for each item in the model, thus compensating for individual differences between participants and items and improving the generalizability of the results (Barr et al. 2013). One of the known difficulties of using noise-band–vocoder stimuli is training effects, which improve performance over time, while another concern could be fatigue, which reduces performance over the course of the experiment. Including a fixed factor to account for such effects associated with presentation order of the conditions could improve the model (Baayen et al. 2008).

In this study, the data were analyzed using the lme4-package (version 1.1–7) in R. The models were constructed starting with the simplest model possible and consecutively adding fixed factors in a manner that followed the experimental design. Each new model was compared with the previous model for improved fit using $\chi^2$ tests, and fixed factors were only included if the fit of the model improved significantly. In our models, we have chosen to include the random effects for participant, to factor out individual differences, and for the sentences presented in the primary task that was performed simultaneously with the secondary RT task. If some of the sentences were inherently more difficult to understand than other sentences, this could result in an increase in RT for the simultaneous secondary task trials due to the specific stimulus rather than the experimental condition. Including the random factor "sentence ID" referring to the specific auditory stimulus could factor out these effects of individual stimuli. The $p$ values reported were obtained using the Satterthwaite approximation reported by the lmerTest package.

## Results

Figure 1 shows all data, averaged over participants, for all three experiments. The columns show, from left to right, the results for Experiments 1 to 3, respectively. The rows, from top to bottom, show sentence intelligibility scores, RTs, and NASA-TLX scores. The average speech intelligibility scores for Experiment 1 (Fig. 1, top-left panel), shown in percentage of sentences correctly repeated, were comparable across all conditions, at just below ceiling as expected. These data were used only to confirm that the desired intelligibility level was reached, as planned, across all conditions.

Visual inspection of the RTs averaged across all participants (Fig. 1, middle-left panel) revealed small differences in RTs between some of the experimental conditions. The RTs were analyzed within subject using LME, as described earlier. Incorrect trials for the visual rhyme judgment task were excluded from analysis of the RTs; they accounted for about 4% of the responses. Including presentation order as a factor in the model to account for learning effects over the course of the experiment significantly improved the fit of the model [$\chi^2(1) = 83.55$; $p < 0.001$]. The factors of interest were "listening mode" (monaural vocoder, binaural vocoder, monaural vocoder with 300 Hz low-pass–filtered acoustic speech presented contralaterally, and monaural vocoder with 600 Hz

low-frequency acoustic speech presented contralaterally) and "spectral resolution" (six-channel and eight-channel vocoder). However, including spectral resolution in the model did not show a significant main effect of spectral resolution, no significant interactions, and did not improve the fit of the model [$\chi^2(1) = 2.636$; $p = 0.621$]. Spectral resolution was therefore not included in the model.

To see whether individual differences in intelligibility scores per condition can explain some of the observed differences in RT, a model was constructed including the intelligibility scores as a factor. However, including speech intelligibility in the model did not improve the fit [$\chi^2(1) = 3.546$; $p = 0.060$] and was therefore not included.

The preferred model, therefore, included the factor "listening mode," the numeric factor "presentation order," and random intercepts for each participant and for each individual sentence among the auditory stimuli. In case of a nonnumeric factor such as "listening mode," the summary of a linear model estimates the value of the reference level and lists the estimated differences between each of the other levels and the reference level. In our design, both the monaural and binaural vocoder conditions were included as control conditions: to investigate the effects of low-pass–filtered speech presented contralaterally to the vocoder signal and whether these effects differ from presenting the vocoder binaurally. Therefore, it makes sense to compare the conditions with low-pass–filtered speech to both the monaural vocoder condition and the binaural vocoder condition. Two versions of the model were, therefore, generated, one using the monaural vocoder condition as the reference level and the other using the binaural vocoder condition as the reference level (Table 2).

The model with the "monaural vocoder" listening mode as reference level is summarized in the top half of Table 2, and the same model with the "binaural vocoder" listening mode as the reference is summarized in the bottom half of Table 2. When comparing with monaural vocoder as the reference, adding either vocoder or low-frequency acoustic signal in the other ear did not significantly change the RTs. The RTs for monaural vocoder were on average halfway between the RTs for binaural vocoder (which are estimated to be 16 msec longer than the RTs for monaural vocoder) and the RTs for both conditions with contralaterally presented low-pass–filtered speech (RTs for "Mon voc + low pass 300 Hz" and "Mon voc + low pass 600 Hz" are estimated to be 17 and 15 msec shorter than monaural vocoder, respectively).

To examine the differences between binaural vocoder and the conditions with contralaterally presented low-pass–filtered speech, the model was also examined using binaural vocoder as the reference level. The intercept of the model corresponds with the listening mode "binaural vocoder" and was estimated at 1.102 sec ($\beta = 1.102$; SE = 0.032; $t = 34.0$; $p < 0.001$). The difference between this estimate and the actual mean RT for the binaural vocoder listening modes as shown in Figure 1 stems from the inclusion of the random intercept for the individual auditory stimuli in the model. The effect of presentation order is significant and estimated at −12 msec ($\beta = -0.012$; SE = 0.001; $t = -9.3$; $p < 0.001$), implying that participants' RTs become 12 msec shorter with each task as the experiment progressed over time. The estimates for the other listening modes are all relative to the intercept, the estimated RT for binaural vocoder. Both listening modes with low-pass–filtered speech resulted in significantly shorter RTs than binaural vocoder: "Mon voc + low
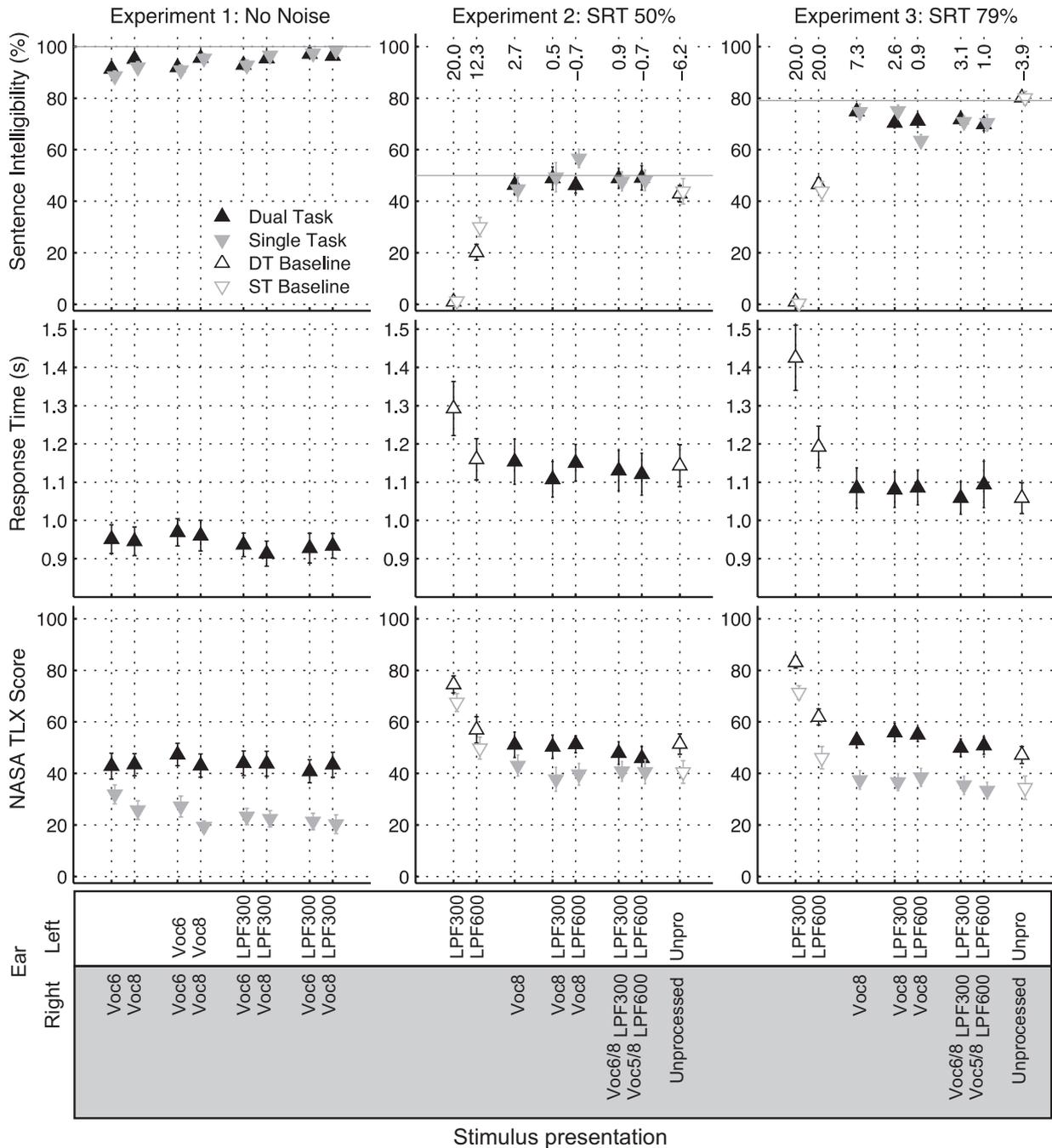
Fig. 1. The group averages of the data for Experiments 1, 2, and 3 (n = 20 for each experiment) are shown in the left, middle, and right columns, respectively. Experimental conditions are listed in the table under the *x* axes, with separate columns for the stimuli presented to the right ear (bottom, gray column) and left ear (top, white column). The top row shows the ST and DT speech intelligibility scores in percentage of sentences correctly repeated. For Experiments 2 and 3, the SNRs at which each of the conditions was presented at the very top of the figure in dB SNR. The middle row shows the DT response times on the secondary task. The bottom row shows the NASA-TLX ratings (higher scores indicate more effort). Up triangles show DT results, and down triangles show ST results; error bars represent 1 SE. Filled symbols show conditions of interest that are included in the analysis, and open symbols show conditions that are tested for reference but not included in the analysis. DT, dual-task; NASA-TLX, NASA Task Load indeX; SNR, signal to noise ratio; SRT, speech reception threshold; ST, single-task.

