Original Article

The Important Role of Contextual Information in Speech Perception in Cochlear Implant Users and Its Consequences in Speech Tests

J. Gertjan Dingemanse[1] (iD) and André Goedegebure[1]

Abstract

This study investigated the role of contextual information in speech intelligibility, the influence of verbal working memory on the use of contextual information, and the suitability of an ecologically valid sentence test containing contextual information, compared with a CNC (Consonant-Nucleus-Consonant) word test, in cochlear implant (CI) users. Speech intelligibility performance was assessed in 50 postlingual adult CI users on sentence lists and on CNC word lists. Results were compared with a normal-hearing (NH) group. The influence of contextual information was calculated from three different context models. Working memory capacity was measured with a Reading Span Test. CI recipients made significantly more use of contextual information in recognition of CNC words and sentences than NH listeners. Their use of contextual information in sentences was related to verbal working memory capacity but not to age, indicating that the ability to use context is dependent on cognitive abilities, regardless of age. The presence of context in sentences enhanced the sensitivity to differences in sensory bottom-up information but also increased the risk of a ceiling effect. A sentence test appeared to be suitable in CI users if word scoring is used and noise is added for the best performers.

Introduction

Cochlear implants (CIs) are currently the treatment of choice for bilateral severe to profound postlingual sensorineural hearing loss, with significant improvements reported in speech intelligibility and quality of life (Gaylor et al., 2013; McRackan et al., 2018). The effect of a CI on speech intelligibility is usually measured with standardized speech tests. However, much variation in used speech materials and scoring methods exists between studies, as reported in Table II of the study of McRackan et al. (2018). Most studies used lists of CNC (Consonant-Nucleus-Consonant) words with a score of either percent correct words or percent correct phonemes. Besides CNC words, several studies reported the use of sentence tests. One of the most important differences compared with word tests is the possibility of using context, because the words in the sentences are related to each other. Although not all words of a sentence may be

perceived correctly, a listener may reconstruct the correct sentence based on a few perceived words. The amount of available contextual information in the sentences of a test has a substantial effect on the score that will be obtained. More context will lead to a better predictability of missing parts and hence to a higher speech score (Boothroyd & Nittrouer, 1988), although the resulting score may depend on the ability of the listener to make use of this contextual information (Grant & Seitz, 2000). However, in the literature, it is reported that sentence tests may be too difficult for use in CI listeners (van

[1]Department of Otorhinolaryngology and Head and Neck Surgery, Erasmus University Medical Center, Rotterdam, the Netherlands

Corresponding author:
J. Gertjan Dingemanse, Department of Otorhinolaryngology and Head and Neck Surgery, Erasmus University Medical Center, Room NT-310, P.O. Box 2040, 3000 CA, Rotterdam, the Netherlands.
Email: g.dingemanse@erasmusmc.nl

Wieringen & Wouters, 2008) or that listening to sentences may require much listening effort (Theelen-van den Hoek, Houben, & Dreschler, 2014). This is not in accordance with the finding of Winn (2016) that understanding of high-context sentences in CI users required less effort than understanding of low-context sentences. Given these observations, it is important to consider whether clinically available sentence tests may be a better choice for evaluating the effect of CI treatment compared with CNC word tests. Especially the effect of contextual information in the sentences needs to be considered.

Several studies that focused on sentence tests for CI users mainly reported on test properties, like the risk of floor or ceiling effects and good reproducibility (test–retest reliability). A floor effect exists if a relatively large proportion of a group of listeners obtains a score on or very nearby the minimum score of a test (in case of a speech test, this is usually 0% intelligibility). A ceiling effect exists if a relatively large proportion of a group of listeners obtains the maximum score of a test. For example, Gifford, Shallop, and Peterson (2008) reported that with the Hearing in Noise Test (HINT) sentence test 28% of 156 adult CI users achieved the maximum score and 71% reached a score above 85% sentence intelligibility in quiet. This makes the HINT not responsive to differences in stimulation strategies or different signal processing options for high-performing CI users. The HINT sentences were selected from the Bamford-Kowal-Bench (BKB) sentences (Bench, Kowal, & Bamford, 1979). These sentences have an easy structure and consist of relatively easy, frequently used words. Words that are unintelligible in the first instance are identified easily, because they are highly predictable. According to Boothroyd and Nittrouer (1988), sentences with high predictability result in higher scores than sentences with low predictability and are therefore more prone to ceiling effects. Ebrahimi-Madiseh, Eikelboom, Jayakody, and Atlas (2016) showed that a ceiling effect also exists in the City University of New York sentence test (Boothroyd, Hanin, & Hnath, 1985) if used in CI recipients. Gifford et al. (2008) recommended the use of the Arizona Biomedical Institute sentence test (Spahr et al., 2012), because this test contains more difficult, less predictable sentences, spoken by different talkers in a casual style. Only 0.7% of the CI users reached the maximum score. The Minimum Speech Test Battery for adult CI users (Luxford, Ad Hoc Subcommittee of the Committee on Hearing, & Equilibrium of the American Academy of Otolaryngology-Head and Neck Surgery, 2001; Minimum Speech Test Battery, 2011) recommends assessment of performance with both CNC word and sentence materials, to increase the probability that a patient's performance will be within the range of at least one test, not confounded by either ceiling or floor effects.

Several studies reported on the reproducibility of sentence tests by describing the test–retest variability (e.g., Firszt et al., 2004; Spahr et al., 2012). The test–retest variability is, among other factors, related to the effective number of statistically independent elements in the speech, which depends on the amount of contextual information within the sentence (Boothroyd et al., 1985; Boothroyd & Nittrouer, 1988; Spahr et al., 2012; Versfeld, Daalder, Festen, & Houtgast, 2000).

Until recently, relatively little attention has been paid in the literature to the ecological validity of a speech test. Ecological validity means that the speech used must be characteristic of everyday speech in different aspects, for example, speaking rate and clarity, sentence structure, and topics. An important aspect of ecological validity is that the speech contains contextual information, as in real speech. The performance on an ecologically valid speech test may better reflect the perceived difficulties with speech intelligibility in real life. A test with sentences could arguably serve as more representative of everyday conversation than a word test. The Arizona Biomedical Institute sentences have relatively good ecological validity (Spahr et al., 2012). Another test that is designed to be more ecologically valid is PRESTO (Perceptually Robust English Sentence Test Open-set), which incorporates variability in words, sentences, talkers, and regional dialects (Gilbert, Tamati, & Pisoni, 2013). In the Netherlands, the VU sentences (Versfeld et al., 2000) have good ecological validity, because they are taken from newspapers, have variation in sentence structure and topics, and are spoken with a normal speaking style and rate.

