ASHA American Speech-Language-Hearing Association
JSLHR

Research Article

Assessment of Binaural Benefits in Hearing and Hearing-Impaired Listeners

Okan Öz,[a,b] Hilal Dinçer D'Alessandro,[c] Merve Özbal Batuk,[b] Gonca Sennaroğlu,[b] and Paul J. Govaerts[a,d]

[a] The Eargroup, Antwerp, Belgium [b] Department of Audiology, Faculty of Health Sciences, Hacettepe University, Ankara, Turkey [c] Department of Audiology, Faculty of Health Sciences, Istanbul University - Cerrahpaşa, Turkey [d] Faculty of Medicine and Health Sciences, Translational Neurosciences, Otorhinolaryngology & Head and Neck Surgery, University of Antwerp, Belgium

ARTICLE INFO

Article History:
Received January 31, 2023
Revision received March 22, 2023
Accepted May 15, 2023

Editor-in-Chief: Peggy B. Nelson
Editor: Vishakha Waman Rawool

https://doi.org/10.1044/2023_JSLHR-23-00077

ABSTRACT

Purpose: The purpose of this study was to (a) investigate which speech material is most appropriate as stimulus in head shadow effect (HSE) and binaural squelch (SQ) tests, (b) obtain normative values of both tests using the material decided to be optimal, and (c) explore the results in bilateral cochlear implant (CI) users.

Method: Study participants consisted of 30 normal-hearing (NH) persons and 34 bilateral CI users. This study consisted of three phases. In the first phase, three different speech materials (1) monosyllabic words, (2) spondee words, and (3) sentences were compared in terms of (a) effect size, (b) test–retest reliability, and (c) interindividual variability. In the second phase, the speech material selected in the first phase was used to test a further 24 NHs to obtain normative values for both tests. In the third phase, tests were administered to a further 23 bilateral CI users, together with localization test and the Speech, Spatial, and Qualities of Hearing scale.

Results: The results of the first phase indicated that spondees and sentences were more robust materials compared with monosyllables. Although the effect size and interindividual variability were comparable for spondees and sentences, sentences had higher test–retest reliability in this sample of CI users. With sentences, the mean (± standard deviation) HSE and SQ in the NH group were 58 ± 14% and 22 ± 11%, respectively. In the CI group, the mean HSE and SQ were 49 ± 13% and 13 ± 14%, respectively. There were no statistically significant correlations between the test results and the interval between the implantations, the length of binaural listening experience, or the asymmetry between the ears.

Conclusions: Sentences are preferred as stimulus material in the binaural HSE and SQ tests. Normative data are given for HSE and SQ with the LiCoS (linguistically controlled sentences) test. HSE is present for all bilateral CI users, whereas SQ is present in approximately seven out of 10 cases.

Binaural hearing can be simply defined as listening with both ears. Binaural hearing improves speech understanding and sound source localization, particularly in background noise (Gantz et al., 2002; Müller et al., 2002; Schleich et al., 2004; Schön et al., 2002). Although binaural hearing has many advantages, defining and assessing it can be challenging. Moreover, clinicians are restricted to a few tests that provide a limited understanding into the true benefits of binaural hearing, and these tests are rarely used in clinical practice.

A wide range of tests has been used in the literature to assess the phenomena associated with binaural hearing, including the head shadow effect (HSE), binaural squelch (SQ), binaural summation (SU), spatial release from masking (SRM), and localization (LOC; Arsenault & Punch, 1999; Cox et al., 1981; Gifford et al., 2014; Tyler et al., 2002, 2003; Van Deun et al., 2010). There are studies showing these benefits in bilateral/bimodal cochlear implant (CI) users and in normal-hearing (NH) people (Buss et al., 2008; Eapen et al., 2009; Litovsky et al., 2006; Schleich et al., 2004; Verhaert et al., 2012).

Correspondence to Okan Öz: oz@eargroup.net. **Disclosure:** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Each of the aforementioned phenomena can be evaluated by psychoacoustic tests that can be performed in a clinical audiological setting. Psychoacoustic measures provide valuable information about the auditory perception of sound. These methods, however, are inevitably influenced by a conscious attentional process and, therefore, can be challenging to use in very young children and older adults (Bertoli et al., 2005; Köse et al., 2022; Luo et al., 2020; Maxon & Hochberg, 1982). Although binaural benefits have traditionally been evaluated utilizing psychoacoustic methods, some recent studies have investigated the binaural benefits also through objective methods (Balkenhol et al., 2020; Rawool, 2016; Rawool & Parrill, 2018). Rawool (2016) used acoustic reflex threshold measurements to assess SU in women. A follow-up study by Rawool and Parrill (2018) found similar results also in men. Balkenhol et al. (2020) assessed HSE, SU, SQ, and SRM in individuals with bimodal hearing using auditory evoked potentials. Thanks to their linguistic independence, objective methods might have the potential to reduce the high variance observed in the results of binaural benefit assessments. However, due to the lack of standards, there are neither universally accepted test setups nor normal values for either psychoacoustic or electrophysiologic methods. The standardized measures of binaural testing become particularly important in the context of CI. Cochlear implantation was traditionally performed unilaterally, which means that binaural hearing was not restored. This has changed in children, since bilateral implantation has become a common treatment in this population, and it is also becoming more common in adults (Brown & Balkany, 2007; Kan & Litovsky, 2015; Litovsky et al., 2006).

The traditional methods to assess the binaural benefits include comparing speech understanding in the binaural listening condition with that in the monaural condition (Eapen et al., 2009; Gantz et al., 2002; Schleich et al., 2004). Here are many factors that have impact on the outcomes of these methods. These factors include both measurement-related variables such as the test setups, the type of noise, the reference ear tested in the monoaural condition, the number and position of the loudspeakers as well as the microphone settings, and participant-related variables such as the individual performance and gender (Bronkhorst & Plomp, 1988; Carhart, 1965; Dieudonné & Francart, 2019; Dirks & Wilson, 1969; Kurz et al., 2021; Laszig et al., 2004; MacKeith & Coles, 1971; Müller et al., 2002; Pyschny et al., 2014; Rawool & Parrill, 2018; Schön et al., 2002; Sheffield et al., 2015). To the best of our knowledge, however, there are no studies evaluating and comparing the effectiveness of different speech materials in binaural benefit assessments. Furthermore, as long as the item lists used in the speech tests are relatively short, 10% or more test–retest differences can be observed (Dillon, 2012, as cited in Avan et al., 2015). However, in most studies in the literature, the test–retest reliability of the applied test material is not measured. In fact, when the individual results are examined, the test results in many conditions may not exceed the test–retest variability.

The first goal of this study was to investigate which speech material (monosyllabic words, spondee words, or sentences) is more effective for assessing the HSE and SQ based on the criteria: (a) the largest effect size in NHs, (b) the highest test–retest reliability in both NHs and CI users, and (c) the smallest interindividual variability in NHs. The most appropriate material was then used to obtain normative values for both tests.

The second goal was to explore the correlation between HSE, SQ, and LOC in bilateral CI users and possible influencing factors, such as the interval between implantations, experience with binaural hearing, etc. Furthermore, a correlation analysis was performed between the results of binaural tests, and the Speech, Spatial, and Qualities of Hearing (SSQ) scale (Falzone et al., 2022).

Method

A prospective study was conducted with NH participants and bilateral CI recipients. The Ethics Committee of Antwerp University Hospital (UZA) approved the study with Project ID 2021–0551 – BUN B3002021000155. Informed consent was obtained from all participants. For participants under 18 years of age, parental consent was obtained in addition to their own.

Participants

A total of 64 people participated in this study. All participants spoke Flemish as their mother tongue. Thirty-four of them were CI users with more than 6 months of binaural listening experience who were being followed in our clinic, and 30 were in the NH group with hearing thresholds of ≤ 20 dB HL at all the octave frequencies tested between 125 and 8000 Hz. The NH group consisted of 14 (47%) male and 16 (53%) female participants, with a median age of 23.5 years (Q1: 22 and Q3: 28.5; range: 19–39). The mean pure-tone average (PTA) of the NH group was 7 ± 4 dB HL in the right ear and 8 ± 4 dB HL in the left ear.

The CI group consisted of 19 (56%) female and 15 (44%) male participants, with a median age of 21 years (Q1: 16 and Q3: 28; range: 14–80). Twelve participants (35%) were implanted postlingually (older than 4 years of age), whereas 22 participants (65%) were implanted before the age of 4 years. Nine (75%) of the 12 participants implanted after 4 years of age had congenital hearing loss and had used hearing aids until implantation. The onset of hearing loss in the remaining three participants (25%) was in adulthood and showed a progressive course. These participants also have a history of hearing aids use before the implantation. Nineteen participants (56%) had their first implant on the right, and 15 participants (44%) had their first implant on the left. None of the participants was implanted simultaneously. The median ages at the first and second implantations were 3 years (Q1: 1 and Q3: 9.3) and 9.5 years (Q1: 4.8 and Q3: 23.3), respectively. The participants had an average of 125 ± 67 months of experience with their 2nd CI.

Procedure

This study consisted of three phases as follows.

1. Investigating the effect of speech material in the HSE and SQ tests:
We compared three different speech materials: monosyllabic words (NVA: Nederlandse Vereniging voor Audiologie; Wouters et al., 1994), spondee words (BLU: Brugge-Leuven-Utrecht; Wouters et al., 1994), and sentences (LiCoS: linguistically controlled sentences; Coene et al., 2018). We envisaged that the optimal speech material should be characterized by three preset criteria: (a) large effect size in the NH group, (b) strong test–retest reliability in both the NH and CI groups, and (c) small interindividual variability in the NH group. This phase of the study was conducted with 11 CI users and six NH individuals.

2. Obtaining normative data of HSE and SQ:
The speech material selected in the first phase was used to test a further 24 NH participants to obtain normative values for both tests.

