SoftVoice improves speech recognition and reduces listening effort in cochlear implant users

Stronks, H.C.; Apperloo, E.; Koning, R.; Briaire, J.J.; Frijns, J.H.M.

Citation

Stronks, H. C., Apperloo, E., Koning, R., Briaire, J. J., & Frijns, J. H. M. (2021). SoftVoice improves speech recognition and reduces listening effort in cochlear implant users. *Ear And Hearing*, *42*(2), 381-392. doi:10.1097/AUD.0000000000000928

SoftVoice Improves Speech Recognition and Reduces Listening Effort in Cochlear Implant Users

H. Christiaan Stronks,[1] Eline Apperloo,[1] Raphael Koning,[2] Jeroen J. Briaire,[1]
and Johan H. M. Frijns[1,3]

Objectives: The ability to perceive soft speech by cochlear implant (CI) users is restricted in part by the inherent system noise produced by the speech processor, and in particular by the microphone(s). The algorithm "SoftVoice" (SV) was developed by Advanced Bionics to enhance the perception of soft speech by reducing the system noise in speech processors. The aim of this study was to examine the effects of SV on speech recognition and listening effort.

Design: Seventeen adult Advanced Bionics CI recipients were recruited and tested in two sessions. The effect of SV on speech recognition was tested by determining the SRT in quiet using the Matrix test. Based on the individual subjects' SRTs, we investigated speech-recognition scores at fixed speech levels, namely SRT –5 dB, SRT +0 dB, SRT +5 dB, and SRT +10 dB, again in quiet and using the Matrix test. Listening effort was measured at each of these speech levels subjectively by using a rating scale, and objectively by determining pupil dilation with pupillometry. To verify whether SoftVoice had any negative effects on speech perception in noise, we determined the SRT in steady state, speech-weighted noise of 60 dBA.

Results: Our results revealed a significant improvement of 2.0 dB on the SRT in quiet with SoftVoice. The average SRT in quiet without SoftVoice was 38 dBA. SoftVoice did not affect the SRT in steady state, speech-weighted noise of 60 dB. At an average speech level of 33 dBA (SRT –5 dB) and 38 dBA (SRT +0 dB) in quiet, significant improvements of 17% and 9% on speech-recognition scores were found with SoftVoice, respectively. At higher speech levels, SoftVoice did not significantly affect speech recognition. Pupillometry did not show significant effects of SoftVoice at any speech level. However, subjective ratings of listening effort indicated a decrease of listening effort with SoftVoice at a speech level of 33 dBA.

Conclusions: We conclude that SoftVoice substantially improves recognition of soft speech and lowers subjective listening effort at low speech levels in quiet. However, no significant effect of SoftVoice was found on pupil dilation. As SRTs in noise were not statistically significantly affected by SoftVoice, we conclude that SoftVoice can be used in noisy listening conditions with little negative impact on speech recognition, if any. The increased power demands of the algorithm are considered to be negligible. It is expected that SoftVoice will reduce power consumption at low ambient sound levels. These results support the use of SoftVoice as a standard feature of Advanced Bionics CI fittings for everyday use.

Key words: Cochlear Implants, Front-end Processing, Intelligibility, Noise, Speech, Sensorineural hearing loss, Pupillometry.

(Ear & Hearing 2021;42;381–392)

INTRODUCTION

Cochlear implantation is the standard of care for people with severe to profound sensorineural hearing loss. It is an effective method to improve speech perception (Forli et al. 2019) and to improve quality of life (Klop et al. 2007; Gaylor et al. 2013). State-of-the-art cochlear implant (CI) devices generally provide good speech recognition in quiet listening conditions, provided that the speech is presented at a normal speech level (Shannon et al. 2004). However, such ideal listening conditions rarely occur in everyday life. Most real-life listening conditions are more adverse, such as when listening to soft speech or listening in noisy conditions. Soft speech of 30 to 45 dB corresponds to the sound level of a quiet whisper and that of a quiet office. It can also be encountered when being spoken to by children, or when listening to someone speaking from another room (HSE Offices 2005; SCENIHR 2008; Holden et al. 2011; Canada Safety Council 2019). These speech levels are challenging for CI users, given that their speech recognition is only 50% at approximately 38 dBA in quiet (this paper). Understanding of soft speech, and speech understanding in adverse listening conditions in general, are challenging and may lead to reduced speech recognition (Spahr et al. 2007) and increased listening effort (Devocht et al. 2017).

Advanced Bionics (Valencia, CA) developed the SoftVoice (SV) algorithm to optimize CI users' speech recognition at low sound levels by removing system noise, and in particular, the noise introduced by the microphone(s) (Holden et al. 2019). The algorithm is implemented after the frequency analysis in the speech processing chain, and it works as a filter within each frequency band. The algorithm determines the signal energy across the last 30 ms in each channel. If this energy is lower than a certain channel-specific threshold level, the algorithm discards the information and the electrode is not stimulated. The threshold values are fixed and have been determined for each band individually based on measurements of system and microphone noise. If the signal energy within a certain band exceeds the threshold value, it will be further processed and become audible to the CI user. In effect, this means that SV reduces noise at low sound levels but does not affect speech processing at normal to high sound levels.

Holden et al. (2019) showed that SV improved the recognition of soft speech. The authors investigated the impact of SV on sentence-recognition scores at fixed speech levels. However, they did not assess the effects of SV on listening effort. Listening effort has been defined as the attention and cognitive resources required to understand speech (Hick & Tharpe 2002; Gosselin & Gagné 2011). CI users experience more listening effort than listeners with normal hearing, especially in challenging listening situations (Hällgren et al. 2005; Ohlenforst et al. 2017; Perreau et al. 2017; Hughes et al. 2018). Effortful

---

[1]Department of Otorhinolaryngology, Leiden University Medical Center, the Netherlands; [2] Advanced Bionics GmbH, European Research Center, Hannover, Germany; and [3] Leiden Institute for Brain and Cognition, Leiden University, Leiden, the Netherlands.

<zdoi; 10.1097/AUD.0000000000000928>

listening may lead to fatigue (Festen & Plomp 1990; Kramer et al. 2006), which in turn might negatively impact speech understanding. Therefore, it is important to minimize listening effort in CI users (Alhanbali et al. 2017). Changes in listening effort can be quantified using subjective, physiological, or behavioral techniques. A commonly used objective measure for listening effort is pupillometry, which makes use of the finding that pupil size increases in more effortful listening conditions and decreases when listening costs less effort (Kramer et al. 1997; Zekveld et al. 2010; Winn 2016; Winn & Moore 2018). Subjectively, listening effort can be quantified using a Likert rating scale, by asking participants to rate the effort it took them to understand the speech material provided (Luts et al. 2010; Zekveld et al. 2010; Brons et al. 2013).

Study Aims

In this study, we extend the study of Holden et al. (2019) by examining the effects of SV on listening effort using pupillometry and a subjective rating scale. In addition, we examined the effects of SV on the speech reception threshold (SRT) in quiet. We also investigated speech-recognition scores at fixed levels of speech. In contrast to Holden et al. (2019), we based these levels on the participant's individual SRT to ensure that every participant was able to perform the task, yet that the task was challenging at soft speech levels. Our hypothesis was that the effects of SV on speech recognition and listening effort are higher at lower speech levels. As a negative control, we assessed the effect of SV on the SRT when determined in steady state, speech-weighted noise of 60 dB. Under these conditions, SV should not affect speech recognition.

MATERIALS AND METHODS

Study Design

This study was a prospective, cross-over study with a single-masked design, in which only the participants were masked from the speech-processing condition being tested. The outcome parameters were measures of speech recognition determined in quiet, namely the SRT (i.e., the speech level required for successfully reproducing 50% of the words) and the speech-recognition score at fixed speech levels. Pupil dilation and a subjective rating scale were used to measure listening effort (Zekveld et al. 2010). SRT in noise was determined to investigate whether SV has any negative effects when listening to louder speech. This study consisted of two test sessions of 2 to 3 hr each.

Participants

The required sample size was based on a power analysis using previously collected data showing that SRTs have a SD of 1.3 dB. We considered an improvement of 1 dB in SRT clinically relevant. Using a conservative SD of 1.4 dB, the effect size *d* was estimated at $d = 1$ dB/1.4 dB = 0.71. An *a priori* power analysis using G*Power 3 (Kiel, Germany) (Faul et al. 2007) indicated that a sample of at least 13 participants would be needed to detect a clinically significant difference of 1 dB in SRT with 80% power and $\alpha = 0.05$. Five additional participants were recruited to account for possible attrition.

Eighteen postlingually deafened adult CI recipients (7 males and 11 females) were recruited from the Leiden University Medical Center. One participant (S09) was excluded from the study, because the participant's SRT in quiet was poor (53 dBA), namely more than 3·SD above the mean speech-recognition score of the other 17 participants (38 dBA, SD, 4.5). The 17 included participants were unilaterally implanted with an Advanced Bionics device unilaterally (n = 16) or bilaterally (n = 1). All used the high-resolution (HiRes Optima) speech-processing strategy with an Input Dynamic Range (IDR) of 60 dB, except for two participants that used an IDR of 55 . All of the participants had experience with the Matrix test before being included in this study. Their residual hearing in the contralateral ear, expressed as the pure-tone average across 500 to 4,000 Hz ($PTA_{500–4000}$), ranged from 57.5 to 130 dB (mean, 91.9 dB; SD, 21.6). Demographics and device information of the participants are shown in Table 1. Participants had no disorders that could affect the study results, other than a hearing impairment. One participant (S05) had Usher's syndrome and was visually impaired, but the pupil response appeared to be normal. All participants provided informed consent. This study was approved by Institutional Review Board of the Leiden University Medical Center and adhered to the tenets of the Declaration of Helsinki (World Medical Association 2013).