However, when testing CI recipients, ecological validity is often secondary to the ease of the test material or properties that are thought to better suit the capabilities of CI users. For example, the Dutch Leuven Intelligibility Sentence Test (LIST) (van Wieringen & Wouters, 2008) uses a relatively low speaking rate of 2.5 syllables/s and clear speech, to make the test easier for CI recipients. Theelen-van den Hoek et al. (2014) investigated if it was possible to reliably measure the speech-reception threshold in noise (SRTn) in CI listeners with the Dutch matrix test. A matrix test generates sentences with a length of five words from a matrix that contains 10 alternatives for each word position. This results in meaningful semantically unpredictable sentences with a fixed grammatical structure. These sentences contain little contextual information and are not very representative for everyday speech. The BKB Speech-in-Noise (BKB-SIN) Test is often used to test CI users because of its easy sentences (Bench et al., 1979). In all these examples, the specific material or test characteristics lead to a reduced ecological validity of the test.

CI recipients have more difficulties with speech perception, because their CI delivers a degraded signal.

The quality of the speech signal is reduced due to limited spectral resolution (Friesen, Shannon, Baskent, & Wang, 2001; Henry & Turner, 2003; Winn, Chatterjee, & Idsardi, 2012) and temporal fine-structure cues (Loizou, 2006; Rubinstein, 2004). In other words, the bottom-up information is limited. Consequently, CI users have to rely more on top-down processing based on linguistic context (Kong, Donaldson, & Somarowthu, 2015; Nittrouer et al., 2014; Oh, Donaldson, & Kong, 2016; Winn et al., 2012).

Therefore, it is reasonable to assume that in CI recipients, speech intelligibility depends also on nonauditory factors like linguistic skills and cognitive abilities. Some studies investigated the relationship between speech intelligibility and linguistic skills or cognitive abilities in adult CI users. Heydebrand, Hale, Potts, Gotter, and Skinner (2007) found that better intelligibility of CNC words 6 months after cochlear implantation was associated with better verbal learning scores and verbal working memory (letter span) but not with general cognitive ability. Holden et al. (2013) reported a significant positive correlation between a composite measure of cognition (including a vocabulary test, a forward and backward digit span tests, and a verbal learning test) with CNC word recognition scores. In contrast, Moberly, Harris, Boyce, and Nittrouer (2017) found no significant correlation between sentence intelligibility in noise (percentage of words correct) and verbal working memory accuracy scores for serial recall of spoken nonrhyming words. Given these inconclusive findings, in the current study, we explored the relation of working memory capacity with sentence intelligibility and word intelligibility within the same group.

Some studies have investigated the use of contextual information in CI users. Amichetti, Atagi, Kong, and Wingfield (2018) reported that CI users made effective use of linguistic context. Older CI users were able to use context to compensate for their initial disadvantage in recognizing words in low-context conditions compared with young CI users but were also more hindered by interference from other words that might also be activated by context. Winn (2016) showed that listening effort as measured by the pupillary response is higher in CI users than in NH listeners, but the listening effort is less for high-context sentences than for low-context sentences. Results from Baskent et al. (2016) suggest that top-down restoration of interrupted speech can only be achieved in a more limited manner in CI listeners compared with NH listeners. Uncertainty still exists about whether CI users make more or less use of contextual information compared with NH listeners.

In summary, contextual information in a speech test is an important factor because of its influence on test scores, reliability, the relation with ecological validity, and the relation with cognitive and linguistic abilities.

In this study, we investigated these aspects of contextual information in an ecological sentence test and a CNC words test in CI users. The purpose was to answer the following questions:

1. What is the effect of contextual information from the speech materials on speech intelligibility in CI users?
2. Are sentence intelligibility and the use of contextual information related to verbal working memory in CI users?
3. To what extent is an ecologically valid sentence test suitable in CI users with respect to a possible ceiling effect, the responsiveness to differences in the CI signal and the reproducibility of the test compared with CNC wordlists?

Materials and Methods

Participants

Fifty adult CI recipients were included in this study, with a mean age of 63 years (*SD*: 14.4; range: 29–89 years), 18 women and 32 men. All participants had postlingual onset of hearing loss and were Dutch native speakers. They were unilateral CI users for at least 1 year with severe hearing loss in the other ear and they did not use a contralateral hearing aid during the test session. Only CI users that had a phoneme score with the CI of at least 60% on clinically used Dutch CNC word lists (Bosman & Smoorenburg, 1995) were included, because participants must have sufficient intelligibility to perform an adaptive speech in noise at a 50% correct level (see later).

Twenty-seven participants had an Advanced Bionics implant with at least 14 active electrode contacts and a Naida Q70 sound processor with all sound enhancement algorithms switched off. Twenty-three participants had a Cochlear Ltd implant with at least 21 active electrode contacts and a Nucleus 5 sound processor with Autosensitivity and ADRO active, as in their daily life program. Volume adjustments were not allowed during the test session.

For the speech-in-noise test, the reference data for normal hearing (NH) was based on 16 subjects, with a mean age of 22 years (*SD* = 3.0; range 20–29 years), 8 women and 8 men. For the reference data (NH) of the CNC word lists, we used the data of Bronkhorst, Bosman, and Smoorenburg (1993) who used the same CNC word material in a group of 20 normal-hearing university students.

Participants signed a written informed consent form, and the Erasmus Medical Center Ethics Committee approved the study protocols of the original studies whose data were taken (as described in the Design and Procedures section).

Speech Intelligibility Tests

Speech intelligibility was measured with Dutch female-spoken, unrelated sentences (Versfeld et al., 2000). These sentences were representative for daily-used communication and mainly selected form a newspaper database. The sentences were pronounced in a natural, clear manner with normal vocal effort and speaking rate. For the estimate of the amount of context, we needed sentences with a fixed number of words (see Context Parameters section). Therefore, we selected sentences with a length of six words and grouped them into lists of 26 sentences. The presentation level of the sentences was fixed at 70 dB (SPL). This speech level is often reached in noisy situations (Pearsons, Bennett, & Fidell, 1977). Participants were instructed to repeat as many words as possible of each sentence and to guess when unsure about any word.

The proportion of correct recognized words in quiet (PCq) was measured at an SNR of 40 dB (i.e., a noise level of 30 dB). This is equivalent to the speech score in quiet, but it has the advantage that it is a distinct point on the psychometrical curve, instead of being the asymptotic value. The SRTn at 50% word intelligibility was measured in steady-state noise with a speech spectrum that corresponds to the long-term spectrum of the sentences. The noise level was varied following an adaptive procedure based on a stochastic approximation method with step size 4 (PC(t − 1) – 0.5), and PC(t − 1) being the proportion correct score of the previous trial. The average of trials in a stochastic approximation staircase with constant step size converges to the target of 50% (Kushner & Yin, 2003). The average proportion correct score was calculated over the last 22 presentation levels. The SRTn was defined as the average SNR over the last 22 presentation levels and the presentation level of the next trial that was calculated from the last response. The starting point was the SRTn of the practice run.