3. Assessing the binaural benefits in bilateral CI users:
Both tests were also administered to a further 23 bilateral CI users. Besides the HSE and SQ tests, the LOC test and the SSQ scale were also applied. The second and third phases were carried out simultaneously after the completion of the first phase.

Outcome Measures

All tests were conducted using the Auditory Speech Sounds Evaluation (A§E) psychoacoustics test suite (Otoconsult NV; Govaerts et al., 2006). A§E is an audiological evaluation tool for detection, discrimination, and identification tasks (Govaerts et al., 2006). The suite includes tests for spectral discrimination, intensity coding, temporal fine structure (TFS), speech audiometry, and binaural integration.

Pure-Tone Audiometry

Pure-tone thresholds at octave frequencies between 250 and 8000 Hz were obtained from all participants using the modified Hughson–Westlake down-up procedure (Carhart & Jerger, 1959; Hughson & Westlake, 1944). NHs were tested in a sound-treated booth using an Aurical audiometer (Natus Medical Incorporated) and a TDH-39P headphone (Telephonics Corporation).

CI users were tested in free-field condition using Otocube (Otoconsult NV). Otocube is a portable desktop box that replaces classic soundproof booths in the testing of CI patients. Otocube has a built-in loudspeaker that allows delivering the stimuli to the patient's sound processor in isolation from the external environment. Before testing in the Otocube, a long coil cable was first connected to the sound processor. The sound processor was then placed in the Otocube, as shown in Figure 1, and the tests were conducted.

Speech Audiometry in Quiet

Before the HSE and SQ tests, speech in quiet test in CI users was performed with their everyday program settings using Flemish monosyllables (Wouters et al., 1994) in Otocube (Otoconsult NV). Speech recognition scores (SRSs) were obtained by presenting two lists of 12 words at four different levels (40, 55, 70, and 85 dB SPL). The weighted average was calculated using the following formula:

$$EaSI = \frac{SRS_{40} + SRS_{55} + (2 \times SRS_{70}) + SRS_{85}}{5}, \quad (1)$$

Figure 1. Portable desktop box (Otocube) used in audiometric examinations of cochlear implant users.

where EaSI stands for Eargroup Speech Index and SRS_x stands for phoneme score at the presentation level of "x" dB SPL.

HSE and SQ

The test setup was created using three Fostex 6301NB Personal loudspeakers (Foster Electric Company Ltd). The speakers were placed at a distance of 1 m from the participant, at 0°, +90°, and −90° azimuth (see Figure 2). The test room was untreated and had an ambient background noise level of 30 dB (A). The reverberation time in the test room varied between 0.40 and 0.52 s for frequencies between 125 and 4000 Hz. The test parameters were as follows.

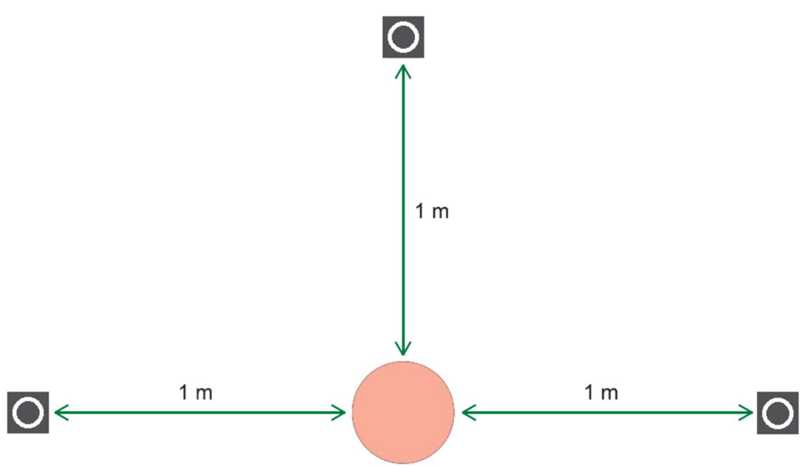
Speech material. SQ and HSE scores were determined by comparing the results of the speech understanding in noise tests performed in binaural and monaural listening conditions. Three different types of speech materials (NVA, BLU, and LiCoS) were used for the first phase of the study.

NVA speech material consists of monosyllabic consonant–vowel–consonant (CVC) words. The test contains 15 lists of 12 words each. Each list has a similar set of initial consonants, vowels, and final consonants. NVA lists are ideal for evaluating speech understanding using phoneme scores based on the percentage of correctly identified phonemes (Wouters et al., 1994). Each correctly repeated phoneme counts about 2.78% since each list has 36 phonemes. Participants had to repeat the words or at least the phonemes they heard in the NVA test. For example, a participant who repeated the word *kop* as *kot* was given a 5.6% phoneme score and 0% word score.

BLU was developed in response to the need for a Flemish speech test based on spondaic words (Wouters et al., 1994). The test contains 15 lists of 10 spondee words each. Therefore, each correctly identified spondee word counts for 10%. Each word has the CVC–CVC structure, and each syllable is a separate existing word. BLU lists are suitable materials for evaluating speech understanding in both quiet and noise based on word scores. The test required participants to repeat every word they heard. Since BLU spondees consist of two syllables that form two separate words, the lists can also be scored based on syllables (Bosman et al., 1995). In such a case, for example, someone who repeated "brood-mes" as "groot-mes" scored 5%.

As a more representative speech material of modern Dutch and Flemish, LiCoS was developed by Coene et al. (2018). The test material consists of sentences articulated by one female and one male speaker. In LiCoS, there are 12 lists of 30 sentences each with two keywords. LiCoS lists comprise sentences with varying syntactic complexity. Different lists are balanced in terms of various linguistic parameters, including lexical, phonological, morphological, and syntactic components of modern Dutch and Flemish (Coene et al., 2018). The number of words in each sentence varies between 6 and 10. The lists are also balanced based on the length of the sentences. However, none of the sentences comprises fixed expressions or proverbs, ensuring low semantic predictability. In LiCoS, the keywords are balanced for eight major word classes (adjectives, adverbs, nouns, prepositions, pronouns, quantifiers, verbs, and conjunctions; Coene et al., 2018). The keywords within the same sentence never belong to the same word class. The participants

Figure 2. Head shadow effect and binaural squelch test setup with three speakers at 0°, –90°, and +90° azimuth, each at a 1-m distance from the participant.



are given 1.67% for each correctly repeated keyword within the same list. The participants had no preknowledge of how the sentences were scored or about the keywords. Their task was to try to repeat all the words in the sentences they heard.

In the first phase of the study, six NH participants underwent both the HSE and the SQ tests. Eleven CI participants were also tested, either with the HSE test ($n = 5$) or the SQ test ($n = 5$) or both ($n = 1$). The tests were done twice (to calculate the test–retest difference) with each material. All participants took a 10-min break between the test and retest sessions. Participants who were tested with HSE only, for example, had to do the test 6 times (twice with NVA, twice with BLU, and twice with LiCoS). In each participant's test, the order of the presentation of the speech materials was determined randomly.

One list from each speech material was played to the participants before the test, and the answers to these lists were not included in the scoring. All subsequent tests used different lists in a random order to minimize learning effects. After the study's first phase, only the speech material that provided the optimal results based on the predefined criteria for the HSE and SQ was used.

Reference ear. The reference ear was the ear that was tested in the monaural listening condition. In CI users, the reference ear was always the first implanted ear. However, the reference ear of each NH participant was randomly assigned. The right ear was tested in half of the participants, and the left ear in the other half.

Locations of the signal and noise. In the HSE test, noise was presented from the speaker at 0° azimuth, while signal was presented from the second ear (nonreference) side (± 90° azimuth). In the SQ test, on the other hand, the signal was presented from 0° azimuth, whereas noise was presented from the second ear side.

Noise. Steady-state speech-shaped noise was used for the measurements. The noise used for each speech material (NVA, BLU, and LiCoS) was different and had the same long-term average spectrum as the target speech.

Adaptive algorithm. Both the SQ and HSE tests began by determining a speech recognition threshold (SRT) using an adaptive algorithm. The adaptive procedure that was followed in this study was based on the simple up–down staircase method. The parameters of the algorithm are listed in Table 1.

Once an SRT was determined using the adaptive algorithm in the binaural listening condition, the test proceeded to the next step. Here, individuals were tested in the monaural listening condition at the SNR determined in the previous step. For example, if a person had an SRT of +5 dB SPL in the first test, the second test was run at a fixed SNR of +5 dB SPL (signal at 70 dB SPL and noise at 65 dB SPL). Since the SRT determined in the first step corresponded to a 70% correct response, a 60% score in the second test, for example, would indicate a 10% binaural benefit compared with monaural listening. Figures 3 and 4 show examples of the display of the HSE and SQ test results.

Table 1. The parameters of the adaptive algorithm used in HSE and SQ tests.

Initial signal level	75 dB SPL
Initial SNR	10 dB SPL
Noise intensity	65 dB SPL (Fixed)
Target	70%
Initial step	10 dB
Minimum step size	2 dB
Step size calculation	$s = S_i \times \left(\frac{1}{2}\right)^R$
Stop criterion	After eight reversals
Threshold estimation	Average of last six reversals

Note. HSE = head shadow effect; SQ = binaural squelch; SPL = sound pressure level; SNR = signal-to-noise ratio; s = step size = S_i, initial step size; R = number of reversals.

CI users were asked to turn off their second CI in the monaural listening condition. The nonreference ears of the NH group were blocked with E-A-RTone 3A insert earphones (Aearo Technologies LLC) and masked via a Madson Itera II audiometer (Natus Medical Incorporated), with a 60-dB HL broadband speech noise.