Experimental Setup

Participants were supplied with a research speech processor (Q90, Advanced-Bionics LLC, Valencia, CA) on the day of testing. It was fitted with their own threshold and maximal comfortable stimulus levels. Participants with bimodal hearing were asked to remove their hearing aid during testing. One participant with substantial residual hearing in the contralateral ear ($PTA_{500–4000}$ of 58 dB) was asked to wear an earplug in the non-implanted ear during testing. Testing was performed in a soundproof booth. Participants were seated in the center of the room and speech was delivered by a single loudspeaker (MSP5A monitor speaker, Yamaha Corp., Japan) placed at a distance of 1 m in front of the participant. Most speech tests were performed without background noise, referred to from here on as speech testing in quiet. For testing in noise, we used a homogeneous field of steady state, speech-weighted noise with a loudness of 60 dBA. The homogeneous noise field was created by eight surround loudspeakers (Control 1, JBL Corp., Los Angeles, CA) distributed symmetrically around the listener in 3D space. Four of the loudspeakers were positioned in the top corners of the room. The others were placed in the middle of the side panels of the booth, approximately 40 cm above the floor. The booth measured 3.4 × 3.2 × 2.4 m (l × w × h). Each of the eight loudspeakers was calibrated individually to yield an identical noise level as the other loudspeakers in the middle of the field, adding up to a noise level of 60 dBA where the subject's head was located. The principle of creating a diffuse, homogeneous noise field as applied in this study has been described in more detail previously (Soede et al. 1993; Van der Beek et al. 2007).

The system noise of the Q90 processor was recorded using a digital oscilloscope (LabNation SmartScope, Antwerp, Belgium). To this end, a manikin (KEMAR) was placed in the test booth at the same position as the participants were normally seated. The manikin was equipped with a speech processor connected to a dummy implant. We measured the effect of SV on the activation pattern of the 16 channels of the dummy implant in quiet, and when playing a random sentence of the Matrix test. Figure 1 demonstrates how the activation pattern of one of the

TABLE 1. Participant Demographic, Device, and Sound Processor Information

Subject	Sex	Age	Listening Mode	CI	CI Use (years)	Duration of deafness CI ear (years)	$\text{PTA}_{500–4000}$	Etiology	Clinical CVC score (%)
S01	F	77	Bimodal	MS	R: 2.9	9.7	85	Progressive HL	93
S02	F	74	Bimodal	MS	R: 5.1	7.7	65	Sudden deafness	86
S03	M	83	Unilateral	MS	L: 2.0	13.8	93	Progressive HL	84
S04	M	75	Bimodal	MS	L: 3.4	9.5	70	DNFA9 mutation	92
S05	F	63	Bimodal	1j	L: 7.9	63	120	Ushers syndrome	93
S06	F	65	Bimodal	MS	R: 2.4	17.7	73	Progressive HL	95
S07	M	69	Unilateral	MS	R: 5.7	8.1	93	DNFA9 mutation	90
S08	M	73	Unilateral	1j	L: 7.8	23.2	68	Meniere’s disease	97
S10	M	72	Unilateral	MS	L: 4.2	34.9	120	Familial	90
S11	F	63	Bimodal	MS	R: 3.8	13.3	115	Progressive HL	85
S12	M	71	Bimodal	1j	L: 6.4	31.1	120	Progressive HL	94
S13	F	68	Unilateral	1j	R: 13.0	68	100	Congenital	90
S14	F	61	Bimodal	1j	R: 7.2	45.9	113	Unknown	93
S15	F	63	Unilateral	1j	R: 6.7	30.3	58	Meniere’s disease	100
S16	F	63	Bimodal	1j	R: 4.5	12.0	60	DNFA9 mutation	96
S17	F	52	Bilateral	AD: CII AS: 1j	R: 17.2 L: 1.8	R: 44.6 L: 44.6	120 120	Meningitis	98
S18	F	67	Bimodal	MS	R: 3.4	14.3	83	Sudden deafness	88
Means	11 F 6 M	68.2			6.1	25.2	93		92

CI devices: MS, HiRes 90K HiFocus Mid-Scala; 1j, HiRes 90K HiFocus 1j; CII Clarion CII HiFocus II with positioner; L, left ear; R, right ear; $\text{PTA}_{500–4000}$, median pure-tone audiogram across 500, 1000, 2000, 4000 Hz; HL, hearing loss; CVC, consonant–vowel–consonant.

channels shows a dramatic reduction in the pulses generated in response to system noise and the background noise in the soundproof booth when SV is activated. Note that the speech was presented at 35 dBA, that is, approximately 30 dB below normal conversational level.

Testing of Speech Recognition

As a measure of speech recognition we determined the SRT, that is, the speech level required for successfully reproducing 50% of the words, using the Dutch/Flemish Matrix test (Luts et al. 2014). Each run consisted of 20 sentences, each sentence consisting of five words with the same fixed grammatical syntax. To determine the SRT, the speech level was adaptively varied using the up-down method (Levitt 1971). If more than 50% of the words were repeated correctly, the speech level of the next sentence was decreased. If the correct score was less than 50%, the sound level of the next sentence was increased. Step size of the speech level was adaptively decreased after each reversal to a minimum of 0.1 dB. The step size reduction depended on the correct score in the previous trial. Typically, the speech level varied several dB in the first few trials of the run, while in the later trials, when the speech level converged onto the 50% word score, the variation was not more than 0.2 dBA. No feedback was given during the tests. The speech levels of the last eight trials (including the level that would have been played on a 21st trial, given the result of the final 20th trial) were averaged for the final SRT. Each session started with two training runs to reduce learning effects (De Jong et al. 2019b). One training run was applied in quiet, the other in steady state noise. The obtained SRTs during training were used to set the starting speech level of the SRT tests. The speech tests were carried out using custom-built software in a MATLAB R2017b programming environment (MathWorks, Inc., Natick, MA, USA). Participants were instructed to repeat every word they heard after the presentation of each sentence. Guessing was allowed. The scoring was performed manually by the experimenter. All participants had prior experience with the Matrix test.

SRTs were acquired four times in each session, namely in quiet and in noise, with SV switched off (SV_{off}) and SV switched on (SV_{on}) in both conditions. A test and re-test of the SRT were obtained across the two sessions. Note that the SRT is typically determined in background noise and hence usually represents the SNR where speech recognition is 50%. Because no background noise was used in this study, other than the system noise of the speech processor, we instead report absolute speech levels as a measure of the SRT obtained in quiet.

The SRT obtained in quiet was subsequently used as basis for determining the speech-recognition scores at four individual, fixed speech levels, namely SRT −5 dB, SRT +0 dB, SRT +5 dB, and SRT +10, again using the Matrix test. Taking the SRTs as a basis for the fixed speech levels assured that the task was challenging, yet manageable for each individual participant. Speech-recognition scores were assessed in the second half of the session. Pupillometry was performed during these tests as a proxy for listening effort. At each speech level, three listening conditions were tested (in randomized order), namely (1) SV_{off} and (2) SV_{on} both at the SRT obtained with SV_{off}, and (3) SV_{on} using the SRT acquired with SV_{on}. The last condition was included to verify if any difference in speech recognition and listening effort could be explained by SV’s effect on the SRT. Suppose a subject has an SRT of 38.2 dB with SV_{off}, and a more favorable SRT of 36.2 dB with SV_{on}, then the applied speech levels at SRT + 0 dB for the three conditions would become: (1) 38.2 dB (SV_{off}), (2) 38.2 dB (SV_{on}), and (3) 36.2 dB (SV_{on}), respectively. Two fixed speech levels were tested per session. In contrast to the SRT, no retest was obtained for the speech-recognition scores.

The test conditions, including speech level, SV_{on}/SV_{off} and Matrix list number (13 lists were available) were randomized

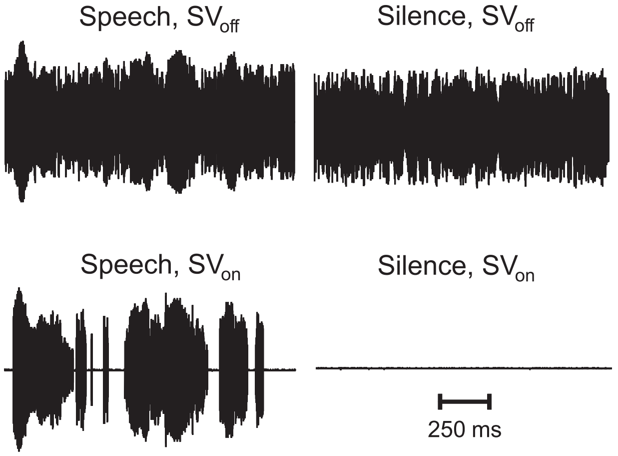

Fig. 1. Activation pattern of electrode channel 4 with and without SoftVoice (SV_{on} resp. SV_{off}) in quiet and during a spoken sentence from the Dutch/Flemish Matrix speech corpus presented at 35 dBA.

across sessions and participants. Per session, lists were only used once, but they were re-used across both test sessions. To limit learning effects during testing (De Jong et al. 2019a), each session was started by two adaptive runs, as recommended by Kollmeier et al. (2015). One of these practice runs was performed in quiet, the other in noise. Learning effects were accounted for in the linear mixed model (LMM), as outlined below.