pass 300 Hz" resulted in 32 msec shorter RTs ($\beta = -0.032$; SE = 0.008; $t = -3.8$; $p < 0.001$) and "Mon voc + low pass 600 Hz" in 30 msec shorter RTs ($\beta = -0.030$; SE = 0.008; $t = -3.6$; $p < 0.001$). RTs for monaural vocoder appear to be slightly shorter than for binaural vocoder; however, this difference is not significant ($\beta = -0.016$; SE = 0.008; $t = -1.8$; $p = 0.064$).

Visual inspection of the across-subject average NASA-TLX scores for Experiment 1 (Fig. 1, left-bottom panel), plotted separately for single-task and dual-task presentation, showed higher self-reported effort for the dual task compared with the single task, as well as some differences between conditions. Because the NASA-TLX scores for the dual-task conditions

**TABLE 2. Summary of linear models for dual-task RT results for Experiment 1**

| Dual-Task RT Results | Estimate (msec) | SE | df | t Value | Pr(>|t|) |
|---|---|---|---|---|---|
| Monaural vocoder (intercept) | 1.086 | 0.032 | 24 | 33.58 | <0.001* |
| Presentation order | −0.012 | 0.001 | 1.365e+04 | 1.85 | <0.001* |
| Binaural vocoder | 0.016 | 0.008 | 1.369e+04 | −1.96 | 0.065 |
| Mon voc + 300 Hz LPF speech | −0.017 | 0.008 | 1.368e+04 | −1.76 | 0.050 |
| Mon voc + 600 Hz LPF speech | −0.015 | 0.008 | 1.362e+04 | −9.32 | 0.078 |
| Bimodal vocoder (intercept) | 1.102 | 0.032 | 24 | 34.0 | <0.001* |
| Presentation order | −0.012 | 0.001 | 1.365e+04 | −9.3 | <0.001* |
| Monaural vocoder | −0.016 | 0.008 | 1.369e+04 | −1.8 | 0.064 |
| Mon voc + 300 Hz LPF speech | −0.032 | 0.008 | 1.362e+04 | −3.8 | <0.001* |
| Mon voc + 600 Hz LPF speech | −0.030 | 0.008 | 1.364e+04 | −3.6 | <0.001* |

*Both models included the factor "listening mode" (four levels) and the numeric factor "presentation order." The top half of the table shows the results for the model using the listening mode "monaural vocoder" as the reference level, and the bottom half of the table shows the results for the model using listening mode "bimodal vocoder" as reference level.*
*LPF, low-pass filtered; RT, response time. Note: *denotes a significant effect at the 0.001 level.*

can be interpreted as an effort rating for the combined listening and secondary rhyme judgment task rather than the listening task alone, the analysis of the NASA-TLX results focused on the single-task NASA-TLX scores. The analysis of the NASA-TLX results was also performed using LME models. A random intercept for participant was included in the model; however, because the NASA-TLX scores consisted of one value per participant per condition, no random intercept per sentence could be included. Including the single-task speech intelligibility significantly improved the model [$\chi^2(1) = 20.923$; $p < 0.001$]. Including presentation order [$\chi^2(1) = 0.384$; $p = 0.536$] or spectral resolution [$\chi^2(1) = 6.108$; $p = 0.191$] in the model did not significantly improve the fit (Table 3).

The best model for the NASA-TLX data included the factors "speech score" and "listening mode" and random intercepts for "participant"; this model is summarized in Table 3. The intercept corresponds to the estimated NASA-TLX score for monaural vocoder, for a speech score of 100% sentence correct, this is estimated at a score of 22.7 out of 100 ($\beta = 22.678$; SE = 2.949; $t = 7.689$; $p < 0.001$). The effect of speech score is significant and estimated at −0.63 ($\beta = -0.630$; SE = 0.119; $t = 05.316$; $p < 0.001$), meaning that for each 1% point drop in speech score, the participants rate the task as 0.63 points out of 100 more effortful on the NASA-TLX multidimensional self-report scales. None of the listening modes differed significantly from the reference-level monaural vocoder (Fig. 1).

To summarize, speech intelligibility was near-ceiling for all conditions, although exact speech scores varied slightly across participants and conditions. The dual-task results of Experiment 1 showed a significant benefit of low-frequency acoustic speech presented contralaterally to the vocoder signal compared with binaural vocoded speech (i.e., shorter RTs), for both 300 and 600 Hz low-pass–filtered speech. However, monaural vocoded speech did not differ significantly from either binaural vocoder or vocoder plus contralateral low-frequency acoustic speech. The subjective measure of listening effort, the NASA-TLX, showed no significant effect of listening mode. Any difference in NASA-TLX ratings between conditions and participants could be entirely attributed to effects of small individual differences in intelligibility.

## EXPERIMENT 2: SPEECH IN NOISE AT 50% INTELLIGIBILITY

### Motivation

In Experiments 2 and 3, we examined the effect of low-frequency acoustic sound in addition to vocoded speech on listening effort in interfering noise at equal intelligibility levels, away from ceiling and at different parts of the psychometric function. In Experiment 2, 50% sentence intelligibility was used. Equal intelligibility across conditions was achieved by presenting the different processing conditions at different signal to noise ratios (SNRs). We hypothesized that even with intelligibility fixed at 50% by varying the SNRs, the added low-frequency speech may still provide an additional benefit in reduced listening effort.