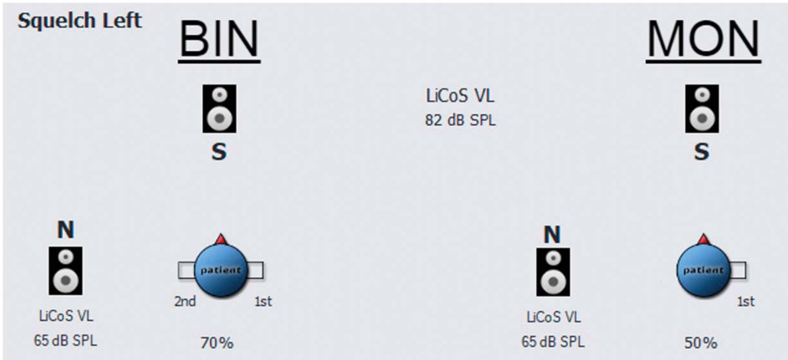
Azimuth LOC

An azimuth LOC test was carried out using seven Fostex 6301B loudspeakers (Foster Electric Company Ltd). This test was performed in the same room used for HSE and SQ tests. The test stimulus was broadband speech noise presented at 70 dB SPL. A ± 3 dB level rove was applied, which means that the presentation level for each loudspeaker randomly varied between 67 and 73 dB SPL, to avoid any additional cues caused by the characteristics of the loudspeakers.

In the test, speakers were placed at 20° angle intervals from −60° (left side) to +60° (right side) and numbered from −3 to +3. Figure 5 illustrates the test setup.

Participants received a brief training before starting the test with feedback on their answers. In the test, stimuli were presented from the seven speakers in random order, and individuals were asked to point to the speaker from which they heard the sound. The stimulus was presented 5 times from each speaker, and at the end of the test, root-mean-square (RMS) was calculated.

Figure 3. BIN: Binaural squelch (SQ) test with both ears available with the signal coming from the front and the noise coming from −90° azimuth when the right ear is selected as the first ear. The speech recognition threshold obtained at 70% correct response rate in this example is 17 dB SPL (82–65 dB SPL). MON: SQ test with only the right ear available. The score at 17 dB SNR is 50%, so the SQ = 20% (70%–50%). LiCoS = linguistically controlled sentences.



The RMS represents the mean test error. An RMS value closer to zero indicates good LOC ability. The RMS was calculated as follows:

$$RMS = \sqrt{\frac{\sum_{S=1}^{N}(Rs - NORMs)^2}{N}}, \qquad (2)$$

where Rs denotes the median response for Speaker S, N denotes the number of speakers, and NORMs denotes the normal value for Speaker S.

SSQ Scale

We used the 12-item version of the SSQ, which was translated into Flemish by KU Leuven University (Noble et al., 2013). The SSQ is a subjective self-assessment tool that mainly focuses on speech, spatial listening, and other qualities of hearing (Gatehouse & Noble, 2004). Since its development, SSQ has been translated into many languages (Falzone et al., 2022; Kılıç et al., 2021; Moulin et al., 2015) and has been widely used in many other studies (Jang et al., 2019; Laske et al., 2009; Mertens et al., 2015; Vermeire & Van de Heyning, 2009). In this scale, respondents rate their own performance on each item from 0 (*not at all*) to 10 (*perfect*). Higher score always indicates a greater ability since none of the items is worded in a negative direction.

Statistical Analyses

Descriptive statistics (minimum, first quartile, median, third quartile, maximum, and interquartile range) were used to summarize data. The independent-samples *t* test was used to compare the CI and NH groups. The paired-samples *t* test and the Wilcoxon signed-ranks test, depending on

Figure 4. BIN: Head shadow effect (HSE) test with both ears available with the signal coming from −90° azimuth and the noise coming from front when the right ear is selected as the first ear. The speech recognition threshold obtained at 70% correct response rate in this example is 9 dB SPL (74–65 dB SPL). MON: HSE test with only the right ear available. The score at 9 dB SNR is 30%, so the HSE = 40% (70%–30%). VL = Vlaams (English translation: Flemish).

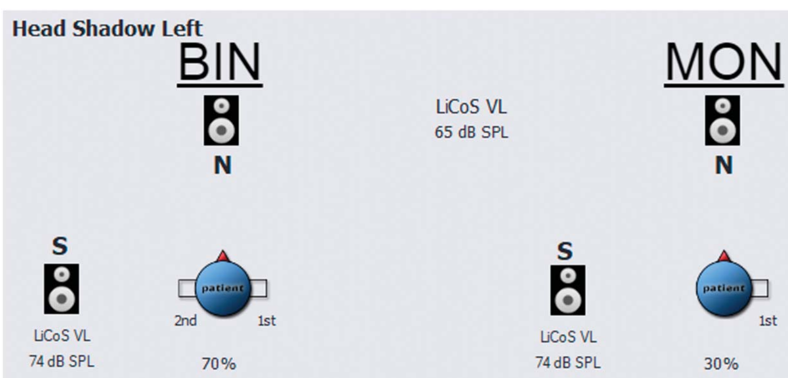
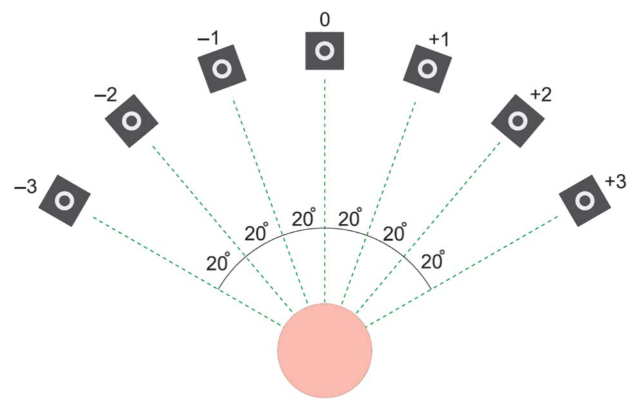


Figure 5. Azimuth localization test setup, with seven loudspeakers each at a 20° angle from the other.



whether parametric test assumptions were met, were used to compare the performance of the first and second implanted ears in the CI group. The Friedman test was used to compare the results of three different speech materials. During the analysis, only the first results from each speech test were used. Retest results were not taken into account in this analysis. The Spearman's rank correlation coefficient was used for correlation analysis. The Shapiro–Wilk test was used to test for normality. The significance level was always set at 0.05. IBM SPSS Statistics for Windows, Version 20.0 (IBM Corp.) was used for data analysis.

At the end of the first phase, the following data were available for each speech material: (a) test results (= effect values), (b) test–retest results, and (c) interindividual variability. The performance of a diagnostic test is judged by how accurately the test result can identify a diseased (sensitivity) or a nondiseased (specificity) person (Hui & Zhou, 1998). To ensure adequate sensitivity and specificity in the HSE and SQ tests, the NH and CI groups must be distinguishable from each other. It is therefore reasonable to choose speech material that produces the largest difference between the groups. Interindividual variability is called random, unexplained, residual, or variation due to error. That's why it is expected to be as low as possible in control groups (in this case the NH group; Drummond & Vowler, 2012). Then, finally, the test–retest reliability is an indication of the internal validity of the test and of its ability to produce stable and reproducible results over time. For these reasons, we determined that the optimal speech material would be characterized by (a) the largest effect size in NHs, (b) the highest test–retest reliability in NHs and CIs, and (c) the smallest interindividual variability in NHs.

Effect size calculations used Cohen's d for independent-samples and paired-samples t tests, effect size r for Wilcoxon signed-ranks test, and Kendall's W for the Friedman test.

Test–retest reliability was calculated using the two-way mixed intraclass correlation coefficient (ICC; Koo & Li, 2016; McGraw & Wong, 1996). ICC is a widely used reliability index in test–retest analysis (Koo & Li, 2016; McGraw & Wong, 1996). ICC has advantages over other methods that can be used for test–retest analysis such as paired t test and Pearson correlation. For example, the paired t test is a method for analyzing agreement between two measurements, and Pearson correlation is a way to measure correlation that does not take systematic differences into account. ICC, however, reflects both degree of correlation and agreement (Koo & Li, 2016). ICC values less than 0.5 indicate poor test–retest reliability, those between 0.5 and 0.75 indicate moderate reliability, those between 0.75 and 0.9 indicate good reliability, and those greater than 0.9 indicate excellent reliability (Koo & Li, 2016; McGraw & Wong, 1996; Portney & Watkins, 2009).

Interindividual variability was calculated using the coefficient of variation (CV). It is calculated by dividing the standard deviation by the mean.

Results

In the first implanted ears of the CI recipients, the mean aided PTA was 18 ± 5 dB HL, whereas in the second implanted ears, it was 22 ± 6 dB HL. The median aided EaSI score was 89% (Q1: 79.8 and Q3: 93) on the first implanted side and 82.5% (Q1: 76.8 and Q3: 89) on the second implanted side. The PTA and the EaSI scoreson the first CI side were 4 ± 6 dB HL (parametric paired-samples t test; $t[33] = -3.812$, $p = .001$, Cohen's $d = 0.65$) and 4% (Q1: −0.3 and Q3: 8.3;nonparametric Wilcoxon signed-ranks test; $Z = -2.602$, $p = .009$, effect size $r = .32$) significantly better than those on the second CI side.

The Effect of Speech Material in the HSE and SQ Tests

Effect Size

Both for the HSE and SQ tests, the BLU and LiCoS test results in the NH group ($n = 6$) were higher than those of the NVA, although the differences were not statistically significant (nonparametric Friedman test for HSE: $\chi^2[2] = 4.364$, $p = .113$, Kendall's $W = 0.36$, and for SQ: $\chi^2[2] = 4.333$, $p = .115$, Kendall's $W = 0.36$). Table 2 presents the descriptive statistics for the results.

Test–Retest Reliability

For both tests, LiCoS achieved the highest $ICC_{agreement}$ in NH individuals (83.4% and 82.6%; see Table 3). The median test–retest differences in HSE for NH individuals

Table 2. Descriptive statistics of the HSE and SQ test results in NHs (Phase I, n = 6).