Phoneme perception in quiet was measured with the clinically used Dutch word lists for speech audiometry of the Dutch Society of Audiology (Bosman & Smoorenburg, 1995), which consist of 11 phonetically balanced CNC words. Data were obtained from a participant's clinical record if it was measured within 6 months before the visit or measured just before the experiment otherwise. The phoneme perception score was measured at 65 and 75 dB (SPL). These scores were averaged to reduce measurement variability and to obtain an estimate of the score at 70 dB (SPL).

For the reference data of the speech-in-noise test in the NH group, the SRTn was measured along with the proportion of correct words at four SNRs around the individual SRTn.

Context Parameters

There are several approaches to quantify the use of context information in speech perception. In this study, we used the approaches of Boothroyd and Nittrouer (1988) and Bronkhorst et al. (1993). Boothroyd and Nittrouer (1988) described two equations to quantify the role of context. The first equation describes the relationship between the recognition probability $p_{e,c}$ of speech elements (e.g., words) presented in context (e.g., sentences) and the recognition probability of wholes p_{wh} (i.e., understanding whole sentences completely correctly). This relation is given by:

$$p_{wh} = p_{e,c}^{j} \tag{1}$$

where j is a parameter to quantify the amount of contextual information, giving the "effective" number of statistically independent elements in the whole. If no context information is available, j is equal to n, the number of elements. The j factor is strongly associated with the ability to fill in the last missing element from contextual information (Bronkhorst, Brand, & Wagener, 2002).

The second equation describes the relationship between the recognition probability $p_{e,c}$ of speech elements presented in context and the recognition probability $p_{e,nc}$ of speech elements presented without context (nc = no context), for example, words in sentences versus words in isolation.

$$p_{e,c} = 1 - \left(1 - p_{e,nc}\right)^{k} \quad \text{or} \quad k = \frac{\log\left(1 - p_{e,c}\right)}{\log\left(1 - p_{e,nc}\right)} \tag{2}$$

The parameter k represents the magnitude of the context effect. Due to the context information, the probability to make an recognition error $(1 - p_e)$ is reduced. The k factor expresses the context effect in terms of the proportional increase of channels of information that would be required to produce the same change of proportion correct (PC) recognition in the absence of context. A k factor >1 means that context information is used to recognize speech elements. If p_e approaches 1, k reduces to 1. The parameter k is a good overall measure of context effects.

We calculated a j factor for the CNC words, for sentences in quiet, and for sentences in noise following Equation 1. A k factor for sentences was calculated according to Equation 2. The PC CNC words was used to estimate the $p_{e,nc}$ values, as explained in more detail later. For six individuals having a value of one on any of the proportions correct in Equations 1 or 2, the factor was not calculated because it reached its asymptotic and, thus, did not accurately reflect the use of context.

Bronkhorst et al. (1993) developed a more extensive model for context effects in speech recognition. Their model gives predictions of the probabilities $p_{wh,m}$ that m ($m = 0, \ldots, n$) elements of wholes containing n elements are recognized. These probabilities $p_{wh,m}$ are a function of the recognition probabilities of the elements

if presented in isolation (no context) and a set of context parameters c_i $(i = 0, \ldots, n)$.

$$p_{\text{wh},n} = f(c_i, \ p_{\text{e,nc}}), \quad 0 \leqslant c_i \leqslant 1 \qquad (3)$$

The context parameters c_i give the probabilities of correctly guessing a missing element given that i of the n elements are missing. They quantify the amount of contextual information used by the listener. The maximum value of 1 means that a missing element is available from context information without uncertainty. If the whole contains no context information, the value of c_i is zero. It should be noted that the c_i values quantify the added effects of all contextual cues from a priori knowledge, coarticulation, word frequency, syntactic constraints, semantic congruency, with the ability to use these cues included. Actually, the model is a set of equations that result in an array of probabilities $p_{\text{wh},m}$ with length n for each value of $p_{\text{e,nc}}$. For details of the model, we refer to (Bronkhorst et al., 1993, 2002). From the array $p_{\text{wh},m}$, we can calculate the average element recognition probability for elements in context:

$$p_{\text{e,c}} = p_{\text{wh},n} + \frac{(n-1)}{n} p_{\text{wh},n-1} + \frac{(n-2)}{n} p_{\text{wh},n-2} + \cdots + \frac{1}{n} p_{\text{wh},1}$$
$$(4)$$

The model prediction of the j factor can be calculated from $p_{\text{e,c}}$ and $p_{\text{wh},n}$, and the prediction of parameter k from $p_{\text{e,c}}$ and $p_{\text{e,nc}}$. A short description of the different context parameters is given in Table 1.

The context model of Bronkhorst et al. was fitted to the data of this study, resulting in a set of context parameters c_i that give the amount of context use at a group level (CI users or NH participants). The fitting process consisted of five steps: (a) Set estimates of the parameters c_i $(i = 1 \ldots n)$. (b) Sampling of the model with values of $p_{\text{e,nc}}$ between 0 and 1 in steps of 0.005, resulting in a calculated $p_{\text{e,c}}$ for each $p_{\text{e,nc}}$ from Equations 3 and 4. (c) Determination of the $p_{\text{e,nc}}$ values that correspond to the measured phoneme scores $p_{\text{e,c}}$ based on linear interpolation. (d) Calculation of $p_{\text{w,m}}$ for these $p_{\text{e,nc}}$ values. (e) Calculation of the root mean square difference between measured and predicted $p_{\text{w,m}}$. The optimal set c_i was obtained by minimizing the root mean square difference using MATLAB routine fminsearch, an unconstrained nonlinear optimization procedure. Confidence intervals (95%) of c_i were obtained by bootstrapping. The parameter c_0 was set to zero, because participants were not forced to make a guess if they did not understand any of the phonemes or words.