Because the results of Experiment 1 revealed no effect of spectral resolution, the six-channel vocoder conditions were dropped in favor of including additional listening configurations based on the eight-channel vocoder conditions. In Experiments 2 and 3, we chose to compare the following simulated device configurations: (1) monaural vocoder with low-pass–filtered speech presented to the contralateral ear (the same as in Experiment 1); (2) the upper six or five channels of an eight-channel vocoder signal presented monaurally, combined with bilaterally presented

**TABLE 3. Summary of the linear model for the NASA-TLX results for Experiment 1**

| ST NASA-TLX Results | Estimate (msec) | SE | df | t Value | Pr(>|t|) |
|---|---|---|---|---|---|
| Monaural vocoder (intercept) | 22.678 | 2.949 | 38.150 | 7.689 | <0.001* |
| Speech score | −0.630 | 0.119 | 150.890 | −5.316 | <0.001* |
| Binaural vocoder | −3.741 | 1.988 | 136.320 | −1.882 | 0.062 |
| Mon voc + 300 Hz LPF speech | −3.171 | 2.027 | 137.020 | −1.564 | 0.120 |
| Mon voc + 600 Hz LPF speech | −3.128 | 2.160 | 139.040 | −1.448 | 0.150 |

*The model included the factor "listening mode" (four levels: monaural vocoder, binaural vocoder, vocoder plus 300 Hz LPF speech, and vocoder plus 600 Hz LPF speech) and the numeric factor "speech score." The model used the listening mode "monaural vocoder" as the reference level.*
*LPF, low-pass filtered; NASA-TLX, NASA Task Load indeX; ST, single-task. Note: *denotes a significant effect at the 0.001 level.*

**TABLE 4.  Summary of listening conditions for Experiments 2 and 3**

| No. | Left Ear | Label | Right Ear | Label | Experiment 2 SRT 50% SNR (SDev) | Experiment 3 SRT 79% SNR |
|-----|----------|-------|-----------|-------|--------------------------------|--------------------------|
| 1 | 300 Hz LPF speech | LPF300 | — | | 20.0* | 20.0* |
| 2 | 600 Hz LPF speech | LPF600 | — | | 12.3 (3.71) | 20.0* |
| 3 | — | | 8-channel vocoder | Voc8 | 2.7 (1.76) | 7.3 |
| 4 | 300 Hz LPF speech | LPF300 | 8-channel vocoder | Voc8 | 0.5 (1.40) | 2.7 |
| 5 | 600 Hz LPF speech | LPF600 | 8-channel vocoder | Voc8 | −0.7 (1.07) | 0.9 |
| 6 | 300 Hz LPF speech | LPF300 | 300 Hz LPF + 6/8-channel vocoder | Voc6/8 LPF300 | 0.9 (1.47) | 3.2 |
| 7 | 600 Hz LPF speech | LPF600 | 600 Hz LPF + 5/8-channel vocoder | Voc5/8 LPF600 | −0.7 (0.99) | 1 |
| 8 | 80–6000 Hz unprocessed | Unpro | 80–6000 Hz unprocessed | Unpro | −6.2 (0.73) | −3.9 |

Columns 1 and 3 show the stimuli that were presented to the left and to the right ear in each of the conditions, respectively, followed by columns 2 and 4 with the stimulus label. The last two columns show the average SNRs at which the desired SRTs were obtained. Values in brackets indicate standard deviations.
*Conditions where the target intelligibility level could not be reached, and therefore, the SNR was set to a nominal value of 20 dB.
LPF, low-pass filtered; SNR, signal to noise ratio; SRT, speech reception threshold.

low-pass–filtered speech, thus roughly approximating a shallow inserted CI combined with residual low-frequency acoustic hearing in both ears (new compared with Experiment 1).

Research with hybrid CI users shows that overlap between the electric and acoustic signals in the same ear can be detrimental for speech understanding in babble noise (Karsten et al. 2013). We, therefore, chose to prevent overlap between the low-pass–filtered speech signal and the vocoder signal. When combined with 300 Hz low-pass–filtered speech, the lower two vocoder channels, which would overlap with the low-pass–filtered speech, were removed and only the higher six out of eight vocoder channels were presented. When combined with 600 Hz low-pass–filtered speech, only the higher five out of eight vocoder channels were presented.

Research shows that CI users can benefit from bilateral low-frequency hearing compared with contralateral low-frequency hearing alone (Dorman & Gifford 2010; Gifford et al. 2013), especially for speech understanding in noise. The magnitude of this benefit most likely depends on the insertion depth of the CI and degree of hearing preservation (Gifford et al. 2013). We, therefore, hypothesized that (1) monaural vocoder combined with bilateral low-frequency speech will require less listening effort than with contralateral low-frequency speech, and (2) five vocoder channels combined with 600 Hz low-pass–filtered speech will be less effortful to understand than six vocoder channels combined with 300 Hz low-pass–filtered speech.

**Methods**

The procedure for Experiment 2 was similar to Experiment 1, therefore, only the differences will be described.

**Participants** • Twenty new participants were recruited for participation in Experiment 2. All were normal-hearing, native Dutch-speaking, young adults (age range, 18 to 33 years; mean, 20 years; 11 female). The results of 1 participant were excluded from the analysis of the NASA-TLX because the questionnaire was not filled out completely.

**Stimuli** • The same auditory and visual stimuli as in Experiment 1 were used. The experimental processing conditions are summarized in Table 4. The eight-channel simulations were chosen over the six-channel simulations to ensure that the desired speech reception thresholds (SRTs) would be attainable at reasonable SNRs. A baseline, unprocessed speech condition was also added for comparison.

The noise used in both the speech-in-noise test and the actual experiment was a speech-shaped steady-state noise that was provided with the VU speech corpus (Versfeld et al. 2000) (Table 4).

**Presentation Levels** • The noise was presented continuously throughout each task and at the same level (50 dBA) for all participants and all conditions. The presentation levels of the auditory stimuli for each condition were determined for each participant individually, before the experiment, by means of a speech-in-noise test using a 1-down-1-up adaptive procedure. The speech-in-noise test procedure used to determine the participants' individual SRTs was similar to the speech audiometric test used in clinics in the Netherlands (Plomp 1986). Each test used one list of 13 sentences. The first sentence was used to quickly converge on the approximate threshold of intelligibility. Starting at 8 dB below the noise and increasing the level in steps of 4 dB, the sentence was repeatedly played until the entire sentence was correctly reproduced. From this level, the adaptive procedure started, where the SNR was increased or decreased by 2 dB after an incorrect or correct response, respectively. A list of 13 sentences was thus sufficient for at least six reversals (often about eight), which is generally accepted to result in a reliable estimate of the 50% SRT (Levitt 1971). The average SRTs (in dB SNR) for all 20 participants are listed in Table 4, second column from right.

Attaining the desired 50% intelligibility levels was not possible for 300 Hz low-pass–filtered speech. Therefore, we chose to present sentences for this condition at 20 dB SNR.

**Procedure** • The adaptive speech-in-noise test, used to determine the presentation levels for the auditory stimuli, at the start of the experiment, required the participant to listen to a minimum of 10 sentences per experimental condition. This provided some initial familiarization with the sentence material and stimulus processing for the participants, and increased testing time by about 15 min. Further training with the sentence material was still provided, although in the interest of time, without feedback. This training session lasted around 10 min. For the rest, the procedure was identical to Experiment 1. The entire session lasted around 2 hr.

**Results**

The speech intelligibility results for Experiment 2 are shown in the top-middle panel of Figure 1. The conditions in which only low-pass–filtered speech was presented were included as

**TABLE 5. Summary of the linear model for the dual-task RT results for Experiment 2**

| Dual-Task RT Results | Estimate (msec) | SE | df | t Value | Pr(>\|t\|) |
|---|---|---|---|---|---|
| Monaural vocoder (intercept) | 1.238 | 0.049 | 26 | 25.259 | <0.001* |
| Speech score (±50%) | −0.002 | 0.000 | 7968 | −6.207 | <0.001* |
| Presentation order | −0.014 | 0.003 | 7976 | −5.360 | <0.001* |
| +LPF speech | −0.030 | 0.013 | 7956 | −2.243 | 0.025† |
| +LPF:mode | −0.002 | 0.012 | 7958 | 0.131 | 0.896 |
| +LPF:cutoff | −0.017 | 0.012 | 7954 | 1.412 | 0.158 |
| +LPF:mode:cutoff | −0.017 | 0.024 | 7958 | 0.719 | 0.472 |

The model included the factors "speech score" and "presentation order," +LPF speech (the contrast between vocoder alone and vocoder plus LPF speech regardless of configuration or LPF cutoff frequency), and within the +LPF conditions: the factor "listening mode" (two levels: contralateral LPF speech and binaural LPF speech) and the factor LPF cutoff frequency (two levels: 300 and 600 Hz cutoff frequency).

LPF, low-pass filtered; RT, response time. Note: *denotes a significant effect at the 0.001 level, †denotes a significant effect at the 0.05 level.

a reference, and to show that low-pass–filtered speech by itself produced limited intelligibility. The unprocessed speech condition was included as a normal-hearing reference point. In Experiment 2, the desired intelligibility level of 50% sentence recognition was achieved by determining the appropriate SNRs for each condition using an adaptive procedure at the start of the experiment, as explained earlier. The across-subject average SNRs are included in the figure. On average, the intelligibility scores were indeed close to 50% for the conditions of interest in this experiment.