Descriptive statistics		Min	Q1	Median	Q3	Max	IQR	95% CI
HSE (%)	NVA	24	26	44	64	70	38	[26, 66]
	BLU	40	49	61	70	70	21	[46, 70]
	LiCoS	37	51	62	70	70	20	[46, 70]
SQ (%)	NVA	−15	−8	7	18	21	26	[−11, −19]
	BLU	5	13	19	54	58	41	[10, 55]
	LiCoS	2	4	19	27	37	22	[4, 30]

Note. HSE = head shadow effect; SQ = binaural squelch; NH = normal hearing; Min = minimum; Q1 = first quartile; Q3 = third quartile; Max = maximum; IQR = interquartile range; 95% CI = 95% confidence interval for the median; LiCoS = linguistically controlled sentences.

(n = 6) were 12% (95% CI [5, 19]), 4% (95% CI [0, 11]), and 8% (95% CI [1, 10]) for NVA, BLU, and LiCoS, respectively. These values were 9% (95% CI [1, 19]), 5% (95% CI [0, 21]), and 5% (95% CI [0, 12]) for SQ. The test–retest differences obtained with NVA, BLU, and LiCoS in the NH group were not statistically different from each other for either the HSE (nonparametric Friedman test; $\chi^2[2] = 1$, $p = .607$, Kendall's $W = 0.08$) or SQ ($\chi^2[2] = 0.09$, $p = .956$, Kendall's $W = 0.008$) tests.

In the CI group (n = 11), NVA's $ICC_{agreement}$ was the highest in HSE (91.7%), whereas LiCoS's was the highest in SQ (69.9%; see Table 3). The median test–retest differences for HSE were 8% (95% CI [2, 15]), 16% (95% CI [8, 25]), and 12% (95% CI [4, 22]) for NVA, BLU, and LiCoS, respectively. For SQ, they were 11% (95% CI [5, 19]), 18% (95% CI [6, 30]), and 10% (95% CI [6, 20]). The test–retest differences obtained with NVA, BLU, and LiCoS in the CI group were also not statistically different from each other for either the HSE (nonparametric Friedman test; $\chi^2[2] = 3.273$, $p = .195$, Kendall's $W = 0.27$) or SQ (Friedman test; $\chi^2[2] = 1.455$, $p = .483$, Kendall's $W = 0.12$) tests.

Interindividual Variability

BLU had the lowest CV among NH individuals (n = 6, 20%) for HSE, and LiCoS had the lowest CV for SQ (74%); see Table 3.

The interquartile range in the NH group in the HSE test was 37.8%, 21%, and 19.5% for NVA, BLU, and LiCoS, respectively. For the SQ, the values were 26.3%, 41%, and 22.3%.

In summary, for the criteria tested, there was no statistical difference between the three different speech materials for the HSE and SQ tests. However, on the basis of descriptive statistics, results indicated that BLU and LiCoS were more robust materials compared with NVA. Although BLU and LiCoS showed similar results in terms of effect size and interindividual variability, we still judged LiCoS to be more effective than BLU because of its higher test–retest reliability, especially in CI users. All the results of this phase are summarized in Table 3.

Normative Data of HSE and SQ

Thirty NHs were tested to obtain normative data for the HSE and SQ tests with LiCoS sentences. The mean HSE (± *SD*) was 58 ± 14% (95% CI [53, 64]), and the mean SQ was 22 ± 11% (95% CI [17, 26]) in the NH group.

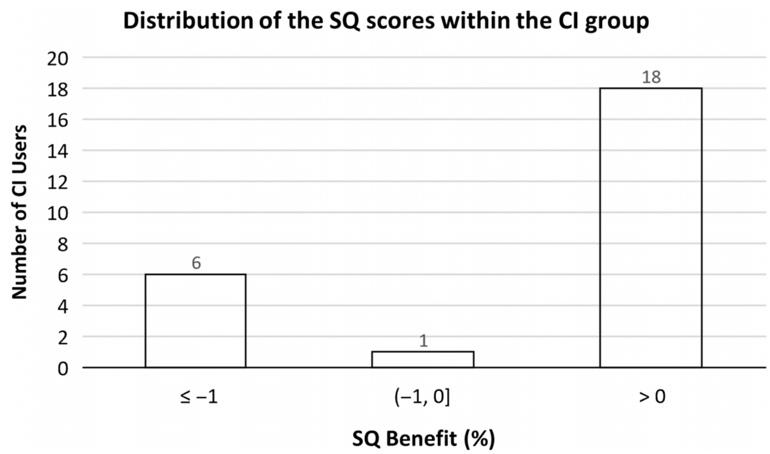
The initial phase of this study revealed that the median test–retest differences in NH individuals with the LiCoS sentence test for HSE and SQ were 8% (Q1: 2 and Q3: 10; range: 0–10) and 5% (Q1: 0 and Q3: 11; range: 0–13), respectively.

Table 3. Summary of all the results in the first phase of the study.

Measure	Test	NVA	BLU	LiCoS
Effect size (NH) (%) (n = 6)	HSE	44	61	62
	SQ	7	19	19
Test–retest reliability (NH and CI) (%) (n = 17)	HSE (NH)	83	82	83
	HSE (CI)	92	58	69
	SQ (NH)	55	82	83
	SQ (CI)	1	44	70
Interindividual variability (NH) (%) (n = 6)	HSE	41	20	21
	SQ	300	77	74

Note. LiCoS = linguistically controlled sentences; NH = normal hearing; HSE = head shadow effect; SQ = binaural squelch; CI = cochlear implant.

Figure 6. Histogram showing the distribution of the binaural squelch (SQ) test results within the cochlear implant (CI) group.

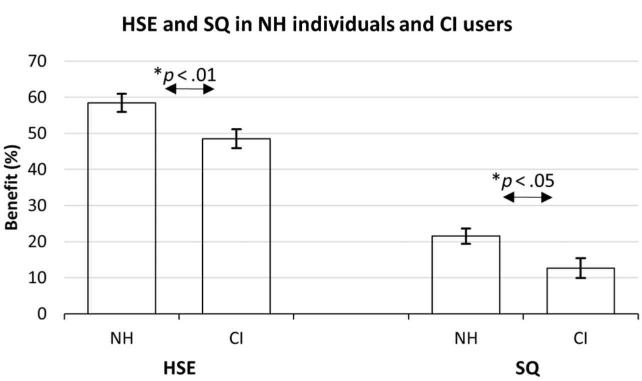


Binaural Benefits in Bilateral CI Users

The mean HSE was 49 ± 13% (n = 25, 95% CI [42, 54]), and the SQ was 13 ± 14% (n = 25, 95% CI [8, 20]) in the CI group. There were six out of 25 participants (24%) with a negative SQ in the CI group (see Figure 6 for the histogram). Furthermore, the HSE and SQ scores of the CI group were significantly lower than those of the NH group, as illustrated in Figure 7 (parametric independent-samples t test for HSE: $t[53] = 2.73$, $p = .009$, Cohen's $d = 0.74$, and for SQ: $t[53] = 2.64$, $p = .011$, Cohen's $d = 2.48$).

The median test–retest differences in the CI users with the LiCoS sentence test for HSE and SQ were 12% (Q1: 5 and Q3: 20; range: 2–25) and 10% (Q1: 6 and Q3: 18; range: 5–25), respectively. The average RMS error in the azimuth LOC test was 15° ± 5° (n = 21, 95% CI [13, 17]); see Figure 8.

Figure 7. Head shadow effect (HSE) and binaural squelch (SQ) test results (mean ± standard error of the mean) in normal-hearing (NH) individuals (n = 30) and cochlear implant (CI) users (n = 25). Independent-samples t test was used to obtain the reported p values.



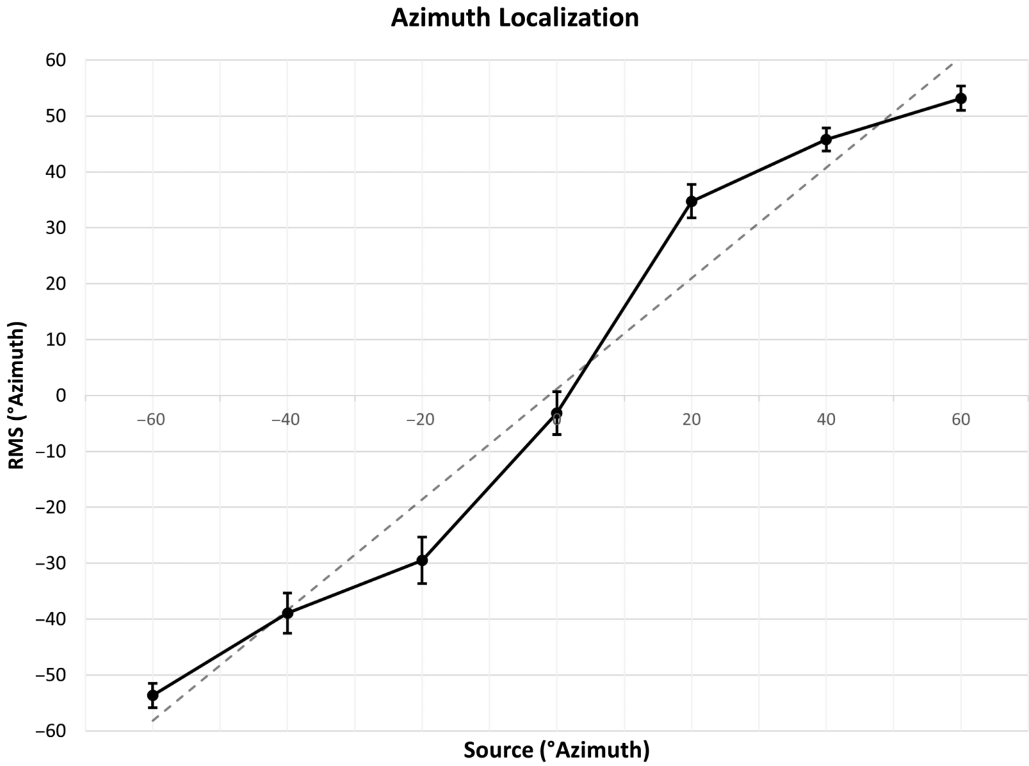
The average score on the SSQ-12 (n = 16) was 5 ± 2 (95% CI [4, 6]). There were no statistically significant correlations between the HSE, SQ, LOC, and SSQ results (nonparametric Spearman's rank correlation coefficient for HSE-LOC: $r(17) = -.061$, $p = .809$; HSE-SSQ: $r(12) = .229$, $p = .452$; SQ-LOC: $r(17) = .120$, $p = .636$; SQ-SSQ: $r(12) = -.237$, $p = .435$; LOC-SSQ: $r(15) = -.162$, $p = .549$).