Subjective Listening Effort and Performance Rating

After completing each run, participants were asked to subjectively rate their listening effort and to estimate their performance on a 9-point Likert scale, based on the rating labels provided by Zekveld et al. (2011). The following are English translations of the Dutch labels used for rating the performance in terms of speech recognition. The corresponding numbers on the scale are given between parentheses. "*None of the words were correct*" (1), "*approximately 25% of the words were correct*" (3), "*approximately 50% of the words were correct*" (5), "*approximately 75% of the words were correct*" (7), "*all words were correct*" (9). For the listening effort rating a similar scale was presented with the following labels (Zekveld et al. 2011): "*no effort"* (1), "*low effort"* (3), "*moderate effort*" (5), "*high effort*" (7), and "*very high effort*" (9). The subjects were instructed to rate listening effort, but to ignore overall task difficulty, fatigue, or concentration effort.

Pupillometry

Pupillometry was performed using eye-tracking glasses (ETG 2.6, SensoriMotor Instruments, Berlin, Germany). The ETG system uses infrared, video-based tracking technology to follow the gaze and measure pupil diameter at a spatial resolution of 1280 × 960 pixels, at a sampling rate of 120 Hz. To assure that the room lighting was appropriate to support the pupil to both dilate and constrict, the pupil diameter was measured at five different light intensities between 40 and 540 lx using dimmable fluorescent lights. An illuminance was selected at which the pupil response was approximately in the middle of its dynamic range and preferably in the reported mid-dynamic range of 4.5–5.0 mm (Watson & Yellott 2012).

Our method of recording the pupil response was based on the method reported by Zekveld et al. (2010). Each of the 20 trials started with a relatively long baseline recording of 8 sec. This relatively long baseline was chosen based on our observation that some pupils need sufficient time to fully recover (unpublished observations). Subsequently, the sentence was presented, which lasted approximately 2 sec. After the sentence, another 3 sec of silence was introduced. A probe tone was then presented after which the participant was allowed to respond. Both pupils were recorded and analyzed. Signal processing of the pupillometry data was based on previously published methodology (Zekveld et al. 2010, 2011). Eye blinks were defined as recorded pupil diameters smaller than 2 mm. Pupil sizes larger than 8 mm were similarly treated as recording artifacts, as these values are outside the normal dynamic range of pupil size (Watson & Yellott 2012). Eye blinks were removed by linear interpolation between 10 samples before the start of the blink, to 10 samples after the end of the detected blink. The last second of the 8 sec of baseline was used to determine the baseline pupil diameter by averaging the available 120 samples. After deblinking, the signal was low-pass filtered using a sixth-order Butterworth filter with a cutoff frequency of 5 Hz.

Three criteria were built into the signal-processing pipeline to discard traces with excess blink contaminations. These were (1) when 50% of the baseline samples were interpolated, that is, ≥ 0.5 sec from the 1 sec available; (2) when 33% or more of the postbaseline recording was interpolated (≥ 4 s from the available 12 sec); and (3) when there was an uninterrupted stretch of interpolation present that exceeded 0.5 sec, as it was regarded as a nap. The remaining accepted traces were baseline corrected to reduce signal drift. At this point, those traces with an overall amplitude (absolute maximum minus absolute minimum of the entire sweep) larger than the mean amplitude + 2·SD of the ensemble average were discarded. This step rejected those traces that showed gross residual artifacts, for example, from partially removed eye blinks. Subsequently the remaining traces were ensemble-averaged to reduce background noise. If the total number of included waveforms from both eyes was less than five, the recording was discarded. This resulted in two instances of missing data, namely one condition in two different subjects. Because we used linear mixed modeling for the analysis of the pupillometry, these missing data were accounted for. On average, 18 sweeps per eye were included, adding up to 36 traces per pupil response.

Peak amplitudes were determined in the resulting bilateral response between 1 sec and 4 sec postbaseline (P_1) and between 5 sec and 12 sec postbaseline (P_2). P_1 reflects the maximal listening effort during sentence presentation, while P_2 corresponds to the maximum effort during the response phase of the trial. P_2, hence, is expected to present a combination of speech processing and response preparation (Winn et al. 2018). Only P_1 was ultimately used, as P_2 showed a high within- and between-participant variability (results not shown). Outcome measures were P_1 peak amplitude relative to baseline, area under the curve (auc) from sentence start to P_1, and P_1 latency (Figure 2). Last, the outcome measures from both pupils were averaged, such that each pupil had an equal weight (i.e., no weighting factor was included to correct for the number of sweeps obtained in the waveforms).

Statistical Analyses

Statistical analyses were performed using GraphPad Prism version 8.0.1 for Windows (GraphPad Software, San

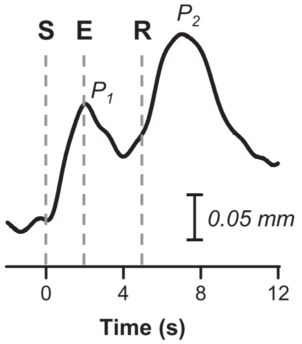

Fig. 2. Sample average pupil response (N = 17) obtained at each subject's own SRT (38 dBA on average). S indicates sentence start; E, sentence end; R, response start; P_1, peak 1; P_2, peak 2; SRT, speech recognition threshold.

Diego, CA), except the construction of LMMs, which was performed with IBM SPSS Statistics for Windows (version 25.0, IBM Corp. Armonk, NY). D'Agostino and Pearson's normality tests were carried out for SRTs, speech-recognition scores, and speech levels. LMMs were used to evaluate the effects of SV on the SRT in quiet and in noise, on speech recognition in quiet, and on pupil dilation in quiet. LMM allowed for the inclusion of potential confounding factors, namely the subject's age, learning effects (session and trial number), fatigue and motivation (trial number), and the average threshold level (T-level) of all the active electrodes in clinical units. Age is known to affect speech understanding and the pupil response (Plomp & Mimpen 1979; Zekveld et al. 2011). The Matrix test has been associated with learning effects across trials and between sessions (De Jong et al. 2019a). Fatigue and motivation can affect the pupil response (Winn et al. 2018), and T-levels have been implicated in the effectiveness of SV (Holden et al. 2019). Categorical, nominal variables ($\mathrm{SV}_{on}/\mathrm{SV}_{off}$, and $\mathrm{noise}_{on}/\mathrm{noise}_{off}$) and discrete, ordinal variables (trial and session number) were entered as fixed-effects factors, while continuous variables (age, mean T-level) were entered as fixed-effects covariates. These parameters were entered as fixed effects and not as random variables, according to the guidelines of Brauer and Curtin (2018). A fixed-effects intercept was included. Subject ID was entered as a random, nominal, categorical variable, including a random intercept (Brauer & Curtin 2018). The fixed-effects covariance matrix type was unstructured, as it represents the covariance structure with the least assumptions (Littell et al. 2000). The covariance structure for the random effect was set at identity. Differences between the subjective ratings of listening effort were tested for significance using a nonparametric test, namely Wilcoxon matched-pairs, signed-rank tests. Multiple comparisons testing was performed using post hoc Bonferroni-corrected *p* values, by multiplying the resulting *p* values with the number of conditions tested. Correlations were examined using linear regression, or nonparametrically with a Spearman test.

RESULTS

Speech Recognition

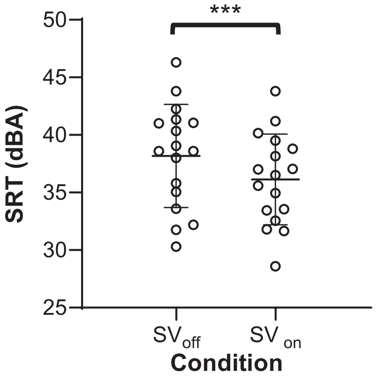

Fig. 3. Individual SRTs determined in quiet without and with SV. SRTs are the speech levels at which 50% speech recognition was reached. Horizontal line: mean; error bars: SD. $^{****}p < 0.001$. SRT indicates speech recognition thresholds; SV, SoftVoice.