The center panel of Figure 1 shows the RTs on the secondary rhyme judgment task for Experiment 2. Incorrect trials for the visual rhyme judgment task were excluded from analysis of the RTs; they accounted for about 5% of the trials. As the goal of this study was to examine the effect of providing low-pass–filtered speech to complement vocoded speech, the conditions of interest are the monaural vocoder and the combined vocoder and low-pass–filtered speech conditions; the analysis, therefore, focuses on these five conditions. Visual inspection of Figure 1 center panel shows that the group average RT for monaural vocoder appears slightly longer than most, although not all, of the conditions with combined vocoder and low-pass–filtered speech. The RTs were analyzed for within-subject effects using LME.

The results were modeled in a design that most closely resembled the contrasting dimensions in this design. Included in the model were the effect of added low-pass–filtered speech on average compared with monaural vocoder alone, the contrast between contralaterally and bilaterally presented low-pass–filtered speech, and the contrast between 300 and 600 Hz low-pass–filtered speech. Including task order in the model significantly improved the fit [$\chi^2(1) = 27.258$; $p < 0.001$]. Speech scores were included in the model to account for differences in speech scores between participants and conditions and to investigate how much of the observed differences in RT can be attributed to differences in intelligibility. Including speech scores did significantly improve the model [$\chi^2(1) = 38.418$; $p < 0.001$]. Each condition was presented at an individually determined SNR that differed for each participant; however, including presentation SNR in the model was not warranted [$\chi^2(1) = 0.604$; $p = 0.437$] (Table 5).

Table 5 summarizes the model. The intercept of the model corresponds to the RT for monaural vocoder alone at 50% sentence intelligibility and is estimated at 1.238 sec ($\beta = 1.238$; SE = 0.049; $t = 25.259$; $p < 0.001$). The effect of speech score is significant and estimated at −2 msec ($\beta = -0.002$; SE = 0.000;

$t = -6.207$; $p < 0.001$), suggesting a decrease in RT of 2 msec for each 1% point increase in intelligibility. The model shows a significant effect of presentation order, estimated at −14 msec ($\beta = -0.014$; SE = 0.003; $t = -5.360$; $p < 0.001$), implying that the RTs are 14 msec shorter RTs for each task compared with the preceding task. The effect of low-frequency acoustic speech in addition to vocoded speech compared with monaural vocoder was significant and estimated at −30 msec ($\beta = -0.030$; SE = 0.013; $t = -2.243$; $p = 0.025$) suggesting on average 30 msec shorter RTs for conditions including low-pass–filtered speech (i.e., for the RTs of all those conditions in which low-pass–filtered speech was presented pooled together) than for simulated monaural vocoder alone. Among the four different conditions with low-pass–filtered speech, no significant differences were found.

The average NASA-TLX ratings for Experiment 2, for both dual and single tasks, are shown in the bottom-middle panel of Figure 1. Visual inspection of the single-task NASA-TLX score across-subject averages shows fairly similar effort ratings for all conditions of interest. The NASA-TLX results were analyzed for within-subject effects in the same manner as the RT results. Adding presentation order to the model was not warranted [$\chi^2(1) = 0.1712$; $p = 0.679$]. Including presentation speech scores did significantly improve the fit of the model [$\chi^2(1) = 46.427$; $p < 0.001$] (Table 6).

The model is summarized in Table 6. The intercept corresponds to the estimated NASA-TLX score for monaural vocoder alone at 50% intelligibility and is estimated at a score of 41 out of 100 ($\beta = 41.004$; SE = 3.946; $t = 10.393$; $p < 0.001$). There is a significant effect of speech score estimated at –0.378 ($\beta = -0.378$; SE = 0.049; $t = -7.675$; $p < 0.001$), implying a 0.378 decrease in NASA-TLX score for each 1% point increase in speech intelligibility. For the NASA-TLX results, none of the effects of additional low-pass–filtered speech, and the different configurations in which low-pass–filtered speech was added, were significant.

In short, speech intelligibility was successfully fixed at 50% sentence recognition for the conditions of interest, at different SNRs for each condition (Table 4). The dual-task results for Experiment 2 showed a significant benefit (i.e., shorter RTs) of additional low-pass–filtered speech compared with monaural vocoder for all four low-pass filtered speech conditions grouped together. No difference was found between the four different low-pass–filtered speech configurations. The NASA-TLX results showed no significant difference in ratings between monaural vocoder alone and with additional low-pass–filtered

**TABLE 6. Summary of the linear model for the NASA-TLX results for Experiment 2**

| ST NASA-TLX Results | Estimate (msec) | SE | df | t Value | Pr(>|t|) |
|---|---|---|---|---|---|
| Monaural vocoder (intercept) | 41.004 | 3.946 | 26.41 | 10.393 | <0.001* |
| Speech score (±50%) | −0.378 | 0.049 | 72.66 | −7.675 | <0.001* |
| +LPF speech | −0.805 | 2.089 | 71.06 | −0.385 | 0.701 |
| +LPF:mode | −0.390 | 1.870 | 71.07 | 0.209 | 0.835 |
| +LPF:cutoff | 2.484 | 1.856 | 71.04 | 1.338 | 0.185 |
| +LPF:mode:cutoff | −2.906 | 3.690 | 71.02 | −0.787 | 0.434 |

*The model included the factors "speech score," +LPF speech (the contrast between vocoder alone and vocoder plus LPF speech regardless of configuration or LPF cutoff frequency), and within the +LPF conditions: the factor "listening mode" (two levels: contralateral LPF speech and binaural LPF speech) and the factor LPF cutoff frequency (two levels: 300 and 600 Hz cutoff frequency).*
*LPF, low-pass filtered; NASA-TLX, NASA Task Load indeX; ST, single-task. Note: *denotes a significant effect at the 0.001 level.*

speech, suggesting that monaural vocoded speech and each of the four low-pass–filtered speech conditions in noise were rated as equally effortful.

## EXPERIMENT 3: SPEECH IN NOISE AT 79% INTELLIGIBILITY

### Motivation

Similar to Experiment 2, listening effort was evaluated for speech in noise. However, in Experiment 3, speech intelligibility level was fixed at 79% to compare effects on listening effort at fixed intelligibility level at a different, shallower point in the psychometric function. The same simulated device configurations as in Experiment 2 were tested in this experiment. The conditions, as well as the SNRs to achieve the 79% sentence intelligibility level, are listed in Table 4.

### Methods

The procedure for Experiment 3 was similar to Experiment 2, therefore, only the differences will be described.

**Participants** • Twenty new participants were recruited for participation in Experiment 3. All were normal-hearing, native Dutch-speaking, young adults (age range, 19 to 26 years; mean, 21 years; 8 female).

Furthermore, 10 additional new participants were recruited for a short test to determine the SRTs for 79% sentence intelligibility. All were normal-hearing, native Dutch-speaking, young adults (age range, 19 to 24 years; mean, 22 years; 6 female).

**Presentation Levels** • Presentation levels were determined with a 3-down-1-up adaptive procedure (Levitt 1971), similar to Experiment 2, except that the SNR was decreased by 2 dB after three consecutive correct responses instead of after each correct response. This procedure requires a substantial amount of time and a large number of sentences to obtain six to eight reversals. Therefore, it was not feasible to determine SRTs for each participant individually before the experiment. Thus, for this experiment, SRTs were determined beforehand with 10 new participants, similar in age and hearing levels to the participants of the experiment. The average SRTs, listed in the rightmost column of Table 4, were used in the experiment.

Attaining the desired 79% sentence recognition with 300 and 600 Hz low-pass–filtered speech was not feasible. Therefore, we chose to present sentences during these conditions at 20 dB SNR.

**Procedure** • As the presentation levels were determined with a different participant group, there was no concern of additional testing time (as was the case in Experiment 2). The participants

of Experiment 3, therefore, received the same 20-min training (with feedback) as participants in Experiment 1 and were tested in an identical procedure to Experiment 1. The entire session lasted around 2 hr.