Further analysis did also not reveal any significant correlations between the test results and the demographic variables such as the interval between the implantations, duration of binaural experience, and asymmetry between the ears (nonparametric Spearman's rank correlation coefficient for HSE-Interval between the CI's: $r(24) = -.240$, $p = .248$; HSE-Binaural experience: $r(24) = -.063$, $p = .766$; HSE-PTA asymmetry: $r(24) = -.148$, $p = .480$; HSE-EaSI asymmetry: $r(24) = .032$, $p = .879$; SQ-Interval between the CI's: $r(24) = -.081$, $p = .702$; SQ-Binaural experience: $r(24) = .132$, $p = .530$; SQ-PTA asymmetry: $r(24) = -.230$, $p = .268$; SQ-EaSI asymmetry: $r(24) = -.154$, $p = .462$).

Discussion

The primary purpose of this study was to identify which speech material is most effective for the HSE and SQ tests. After that, normal values in NH individuals were obtained using the speech material decided to be optimal. In addition, LOC, SSQ, HSE, and SQ tests were used to test bilateral CI users.

Figure 8. Azimuth localization test results (mean ± standard error of the mean; black solid line) in cochlear implant users (n = 21), and the reference curve (gray dashed line). RMS = root-mean-square.



The benefits of hearing with two ears compared to hearing with one ear can be divided into two groups: physical effects and central effects. HSE is defined as a purely physical effect (Brown & Balkany, 2007; Van Hoesel & Litovsky, 2011), whereas SU, SQ, and LOC result from binaural processing in the central auditory system (Tyler et al., 2002, 2003). SU does not originate from interaural cues and is solely dependent on binaural redundancy (Laszig et al., 2004; Van Hoesel, 2012). Hence, a more redundant material in the SU test could result in a larger effect size, which would result in lower test–retest differences. Since HSE and SQ do not rely on redundancy, but rather originate from interaural differences, it is not possible to make such an assumption for them. As for SRM, its origins are controversial. While some researchers suggest it is a combination of HSE and SQ (Aronoff et al., 2011; Dieudonné & Francart, 2019), others say it is only the result of HSE (Gifford et al., 2014; Sheffield et al., 2015). However, in any case, its origins are not different from HSE and/or SQ. For these reasons, only HSE and SQ were investigated in this study. Even though we believe that the results of this study can be generalized to SU and SRM, the fact they were not included in this study can be considered a shortcoming.

The Effect of Speech Material in the HSE and SQ Tests

The materials tested in the initial phase included monosyllabic NVA words, spondaic BLU words, and LiCoS sentences.

The small sample size in the first phase was the most likely reason for the lack of statistical differences between the speech materials. Due to the absence of statistically significant differences based on inferential statistics, descriptive statistics were used to interpret the results of the first phase. This might be considered one of the limitations of this study. However, descriptive statistics indicated that the sentences provided higher test–retest reliability than the monosyllables and spondees.

One might question the applicability of linguistically controlled sentences in children with CI. Existing research suggests that speech intelligibility measured by predictable linguistically controlled sentences varies mainly by age and cognitive abilities (Uslar et al., 2011, 2013). Uslar et al. (2011) showed that young listeners have much more difficulty with complex speech materials than with simple materials in comparison to adults. Pichora-Fuller (2008) and Uslar et al. (2011) attributed the advantageous effects of older age to (a) greater experience with language and speech processing in difficult circumstances; (b) improved benefits from supportive context and redundancies, such as syntactic cues; and (c) their expert knowledge of structures when signal quality is poor. However, when the sentences are unpredictable with varying degrees of syntactic complexity (like in LiCoS), older adults have no advantage over young listeners anymore because the strategies used by experienced listeners do not work for unpredictable sentences (Uslar et al., 2011, 2013). Uslar et al. (2013) reported also that cognitive skills play a greater role in speech understanding of complex sentences in the presence of fluctuating noise than in the presence of stationary noise. In other words, listening in fluctuating noise requires a higher cognitive contribution than, for instance, listening in quiet or in stationary noise. It is important to note that the noise used in this study was stationary. Cognitive skills are, therefore, not expected to have a significant impact on the outcomes of the LiCoS test.

Effect Size

To the best of our knowledge, there are no other studies in the literature evaluating and comparing the effectiveness of different speech materials in HSE and SQ tests. Nonetheless, when the results of a variety of studies, each of which used a different material, are analyzed together, it becomes evident that the choice of speech material could influence the outcomes. As an example, monosyllabic words have the lowest SQ effect (1.9–3.7 dB SRT), whereas spondee words and closed-set sentences produce similar effects (4–7 dB and 4.5–7 dB SRT, respectively; Bronkhorst & Plomp, 1988; Carhart, 1965; Dirks & Wilson, 1969; MacKeith & Coles, 1971). In this study, we also found that spondee words and sentences had the largest effect sizes both in the HSE and SQ tests. Thus, it becomes apparent that binaural benefits increase with the redundancy of the speech material of interest. In the test setup, we create a difficult listening situation that creates gaps in the incoming acoustic signal. The test task is to fill those gaps by adding the second ear, and it turns out that this addition does so better if the acoustic material also contains redundant information.

Test–Retest Reliability

Studies conducted so far have used a variety of speech materials. Although many researchers used sentence stimuli, some also used other materials. For example, Morera et al. (2012) and Verhaert et al. (2012) used spondee words, and Arsenault and Punch (1999) used nonsense syllables. In their studies, Aronoff et al. (2011) and Devocht et al. (2017) evaluated test–retest variations of the sentences they used but made no comparison with other materials. Aronoff et al. (2011) found a statistically significant correlation between test and retest results with a correlation coefficient of .9. In this study, however, the ICC method was preferred over correlation analysis. In both the HSE and SQ tests in NHs, LiCoS had an ICC value of > 80%, indicating good reliability (Koo & Li, 2016). For CI users, the ICC values were 69% and 70%, respectively, indicating moderate reliability, which is also in line with the results reported by Devocht et al. The ICC values reported by Spyridakou et al. (2020) for the right and left ears in speech understanding in noise test using monosyllabic words in NH individuals were 25% and 39%, respectively. The ICC values for NH individuals obtained using monosyllabic words in this study were 82.6% and 55.3% for the HSE and SQ, respectively (for LiCoS sentences the values were even higher: 83.4% and 82.6%). However, the test–retest variability observed among NHs was lower than that among CI users, possibly due to ceiling effects as indicated by other studies demonstrating that test–retest variation decreases with the increasing SNR of the test, with the increasing performance level of the listeners, and with the decreasing linguistic complexity of the speech material (Grange, 2013; Hey et al., 2014; Kim et al., 2015; Uslar et al., 2011, 2013). The number of items in a speech list is another important factor affecting test–retest differences. Kim et al. (2015) demonstrated that there was no significant difference in test–retest differences between 25- and 50-item lists, but there was a significant difference when a 10-item list was included in the comparison. In each speech list, NVA and BLU had 12 items; LiCoS had 30. However, the participants in this study were presented with two lists of NVA and BLU and one list of LiCoS to establish an equivalence. As a result, 24 items from each of the NVA and BLU tests and 30 items from LiCoS were presented to the participants, ensuring that the number of items would not affect the test–retest scores.

Even though the test–retest reliability reported in this study is consistent with those reported in the literature, and ICC analysis shows that these tests have moderate-to-good test–retest reliability depending on the target population, the results should be interpreted with caution. As discussed above, test–retest differences in speech audiometry can reach up to 10% in NH individuals and up to 15% in CI users. This makes it particularly difficult or risky to interpret such results in individual test subjects. It is only by analyzing group results that the relatively large test–retest differences get averaged out. Therefore, the authors believe that interpreting speech audiometry results, in this case especially the binaural test results, would yield more reliable results on a group basis.

Interindividual Variability

In line with previous research, the results showed considerable variation among individuals. As expected, the variability among individuals in the NH group was lower than that in the CI group. One might argue that the higher variation in the CI group was due to the wider age range of the individuals. However, there were no outliers in the data of the participants aged > 50 years or < 16 years (a total of seven participants). In HSE, five out of seven participants (71%), and in SQ, all seven participants had data within the mean ± 1 *SD*. Other possible reasons for the higher variation in the CI group's results include different CI signal processing strategies, the different spread of the electrical field generated by the implant, and the preservation of spiral ganglion cells (Williges et al., 2015). Future research can focus on the development of test methods (e.g., different test setups or implementation of electrophysiological methods) that show less variable results in the assessment of binaural benefits.