To model the relation between scores for the CNC speech material and the VU sentence material, we regarded the CNC word scores as proportions correct of isolated words (without context) that could be used

Table 1. Definition of Three Different Context Measures.

j factor $(1 \leqslant j \leqslant n)$	Measure of the use of context, expressed as a number of "effective" independent elements from n elements. The lower j the more use of context. For a sentence of six words, $j = 6$ in case of no context use (all words of the sentence must be recognized independently to correctly repeat the sentence) and $j = 1$ in case of maximal context use (recognizing only one word is enough to repeat the complete sentence).
k factor $(k \geqslant 1)$	Measure of the effect of context, expressed as the increase of proportion correct recognized elements due to context, compared with recognition of isolated speech elements. The k factor can be interpreted as the proportional increase of channels of information that would be required to produce the same change of proportion correct recognition in the absence of context.
c_i $(i = 1 \ldots n)$ $(0 \leqslant c_i \leqslant 1)$	Context parameters that give the probability of correctly guessing a missing element given that i of the n elements were missed. The higher these probabilities, the more use of context. For example, c_1 is the probability to guess the missing word if only one word of a sentence is missing.

as input in the context model of the sentences. However, the words in the sentences have different lengths, varying from 2 phonemes to 10 phonemes (mean: 4.4), while CNC words have 3 phonemes. Therefore, we designed a transform of the CNC word scores to scores for words of five phonemes (as the first integer value above the mean phoneme length of 4.4). This transform is a simplification, because in fact the transform should be the weighted sum of the transforms for each number of phonemes. However that would result in too many parameters. Because we only fit the relation between the score of isolated phonemes and *average* word score from the sentence test, a transfer function with five parameters appeared to be sufficient to achieve an acceptable fit. We hypothesized that participants make more use of contextual information for words that consist of more than three phonemes, because if they initially understood more than half the elements of the word, the number of words that fits with the already perceived elements is often very limited. On the other hand, if they perceived only one or two phonemes of a long word, the chance to guess the whole word correctly is low, because there are still many alternatives. In the context model, this means that c_1 and c_2 are relatively high, but the c_i for

$i \geqslant 3$ are relatively low. We modeled a transform of the PC words $p_{\mathrm{wh},3}$ from the CNC context model to word scores $p_{\mathrm{wh},5}$ of a five-elements (phonemes) contextual model, with c values to be fitted to the data. The measured $p_{\mathrm{wh},3}$ was converted to $p_{\mathrm{e,nc}}$, using the known relationship of $p_{\mathrm{wh},3}$ and $p_{\mathrm{e,nc}}$ from the CNC context model. Next the $p_{\mathrm{e,nc}}$ was converted to $p_{\mathrm{wh},5}$, using a 5-phonemes context model. The obtained $p_{\mathrm{wh},5}$ values were used as input in the context model of sentences as the PC scores for words without sentence context. Figure 1 demonstrates the steps of the transform: PC CNC words of 0.7 are transformed to PC words in sentences of 0.97 by following the arrows. In the left panel, the value is transformed to the value for 5-phoneme words (0.79). The 5-phoneme words are isolated words that serve as input in the context model of sentence intelligibility (Equation 3). The use of context leads to a PC of 0.97. The output variables of the context model of sentences are the word and sentence scores. We fitted the 5-phonemes model by minimizing the summed squares error of the calculated word and sentence scores and the measured scores. The CNC words and the VU sentences were both spoken by a female talker with a clear articulation. Therefore, talker differences were expected to be small.

Responsiveness and Reproducibility

We defined the responsiveness to bottom-up differences as the change in a speech score in reaction on a change in the PC of isolated phonemes ($\Delta p_{\mathrm{isol\_ph}}$). We regarded the last as an adequate measure of sensory bottom-up information in accordance with Boothroyd and Nittrouer (1988). It was not possible to measure these proportions correct, because no recordings of isolated phonemes were available. However, measured values were not needed, because the context model provided us with the relations between the PC-isolated phonemes and the other speech measures $p_{\mathrm{e,c}}$ and $p_{\mathrm{wh,n}}$ for both CNC words and sentences. The responsiveness to changes in the bottom-up information was defined as

$$\Delta p_{\mathrm{e,c}} / \Delta p_{\mathrm{isol\_ph}} \quad \text{and} \quad \Delta p_{\mathrm{wh,n}} / \Delta p_{\mathrm{isol\_ph}}$$

For example, in Figure 1, the slope of the curve for CNC words (left panel) is almost one. This slope is the responsiveness for CNC words. For sentences, the transform of Figure 1 was used to obtain the responsiveness.

We also defined a measure of reproducibility with the influence of context included. As already described by Thornton and Raffin (1978), each score from trials having two response options ("true" or "false") can be modeled according to a binomial distribution. In a sentence test with word scoring, the recognition of each word can be true or false. However, in a sentence, the recognition of each word is not independent from the recognition of the other words. According to Equation 1, there are only j independent elements. From the binomial distribution, the standard deviation (SD) is given with the equation:

$$SD(p) = \sqrt{\frac{p(1-p)}{jT}} \tag{5}$$

with T the number of trials. The total number of the "effective" independent elements (as if context effects were removed) in a test is $j \cdot T$.
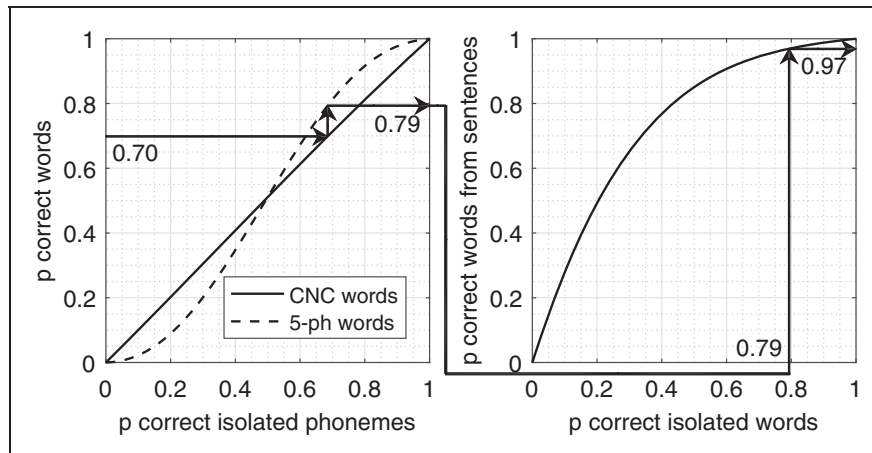

Figure 1. Illustration of the transform of a CNC word score to a word from sentence score, using the context model of CNC words (solid line in left panel), a context model of words with five phonemes (dashed line in left panel), and the context model of sentences (solid line in right panel).
CNC = consonant-nucleus-consonant.

We calculated also responsiveness-reliability ratios. Use of context may enhance the responsiveness, but may also enlarge the *SD*, because j is lower for more use of context. The ratio of responsiveness and reliability is a measure of the sensitivity of a speech test to reliably measure a change between different test conditions that differ in the amount of sensory bottom-up information.