### Results

The speech intelligibility scores for Experiment 3 are shown in the top-right panel of Figure 1. As in Experiment 2, the conditions in which only low-pass–filtered speech was presented, as well as the unprocessed speech condition, were included as a reference and therefore excluded from the analysis. In Experiment 3, the desired intelligibility level of 79% sentence recognition was achieved by presenting the conditions at SNRs determined with a group of 10 participants similar in age and hearing level to the participants in this experiment. These SNRs are included in the figure. On average, the intelligibility scores were around 75%, and speech intelligibility in the dual task did not vary significantly across the conditions of interest.

The middle-right panel shows the RTs on the secondary rhyme judgment task for Experiment 3. Incorrect trials for the visual rhyme judgment task were excluded from analysis of the RTs; they accounted for about 4% of the responses for Experiment 3. Including presentation order in the model significantly improved the fit [$\chi^2(1) = 50.084$; $p < 0.001$], as did including speech score [$\chi^2(1) = 29.189$; $p < 0.001$] (Table 7).

The model is summarized in Table 7. The intercept corresponds to RTs to monaural vocoded speech alone in noise at 79% intelligibility and is estimated at 1.238 sec ($\beta = 1.238$; SE = 0.049; $t = 24.600$; $p < 0.001$). The effect of speech score is significant and estimated at −4 msec ($\beta = -0.004$; SE = 0.001; $t = -5.404$; $p < 0.001$), implying a 4-msec reduction in RT for each 1% point increase in speech score. Presentation order has a significant effect on RT and is estimated at −16 msec ($\beta = -0.016$; SE = 0.002; $t = -6.430$; $p < 0.001$), suggesting a 16-msec decrease in RT for each consecutive task. None of the modeled contrasts between vocoded speech with versus without low-pass–filtered speech, 300 versus 600 Hz, and monaural versus binaural low-pass–filtered speech conditions revealed any significant differences (Table 8).

The average NASA-TLX ratings for Experiment 3 are shown in the bottom-right panel of Figure 1. The NASA-TLX data were modeled in a similar manner as for Experiment 2. Adding presentation order to the model was not warranted [$\chi^2(1) = 1.354$; $p = 0.245$]. Including speech score in the model did significantly improve the fit [$\chi^2(1) = 7.411$; $p = 0.006$]. The model is summarized in Table 8. The NASA-TLX score for monaural vocoder alone at 79% intelligibility is estimated at 36 out of 100

**TABLE 7. Summary of the linear model for the dual-task RT results for Experiment 3**

| Dual-Task RT Results | Estimate (msec) | SE | df | t Value | Pr(>|t|) |
|---|---|---|---|---|---|
| Monaural vocoder (intercept) | 1.214 | 0.049 | 26 | 24.600 | <0.001* |
| Speech score (±79%) | −0.004 | 0.001 | 8131 | −5.404 | <0.001* |
| Presentation order | −0.016 | 0.002 | 8256 | −6.430 | <0.001* |
| +LPF speech | −0.011 | 0.013 | 8207 | −0.838 | 0.402 |
| +LPF:mode | −0.011 | 0.012 | 8216 | −1.010 | 0.312 |
| +LPF:cutoff | 0.017 | 0.012 | 8224 | 1.521 | 0.128 |
| +LPF:mode:cutoff | −0.005 | 0.023 | 8247 | −0.220 | 0.826 |

The model included the factors "speech score" and "presentation order," +LPF speech (the contrast between vocoder alone and vocoder plus LPF speech regardless of configuration or LPF cutoff frequency), and within the +LPF conditions: the factor "listening mode" (two levels: contralateral LPF speech and binaural LPF speech) and the factor LPF cutoff frequency (two levels: 300 and 600 Hz cutoff frequency).
LPF, low-pass filtered; RT, response time. Note: *denotes a significant effect at the 0.001 level.

($\beta$ = 36.534; SE = 3.443; $t$ = 10.560; $p$ < 0.001). The effect of speech score was significant and estimated at −0.25, implying a decrease in NASA-TLX score of 0.25 per 1% point increase in speech intelligibility. Between the different listening conditions of interest, monaural vocoder and the four conditions with additional low-pass−filtered speech, effort was not rated any differently.

To summarize, speech intelligibility was successfully fixed at, on average, 75% for the conditions of interest, at different SNRs for each condition (Table 4). The dual-task results for Experiment 3 showed no difference in listening effort for any of the conditions of interest. The NASA-TLX showed no benefits in listening effort between any of the simulated CI and EAS conditions.

## DISCUSSION

In this study, we aimed to examine how the addition of low-frequency acoustic speech affects listening effort for normal-hearing listeners, when interpreting spectrotemporally degraded, noise-band–vocoded speech in quiet or in background noise, specifically when intelligibility is held constant across conditions. Three dual-task experiments were conducted at three different intelligibility levels: at near-ceiling intelligibility (in quiet) and at 50% and 79% sentence intelligibility (in background noise). The outcome measure of interest in this study was the RT on the secondary task, which was used as a behavioral measure of listening effort. For comparison, we included the NASA-TLX rating scale as a subjective self-report measure of listening effort; however, in line with the results of our earlier study (Pals et al. 2013), the NASA-TLX could not distinguish between the experimental conditions of interest at equal intelligibility levels. The dual-task RTs did show some

effects, but only between a limited number of conditions. On the basis of the results from these three experiments, we have to reject hypothesis 1: the RT results from Experiment 1 showed no significant main effect of spectral resolution. The RT results from the three experiments provided mixed, inconclusive evidence in support of hypothesis 2a, namely that the presence of low-pass–filtered speech will improve listening effort. We will address the specific findings and their implications in more detail later in the discussion. Purely based on the comparison of binaural vocoder RTs and the conditions including low-pass–filtered speech in Experiment 1, hypothesis 2b appears to be supported, as the binaural vocoder condition resulted in significantly longer RTs. However, a counter-intuitive result for the monaural vocoder RTs, which will be elaborated on later in the discussion, makes these results difficult to interpret. Hypothesis 3 is rejected: the RT results for none of the experiments show a significant difference between conditions with 300 versus 600 Hz low-pass–filtered speech. Hypothesis 4 is supported: the NASA-TLX results revealed no significant differences between any of the experimental conditions at fixed intelligibility levels; however, the NASA-TLX results in all three experiments showed a significant main effect of intelligibility.

One of the challenges when investigating listening effort is to disentangle the effects of intelligibility and background noise. Research shows that both intelligibility and SNR can affect listening effort (e.g., Wu et al. 2016; Zekveld et al. 2010). Zekveld et al. (2010) conducted a pupillometry study to investigate the effect of intelligibility on listening effort, in which intelligibility was manipulated using SNR. Higher, that is, more favorable SNRs, produced higher intelligibility and also resulted in lower listening effort. It is interesting that, Zekveld et al. observed that, even for sentences presented at the same SNR, those sentences that were not heard correctly elicited higher listening

**TABLE 8. Summary of the linear model for the NASA-TLX results for Experiment 3**

| ST NASA-TLX results | Estimate (msec) | SE | df | t value | Pr(>|t|) |
|---|---|---|---|---|---|
| Monaural vocoder (intercept) | 36.354 | 3.443 | 37.89 | 10.560 | <0.001* |
| Speech score (±79%) | −0.250 | 0.092 | 81.20 | −2.707 | 0.008† |
| +LPF speech | −2.649 | 2.385 | 75.25 | −1.111 | 0.270 |
| +LPF:mode | −2.838 | 2.101 | 75.06 | −1.351 | 0.181 |
| +LPF:cutoff | −1.532 | 2.168 | 75.45 | −0.707 | 0.482 |
| +LPF:mode:cutoff | −1.094 | 4.319 | 75.40 | −0.253 | 0.800 |

The model included the factors "speech score," +LPF speech (the contrast between vocoder alone and vocoder plus LPF speech regardless of configuration or LPF cutoff frequency), and within the +LPF conditions: the factor "listening mode" (two levels: contralateral LPF speech and binaural LPF speech) and the factor LPF cutoff frequency (two levels: 300 and 600 Hz cutoff frequency).
LPF, low-pass filtered; NASA-TLX, NASA Task Load indeX; ST, single-task. Note: *denotes a significant effect at the 0.001 level, †denotes a significant effect at the 0.01 level.

effort than sentences that were repeated correctly. Despite our initial intention to present each of the conditions at three specific, fixed intelligibility levels for the three experiments, in practice, the intelligibility scores did vary somewhat across participants and conditions. In line with the findings by Zekveld et al., we observed that speech intelligibility scores affected the RT measure of listening effort; around the 50% intelligibility level, a 1% point increase in intelligibility resulted in a 2 msec reduction of the RTs (Experiment 2), and around the 79% intelligibility level, a 1% point increase in intelligibility resulted in a 4-msec reduction of the RTs (Experiment 3). These effects of intelligibility were accounted for in the models of our results, and thus, any difference in RTs between conditions reported in this study should not be attributed to differences in intelligibility alone.