Normative Data of HSE and SQ

The mean HSE was 58 ± 14%, and the mean SQ was 22 ± 11% in the NH group. In the literature, HSE varies between 8.9 and 10.7 dB in individuals with NH (Arsenault & Punch, 1999; Bronkhorst & Plomp, 1988), and the SQ ranges between 1.9 and 4.9 dB (Arsenault & Punch, 1999; Bronkhorst & Plomp, 1988; Carhart, 1965; Cox et al., 1981; MacKeith & Coles, 1971). In test setups run at fixed SNRs, HSE ranges between 20% and 30% (Arsenault & Punch, 1999) and SQ ranges between 10% and 26% (Arsenault & Punch, 1999; Cox et al., 1981).

Binaural Benefits in Bilateral CI Users

Bilateral CI users in this study had 49 ± 13% HSE and 13 ± 14% SQ. These results are in line with previous studies in literature in which HSE ranged from 22% to 49% and SQ from 1.7% to 18% (Buss et al., 2008; Eapen et al., 2009; Gantz et al., 2002; Laszig et al., 2004; Tyler et al., 2002; Van Hoesel, 2012; Van Hoesel et al., 2002; Verhaert et al., 2012).

HSE had the most robust results. In this study, all CI users (n = 25) showed HSE, but 72% were able to have a positive SQ. Tyler et al. (2002) and Gantz et al. (2002) demonstrated that 80% of bilateral CI users had an HSE, whereas a significant SQ was reported only in three out of nine participants. Laszig et al. (2004), on the other hand, reported significant HSE and SU but no significant SQ even 6 months after the implantation. Litovsky et al. (2006) reported that 94% of 34 simultaneously implanted bilateral CI users had HSE, whereas only 47% had SQ. Laske et al. (2009) reported significant results after bilateral implantation only in HSE. Thus, in most studies, less than half of bilateral CI users have a measurable SQ benefit from their second CI. This finding seems important to us. This means that more than half of the participants do not experience a significant SQ after the second CI, and some even have a negative SQ, indicating a deterioration of speech understanding in noise with the second CI on compared with the first CI alone. In our sample, 24% of the participants also had a negative SQ. However, better speech understanding in noise is actually often cited as an important argument for a second CI. It turns out that we should be careful with this argument. Therefore, the experience that most bilateral CI users are satisfied with the second implant is mainly attributable to the HSE that is observed for a high proportion.

Overall, the binaural test results of CI users were significantly lower than those of NH individuals. Other studies also reported similar results (Arsenault & Punch, 1999; Bronkhorst & Plomp, 1988; Kokkinakis & Pak, 2014). A functional hearing system is able to process and integrate bilateral acoustic cues smoothly. However, it is not always possible to achieve a similar success in artificial hearing provided by electrical stimulation in CI users. CI users perform less well in binaural hearing tasks than NH people for a variety of reasons. Most notably, CI users have a reduced sensitivity for interaural time and level differences (ITD and ILD) in addition to inadequately encoded TFS information with current sound processing strategies (Brown & Balkany, 2007; Kokkinakis & Pak, 2014; Litovsky et al., 2006; Tyler et al., 2003). In their study, Litovsky et al. (2006) outlined three main reasons for reduced binaural advantages in CI users: "hardware- and software-related, surgical-based, and pathology-related." In summary, CI users have two independent monaural hearing systems. Thus, the time base of each processor can differ slightly, resulting in random jitters in the ITD of the envelope and the carrier pulses, disrupting the ITD cues (Litovsky et al., 2006). Most CI systems process incoming signals by extracting only the temporal envelope and amplitude, modulating it to a fixed-rate pulsatile carrier (Ching et al., 2007). They do not provide the TFS information that is critically important for detecting ITDs. Despite the potential for the temporal envelope to convey timing information, ITDs would not be consistent because of the variations in detection thresholds across different electrodes (Ching et al., 2007). Furthermore, although CI users' speech understanding improves with higher pulse rates, their ITD sensitivities drop significantly (Dunn et al., 2006; Thakkar et al., 2018; Van Hoesel et al., 2009). Additional factors that may distort binaural cues include different microphone characteristics, independent automatic gain control and compression algorithms, and different signal processing strategies between the two implants (Brown & Balkany, 2007; Ching et al., 2007; Litovsky et al., 2012; Snik et al., 2015; Tyler et al., 2002, 2003, 2006). For instance, when one of the processors compresses its input more than the other, the brain perceives the sound as moving from one side to the other, which may negatively affect spatial hearing (Tyler et al., 2003). Another problem is the spectral mismatch between the electrodes due to the different surgical insertion depths. Accordingly, Yoon et al. (2013) found that increased spectral mismatch caused by different insertion depths affected SQ negatively but not HSE. A binaural CI may eliminate the aforementioned problems of two independent CI systems. A binaural CI has two different electrode arrays protruding from a single internal device, and these electrodes are placed in both cochleae. While there is a sound processor on the side of the internal device, the contralateral ear only has a microphone connected to the sound processor by a cable. Verhaert et al. (2012) investigated the effects of binaural cochlear implantation after 12 months of use in 14 adults with postlingual hearing loss. There was a significant difference between participants' SRSs in silence and noise in the binaural condition compared with the unilateral condition. Significant binaural advantages were also present in HSE, SU, and SQ tests. In addition, a significant improvement of 35° RMS was observed in the LOC task. These results confirm that pseudosynchronous stimulation of binaural CIs might have positive effects on binaural hearing.

There was no significant correlation between HSE/SQ test results and LOC. Similarly, Schleich et al. (2004) and Tyler et al. (2006) also could not find a significant correlation between HSE/SQ and horizontal LOC in bilateral CI users. While HSE is not directly related to spatial listening, it is believed that SQ relies on the same binaural cues (ITD) that also allow the LOC of sound sources (Eapen et al., 2009). It was therefore expected that LOC and SQ results could have a significant correlation. Lack of correlation between them raises the question of whether they represent the ability to use the same cues. Similarly, Cox and Bisset (1984) concluded in their study that SQ tested with traditional methods does not reflect the ability to exploit interaural differences. These findings support the idea that future research should focus on the development of different test methods to assess SQ.

The SSQ was used in this study to examine how well the psychoacoustic test results matched patients' daily life experiences. However, SSQ scores were not significantly correlated with HSE/SQ or LOC results. Using a similar sample of 34 bilateral CI users with at least 6 months of binaural experience, Laske et al. (2009) found that the LOC results were significantly correlated with the "spatial hearing" subcategory of the SSQ. Despite similar patient populations, there were several methodological differences between the two studies, including the type of noise (broadband vs. speech) and the total number of speakers (12 vs. 7) used in the LOC test, and the number of items in the SSQ (25-item vs. 12-item). These methodological differences may have led to different results in the studies.

The results of this study showed that none of the variables investigated correlated with the binaural hearing advantages. Similarly, Schleich et al. (2004), Tyler et al. (2006), and Laske et al. (2009) could not find a relationship between binaural benefits and other variables such as the duration of deafness of the ears or the asymmetry in the performance of the ears. These findings support the argument of Gantz et al. (2002) that there is no parameter for predicting the postoperative binaural benefits. We believe that the main reason for this is the high variation in both individual and group results observed in the HSE and SQ tests.

Conclusions

The primary purpose of this study was to identify which speech material is most effective for the HSE and SQ tests. The results showed that although there were no significant differences between spondee words and sentences in terms of effect size and interindividual variability, test–retest reliability was higher with sentence stimuli, especially in CI users. With the LiCoS sentence test as optimal speech material selected, normative data were obtained in NH individuals. All CI users in the study sample had an HSE, whereas 72% had SQ. We conclude that the benefit of the second CI is mainly attributable to HSE whereas SQ benefit is not achieved in all cases.

Data Availability Statement

All data generated and/or analyzed in this study are available from the corresponding author on reasonable request.

Author Contributions

Okan Öz: Conceptualization (Equal), Data curation (Lead), Formal analysis (Lead), Investigation (Lead), Methodology (Equal), Project administration (Equal), Validation (Equal), Visualization (Lead), Writing – original draft (Lead). **Hilal Dinçer D'Alessandro:** Conceptualization (Equal), Methodology (Equal), Project administration (Equal), Supervision (Equal), Validation (Equal), Writing – review & editing (Equal). **Merve Özbal Batuk:** Conceptualization (Equal), Methodology (Equal), Supervision (Supporting), Writing – review & editing (Supporting). **Gonca Sennaroğlu:** Conceptualization (Equal), Methodology (Equal), Supervision (Supporting), Writing – review & editing

(Supporting). **Paul J. Govaerts:** Conceptualization (Equal), Data curation (Supporting), Formal analysis (Supporting), Methodology (Equal), Project administration (Equal), Resources (Lead), Software (Lead), Supervision (Equal), Validation (Equal), Writing – review & editing (Equal).

References

Aronoff, J. M., Freed, D. J., Fisher, L. M., Pal, I., & Soli, S. D. (2011). The effect of different cochlear implant microphones on acoustic hearing individuals' binaural benefits for speech perception in noise. *Ear and Hearing, 32*(4), 468–484. https://doi.org/10.1097/AUD.0b013e31820dd3f0

Arsenault, M. D., & Punch, J. L. (1999). Nonsense-syllable recognition in noise using monaural and binaural listening strategies. *The Journal of the Acoustical Society of America, 105*(3), 1821–1830. https://doi.org/10.1121/1.426720

Avan, P., Giraudet, F., & Büki, B. (2015). Importance of binaural hearing. *Audiology and Neurotology, 20*(Suppl. 1), 3–6. https://doi.org/10.1159/000380741

Balkenhol, T., Wallhäusser-Franke, E., Rotter, N., & Servais, J. J. (2020). Cochlear implant and hearing aid: Objective measures of binaural benefit. *Frontiers in Neuroscience, 14,* Article 586119. https://doi.org/10.3389/fnins.2020.586119

Bertoli, S., Smurzynski, J., & Probst, R. (2005). Effects of age, age-related hearing loss, and contralateral cafeteria noise on the discrimination of small frequency changes: Psychoacoustic and electrophysiological measures. *Journal of the Association for Research in Otolaryngology, 6*(3), 207–222. https://doi.org/10.1007/s10162-005-5029-6

Bosman, A. J, Wouters, J., & Damman, W. (1995). Realisatie van een cd voor spraakaudiometrie in Vlaanderen [Realization of a CD for speech audiometry in Flanders]. Logopedie en foniatrie: maandblad van de Nederlandse vereniging voor logopedie en foniatrie, *67*(9), 218–225.