To investigate the effect of SV on recognition of speech, we determined SRTs in quiet (Fig. 3) and in background noise (Fig. 4). All these SRT estimates, both with and without SV, passed D'Agostino and Pearson's normality tests ($p > 0.05$). Therefore, parametric statistics were used for the analysis of SRT. To this end, we constructed the following LMM:

$$SRT(dB) = Intercept + SV + N + (SV \cdot N) + Session + Trial + (a \cdot Age) + (b \cdot T)$$

where *SRT* was the dependent variable; *SV* and *N* the categorical variables SoftVoice and noise ($\mathrm{SV}_{on/}\mathrm{SV}_{off}$ and $\mathrm{noise}_{on}/\mathrm{noise}_{off}$); (*SV·N*) the interaction factor between *SV* and *N*; Session and Trial, the ordinal variables session number and trial number; age and *T* (T-level in clinical units), the continuous covariates; *a, b* the covariate coefficients; and Intercept, the fixed-effects intercept. SV had a significant main effect on the SRT ($F = 4.550$, $p = 0.035$), as had noise ($F = 165.9$, $p < 0.001$). There was a significant interaction between SV and noise ($F = 12.234$, $p = 0.001$), and session number and age had significant main effects as well ($F = 26.622$, $p < 0.001$ and $F = 20.058$, $p = 0.001$, respectively). The SRT was 1.8 dB lower, that is, more favorable in the second session, and age negatively impacted the SRT with 0.34 dB increase per year, on average. Trial number and T-level had no significant main effect on SRT ($F = 1.252$, $p = 0.294$ and $F = 0.014$, $p = 0.908$, respectively). The mean SRT in quiet without SV was 38.2 dBA. Estimation of the fixed effects showed that there was a significant improvement of the SRT with SV of 2.0 dB in quiet ($t = 3.982$, $p \leq 0.001$). As a negative control, the effect of SV

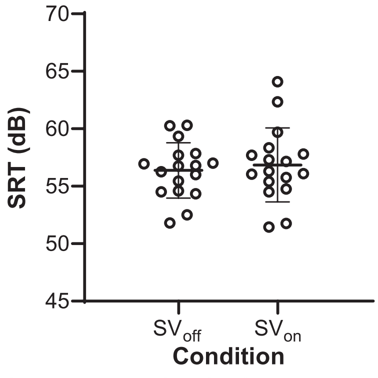

Fig. 4. Individual SRTs in noise with and without SV. Horizontal line: mean; error bars: SD. SRT indicates speech recognition threshold; SV, SoftVoice.

on the SRT in background steady state noise of 60 dBA was assessed. The mean SRT in noise was −3.6 dB, corresponding to a speech level of 56.4 dBA. SV had no significant effect on the SRT under this condition ($t = 0.965$, $p = 0.337$). Table 2 shows the parameter estimates of the LMM.

Having obtained the SRTs, we determined the effect of SV on speech recognition at four different, fixed speech levels based on the SRT in quiet, namely at SRT −5 dB, SRT +0 dB, SRT +5 dB, and SRT +10 dB. Figure 5 shows the mean speech-recognition scores in quiet at different speech levels. The speech-recognition scores passed D'Agostino and Pearson's normality tests. An LMM similar to the one used for the statistical significance testing of the SRT was built for the analysis of the speech- scores at fixed speech levels, namely:

$$\begin{aligned} Speech\,recognition(\%) = Intercept + SV + L + (SV \cdot L) + \\ Session + Trial + (a \cdot Age) + (b \cdot T) \end{aligned}$$

where speech recognition was the outcome variable; *SV* the categorical variable SoftVoice (SV_{on}/SV_{off}); *L* the continuous variable speech level relative to the individual SRT (−5, 0, +5, or +10 dB); ($SV{\cdot}L$) the interaction factor of *SV* and *L*. The remaining variables were identical to those in Eq. 1. The LMM showed significant main effects of SV and speech level ($F = 26.203$, $p < 0.001$ and $F = 222.7$, $p < 0.001$, respectively) and a significant interaction between the two factors ($F = 5.632$, $p = 0.001$). Age also significantly affected the speech-recognition score ($F = 5.067$, $p = 0.041$) and was negatively correlated to speech recognition (0.3% decrease in speech recognition per year). The remaining variables (session number, trial number, and T-level) did not ($F = 2.507$, $p = 0.116$; $F = 1.469$, $p = 0.206$; and $F = 2.120$, $p = 0.168$, respectively).

To test our hypothesis that SV's effects are greatest at low speech levels, we compared the fixed-effect estimates of SV on speech-recognition scores at the different speech levels using post hoc Bonferroni-corrected multiple comparisons *t* testing. SV significantly improved speech-recognition scores by 17.4% and 8.5% at SRT −5 dB and SRT +0, respectively ($t = 5.762$, adjusted $p < 0.001$, and $t = 2.842$, adjusted $p = 0.022$, respectively). Insignificant effects of SV were found at higher speech levels of SRT +5 dB and +10 dB ($t = 0.543$, adjusted $p = 1.0$ and $t = 1.068$, adjusted $p = 1.0$, respectively). We did not ask explicitly if participants noticed changes with the different programs, but three participants reported to perceive less background noise (air conditioning) when SV was turned on. Model estimates are provided in Table 3. Note that the values in the table differ slightly from the effects of SV in Figure 5. This difference is caused by the fact that the LMM values are corrected for multiple variables, whereas the figure shows simple averages.

To test whether SV's effects linearly decline with speech level, we performed a correlation analysis using linear regression between the benefit of SV and speech level at SRT –5 dB (Fig. 6). The benefit of SV was calculated by subtracting the speech-recognition score obtained with SV_{off} from the score obtained with SV_{on}. Note that SRTs were participant-specific and hence resulted in a range of different speech levels within the SRT –5 dB condition (and the other conditions), allowing for a correlation analysis. The correlation between the effect of SV and speech level was significant (*F-test*, $F(1,15) = 4.916$, $p = 0.043$, $r^2 = 0.247$).

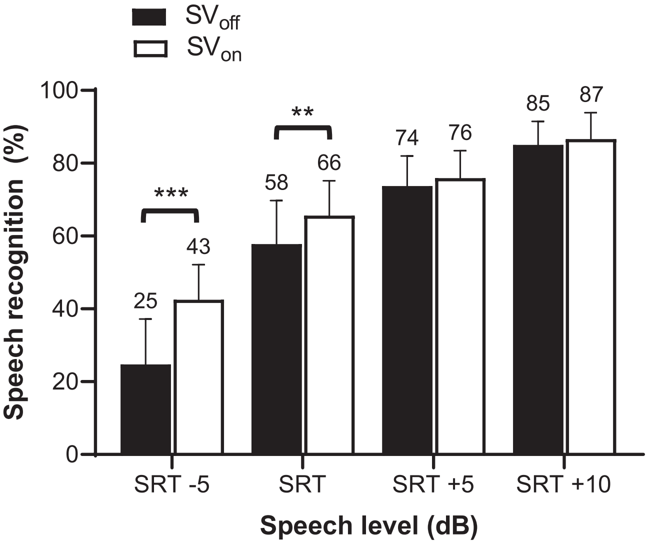

Fig. 5. Average speech-recognition scores obtained at different speech levels (relative to the individuals' SRT) without background noise with SV switched off (SV_{off}, black bars) and on (SV_{on}, white bars). The average speech-recognition scores in this figure (above error bars) differ slightly from the linear mixed model outcomes mentioned in the main text, as no additional parameters (age, learning effects, etc.) were taken into account here. Error bars: SD. **$p < 0.01$; ***$p < 0.001$. SRT indicates speech recognition threshold; SV, SoftVoice.

To assess whether SV improves speech recognition solely by improving the signal-to-noise ratio (SNR), or if other factors were in play as well (e.g., esthetic sound quality), we additionally determined the effect of SV_{on} on speech-recognition scores at speech levels corrected for the improved SRT obtained with SV_{on} (approximately 2 dB on average). The speech-recognition scores determined in this SRT-corrected condition could then be compared with SV_{off} at speech levels based on the SRT obtained with SV_{off}. Note that the effect of SV on speech recognition described earlier was tested by comparing SV_{off} at speech levels based on the SRT of SV_{off} with SV_{on} at speech levels based on the SRT of SV_{off}. An LMM was built identical to Eq. 2 where SV_{off} was compared with SV_{on} using speech levels corrected for the improved SRT at SV_{on}. The LMM showed that SV did not affect speech recognition significantly any longer ($F = 0.277$, $p = 0.600$), and the ($SV{\cdot}L$) was not significant either ($F = 0.270$, $p = 0.847$). Similar to the other analyses, session number, trial number, and T-level were not significant ($F = 0.143$, $p = 0.706$; $F = 0.581$, $p = 0.715$; $F = 2.152$, $p = 0.145$, respectively) while the speech level ($F = 214$, $p < 0.001$) and age ($F = 7.121$, $p = 0.009$) significantly affected speech recognition.

We hypothesized that any effects of SV may be decreased with increased residual hearing in the contralateral ear, under the assumption that ipsilateral acoustic hearing was negligible after implantation. Ipsilateral audiograms are not recorded as part of the clinical protocol at our hospital, and were hence, unfortunately, not available to us. The benefit of SV was expressed as the speech-recognition score at SV_{on} minus SV_{off} at SRT –5 dB. As a measure of residual hearing we determined the median, pure-tone threshold in the audiogram across 500, 1,000, 2,000, and 4,000 Hz ($PTA_{500–4000}$). We excluded two participants from this analysis. One of them was a bilateral CI user with negligible residual hearing in both ears. The other excluded participant wore an earplug during testing. After excluding these two subjects, no significant correlation was found between

TABLE 2. LMM Estimates of the Fixed Effects of SV and Other Parameters on the SRT (in dB) in Quiet and in Steady State Speech-Weighted noise (60 dBA)

Parameter	Effect Type	Estimate (dB)	*df*	*t*	*p*
SV_{on} × $Noise_{off}$	Interaction	–2.0	112	3.982	< 0.001
SV_{on} × $Noise_{on}$	Interaction	+0.5	112	0.965	0.337
Session 2	Main	–1.8	112	5.160	< 0.001
Trial 2	Main	–0.1	112	0.258	0.797
Trial 3	Main	–1.2	125	0.821	0.413
Trial 4	Main	–2.0	125	1.346	0.181
a (Age)	Covariate	+0.3	14	4.479	0.001
b (T-level)	Covariate	–0.003	14	0.117	0.908

LMM, linear mixed model; SV, SoftVoice; N, noise; SRT, speech recognition threshold; a, b, coefficients; T-level, average threshold level (CU) of all the active electrodes.
The interaction factor estimates are constants. The covariates represent coefficients. SRT(dB) = Intercept + SV + N + (SV × N) + Session + Trial + (a × Age) + (b × T).
p values for the interaction factors (SV_{on} × $Noise_{on/off}$) were determined relative to their corresponding SV_{off} × $Noise_{on/off}$ condition (estimates set at 0 set at 0 in the model). Parameter estimates of SV, N, session 1 and trial 1 were set to 0 in the LMM. Trials 1 and 2 were speech-in-noise tests, trials 3 and 4 were performed in quiet. Two practice runs preceded these four trials. Degrees of freedom (df) were determined with the Satterthwaite approximation (Satterthwaite 1946).