As mentioned earlier, some of the RT results appeared to support hypothesis 2a: in some cases, the additional low-frequency acoustic speech resulted in faster dual-task RTs. In Experiment 1, for speech in quiet at near-ceiling intelligibility, the RTs were significantly shorter for the conditions with low-pass–filtered speech compared with binaural vocoder, although not compared with monaural vocoder. In Experiment 2, for speech in noise at 50% intelligibility, the RTs were significantly shorter for vocoder plus low-pass–filtered speech than for monaural vocoder alone, even though the vocoded plus low-pass–filtered speech was presented at less favorable SNRs. A possible interpretation of these results is that the additional speech cues represented in the low-frequency acoustic sound facilitate easier speech understanding. An important carrier of low-frequency speech cues is the fundamental frequency (F0; Başkent et al., Reference Note 1), which has been shown to improve speech perception in noise when combined with envelope cues (Brown & Bacon 2009; Cullington & Zeng 2010; Qin & Oxenham 2006), and even alone (Binns & Culling 2007). The availability of F0 information can help segregate the target speech from the background (Cullington & Zeng 2010; Oxenham 2008). Furthermore, F0 carries phonetic information, such as consonant voicing (Brown & Bacon 2009; Kong & Carlyon 2007), and the associated envelope provides information about manner of articulation (Zhang et al. 2010a). Access to onset-of-voicing and prosodic information in the low-frequency acoustic signal provides the listener with information about word boundaries as well as stress patterns, which can facilitate word and sentence recognition (Tyler & Cutler 2009; Zhang et al. 2010a). While in English, listeners can rely on other cues for word stress, in Dutch, the language used in our experiments, these cues are less prominent and access to pitch information carried by F0, therefore, provides a larger benefit than for English (Cutler et al. 2006). We should note, however, that the filters used to obtain the low-pass–filtered stimuli had a 36 dB per octave roll-off. Even the 300 Hz low-pass–filtered speech stimuli, when presented at 60 dB, would still be audible up to frequencies around 600 Hz, and the 600 Hz low-pass–filtered stimuli up to 1200 Hz. This is a significantly wider range of frequencies than the 125 to 500 Hz available to most bimodal CI users implanted in recent years (e.g., Gantz et al. 2016), and our results should, therefore, not be interpreted as generalizable to the CI population.

As mentioned earlier, low-frequency acoustic speech improves speech understanding in noise, both for normal-hearing listeners presented with vocoded speech (Brown & Bacon 2009; Dorman et al. 2005; Kong & Carlyon 2007) and for CI

users with residual hearing (Kiefer et al. 2005; Kong et al. 2005). Conversely, available low-pass–filtered speech allows a specific desired intelligibility level to be achieved at less favorable, that is, lower SNRs (Büchner et al. 2009; Dorman & Gifford 2010; Qin & Oxenham 2006). Our results are in line with this trend. In Experiment 3, the vocoded speech was presented at 7.3 dB SNR to achieve 79% intelligibility, and the speech stimuli with low-pass–filtered speech at, on average, 1.9 (range, 0.9 to 3.1) dB SNR, a 5.4 dB lower SNR. In Experiment 2, the vocoded speech was presented at 2.7 dB SNR to achieve 50% intelligibility, and the stimuli with added low-pass–filtered speech at, on average, 0 (range, −0.7 to 0.9) dB SNR, a 2.7 dB lower SNR. These values are very similar to between-group values reported for CI users; Dorman and Gifford (2010) showed that speech reception thresholds (at 50% intelligibility) were on average 2.62 dB SNR lower for EAS listeners than for unilateral CI users. On the basis of research that has shown that less favorable, lower SNR can result in higher listening effort (Wu et al. 2016; Zekveld et al. 2010), we would expect increased listening effort at these lower SNRs. Our results, however, suggest that added low-frequency acoustic signal can offset these adverse effects of increased noise interference; in Experiment 2, low-pass–filtered speech improved listening effort compared with monaural vocoder, despite being presented at less favorable SNRs. In Experiment 3, while the results did not show an improvement in listening effort for added low-pass–filtered speech, neither did they show an increase in listening effort due to the 5.4 dB lower SNR for these conditions.

While our results did show some effects of low-frequency acoustic speech on the behavioral, dual-task measure of listening effort, our subjective self-report measure of listening effort, the NASA-TLX, showed no difference in perceived effort between any of the conditions of interest. These findings are in line with hypothesis 4, as well as our previous research (Pals et al. 2013). In the aforementioned previous study, the NASA-TLX revealed no significant differences in perceived effort between conditions that resulted in near-ceiling intelligibility, while the dual-task RTs captured differences between two of those conditions (Pals et al. 2013). In the present study, we specifically investigated listening effort at equal levels of intelligibility, and while the NASA-TLX results did not reveal any effects of processing, they did show a significant effect of intelligibility; participants rated slightly less intelligible speech as more effortful. Similarly, in our previous study, the NASA-TLX did show significant differences in perceived effort between conditions that also differed significantly in intelligibility (Pals et al. 2013). This might lead to the conclusion that the dual-task measure is more sensitive to listening effort while the subjective measure more closely reflects intelligibility.

Other research shows a similar lack of correspondence between objective and subjective measures of listening effort (e.g., Feuerstein 1992; Fraser et al. 2010; Gosselin & Gagné 2011a; Zekveld et al. 2010). However, not all of these support the suggestion that an objective measure may be more sensitive to changes in listening effort; in some cases, it appears the other way around. Feuerstein (1992), for example, reported that listening conditions that were less intelligible were also perceived as less "easy" to understand, while the dual-task RT measure did not distinguish between some of these conditions. Feuerstein suggests that the conscious awareness of reduced performance, that is, reduced intelligibility, may play a role in perceived

effort. Other objective measures of listening effort also appear not to correlate with perceived effort. Zekveld et al. showed that, on a group-level, the subjective and pupillometric measure of listening effort both indicated increased effort with decreasing intelligibility; however, individual participants' perceived effort did not correlate with their pupil dilation. These findings suggest that, rather than one measure being more sensitive than the other, perhaps, objective and subjective measures reflect different aspects of listening effort. Lemke and Besser (2016) suggest that listening effort encompasses both "processing effort" and "perceived effort," and we should take care to differentiate between the two. In the case of our results, we could conclude that the available low-frequency speech cues can, in some cases, improve processing effort, even if intelligibility and perceived effort are unaffected.

A limitation of the present study is that the reported effects of low-frequency acoustic speech on dual-task RTs are undeniably small: approximately 30 msec on RTs ranging from around 900 msec (for quiet listening conditions and speech intelligibility near-ceiling) to 1250 msec (for listening conditions around 0 dB SNR and speech intelligibility around 50%). At the outset of this study, we had decided against including baseline RT measures because our main interest was in comparing between the different listening conditions at similar speech intelligibility levels. However, in hindsight, it appears difficult to interpret the size of the effects reported in this study without baseline RTs. In our previous study, in which the same secondary visual RT task was used, a single-task visual RT task without auditory stimuli present was included as a baseline RT measure (Pals et al. 2013). These baseline RTs were on average around 900 msec, comparable to the fastest recorded RTs for speech at near-ceiling intelligibility in the present study. Considering a 30-msec reduction of RTs that are approximately 200 msec (Experiment 1, in quiet at near-ceiling intelligibility), or 350 msec (Experiment 2, at 0 dB SNR and 50% intelligibility) above baseline, may seem less insubstantial than a 30-msec reduction of a 1250 msec absolute RT. Dual-task paradigms using a RT task as secondary task often include one baseline measure: performance on the RT task without any of the auditory stimuli present (e.g., Hughes & Galvin 2013; Pals et al. 2014; Seeman & Sims 2015; Tun et al. 1991; Ward et al. 2017). Howard et al. (2010), however, included several baseline measures: one for each listening condition, with auditory stimulus present and the participant instructed to ignore the sound. Their results revealed a nonlinear relationship between SNR and baseline secondary task performance. While their secondary task was a memory task rather than a RT task, baseline RT task performance might similarly vary with SNR. A suggestion for future research is therefore to include baseline RT scores, not only in quiet, without speech present, but also with speech present, for a number of reference conditions relevant for the experiment.