Bronkhorst, A. W., & Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *The Journal of the Acoustical Society of America, 83*(4), 1508–1516. https://doi.org/10.1121/1.395906

Brown, K. D., & Balkany, T. J. (2007). Benefits of bilateral cochlear implantation: A review. *Current Opinion in Otolaryngology & Head and Neck Surgery, 15*(5), 315–318. https://doi.org/10.1097/MOO.0b013e3282ef3d3e

Buss, E., Pillsbury, H. C., Buchman, C. A., Pillsbury, C. H., Clark, M. S., Haynes, D. S., Labadie, R. F., Amberg, S., Roland, P. S., Kruger, P., Novak, M. A., Wirth, J. A., Black, J. M., Peters, R., Lake, J., Wackym, P. A., Firszt, J. B., Wilson, B. S., Lawson, D. T., & Barco, A. L. (2008). Multicenter U.S. bilateral MED-EL cochlear implantation study: Speech perception over the first year of use. *Ear and Hearing, 29*(1), 20–32. https://doi.org/10.1097/AUD.0b013e31815d7467

Carhart, R. (1965). Monaural and binaural discrimination against competing sentences. *The Journal of the Acoustical Society of America, 37*(6), 1205–1205. https://doi.org/10.1121/1.1939552

Carhart, R., & Jerger, J. F. (1959). Preferred method for clinical determination of pure-tone thresholds. *Journal of Speech and Hearing Disorders, 24*(4), 330–345. https://doi.org/10.1044/jshd.2404.330

Ching, T. Y. C., Van Wanrooy, E., & Dillon, H. (2007). Binaural-bimodal fitting or bilateral implantation for managing severe to profound deafness: A review. *Trends in Amplification, 11*(3), 161–192. https://doi.org/10.1177/1084713807304357

Coene, M., Krijger, S., Van Knijff, E., Meeuws, M., De Ceulaer, G., & Govaerts, P. J. (2018). LiCoS: A new linguistically controlled sentences test to assess functional hearing performance. *Folia Phoniatrica et Logopaedica, 70*(2), 90–99. https://doi.org/10.1159/000490050

Cox, R. M., & Bisset, J. D. (1984). Relationship between two measures of aided binaural advantage. *Journal of Speech and Hearing Disorders, 49*(4), 399–408. https://doi.org/10.1044/jshd.4904.399

Cox, R. M., DeChicchis, A. R., & Wark, D. J. (1981). Demonstration of binaural advantage in audiometric test rooms. *Ear and Hearing, 2*(5), 194–201. https://doi.org/10.1097/00003446-198109000-00003

Devocht, E. M., Janssen, A. M. L., Chalupper, J., Stokroos, R. J., & George, E. L. (2017). The benefits of bimodal aiding on extended dimensions of speech perception: Intelligibility, listening effort, and sound quality. *Trends in Hearing, 21*. https://doi.org/10.1177/2331216517727900

Dieudonné, B., & Francart, T. (2019). Redundant information is sometimes more beneficial than spatial information to understand speech in noise. *Ear and Hearing, 40*(3), 545–554. https://doi.org/10.1097/AUD.0000000000000660

Dirks, D. D., & Wilson, R. H. (1969). The effect of spatially separated sound sources on speech intelligibility. *Journal of Speech and Hearing Research, 12*(1), 5–38. https://doi.org/10.1044/jshr.1201.05

Drummond, G. B., & Vowler, S. L. (2012). Analysis of variance: Variably complex. *Advances in Physiology Education, 36*(2), 85–88. https://doi.org/10.1152/advan.00025.2012

Dunn, C. C., Tyler, R. S., Witt, S. A., & Gantz, B. J. (2006). Effects of converting bilateral cochlear implant subjects to a strategy with increased rate and number of channels. *Annals of Otology, Rhinology & Laryngology, 115*(6), 425–432. https://doi.org/10.1177/000348940611500605

Eapen, R. J., Buss, E., Adunka, M. C., Pillsbury, H. C., III, & Buchman, C. A. (2009). Hearing-in-noise benefits after bilateral simultaneous cochlear implantation continue to improve 4 years after implantation. *Otology & Neurotology, 30*(2), 153–159. https://doi.org/10.1097/MAO.0b013e3181925025

Falzone, C., Guerzoni, L., Pizzol, E., Fabrizi, E., & Cuda, D. (2022). An adaptation and validation study of the Speech, Spatial, and Qualities of Hearing Scale (SSQ) in Italian normal-hearing children. *Audiology Research, 12*(3), 297–306. https://doi.org/10.3390/audiolres12030031

Gantz, B. J., Tyler, R. S., Rubinstein, J. T., Wolaver, A., Lowder, M., Abbas, P., & Preece, J. P. (2002). Binaural cochlear implants placed during the same operation. *Otology & Neurotology, 23*(2), 169–180. https://doi.org/10.1097/00129492-200203000-00012

Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology, 43*(2), 85–99. https://doi.org/10.1080/14992020400050014

Gifford, R. H., Dorman, M. F., Sheffield, S. W., Teece, K., & Olund, A. P. (2014). Availability of binaural cues for bilateral implant recipients and bimodal listeners with and without preserved hearing in the implanted ear. *Audiology and Neurotology, 19*(1), 57–71. https://doi.org/10.1159/000355700

Govaerts, P. J., Daemers, K., Yperman, M., De Beukelaer, C., De Saegher, G., & De Ceulaer, G. (2006). Auditory Speech Sounds Evaluation (A§E®): A new test to assess detection, discrimination and identification in hearing impairment. *Cochlear Implants International, 7*(2), 92–106. https://doi.org/10.1179/146701006807508106

Grange, M. E. (2013). *Test-retest reliability in word recognition testing in subjects with varying levels of hearing loss* [Master's thesis, Brigham Young University]. ScholarsArchive. http://hdl.lib.byu.edu/1877/etd6109

Hey, M., Hocke, T., Hedderich, J., & Müller-Deile, J. (2014). Investigation of a matrix sentence test in noise: Reproducibility and discrimination function in cochlear implant patients. *International Journal of Audiology, 53*(12), 895–902. https://doi.org/10.3109/14992027.2014.938368

Hughson, W., & Westlake, H. (1944). Manual for program outline for rehabilitation of aural casualties both military and civilian. *Transactions—American Academy of Ophthalmology and Otolaryngology, 48*(Suppl.), 1–15.

Hui, S. L., & Zhou, X. H. (1998). Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research, 7*(4), 354–370. https://doi.org/10.1177/096228029800700404

Jang, J. H., Roh, J. M., Choo, O. S., Kim, Y. J., Kim, H., Park, H. Y., & Choung, Y. H. (2019). Critical factors for binaural hearing in children with bilateral sequential cochlear implantation: First implant performance and inter-implant interval. *Audiology and Neurotology, 24*(4), 174–182. https://doi.org/10.1159/000500700

Kan, A., & Litovsky, R. Y. (2015). Binaural hearing with electrical stimulation. *Hearing Research, 322,* 127–137. https://doi.org/10.1016/j.heares.2014.08.005

Kılıç, N., Kamışlı, G. İ. Ş., Gündüz, B., Bayramoğlu, İ., & Kemaloğlu, Y. K. (2021). Turkish validity and reliability study of the Speech, Spatial and Qualities of Hearing Scale. *Turkish Archives of Otorhinolaryngology, 59*(3), 172–187. https://doi.org/10.4274/tao.2021.2021-4-3

Kim, J., Lee, J., Lee, K. W., Bahng, J., Lee, J. H., Choi, C. H., & Park, J. (2015). Test-retest reliability of word recognition score using Korean standard monosyllabic word lists for adults as a function of the number of test words. *The Journal of Audiology & Otology, 19*(2), 68–73. https://doi.org/10.7874/jao.2015.19.2.68

Kokkinakis, K., & Pak, N. (2014). Binaural advantages in users of bimodal and bilateral cochlear implant devices. *The Journal of the Acoustical Society of America, 135*(1), EL47–EL53. https://doi.org/10.1121/1.4831955

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Köse, B., Karaman-Demirel, A., & Çiprut, A. (2022). Psychoacoustic abilities in pediatric cochlear implant recipients: The relation with short-term memory and working memory capacity. *International Journal of Pediatric Otorhinolaryngology, 162,* Article 111307. https://doi.org/10.1016/j.ijporl.2022.111307

Kurz, A., Zanzinger, M., Hagen, R., & Rak, K. (2021). The impact of cochlear implant microphone settings on the binaural hearing of experienced cochlear implant users with single-sided deafness. *European Archives of Oto-Rhino-Laryngology, 278*(6), 2067–2077. https://doi.org/10.1007/s00405-020-06450-5

Laske, R. D., Veraguth, D., Dillier, N., Binkert, A., Holzmann, D., & Huber, A. M. (2009). Subjective and objective results after bilateral cochlear implantation in adults. *Otology & Neurotology, 30*(3), 313–318. https://doi.org/10.1097/MAO.0b013e31819bd7e6

Laszig, R., Aschendorff, A., Stecker, M., Müller-Deile, J., Maune, S., Dillier, N., & Doering, W. (2004). Benefits of bilateral electrical stimulation with the nucleus cochlear implant in adults: 6-month postoperative results. *Otology & Neurotology, 25*(6), 958–968. https://doi.org/10.1097/00129492-200411000-00016