$PTA_{500–4000}$ and the benefit of SV on speech recognition (*F-test*, $F(1,13) = 0.027$, $p = 0.873$, $r^2 = 0.002$).

Descriptive statistics of the speech-recognition scores at the different speech levels used are shown in Table 4. Note that the difference between the means of the speech levels may not be exactly 5 dB, as we determined the SRT again in the second session and hence the speech levels were based on a different SRT in both sessions. Two of the four speech levels were tested in session one, the other two in the second session.

Listening Effort

To measure the effect of SV on listening effort, participants were asked to rate the effort after each run performed at fixed speech levels. Figure 7A compares the mean subjective effort ratings per speech condition with and without SV. Ratings were performed on an ordinal scale; hence a nonparametric test was used, namely a Wilcoxon matched-pairs, signed rank test, to evaluate ratings per speech level. Post hoc, Bonferroni-corrected *p* values were calculated to compare the outcomes at different speech level levels. At the lowest speech level tested (SRT –5 dB), the median rating of listening effort was significantly lowered by 1 point on a 9-point Likert scale when SV was switched on ($W(17) = –55$, corrected $p = 0.008$). No statistically significant effects of SV on subjective listening effort were found at higher speech levels (SRT +0 dB, SRT +5 dB, and SRT +10 dB; $W(17) = −24$, corrected $p = 0.250$, $W(17) = −16$, corrected $p > 1$, and $W(17) = –57$, corrected $p = 0.195$, respectively). Similar to the perceived listening effort, the medians of subjectively perceived speech-recognition scores (Fig. 7B) were determined and were shown to improve significantly by 1 point on the 9-point scale at the lowest speech level tested (SRT –5 dB, $W(17) = 153$, corrected $p < 0.001$). No statistically significant effect of SV on subjective speech recognition was found at SRT +0 dB, SRT +5 dB, and SRT +10 dB ($W(17) = 60$, corrected $p = 0.238$, $W(17) = 31$, corrected $p > 1$, and $W(17) = 48$, corrected $p = 0.247$, respectively). Figure 7 shows that high median subjective effort was generally accompanied by lower perceived speech recognition. We tested this observation for significance by pooling the data of SV_{off} and SV_{on} across four speech levels to cover the full range of ratings and performing a nonparametric Spearman correlation test. Subjective effort and speech recognition were indeed negatively correlated (Spearman $r = −0.796$, $p < 0.001$).

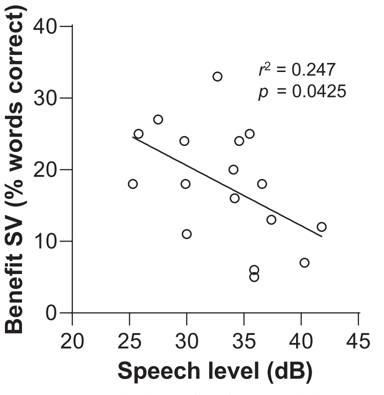

Fig. 6. Correlation between the benefit of SV and the speech level in the SRT—5 dB condition. Linear regression showed a significant correlation. Benefit SV: speech-recognition score with SV_{on} minus the score with SV_{off}. SRT indicates speech recognition threshold; SV, SoftVoice.

To assess whether SV reduced subjective listening effort and enhanced subjectively perceived speech-recognition scores solely by improving the SNR, as observed for the objective speech-recognition scores above, we compared subjective listening effort and subjective speech-recognition scores between two conditions, namely SV_{off} at speech levels based on the SRT of SV_{off}, versus SV_{on} at speech levels based on the SRT of SV_{on}, identical to the analysis for the objective speech-recognition scores described earlier. Bonferroni-corrected, Wilcoxon matched-pairs, signed rank tests did not reveal any statistically significant differences between subjective listening effort with and without SV using SV-corrected speech levels (corrected $p = 0.738$, or higher). Compared with SV_{off}, subjectively perceived speech-recognition scores were significantly elevated by 0.5 point in the SV-corrected speech level condition ($W(17) = 91$, corrected $p = 0.011$, or smaller). At higher speech levels, no significant differences were found ($p = 0.613$ or higher).

Objective measurement of the effect of SV on listening effort was performed by pupillometry at fixed speech levels. Sample waveforms are shown in Fig. 8A. In two subjects, one condition was discarded, because the number of sweeps was less than 5. Because of these missing data, and the presence of confounding factors, we performed an LMM similar to the one used to analyze speech recognition, namely:

$$P_1\,difference = Intercept + SV + L + (SV \cdot L) + session + trial + (a \cdot Age) + (b \cdot T)$$

TABLE 3. LMM Estimates of the Fixed Effects of SV and Other Parameters on Word Recognition in Quiet

Parameter	Effect Type	Estimate (%)	*df*	*t*	*p*
SV_{on} × L = −5 dB	Interaction	+17.4	106	5.762	<0.001*
SV_{on} × L = 0 dB	Interaction	+8.5	106	2.842	0.0215*
SV_{on} × L = +5 dB	Interaction	+1.6	106	0.543	1.0*
SV_{on} × L = +10 dB	Interaction	+3.3	107	1.068	1.0*
Session	Covariate	+2.4	106	1.583	0.116
Trial 2	Main	+4.3	113	1.631	0.528*
Trial 3	Main	+6.1	114	2.251	0.131*
Trial 4	Main	+6.7	113	2.370	0.098*
Trial 5	Main	+4.4	111	1.696	0.464*
Trial 6	Main	+3.4	109	1.272	1.0*
a (age)	Covariate	−0.3	14	2.251	0.041*
d (T-level)	Covariate	+0.07	14	1.456	0.168*

LMM, linear mixed model; SV, SoftVoice; L, speech level; a, b, coefficients; T-level, average threshold level (CU) of all the active electrodes.

Word recognition (%) = Intercept + SV + L + (SV × L) + Session + Trial + (a × Age) + (b × T). P levels for the interaction factors (SV_{on} × L) were determined relative to their corresponding SV_{off} × L condition (estimate set at 0 in the model). P levels of L were determined relative to the SRT +1 db condition (set at 0).

[*]P levels were Bonferroni-corrected for multiple comparisons testing. Degrees of freedom (df) were determined with the Satterthwaite approximation (Satterthwaite 1946).

where P_1 difference was the peak pupil dilation of the first peak. The (SV·L) interaction factor included SV_{on}, SV_{off}, and the SV_{on} condition with the corrected SRT. The LMM did not show significant main effects of SV ($F = 0.412$, $p = 0.663$), (SV·L) ($F = 0.506$, $p = 0.803$), session number ($F = 0.594$, $p = 0.442$), trial number ($F = 0.719$, $p = 0.610$), age ($F = 0.231$, $p = 0.638$), or mean T-level ($F = 0.077$, $p = 0.785$) on the peak amplitude P_1. Speech level was the only significant variable in the analysis ($F = 5.281$, $p = 0.002$). A Bonferroni-corrected, multiple pair-wise comparisons post hoc test between the speech levels showed that the pupil response significantly differed between SRT −5 dB and SRT +5 dB (adjusted $p = 0.002$). The difference between the pairs SRT −5 and SRT +10 dB, and SRT +0 and SRT +5 dB approached significance (adjusted $p = 0.052$ and 0.083, respectively). The other pairs did not differ significantly (adjusted $p > 0.9$). Figure 8B shows the P_1 amplitude per speech level with and without SV. The effect of SV on the pupil response was additionally analyzed using the same LMM with different outcome measures, namely the area under the curve between $t = 0$ and P_1, and P_1 latency. Both measures were not significantly affected by SV ($F = 0.708$, $p = 0.494$ and $F = 0.316$, $p = 0.729$, respectively).