Comparing dual-task RTs and significant effects reported across studies can pose another challenge. Across studies, the reported dual-task RTs vary greatly with different types of secondary RT tasks used. A probe RT task (e.g., Downs 1982; baseline RTs approximately 335 msec, dual-task RTs 439 to 475 ms) or visual identification task (e.g., Seeman & Sims 2015; baseline RTs 411 msec, dual-task RTs 427 and 477 msec at +15 and +5 dB SNR, respectively) typically result in short baseline RTs, while more complex secondary tasks result in longer RTs (e.g., Wu et al. 2016; easy RT task 300 to 400 msec,

difficult RT task approximately 750 to 1000 msec for a range of SNRs). The effects reported in these studies are substantially larger than the 30 msec reported in the present study. However, it is important to realize that in the studies referred earlier, conditions differed not only in RT but also in two other factors that are known to affect dual-task performance, namely SNR (Seeman & Sims 2015; Wu et al. 2016) and intelligibility (Payne et al. 1994). In the present study, on the other hand, intelligibility was controlled for and fixed with only slightly varying SNRs to achieve similar intelligibility across conditions. Dual-task studies comparing different types of processing at the same SNR, with slightly varying intelligibility across conditions, typically report shorter RTs. Sarampalis et al. (2009), for example, showed that a hearing-aid–like noise reduction algorithm can reduce dual-task RTs by approximately 50 msec at −6 dB SNR from about 740 to 690 msec. Gustafson et al. (2014) describe similar results using verbal response latencies rather than dual-task RTs; average verbal RTs were around 1300 to 1350 msec, and noise reduction provided a reduction of up to 40 msec. Our 30-msec reduction in RTs is, while still small, at least within a similar range when compared with studies with similar materials and designs. Considering a large part of our daily lives involves listening, even small changes in listening effort might have substantial effects on the listener in the long run, such as listening-related fatigue (Hick & Tharpe 2002). A suggestion for future research would be to systematically evaluate the relationship between instantaneous measures of listening effort, such as dual-task RTs, and effects of sustained listening, such as fatigue.

One curious finding in this study was the following: for speech in quiet, at near-ceiling intelligibility (Experiment 1), the dual-task RT results showed a significant benefit of additional low-frequency acoustic speech compared with binaural vocoder alone, but not compared with monaural vocoder. Although the RTs for monaural vocoded speech were on average longer than the RTs for vocoder plus low-pass–filtered speech, they were shorter than the average RTs for binaural vocoded speech and neither difference was significant. Intuitively, one would expect monaural, rather than binaural, vocoded speech to be the more effortful to understand; thus, we would expect longer, rather than shorter, RTs for monaural vocoder. What might have affected the RTs for the monaural vocoded speech is a difference in presentation level; to account for binaural loudness summation (Epstein & Florentine 2012), monaural vocoder was presented at a slightly higher sound level (65 dBA) than the binaural vocoder, and vocoder plus low-pass–filtered speech (60 dBA in each ear). Whether this resulted in exactly equal perceived loudness for the monaural compared with the other conditions is not certain. Differences in frequency content between the vocoder and low-pass–filtered speech signals may have affected perceived loudness as well. Therefore, it is possible that differences in level and perceived loudness have affected the dual-task RTs for monaural vocoder compared with the other conditions.

In summary, from the results of this study, we conclude that at equal levels of intelligibility, low-frequency acoustic speech can, in some cases, provide a benefit in listening effort, as reflected by the dual-task RTs, for the understanding of noise-band vocoded speech, at least in conditions in which the effect of the additional low-pass–filtered speech on listening effort is not overshadowed by the counter-directional effect of background noise on listening

effort. This conclusion is specifically based on our study with normal-hearing listeners and noise-band–vocoded speech complemented with low-pass–filtered speech. Although we intended to approximate CI and EAS hearing with our choice of stimuli, we should be careful to generalize to CI users. The low-frequency hearing available to CI users is often impaired and not comparable to low-pass–filtered speech perceived with normal hearing. This could mean that the speech cues that are available in the low-frequency acoustic speech may not be accessible to some EAS listeners depending on the magnitude of hearing loss and the health of the cochlea (Dorman & Gifford 2010). Therefore, EAS listeners may not experience the same benefits in listening effort as we found in normal-hearing listeners even from similar levels of low-frequency speech. However, as described earlier, Dorman and Gifford (2010) did report a benefit of EAS in terms of tolerance to masking noise for CI users, similar in magnitude to the improved SRTs reported for normal-hearing listeners in the present study. Gifford et al. (2013) described that CI listeners anecdotally report that speech understanding with additional binaural acoustic hearing to be much easier than with only the contralateral acoustic hearing available, despite providing only a small benefit in SRTs. This suggests that low-frequency residual hearing may provide a benefit in listening effort for CI users. At this point, we can only speculate. Whether EAS can provide a benefit in listening effort for CI users should be addressed in future research.

## REFERENCES

Baayen, R., Davidson, D. J., Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang*, *59*, 390–412.

Baayen, R., Piepenbrock, R., Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Barr, D. J., Levy, R., Scheepers, C., et al. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J Mem Lang*, *68*, 255–278.

Başkent, D. (2012). Effect of speech degradation on top-down repair: Phonemic restoration with simulations of cochlear implants and combined electric-acoustic stimulation. *J Assoc Res Otolaryngol*, *13*, 683–692.

Başkent, D., & Chatterjee, M. (2010). Recognition of temporally interrupted and spectrally degraded sentences with additional unprocessed low-frequency speech. *Hear Res*, *270*, 127–133.

Başkent, D., Gaudrain, E., Tamati, T., et al. (2016). Perception and psychoacoustics of speech in cochlear implant users. In A. Cacace, E. de Kleine, A. Genene Holt, P. van Dijk (Eds.) *Scientific Foundations of Audiology: Perspectives From Physics, Biology, Modeling, and Medicine* (p. 285). San Diego, CA: Plural Publishing.

Benard, M. R., & Başkent, D. (2013). Perceptual learning of interrupted speech. *PLoS One*, *8*, e58149.

Binns, C., & Culling, J. F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *J Acoust Soc Am*, *122*, 1765–1776.

Brown, C. A., & Bacon, S. P. (2009). Low-frequency speech cues and simulated electric-acoustic hearing. *J Acoust Soc Am*, *125*, 1658–1665.

Büchner, A., Schüssler, M., Battmer, R. D., et al. (2009). Impact of low-frequency hearing. *Audiol Neurootol*, *14*(Suppl 1), 8–13.

Classon, E., Rudner, M., Rönnberg, J. (2013). Working memory compensates for hearing related phonological processing deficit. *J Commun Disord*, *46*, 17–29.

Cullington, H. E., & Zeng, F. G. (2010). Bimodal hearing benefit for speech recognition with competing voice in cochlear implant subject with normal hearing in contralateral ear. *Ear Hear*, *31*, 70–73.

Cutler, A., Pasveer, D. (2006). Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition. In R. Hoffmann & H. Mixdorff(Eds.), *3rd International Conference on Speech Prosody*. Dresden, Germany: TUD Press.

Dorman, M. F., & Gifford, R. H. (2010). Combining acoustic and electric stimulation in the service of speech recognition. *Int J Audiol*, *49*, 912–919.

Dorman, M. F., Spahr, A. J., Loizou, P. C., et al. (2005). Acoustic simulations of combined electric and acoustic hearing (EAS). *Ear Hear*, *26*, 371–380.

Downs, D. W. (1982). Effects of hearing and use on speech discrimination and listening effort. *J Speech Hear Disord*, *47*, 189–193.

Dudley, H. (1939). The automatic synthesis of speech. *Proc Natl Acad Sci U S A*, *25*, 377.