Litovsky, R., Parkinson, A., Arcaroli, J., & Sammeth, C. (2006). Simultaneous bilateral cochlear implantation in adults: A multicenter clinical study. *Ear and Hearing, 27*(6), 714–731. https://doi.org/10.1097/01.aud.0000246816.50820.42

Litovsky, R. Y., Goupell, M. J., Godar, S., Grieco-Calub, T., Jones, G. L., Garadat, S. N., & Misurelli, S. (2012). Studies on bilateral cochlear implants at the University of Wisconsin's Binaural Hearing and Speech Laboratory. *Journal of the American Academy of Audiology, 23*(06), 476–494. https://doi.org/10.3766/jaaa.23.6.9

Luo, X., Kolberg, C., Pulling, K. R., & Azuma, T. (2020). Psychoacoustic and demographic factors for speech recognition of older adult cochlear implant users. *Journal of Speech, Language, and Hearing Research, 63*(6), 1712–1725. https://doi.org/10.1044/2020_JSLHR-19-00225

MacKeith, N. W., & Coles, R. R. A. (1971). Binaural advantages in hearing of speech. *The Journal of Laryngology & Otology, 85*(3), 213–232. https://doi.org/10.1017/S0022215100073369

Maxon, A. B., & Hochberg, I. (1982). Development of psychoacoustic behavior: Sensitivity and discrimination. *Ear and Hearing, 3*(6), 301–308. https://doi.org/10.1097/00003446-198211000-00003

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

Mertens, G., Punte, A. K., De Bodt, M., & Van de Heyning, P. (2015). Binaural auditory outcomes in patients with postlingual profound unilateral hearing loss: 3 years after cochlear implantation. *Audiology and Neurotology, 20*(Suppl. 1), 67–72. https://doi.org/10.1159/000380751

Morera, C., Cavalle, L., Manrique, M., Huarte, A., Angel, R., Osorio, A., & Morera-Ballester, C. (2012). Contralateral hearing aid use in cochlear implanted patients: Multicenter study of bimodal benefit. *Acta Oto-Laryngologica, 132*(10), 1084–1094. https://doi.org/10.3109/00016489.2012.677546

Moulin, A., Pauzie, A., & Richard, C. (2015). Validation of a French translation of the Speech, Spatial, and Qualities of Hearing Scale (SSQ) and comparison with other language versions. *International Journal of Audiology, 54*(12), 889–898. https://doi.org/10.3109/14992027.2015.1054040

Müller, J., Schon, F., & Helms, J. (2002). Speech understanding in quiet and noise in bilateral users of the MED-EL COMBI 40/40+ cochlear implant system. *Ear and Hearing, 23*(3), 198–206. https://doi.org/10.1097/00003446-200206000-00004

Noble, W., Jensen, N. S., Naylor, G., Bhullar, N., & Akeroyd, M. A. (2013). A short form of the Speech, Spatial and Qualities of Hearing Scale suitable for clinical use: The SSQ12. *International Journal of Audiology, 52*(6), 409–412. https://doi.org/10.3109/14992027.2013.781278

Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology, 47*(Suppl. 2), 144–154. https://doi.org/10.1080/14992020802307404

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: Applications to practice* (Vol. 892). Pearson/Prentice Hall.

Pyschny, V., Landwehr, M., Hahn, M., Lang-Roth, R., Walger, M., & Meister, H. (2014). Head shadow, squelch, and summation effects with an energetic or informational masker in bilateral and bimodal CI users. *Journal of Speech, Language, and Hearing Research, 57*(5), 1942–1960. https://doi.org/10.1044/2014_JSLHR-H-13-0144

Rawool, V. W. (2016). *Auditory processing deficits: Assessment and intervention*. Thieme.

Rawool, V. W., & Parrill, M. (2018). Objective evaluation of binaural summation through acoustic reflex measures. *International Journal of Audiology, 57*(7), 493–501. https://doi.org/10.1080/14992027.2018.1435914

Schleich, P., Nopp, P., & D'haese, P. (2004). Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBI 40/40+ cochlear implant. *Ear and Hearing, 25*(3), 197–204. https://doi.org/10.1097/01.AUD.0000130792.43315.97

Schön, F., Müller, J., & Helms, J. (2002). Speech reception thresholds obtained in a symmetrical four-loudspeaker arrangement from bilateral users of MED-EL cochlear implants. *Otology & Neurotology, 23*(5), 710–714. https://doi.org/10.1097/00129492-200209000-00018

Sheffield, S. W., Haynes, D. S., Wanna, G. B., Labadie, R. F., & Gifford, R. H. (2015). Availability of binaural cues for pediatric bilateral cochlear implant recipients. *Journal of the American Academy of Audiology, 26*(03), 289–298. https://doi.org/10.3766/jaaa.26.3.8

Snik, A., Agterberg, M., & Bosman, A. (2015). How to quantify binaural hearing in patients with unilateral hearing using hearing implants. *Audiology and Neurotology, 20*(Suppl. 1), 44–47. https://doi.org/10.1159/000380747

Spyridakou, C., Rosen, S., Dritsakis, G., & Bamiou, D. E. (2020). Adult normative data for the Speech in Babble (SiB) test. *International Journal of Audiology, 59*(1), 33–38. https://doi.org/10.1080/14992027.2019.1638526

Thakkar, T., Kan, A., Jones, H. G., & Litovsky, R. Y. (2018). Mixed stimulation rates to improve sensitivity of interaural timing differences in bilateral cochlear implant listeners. *The Journal of the Acoustical Society of America, 143*(3), 1428–1440. https://doi.org/10.1121/1.5026618

Tyler, R. S., Dunn, C. C., Witt, S. A., & Preece, J. P. (2003). Update on bilateral cochlear implantation. *Current Opinion in Otolaryngology & Head and Neck Surgery, 11*(5), 388–393. https://doi.org/10.1097/00020840-200310000-00014

Tyler, R. S., Gantz, B. J., Rubinstein, J. T., Wilson, B. S., Parkinson, A. J., Wolaver, A., Preece, J. P., Witt, S., & Lowder, M. W. (2002). Three-month results with bilateral cochlear implants. *Ear and Hearing, 23*(1), 80S–89S. https://doi.org/10.1097/00003446-200202001-00010

Tyler, R. S., Noble, W., Dunn, C., & Witt, S. (2006). Some benefits and limitations of binaural cochlear implants and our ability to measure them. *International Journal of Audiology, 45*(Suppl. 1), 113–119. https://doi.org/10.1080/14992020600783095

Uslar, V., Ruigendijk, E., Hamann, C., Brand, T., & Kollmeier, B. (2011). How does linguistic complexity influence intelligibility in a German audiometric sentence intelligibility test? *International Journal of Audiology, 50*(9), 621–631. https://doi.org/10.3109/14992027.2011.582166

Uslar, V. N., Carroll, R., Hanke, M., Hamann, C., Ruigendijk, E., Brand, T., & Kollmeier, B. (2013). Development and evaluation of a linguistically and audiologically controlled sentence intelligibility test. *The Journal of the Acoustical Society of America, 134*(4), 3039–3056. https://doi.org/10.1121/1.4818760

Van Deun, L., Van Wieringen, A., & Wouters, J. (2010). Spatial speech perception benefits in young children with normal hearing and cochlear implants. *Ear and Hearing, 31*(5), 702–713. https://doi.org/10.1097/AUD.0b013e3181e40dfe

Van Hoesel, R., Ramsden, R., & O'Driscoll, M. (2002). Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user. *Ear and Hearing, 23*(2), 137–149. https://doi.org/10.1097/00003446-200204000-00006

Van Hoesel, R. J. (2012). Contrasting benefits from contralateral implants and hearing aids in cochlear implant users. *Hearing Research, 288*(1–2), 100–113. https://doi.org/10.1016/j.heares.2011.11.014

Van Hoesel, R. J., Jones, G. L., & Litovsky, R. Y. (2009). Interaural time-delay sensitivity in bilateral cochlear implant users: Effects of pulse rate, modulation rate, and place of stimulation. *Journal of the Association for Research in Otolaryngology, 10*(4), 557–567. https://doi.org/10.1007/s10162-009-0175-x

Van Hoesel, R. J., & Litovsky, R. Y. (2011). Statistical bias in the assessment of binaural benefit relative to the better ear. *The Journal of the Acoustical Society of America, 130*(6), 4082–4088. https://doi.org/10.1121/1.3652851

Verhaert, N., Lazard, D. S., Gnansia, D., Bébéar, J. P., Romanet, P., Meyer, B., & Truy, E. (2012). Speech performance and sound localization abilities in Neurelec Digisonic SP binaural cochlear implant users. *Audiology and Neurotology, 17*(4), 256–266. https://doi.org/10.1159/000338472

Vermeire, K., & Van de Heyning, P. (2009). Binaural hearing after cochlear implantation in subjects with unilateral sensorineural deafness and tinnitus. *Audiology and Neurotology, 14*(3), 163–171. https://doi.org/10.1159/000171478

Williges, B., Dietz, M., Hohmann, V., & Jürgens, T. (2015). Spatial release from masking in simulated cochlear implant users with and without access to low-frequency acoustic hearing. Trends in Hearing, *19*. https://doi.org/10.1177/2331216515616940

Wouters, J., Damman, W., & Bosman, A. J. (1994). Vlaamse opname van woordenlijsten voor spraakaudiometrie [Flemish recording of word lists for speech audiometry]. *Logopedie: Informatiemedium van de Vlaamse Vereniging Voor Logopedisten, 7*(6), 28–34.

Yoon, Y. S., Shin, Y. R., & Fu, Q. J. (2013). Binaural benefit with and without a bilateral spectral mismatch in acoustic simulations of cochlear implant processing. *Ear and Hearing, 34*(3), 273–279. https://doi.org/10.1097/AUD.0b013e31826709e8