Reading Span Task

We used a computerized Dutch version of the Reading Span Task as a measure of verbal working memory capacity (van den Noort, Bosch, Haverkort, & Hugdahl, 2008). Participants had to read sentences aloud, which appeared on a computer screen for 6.5 s, and to remember the final word of each sentence. After reading the sentence, they had to press the space bar to go to the next item. If participants could not finish the sentence within this time, the next sentence was shown automatically. Sentences were presented in different set sizes of 2, 3, 4, 5, or 6 sentences in random order. After a set, the word "recall" appeared, and the participants had to recall the final word of each sentence in the set (in free order). The reading span (Rspan) score was the average of the number of correctly recalled words for three sets of 20 sentences, giving a Rspan score range from 0 to 20.

Design and Procedures

The speech intelligibility and reading span data were available from three recent studies of our Department of Otorhinolaryngology: data of Vroegop, Dingemanse, van der Schroeff, Metselaar, and Goedegebure (2017) and data of Dingemanse and Goedegebure (2018) and Dingemanse, Vroegop, and Goedegebure (2018). From Dingemanse and Goedegebure (2018), we included only 11 participants, because the other participants were already included from Dingemanse et al. (2018). In all studies, each participant was tested in one test session following partly the same protocol. First, a practice run of the sentence-in-noise test was done to make the participants familiar with the voice and the task and to obtain a first estimation of a participant's SRTn. Second, sentence tests in quiet and in noise were performed. Next, tests were performed that were specific of the aforementioned studies where the data are taken from. At the end of the test session, a Reading Span Task was performed to obtain a measure of the verbal working memory span.

Equipment

All testing was performed in a sound-treated room. Participants sat 1 m in front of a Westra Lab 251 loudspeaker that was connected to an external soundcard (MOTU UltraLite mk3 Hybrid and after failure of the MOTU card a Roland Octa-capture UA-1010, calibration was checked) and a computer. The tests were presented in a custom application (cf. Dingemanse & Goedegebure, 2015) running in MATLAB.

Data Analysis

Speech performance scores were transformed to rationalized arcsine unit scores in order to make them suitable for statistical analysis according to Studebaker (1985), but not for use in the context models. In cases of multiple comparisons, we used the Benjamini–Hochberg method to control the false discovery rate at level 0.05 (Benjamini & Hochberg, 1995). Data analysis was performed with MATLAB (MathWorks, v9.0.0).

Results

Table 2 shows the descriptive values of PC for different scoring methods and speech materials, SRTn, calculated j factors, k factor, and Rspan values for the CI group. Lower SRTn indicated better performance. Lower j factors indicate more use of contextual information. As expected, the proportion of completely correctly understood sentences is less than the proportion of correctly recognized words from the sentences. Also, the proportion of correctly recognized CNC words is less than the proportion of correctly recognized phonemes. The j factor for sentences in noise is 2.2, indicating that understanding of a whole sentence of 6 elements is equivalent to recognition of 2.2 statistically independent elements. In quiet, the j factor is 3.9, demonstrating that less contextual information is used at higher PC scores. For the NH group, the mean SRTn value (using word scores) was −5.5 dB with a *SD* of 0.6 dB.

Use of Context

Figure 2 shows the results for each of the three context parameters $c(i)$, j, and k derived from the CNC scores by fitting the context model of Bronkhorst et al. (1993) to the data. The left panel shows the c_i values obtained from the CNC scores in CI users compared with c_i values for normal-hearing subjects (obtained from Bronkhorst et al., 1993). The context parameters for the CI users were significantly higher than the context parameters for the normal-hearing listeners, even for the listening condition with added noise (NHn). For example, the CI users had a 70% chance of correctly guessing the missing phoneme ($i=1$) if they had recognized already two phonemes in quiet, whereas the NH subjects had a chance of only 45% in noise.

The center panel of Figure 2 shows the calculated j factor (note that the y axis is reversed) as a function of

Table 2. Descriptive Values of Mean (*M*), Standard Deviation (*SD*), and Range of Proportion Correct (PC) in Quiet (q) and Noise (n) Using Phoneme Scoring (ph), Word Scoring (wrd) or Sentence Scoring (sen), Speech Reception Thresholds in Noise (SRTn) for Different Scoring Methods, Context Factors (*j* and *k*), and Reading Span (Rspan) Scores for the Group of CI Recipients.

	Speech type	Scoring	Noise	*M*	*SD*	Range
PCq_ph_CNC (rau)	CNC	Phonemes	Quiet	0.82	0.15	0.57–1.21
PCq_wrd_CNC (rau)	CNC	Words	Quiet	0.42	0.074	0.23–0.50
j_CNC	CNC		Quiet	2.1	0.49	1.0–3.2
PCq_wrd (rau)	Sentences	Words	Quiet	0.97	0.18	0.61–1.21
PCq_sen (rau)	Sentences	Sentences	Quiet	0.79	0.25	0.15–1.19
PCn_wrd (rau)	Sentences	Words	Noise	0.51	0.030	0.43–0.58
PCn_sen (rau)	Sentences	Sentences	Noise	0.26	0.078	0.077– 0.42
SRTn (dB)	Sentences	Words	Noise	5.8	4.8	−1.1–19.5
SRTn_sen (dB)	Sentences	Sentences	Noise	6.8	4.8	−0.1–20.7
j_q	Sentences		Quiet	3.9	1.6	1.5–6.4
j_n	Sentences		Noise	2.2	0.59	1.1–3.9
k_q	Sentences		Quiet	2.5	1.1	1.0–5.0
Rspan				9.5	2.8	4.0–18.0

CNC = consonant-nucleus-consonant; SRTn = speech-reception threshold in noise; PCq = proportion of correct recognized words in quiet; PCn: proportion of correct recognized words in noise; CI = cochlear implant.
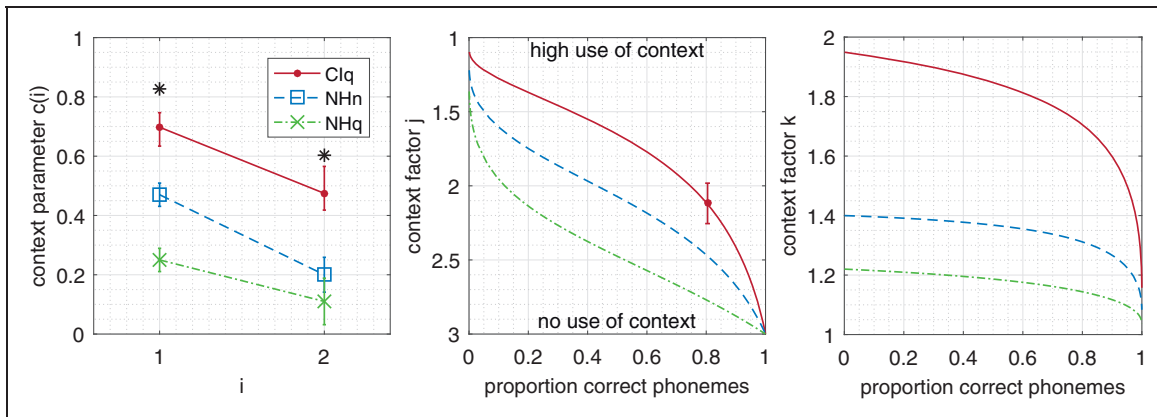