To analyze the interrelationships between the various subjective and objective listening effort outcome measures, we correlated them using nonparametric Spearman tests. To cover the entire range of subjective ratings we pooled the data from the four speech levels tested, as effort ratings were frequently maximal at low speech levels. To increase statistical power, we also pooled the recordings with and without SV (but excluding the data of SV_{on} and corrected SRT, which was statistically not significantly different from the SV_{off} condition). Hence, the correlation analysis was performed on 17 participants each

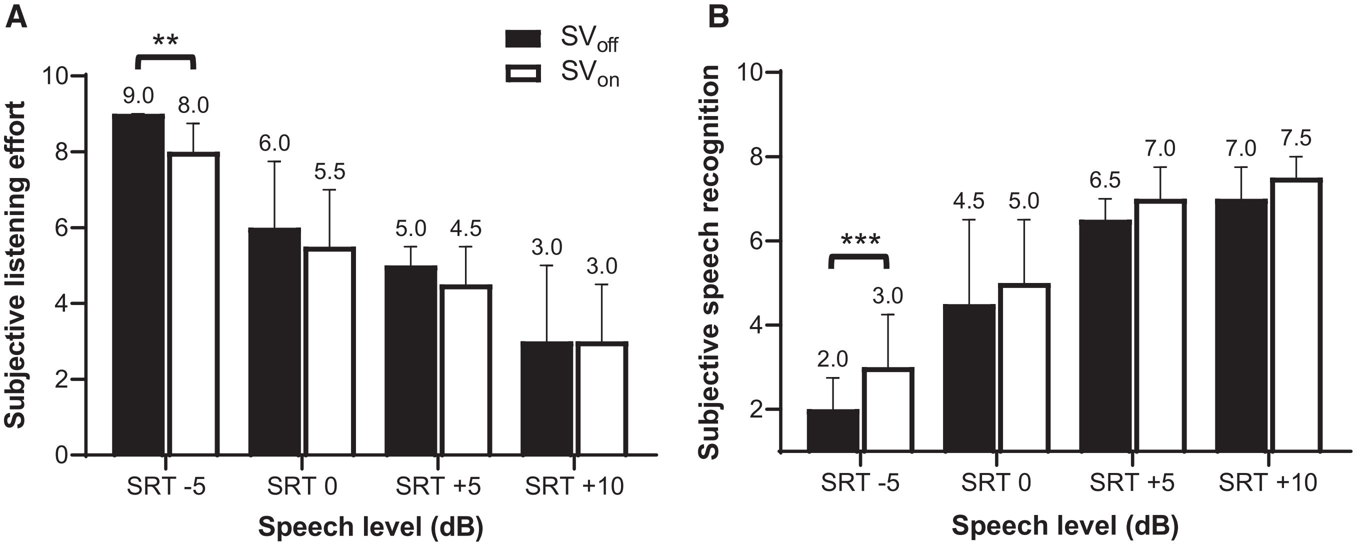

Fig. 7. Subjective ratings at different speech levels with and without SoftVoice. Median subjective ratings of listening effort (A) and subjectively perceived speech-recognition scores (B) at different speech levels (relative to the individuals' SRT) with SoftVoice switched off (SV_{off}, black bars) and on (SV_{on}, white bars). Numbers above error bars: median values. Error bars: interquartile range. Rating 1: no effort; 9: extreme effort (in A), or 1: no words correct; 9: all words correct (in B). $^{**}p < 0.05$; $^{***}p < 0.001$. SRT indicates speech recognition threshold; SV, SoftVoice.

TABLE 4. Mean Speech Level and Correct Scores for the Different Test Conditions

	SRT –5 dB			SRT +0 dB			SRT +5 dB			SRT +10 dB		
	Range	M	SD	Range	M	SD	Range	M	SD	Range	M	SD
SRT (dBA)	25.3–41.8	33.4	4.7	30.3–46.8	38.0	4.7	35.3–50.8	42.6	4.3	40.3–55.8	48.7	4.6
Correct score (%) SV off	6–49	24.8	12.5	42–85	57.8	12.0	61–90	73.8	8.3	69–93	85.1	6.4
Correct score (%) SV on	26–55	42.5	9.6	46–79	65.6	9.6	63–86	76.0	7.5	72–95	86.7	7.2
Benefit SV correct score SV ON–OFF	5–33	17.8	8.0	–7 to 18	7.8	7.9	–4 to 10	1.9	4.2	–9 to 8	1.6	3.9

M, mean; SD, standard deviation; SRT, speech reception threshold; SV, SoftVoice.

contributing eight data points (four speech levels × SV_{on}/SV_{off}). Subjective listening effort proved to be significantly, negatively correlated to the subjectively perceived speech-recognition score (Spearman $r = -0.819$, $p < 0.001$) and the objective speech-recognition score ($r = -0.778$, $p < 0.001$). In other words, listening conditions associated with lower (perceived) speech-recognition scores were experienced to take more effort. Subjective listening effort did not correlate well with pupillometry measures, including P_1 amplitude ($r = -0.153$ $p = 0.078$), P_1 auc ($r = -0.091$, $p = 0.298$), or P_1 latency ($r = 0.059$, $p = 0.495$).

DISCUSSION

In this study, we have investigated the effect of SV on speech recognition and on subjective and objective listening effort. We have shown a statistically significant improvement of speech perception and a significant reduction of subjective listening effort at low speech levels. Specifically, SV decreased the SRT, increased speech-recognition scores, and decreased rated listening effort at soft speech levels. However, we found no effect of SV on pupil dilation.

We report statistically significant improvements of speech-recognition scores with SV at SRT −5 dB (17%) and at SRT +0 dB (9%), but not at higher speech levels (Fig. 5). SRT –5 dB and SRT +0 dB correspond to mean speech levels of 33 and 38 dBA (38 and 43 dB SPL), respectively (Table 4). Holden et al. (2019) examined fixed speech levels in their study that were independent on the individual subjects' SRTs, namely 35, 40, 45, 50, and 60 dB SPL for all their subjects. They report significant improvements of SV on sentence scores at 35 dB up to 50 dB SPL. Their tested speech levels of 40 dB and 45 dB SPL correspond most closely to the levels where we found a significant effect (38 and 43 dB SPL). At these levels, the authors report a significant improvement of mean AzBio sentence scores of 10% and 13%, respectively. We found an insignificant improvement of 2% at 48 dB SPL, while they report a significant improvement of 9% at 50 dB SPL. In Table 5, the data found in this study and the one of Holden et al. (2019) are summarized, and the speech levels as applied in our paper (A-weighted) have been converted into Z-weighting (dB SPL) to match Holden et al. (2019).

Taken together, we have found a larger maximal improvement of speech-recognition scores than the improvement in sentence recognition reported by Holden et al. (2019) at the lowest tested speech level, while they report larger improvements at higher speech levels. We speculate that this discrepancy might be due to differences in overall speech recognition of the participants between studies. Some of their participants performed at floor level at the lowest speech level, as their speech levels were not based on the individual's SRT (personal communication with Laura K. Holden, Washington University School of Medicine, St. Louis,

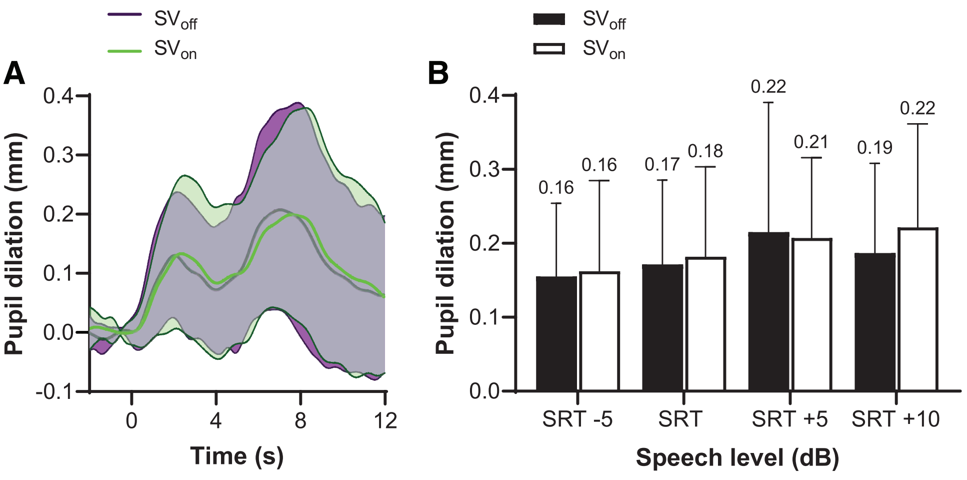

Fig. 8. Pupil responses measured with SV switched off (SV_{off}) and on (SV_{on}). (A) The average waveforms of the pupil responses with SV_{off} (purple) and SV_{on} (green) at the lowest speech level tested (SRT—5 dB). Envelopes: SD. (B) Average pupil responses at different speech levels (relative to the individuals' SRT) with SoftVoice switched off (SV_{off}, black bars) and on (SV_{on}, white bars). Error bars: SD. No significant effects of SV were found. SRT indicates speech recognition thresholds; SV, SoftVoice.

MO, 2019). Hence, adding SV may not have been beneficial for some of their participants as the speech level was too low (floor effect). On the other hand, our participants may have experienced ceiling effects at higher speech levels as seen in Figure 5; the improvement in average speech-recognition scores between speech levels seem to level off at the highest two speech levels (74% and 85% speech recognition at 48 and 54 dB SPL without SV). Holden et al. (2019) show sentence scores of 61% at 50 dB SPL and 81.1% at 60 dB SPL without SV. This left a more favorable range for improvement for their participants at these speech levels. Another factor that might have played a role is that they used English AzBio sentences, whereas we used the Dutch/Flemish Matrix test. AzBio sentence recognition likely requires more language processing, as they represent open-set stimuli consisting of normal English sentences. The speech corpus of the Dutch/Flemish Matrix test is a closed set of 50 words, and hence speech recognition may require less cognitive processing, as the sentences lack lexical meaning and have the same syntax.