Epstein, M., & Florentine, M. (2009). Binaural loudness summation for speech and tones presented via earphones and loudspeakers. *Ear Hear*, *30*, 234–237.

Epstein, M., & Florentine, M. (2012). Binaural loudness summation for speech presented via earphones and loudspeaker with and without visual cues. *J Acoust Soc Am*, *131*, 3981–3988.

Feuerstein, J. F. (1992). Monaural versus binaural hearing: Ease of listening, word recognition, and attentional effort. *Ear Hear*, *13*, (p. 367)80–86.

Fraser, S., Gagné, J. P., Alepins, M., et al. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: The effects of providing visual speech cues. *J Speech Lang Hear Res*, *53*, 18–33.

Friesen, L. M., Shannon, R. V., Baskent, D., et al. (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.

Gagné, J. P., Besser, J., Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm. *Trends Hear*, *21*, 1–25.

Gantz, B. J., Dunn, C., Oleson, J., et al. (2016). Multicenter clinical trial of the Nucleus Hybrid S8 cochlear implant: Final outcomes. *Laryngoscope*, *126*, 962–973.

Gatehouse, S. (1990). The role of non-auditory factors in measured and self-reported disability. *Acta Otolaryngol Suppl*, *476*, 249–256.

Gifford, R. H., Dorman, M. F., Skarzynski, H., et al. (2013). Cochlear implantation with hearing preservation yields significant benefit for speech recognition in complex listening environments. *Ear Hear*, *34*, 413–425.

Gosselin, P. A., & Gagné, J. P. (2011a). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *Int J Audiol*, *50*, 786–792.

Gosselin, P. A., & Gagné, J. P. (2011b). Older adults expend more listening effort than young adults recognizing speech in noise. *J Speech Lang Hear Res*, *54*, 944–958.

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

Gstoettner, W. K., van de Heyning, P., O'Connor, A. F., et al. (2008). Electric acoustic stimulation of the auditory system: Results of a multi-centre investigation. *Acta Otolaryngol*, 128, 968–975.

Gustafson, S., McCreery, R., Hoover, B., et al. (2014). Listening effort and perceived clarity for normal-hearing children with the use of digital noise reduction. *Ear Hear*, 35, 183–194.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA TLX (task load index): Results of empirical and theoretical research. *Human Mental Workload*, 1, 139–183.

Hick, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *J Speech Lang Hear Res*, 45, 573–584.

Howard, C. S., Munro, K. J., Plack, C. J. (2010). Listening effort at signal-to-noise ratios that are typical of the school classroom. *Int J Audiol*, 49, 928–932.

Hughes, K. C., & Galvin, K. L. (2013). Measuring listening effort expended by adolescents and young adults with unilateral or bilateral cochlear implants or normal hearing. *Cochlear Implants Int*, 14, 121–129.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Karsten, S. A., Turner, C. W., Brown, C. J., et al. (2013). Optimizing the combination of acoustic and electric hearing in the implanted ear. *Ear Hear*, 34, 142–150.

Kiefer, J., Pok, M., Adunka, O., et al. (2005). Combined electric and acoustic stimulation of the auditory system: Results of a clinical study. *Audiol Neurootol*, 10, 134–144.

Kong, Y. Y., & Carlyon, R. P. (2007). Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation. *J Acoust Soc Am*, 121, 3717–3727.

Kong, Y. Y., Stickney, G. S., Zeng, F. G. (2005). Speech and melody recognition in binaurally combined acoustic and electric hearing. *J Acoust Soc Am*, 117(3 Pt 1), 1351–1361.

Lemke, U., & Besser, J. (2016). Cognitive load and listening effort: Concepts and age-related considerations. *Ear Hear*, 37(Suppl 1), 77–84.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*, 49(Suppl 2), 467–477.

Neel, A. T. (2009). Effects of loud and amplified speech on sentence and word intelligibility in Parkinson disease. *J Speech Lang Hear Res*, 52, 1021–1033.

Noble, W., Tyler, R., Dunn, C., et al. (2008). Unilateral and bilateral cochlear implants and the implant-plus-hearing-aid profile: Comparing self-assessed and measured abilities. *Int J Audiol*, 47, 505–514.

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

Pals, C., Sarampalis, A., Baskent, D. (2013). Listening effort with cochlear implant simulations. *J Speech Lang Hear Res*, 56, 1075–1084.

Pals, C., Sarampalis, A., Başkent, D. (2014). Listening effort in cochlear implant users. *Association for Research in Otolaryngology Mid Winter Meeting*. San Diego, CA.

Payne, D. G., Peters, L. J., Birkmire, D. P., et al. (1994). Effects of speech intelligibility level on concurrent visual task performance. *Hum Factors*, 36, 441–475.

Pichora-Fuller, M. K., Schneider, B. A., Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J Acoust Soc Am*, 97, 593–608.

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *J Speech Hear Res*, 29, 146–154.

Qin, M. K., & Oxenham, A. J. (2006). Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech. *J Acoust Soc Am*, 119, 2417–2426.

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *Int J Audiol*, 42(Suppl 1), S68–S76.

Rönnberg, J., Lunner, T., Zekveld, A., et al. (2013). The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances. *Front Syst Neurosci*, 7, 31.

Rönnberg, J., Rudner, M., Foo, C., et al. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *Int J Audiol*, 47(Suppl 2), S99–S105.

Sarampalis, A., Kalluri, S., Edwards, B., et al. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *J Speech Lang Hear Res*, 52, 1230–1240.

Seeman, S., & Sims, R. (2015). Comparison of psychophysiological and dual-task measures of listening effort. *J Speech Lang Hear Res*, 58, 1781–1792.

Shannon, R. V., Zeng, F. G., Kamath, V., et al. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303–304.

Steel, M. M., Papsin, B. C., Gordon, K. A. (2015). Binaural fusion and listening effort in children who use bilateral cochlear implants: A psychoacoustic and pupillometric study. *PloS One*, 10, e0117611.

Tun, P. A., Wingfield, A., Stine, E. A. (1991). Speech-processing capacity in young and older adults: A dual-task study. *Psychol Aging*, 6, 3–9.

Turner, C., Gantz, B. J., Lowder, M. W., et al. (2005). Benefits seen in acoustic hearing+ electric stimulation in same ear. *Hear J*, 58, 53–55.

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *J Acoust Soc Am*, 126, 367–376.

Versfeld, N. J., Daalder, L., Festen, J. M., et al. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, 107, 1671–1684.

von Ilberg, C., Kiefer, J., Tillein, J., et al. (1999). Electric-acoustic stimulation of the auditory system. New technology for severe hearing loss. *ORL J Otorhinolaryngol Relat Spec*, 61, 334–340.

Wagner, A. E., Toffanin, P., Başkent, D. (2016). The timing and effort of lexical access in natural and degraded speech. *Front Psychol*, 7, 1–4.

Ward, K. M., Shen, J., Souza, P. E., et al. (2017). Age-related differences in listening effort during degraded speech recognition. *Ear Hear*, 38, 74–84.

Wild, C. J., Yusuf, A., Wilson, D. E., et al. (2012). Effortful listening: The processing of degraded speech depends critically on attention. *J Neurosci*, 32, 14010–14021.

Wingfield, A. (1996). Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory. *J Am Acad Audiol*, 7, 175–182.

Winn, M. B., Edwards, J. R., Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear Hear*, 36, e153–e165.

Wu, Y. H., Stangl, E., Zhang, X., et al. (2016). Psychometric functions of dual-task paradigms for measuring listening effort. *Ear Hear*, 37, 660–670.

Zekveld, A. A., Kramer, S. E., Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear Hear*, 31, 480–490.

Zhang, T., Dorman, M. F., Spahr, A. J. (2010a). Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation. *Ear Hear*, 31, 63–69.

Zhang, T., Spahr, A. J., Dorman, M. F. (2010b). Frequency overlap between electric and acoustic stimulation and speech-perception benefit in patients with combined electric and acoustic stimulation. *Ear Hear*, 31, 195–201.

## REFERENCE NOTE

Başkent, D., Luckmann, A., Ceha, J., et al. (2018). The discrimination of voice cues in simulations of bimodal electro-acoustic cochlear-implant hearing. *JASA Express Letters*. In press.