Figure 2. Left panel: Context parameters c_i that gave the best fit of the context model to the data of the CNC word intelligibility in quiet in CI users (CIq), plotted as a function of index *i*. Higher c_i values indicate more use of context. Also plotted were the parameters c_i obtained in a normal-hearing group for words in noise (NHn) and in quiet (NHq) taken from Bronkhorst et al. (1993). The index *i* represents the number of missing phonemes, and c_i is the probability that one of the missing phonemes is guessed correctly based on contextual information. Error bars give 95% confidence intervals. Significant differences between the CI group and the NHn group are denoted with an *. Center and right panel: The predicted *j* factor and *k* factor from the model as a function of the proportion correct elements (phonemes). The dot in the center panel is a data point (j_CNC) from Table 2. Note that the *y* axis of the center panel is inverted. Lower *j* values and higher *k* values indicate more use of context.

PC phonemes. The average *j* factor from the data (j_CNC from Table 2) is also plotted. The factor *j* is smaller in CI users than in NH users, again indicating more use of context in the CI users group. The *j* factor increases (meaning less use of context) for increasing PC phonemes, as expected. However, the *j* factor remains low (<2) even for a PC phonemes up to 0.8, indicating that CI users rely more on context cues even for more easy listening conditions. The right panel of Figure 2

shows the calculated *k* factor based on the context model. The *k* factor shows the same observation that CI users make more use of context than NH listeners.

The context model was also fitted to the sentence intelligibility data, following the same approach as in the fitting of the CNC words. Both the data of sentences in quiet and in noise were used, because we found that the speech intelligibility in quiet (PCq_wrd) and in noise (SRTn) were highly correlated ($\rho = 0.87$, $p < 10^{-16}$) and both fitted

well in one model (see also Figure 4, Panel b). The left panel of Figure 3 shows the context parameters c_i for the CI users and the NH group of this study. The context parameters were significantly higher for $i = 2$ to 5 in the CI group. The difference was largest for $i = 3, 4$, or 5. This means that if the CI users initially recognized one, two, or three words, they were better in correct prediction of the missing words based on context, than NH subjects.

The center panel of Figure 3 shows the calculated j factor from the model. The average j factors for speech in noise and in quiet from Table 2 (j_q and j_n) were also plotted. For the NH group, we plotted four average j values from the four measurements at fixed SNRs. There was no significant difference between the j factors of CI users and NH listeners. Below PC words of 0.8, the j factor was relatively low for both groups, indicating that much context information is used. For higher PC words, there is less need to use contextual information as reflected by a higher j factor. The k factor from the model was plotted in the right panel. It is apparent from this panel that the use of contextual information is relatively constant over the PC words, until this proportion reaches a value of 0.8. CI users made more use of context than NH listeners, in accordance with the difference in c_i values in the left panel.

Speech Intelligibility and Context Factors in Relation to the Reading Span

Table 3 provides Spearman correlation coefficients for correlations of speech intelligibility measures with age

and the Rspan measure. The PC CNC phonemes were not significantly correlated with the Rspan, but the PC words from sentences and the PC sentences in quiet were positively correlated with Rspan. For the SRTn, we also

Table 3. Spearman Correlation Coefficients of Speech Intelligibility Measures (PC and SRTn), and Context Factors (j and k), with Age and the Reading Span (Rspan) Score.

	Age		Rspan	
	rho	p	rho	p
PCq_ph_CNC(rau)	−0.23	.11	0.18	.24
PCq_wrd_CNC (rau)	−0.29	.043	0.09	.58
PCq_wrd (rau)	−0.31	.030	0.37	.011*
PCq_sen (rau)	−0.31	.026	0.38	.009*
SRTn (dB)	0.34	.016*	−0.30	.042
j_CNC	−0.07	.62	0.17	.28
j_CNC (−PCq_ph_CNC)	−0.03	.82	0.11	.47
j_n	−0.11	.45	0.31	.039
k_q	−0.05	.72	0.24	.13
k_q (−PCq_isol_wrd)	−0.14	.37	0.44	.0055*
k_q (−PCq_isol_wrd, -Age)			0.41	.011*
Rspan	−0.33	.024*		

Note. Rspan = reading span; CNC = consonant-nucleus-consonant; SRTn = speech-reception threshold in noise; PCq = proportion correct recognized words in quiet.
*The correlation is significant (<.05) after correction for multiple testing. Variables that were partialled out were given between brackets.
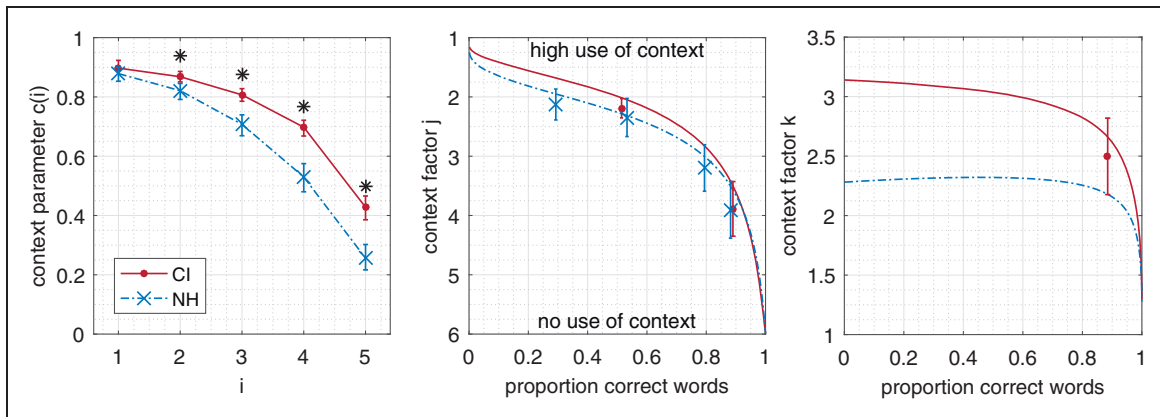

Figure 3. Left panel: Context parameters c_i that gave the best fit of the context model to the data of the sentence intelligibility, plotted as a function of index i. This index represents the number of missing words, and c_i is the probability that one of the missing words is guessed correctly based on contextual information. Higher c_i values indicate more use of context. Error bars give 95% confidence intervals. Significant differences between the CI group and the NH group are denoted with an *. Center and right panel: Predictions of the j factor and k factor from the model as a function of the proportion correct elements (words). The two dots in the center panel are data points from Table 2 (j_n and j_q), the cross markers give mean j factors from the data of the NH group. Note that the y axis of the center panel is inverted. The dot in the right panel is k_q from Table 2. Lower j values and higher k values indicate more use of context. CI = Cochlear implant; NH = normal hearing.

found a correlation with Rspan, but after correction for multiple comparisons, this correlation was not significant.