Holden et al. (2019) did not investigate effects of SV on the SRT in quiet. We supplement their findings by showing a significant improvement of the SRT of 2.0 dB in quiet. No significant difference was observed between SV_{on} and SV_{off} on the SRT in noise, indicating that SV does not deteriorate speech recognition in noise, in line with the notion that SV only acts at soft sound levels (Holden et al. 2019), We expected that residual hearing would negatively impact SV's benefits on speech recognition, as any acoustic hearing will be unaffected by SV and will dilute the effects of SV. However, we did not find such a relation. Because the $\text{PTA}_{500\text{–}4000}$ was not better than 60 dBA in any of the tested participants (Table 1, note that the subject with $\text{PTA}_{500\text{–}4000}$ wore an earplug during testing), and since the range of speech levels where SV was significantly beneficial (25 to 47 dBA, Table 4) the applied speech levels were expectedly too low to be heard with the contralateral residual acoustic hearing.

To investigate if SV's effect on speech recognition can be fully explained by effects on the SNR, we compared speech-recognition scores and listening effort obtained with SV_{on} and SV_{off} using speech levels that were based on the SRT obtained with SV. Correcting for this lower SRT with SV_{on} removed all significant effects of SV on the SRT, as well as on objective speech-recognition scores and subjectively perceived effort at all speech level levels. These results indicate that SV's effects can be fully explained by an improvement of the SNR, thereby lowering the SRT to the same extent. This finding is in agreement with the way SV acts, namely lowering the amount of noise in soft speech (Holden et al. 2019). However, at a speech level of SRT −5 (but not at higher levels), subjects rated the perceived speech recognition 0.5 points better for SV_{on}, even when the speech level was corrected for a better SRT. Since it occurred at a speech level of SRT −5 dB, where the effect of SV was the most pronounced, it may reflect a genuine effect of SV other than merely an improvement of the SNR. For instance, SV may have affected qualitative attributes of speech other than its clarity, such as naturalness, or pleasantness that caused subjects to rate the perceived speech recognition higher.

TABLE 5. SV Benefits Found in This Study and the Report of Holden et al. (2019)

Speech Level This Study (dB SPL eq)	Speech Level Holden et al. (2019) (dB SPL)	SV Benefit This Study (%)	SV Benefit Holden et al. (2019) (%)
	35		8***
38	40	17****	10*
43	45	9*	13***
48	50	2	9**
54	60	2	1

The sound levels used in this study are expressed in dB SPL instead of dBA to match Holden et al. (2019). Numerical benefits corresponding to Figure 2 in Holden et al. (2019) were kindly provided by R. Reeder and L. Holden. Holden et al. measured AzBio sentence correct scores, while we used word recognition scores using the Dutch/Flemish Matrix test.

$^{****}p < 0.0001$; $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

SV, SoftVoice; SPL, sound pressure level.

This is the first study that focused on SV's effect on listening effort. Our results indicate that SV significantly decreases subjective listening effort at a soft speech level of 33 dBA, but not at higher levels. Pupillometry did not reveal a significant effect of SV. This apparent lack of correlation between subjective listening effort and an objective assessment is consistent with findings of Zekveld et al. (2010) and Koelewijn et al. (2012), who also showed that subjectively perceived effort may not be correlated to the pupil response. Koelewijn et al. (2012) proposed that subjectively perceived effort is hard to discern from experienced performance by subjects and hence that objective measures may be the preferred way of measuring listening effort. In line with this, we found a significant correlation between ratings of effort and speech recognition. However, this correlation can also support the opposite notion that performance and effort are correlated and that performance measures in themselves are a valid predictor of listening effort. Indeed, pupil dilation is correlated to higher levels of attentional focus and vigilance, which may not necessarily translate into subjectively perceived listening effort (Alhanbali et al. 2019), as many other factors may influence attention and vigilance. For instance, some of our subjects, and especially the elderly, reported to have difficulty separating listening effort from general task difficulty, fatigue, or loss of concentration, which may have affected the rating results. We, however, did not systematically make note of the number of times this was reported by our subjects. Hence, the impact of the ratings containing an element of task difficulty could not be quantitatively assessed.

The lack of effect of SV on the objective pupillometric measure may additionally be explained by the finding that pupillary responses only correlate to task difficulty when processing demands are within the resource capacity. When subjects are reaching their resource limits, pupil responses change little and may even begin to decline when task difficulty demands processing power exceeding the resources (Granholm et al. 1996). In fact, subjects may even withdraw putting effort into difficult tasks, because of a lack of success (reviewed by Winn et al. 2018). In line with this hypothesis, we indeed found significantly smaller pupil responses at a speech level of SRT −5 dB than at SRT +5 dB. Speech recognition at SRT −5 dB was experienced as markedly, and sometimes even frustratingly, difficult by our subjects. As significant effects of SV on speech recognition and listening effort were confined to low speech levels of SRT –5 dB and SRT +0 dB, we cannot exclude that our pupillometry measures were hampered by ceiling effects, as people were operating at their maximum cognitive load. Moreover, studies have shown that adults with a hearing impairment, and especially the elderly, have smaller peak pupil responses than people with normal hearing, which can be attributed to listening fatigue and other factors (Piquado et al. 2010; Ohlenforst et al. 2017; Wang et al. 2018; Winn et al. 2018). In all, we conclude that subjectively experienced listening effort statistically significantly decreased with SV at soft speech levels, but that an objective pupillometric measure did not support this observation, supporting the notion that the two measures reflect separate dimensions of listening effort in CI users (Alhanbali et al. 2019).

There are a few limitations of this study that deserve mention. First, the Dutch/Flemish Matrix test we used was validated in a group of listeners with normal hearing and was intended to determine the SRT in noise (Luts et al. 2014). We used the test in a group of CI users and applied it in quiet, without background noise. We found an improvement of 7.3% of the speech-recognition score with SV at the SRT. Since SV improved the SRT by 2.0 dB, the steepness at SRT seems to be approximately 3.7%/dB. However, Luts et al. (2014) report a steepness of 13.9%/dB at the SRT. This implies that our SRT improvement of 2.0 dB should equal an increase of approximately 28% instead. Luts et al. (2014) determined the slope of the psychometric function in noise. Using this test in quiet may lead to altered characteristics of the underlying psychometric function, which may have affected the accuracy of the estimated SRT. In addition, none but one of our participants reached a 100% word score with consonant–vowel–consonant stimuli presented at 65 dB SPL in quiet (Table 1). Hence, the psychometric function of our study population expectedly deviates from the ideal curve as described by Luts et al. (2014) that reached 100% correct scores at favorable SNRs. Last, given the closed-set design of the Matrix test it, arguably, assesses word-discrimination ability, rather than speech recognition. This can result in learning effects and it may not be representative of everyday listening to speech. While learning effects were mitigated by training, randomization and the use of an LMM, a lack of context may have affected our pupillometry findings. For instance, the discrimination of a closed set of words in the Matrix test might not reflect the effort needed to understand normal speech (Winn et al. 2018). In addition, the lack of context takes away the chance to “fill in the gaps” retrospectively when words are not heard. The main reason, we used the Matrix test nonetheless, is that it has been specifically designed to reliably determine the SRT (Luts et al. 2014). Further, it can be used repeatedly on the same subject without substantial learning effects (Kollmeier et al. 2015), as seen in other speech tests that consist of a collection of unique sentences. Last, the Dutch/Flemish Matrix sentences are voiced at a natural speed and are challenging for most CI users, yet they are not frustratingly difficult to recognize in noise.

There are at least two fitting adaptations that can potentially enhance SV’s effects beyond what we have shown here. The perception of soft speech can potentially be enhanced by changing the IDR settings of the speech processor when SV is used (Zeng et al. 2002; Firszt et al. 2004; Spahr et al. 2007; Holden et al. 2011). A wider IDR provides lower minimum stimulation levels and improves speech recognition of soft speech in quiet and in noise (Holden et al. 2007; Spahr et al. 2007). Spahr et al. (2007) hence recommended an IDR of 60 dB, which is still the default setting in devices from Advanced Bionics at the time of writing. An even wider IDR may further enhance the perception of soft speech, but can potentially introduce audible system noise (Dawson et al. 2007; Holden et al. 2007). In combination with SV, however, this system noise can be reduced and it is therefore of interest to investigate if SV users will benefit from IDRs wider than 60 dB. In this study, all participants were fitted with their home-use settings, that is, an IDR of 60 dB, except for two subjects who had an IDR of 55 dB.

Threshold (T) levels are another potential variable to enhance SV’s effects. The default T-level as recommended by Advanced Bionics is 10% of the maximal acceptable (M) level (Advanced Bionics Corporation 2003) with an IDR of 60 dB. Holden et al. (2019) suggest that the use of SV can lead to subjective awareness of background noise otherwise not heard, and hence that T-levels should be set higher than 10% of M when SV is used. This way, soft environmental sound may become less audible. In our medical center, the T = 10% of M guideline is not used. Instead, thresholds are independently determined by audiologists. As a result, the mean T-level was 24% of the M-level across electrodes. In fact, none of our subjects had a T-level as low as 10% of M-level. While we performed testing in a soundproof booth and environmental sounds were of little importance, in real-life situations, the audibility of soft background noise may become important. Future research focusing on setting a wider IDR and optimizing threshold settings may potentially enhance SV’s benefits. In terms of battery life, we expect that the added power consumption of SV can be considered negligible (personal communication with Leonid Litvak, Advanced Bionics LLC, Valencia, CA). In fact, at low ambient sound levels SV may increase battery life by reducing electrode stimulations caused by system noise.

CONCLUSION

SV improved speech recognition and decreased subjective listening effort at soft speech levels presented in quiet. These benefits were obtained in a group of experienced CI users, without making any other changes to their existing fitting. Hence, simply activating SV can substantially improve the recognition of soft speech and decrease the perceived effort. Objective measures of listening effort were, however, unaffected by SV. Together with the finding that SV did not significantly affect speech recognition in noise, we recommend activating the SV algorithm by default for users of CIs from Advanced Bionics.

ACKNOWLEDGMENTS

We thank the study participants for their time and dedication. We wish to express our gratitude to Laura K. Holden and Ruth M. Reeder (Washington University School of Medicine, St. Louis, MO) for generously making available their data to us, and to Leonid Litvak (Advanced Bionics LLC, Valencia, CA) for sharing information on the technical aspects of the SoftVoice algorithm. This research was supported by funds from Advanced Bionics, LLC. R. K. is employed by Advanced Bionics, GmbH.

H. C. S. was responsible for designing experiments, IRB approval, data analysis, and critical revision of first drafts; E. A. was responsible for data collection, data analysis, and draft writing; R. K. for designing experiments, and critical revision of first drafts; J. J. B. for critical revision first drafts; and J. H. M. F. for critical revision first drafts.

Address for correspondence: H. Christiaan Stronks, Department of Otorhinolaryngology, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, the Netherlands. E-mail: h.c.stronks@lumc.nl

Received November 14, 2019; accepted June 29, 2020.

REFERENCES

Advanced Bionics Corporation. (2003). New methodology for fitting cochlear implants. White paper.

Alhanbali, S., Dawes, P., Lloyd, S., Munro, K. J. (2017). Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear Hear, 38*, e39–e48.

Alhanbali, S., Dawes, P., Millman, R. E., Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear Hear, 40*, 1084–1097.

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychol Methods, 23*, 389–411.

Brons, I., Houben, R., Dreschler, W. A. (2013). Perceptual effects of noise reduction with respect to personal preference, speech intelligibility, and listening effort. *Ear Hear, 34*, 29–41.

Canada Safety Council. (2019). Office noise and acoustics: Workplace safety. Retrieved September 25, 2019 from https://canadasafetycouncil.org/office-noise-and-acoustics/.

Dawson, P. W., Vandali, A. E., Knight, M. R., Heasman, J. M. (2007). Clinical evaluation of expanded input dynamic range in nucleus cochlear implants. *Ear Hear, 28*, 163–176.

de Jong, M. A. M., Briaire, J. J., Frijns, J. H. M. (2019a). Dynamic current focusing: A novel approach to loudness coding in cochlear implants. *Ear Hear, 40*, 34–44.

de Jong, M. A. M., Briaire, J. J., van der Woude, S. F. S., et al. (2019b). Dynamic current focusing for loudness encoding in cochlear implants: a take-home trial. *Int J Audiol, 58*, 553–564.

Devocht, E. M. J., Janssen, A. M. L., Chalupper, J., et al. (2017). The benefits of bimodal aiding on extended dimensions of speech perception: Intelligibility, listening effort, and sound quality. *Trends Hear, 21*, 2331216517727900.

Faul, F., Erdfelder, E., Lang, A. G., Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods, 39*, 175–191.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am, 88*, 1725–1736.

Firszt, J. B., Holden, L. K., Skinner, M. W., et al. (2004). Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems. *Ear Hear, 25*, 375–387.

Forli, F., Lazzerini, F., Fortunato, S., et al. (2019). Cochlear implant in the elderly: Results in terms of speech perception and quality of life. *Audiol Neurootol, 24*, 77–83.

Gaylor, J. M., Raman, G., Chung, M., et al. (2013). Cochlear implantation in adults: A systematic review and meta-analysis. *JAMA Otolaryngol Head Neck Surg, 139*, 265–272.

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

Granholm, E., Asarnow, R. F., Sarkin, A. J., Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology, 33*, 457–461.

Hällgren, M., Larsby, B., Lyxell, B., Arlinger, S. (2005). Speech understanding in quiet and noise, with and without hearing aids. *Int J Audiol, 44*, 574–583.

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.

Holden, L. K., Firszt, J. B., Reeder, R. M., et al. (2019). Evaluation of a new algorithm to optimize audibility in cochlear implant recipients. *Ear Hear, 40*, 990–1000.

Holden, L. K., Reeder, R. M., Firszt, J. B., et al. (2011). Optimizing the perception of soft speech and speech in noise with the Advanced Bionics cochlear implant system. *Int J Audiol, 50*, 255–269.

Holden, L. K., Skinner, M. W., Fourakis, M. S., et al. (2007). Effect of increased IIDR in the nucleus freedom cochlear implant system. *J Am Acad Audiol, 18*, 777–793.

HSE Offices. (2005). Controlling noise at work. In *Guidance for Employers on the Control of Noise at Work Regulations*. Health and Safety Executive.

Hughes, S. E., Hutchings, H. A., Rapport, F. L., et al. (2018). Social connectedness and perceived listening effort in adult cochlear implant users: A grounded theory to establish content validity for a new patient-reported outcome measure. *Ear Hear, 39*, 922–934.

Klop, W. M., Briaire, J. J., Stiggelbout, A. M., et al. (2007). Cochlear implant outcomes and quality of life in adults with prelingual deafness. *Laryngoscope, 117*, 1982–1987.

Koelewijn, T., Zekveld, A. A., Festen, J. M., et al. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear Hear, 33*, 291–300.

Kollmeier, B., Warzybok, A., Hochmuth, S., et al. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *Int J Audiol, 54*, 3–16.

Kramer, S., S. Kapteyn, T., M. Festen, J., et al. (1997). Assessing aspects of auditory handicap by means of pupil dilatation. *Audiology, 36*, 155–164.

Kramer, S. E., Kapteyn, T. S., Houtgast, T. (2006). Occupational performance: Comparing normally-hearing and hearing-impaired employees using the Amsterdam Checklist for Hearing and Work. *Int J Audiol, 45*, 503–512.

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

Littell, R. C., Pendergast, J., Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Stat Med, 19*, 1793–1819.

Luts, H., Eneman, K., Wouters, J., et al. (2010). Multicenter evaluation of signal enhancement algorithms for hearing aids. *J Acoust Soc Am, 127*, 1491–1505.

Luts, H., Jansen, S., Dreschler, W., et al. (2014). *Development and Normative Data for the Flemish/Dutch Matrix Test*. Katholieke universiteit Leuven, Belgium and Academic Medical Center, Amsterdam, The Netherlands..

Ohlenforst, B., Zekveld, A. A., Jansma, E. P., et al. (2017). Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review. *Ear Hear, 38*, 267–281.

Perreau, A. E., Wu, Y. H., Tatge, B., et al. (2017). Listening effort measured in adults with normal hearing and cochlear implants. *J Am Acad Audiol, 28*, 685–697.

Piquado, T., Isaacowitz, D., Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology, 47*, 560–569.

Plomp, R., & Mimpen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level. *J Acoust Soc Am, 66*, 1333–1342.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, *2*, 110–114.

Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR). (2008). *Potential Health Risks of Exposure to Noise From Personal Music Players and Mobile Phones Including a Music Playing Function. Preliminary report* European Commission.

Shannon, R. V., Fu, Q.-J., Galvin, J., et al. (2004). Speech perception with cochlear implants. In F.-G. Zeng, A. N. Popper, R. R. Fay (Eds.), *Cochlear Implants. Auditory Prostheses and Electric Hearing* (pp. 334–376). Springer-Verlag.

Soede, W., Bilsen, F. A., Berkhout, A. J. (1993). Assessment of a directional microphone array for hearing-impaired listeners. *J Acoust Soc Am*, *94*, 799–808.

Spahr, A. J., Dorman, M. F., Loiselle, L. H. (2007). Performance of patients using different cochlear implant systems: Effects of input dynamic range. *Ear Hear, 28*, 260–275.

Van der Beek, F. B., Soede, W., Frijns, J. H. M. (2007). Evaluation of the benefit for cochlear implantees of two assistive directional microphone systems in an artificial diffuse noise situation. *Ear Hear, 28*, 99–110.

Wang, Y., Kramer, S. E., Wendt, D., et al. (2018). The pupil dilation response during speech perception in dark and light: The involvement of the parasympathetic nervous system in listening effort. *Trends Hear*, *22*, 2331216518816603.

Watson, A. B., & Yellott, J. I. (2012). A unified formula for light-adapted pupil size. *J Vis, 12*, 1–16.

Winn, M. B. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends Hear*, *20*, 2331216516669723.

Winn, M. B., & Moore, A. N. (2018). Pupillometry reveals that context benefit in speech perception can be disrupted by later-occurring sounds, especially in listeners with cochlear implants. *Trends Hear, 22*, 2331216518808962.

Winn, M. B., Wendt, D., Koelewijn, T., et al. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends Hear, 22*, 2331216518800869.

World Medical Association. (2013). World medical association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, *310*, 2191–2194.

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.

Zekveld, A. A., Kramer, S. E., Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear Hear, 32*, 498–510.

Zeng, F. G., Grant, G., Niparko, J., et al. (2002). Speech dynamic range and its effect on cochlear implant performance. *J Acoust Soc Am, 111*(1 Pt 1), 377–386.