None of the j factors was significantly correlated with Rspan. Because the j_CNC factor was also dependent on the PC scores of elements (see Figure 2, center panel), we partialled out this variable, but still no significant relationship was found. The k factor was only available for the sentence material and had a weak, but not significant correlation with the Rspan. But from the right panel of Figure 3, it is clear that the k factor is dependent on the PC words from sentences. From the context model, it follows that this dependence also exists for the PC of isolated words. If this effect is partialled out, the k factor is significantly related to the Rspan, showing that more use of context is related to a better verbal working memory span.

Table 3 provides also Spearman correlation coefficients for correlations of speech intelligibility measures with age.

All speech scores tend to be lower for higher age, but the correlations were not significant, except for the SRTn measure. The j and k factors were not related to age. For the Rspan, a significant negative correlation with age was found. Furthermore age was partialled out from the correlation of the k factor with Rspan, but this did not change this correlation, indicating that age was not a dominant factor in the relation between ability to use contextual information and working memory capacity.

Responsiveness and Reproducibility

We plotted relations between the different scoring methods and the different speech materials in Figure 4 to obtain information about floor and ceiling effects and to get more insight into the suitability of the materials and scoring methods in individual CI users. In Panel a of Figure 4, the CNC word scores (PCq_wrd_CNC) are
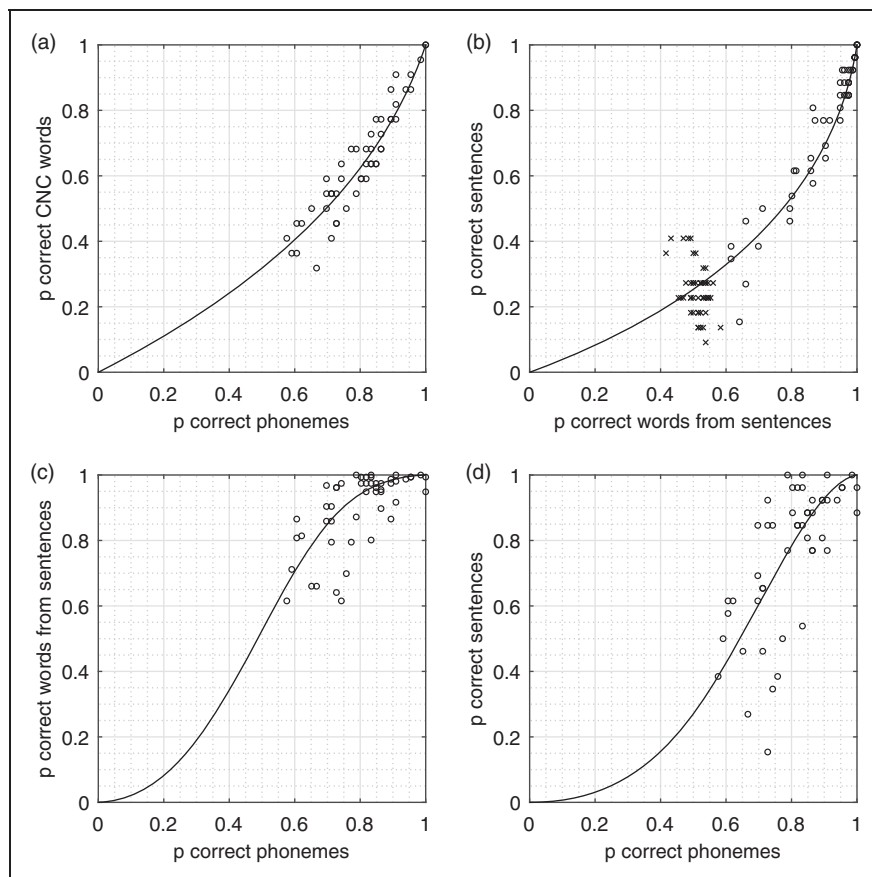

Figure 4. Relations between proportions correct recognition for different scoring methods and different speech materials. Panel a shows the relation between CNC phoneme scores (PCq_ph_CNC) and CNC word scores (PCq_wrd_CNC). Panel b shows the relation of the proportion of correctly recognized words from sentences (PCn_wrd) and the proportion correct recognized sentences (PCn_sen). The curves in Panels a and b are the result of fitting of the context model of Bronkhorst et al. (1993) to the data. Panels c and d show a comparison of CNC phoneme scores with scores from the sentence material. See the text for more information. Data from speech in noise are plotted with a x marker and data from speech-in-quiet conditions with a o marker.

plotted against the CNC phoneme scores (PCq_ph_CNC). The *j* factor from the center panel of Figure 2 was applied to the PC phonemes to obtain the curve in Panel a, showing good agreement with the data. Panel b presents the relation of the PC recognized sentences (PCq_sen and PCn_sen) and the proportion of correctly recognized words from sentences (PCq_wrd and PCn_wrd). The individual data points for the speech in noise condition are plotted together with the data from the speech in quiet condition. The curve in Panel b resulted from the fitting of the context model to the data (for details see Use of Context section) and is in good agreement with the data. From Panels a and b, it is clear that scoring of the elements causes some ceiling effect, most for words from sentence scoring.

Panel c of Figure 4 shows that, on average, the PC words from sentences were higher than CNC phoneme scores for phoneme scores >0.5. Panel c shows an apparent ceiling effect for words from sentences. Panel d shows that the PC sentences were less than the PC phonemes, except for phoneme scores >0.8. For sentence scoring, no ceiling effect was seen, but a floor effect was obvious.

The plotted curves in Panels c and d of Figure 4 are based on a fitted transform of CNC word scores to sentence scores, as described in the Methods section and illustrated in Figure 1. The resulting values of the c_i ($i = 1, \ldots, 5$) from the fit were (0.98, 0.89, 0.20, 0.04, 0). These values show that participants made more use of contextual information for words that consisted of more than three phonemes, if they understood a part of a word initially. On the other hand, if they perceived only one or two phonemes of a long word, the chance to guess the whole word correctly was low.

Interestingly, the sentence scores in Panel d differ largely between subjects in a range of 0.15 to 1 for phoneme scores between 0.5 and 0.8, suggesting that the ability to use contextual information differs between subjects. Therefore, we calculated the correlation between sentence scores and the *k* factor. The sentence scores were significantly correlated with the context factor k_q ($r = 0.41$, $p = .0036$).

The left panel of Figure 5 shows the PC of the different scoring methods and the different speech materials, plotted against the PC for isolated phonemes. From this figure, it is clear that differences in ceiling effects between materials are related to the amount of context within the material. For sentences, the PC score is already near maximum if still not all isolated phonemes were recognized. If the wholes are scored (CNC words or sentences), a larger PC-recognized isolated phonemes is needed for correct understanding of the wholes.

The center panel of Figure 5 shows the *SD* of the different scoring methods and the different speech materials, based on 22 trials (the length of two Dutch NVA CNC word lists). The *x*- and *y* axis were switched, to make the *y* axis of the left panel and the center panel the same. For example, for a PC recognized isolated phonemes of 0.4, the sentence scoring is 0.43 (left panel). The center panel shows the corresponding *SD*. For a value of 0.43 on the *y* axis, the SD of sentence scoring is 0.097. As expected from Equation 5, the SD for element scoring was smaller. The smallest SD was found for sentences with word scoring, because of the fact that the *j* factor for words from sentences was greater than the *j* factor of CNC phoneme scoring in CI users.
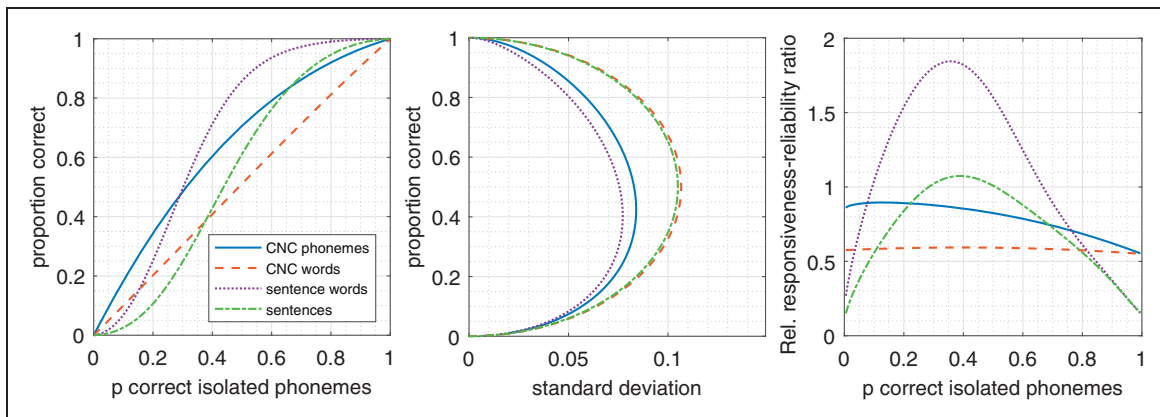

Figure 5. Left panel: Proportion correct values of the different speech materials and scoring methods and plotted against the proportion correct for isolated phonemes as obtained from the context models. Center panel: Standard deviations of the proportion correct values of the different speech materials and scoring methods from Equation 5. Right panel: Responsiveness-reliability ratios for CNC words and sentences with different scoring methods from the CI group relative to the responsiveness-reliability ratio of isolated phonemes. CNC = consonant-nucleus-consonant.

The right panel of Figure 5 presents the relative responsiveness-reliability ratios for CNC words and sentences with different scoring methods. As explained in the Methods section, the slope of the curves of the left panel was divided by the *SD*, relative to the *SD* of isolated phonemes. A higher ratio value is associated with a better sensitivity of the test taking into account the reliability. The ratio of CNC phoneme scoring is below 1, meaning that it was slightly less sensitive to reliably measure a change in sensory bottom-up information than isolated phonemes. CNC word scoring was even less sensitive. It is obvious that scoring the words of sentences gave the best opportunity to reliably measure a change in sensory bottom-up information if the isolated phoneme scores are below 0.75. Above this score of 0.75, word scoring suffered from a ceiling effect and became insensitive to changes in bottom-up information. Between a score of 0.75 and 1, CNC phoneme scoring had the best ratio. If an adaptive procedure is used with a target of 0.5, using words from sentence scoring, the ratio for word scoring is 1.8.

Discussion

Use of Context

This study has shown that contextual information from the speech materials has several effects on speech intelligibility in CI users. First, an important finding of this study was that CI users rely significantly more on contextual information in speech perception than normal-hearing listeners. This was true for both CNC words and sentences. In CNC words, the contextual information comes mainly from phonotactic constraints: the permissible phoneme sequences or syllables in a language. In the recall of sentences, the difference with NH listeners was largest if three, four, or five words were missing, that is, if relatively little information is available initially (see left panel Figure 3). For sentences, the difference between the CI group and the NH group is mainly the difference in the k factor, not the j factor. This reflects that CI users made better use of cues from known morpho-syntactic and semantic restrictions (Boothroyd & Nittrouer, 1988). These findings suggest that CI users are trained in finding correct words based on scarce information. The CI recipients have not had a formal training, but they were all experienced CI users with at least one year CI use. Likely they acquired the speech recognition skills by unintentional learning, because they have to practice the use of contextual information in daily life more than NH listeners.

A second effect of the extensive use of contextual information in CI users is that the variance in performance scores is somewhat increased, especially in CNC phoneme scores. This observation resulted from Equation 5, which shows that a lower j factor (more use of context) results in a higher variance. Figure 2 shows that for CNC words, the j factor was substantial lower in CI users compared with NH listeners. For the sentences, the j factor was not very different for the CI group compared with the NH group (Figure 3, center panel). This result may be explained by the fact that the j factor is mainly related to the c_1 parameter (the probability to guess the last word correctly if one word is missing), as described by Bronkhorst et al. (2002). The c_1 parameter was already high in the NH group, making it difficult to find a significant higher c_1 value in the CI group.

Third, this study showed that the use of contextual information from sentences could enhance the responsiveness of the speech test to changes in sensory bottom-up information on speech scores. This follows from the interpretation of Figure 5 (left panel) that due to the use of contextual information the responsiveness (the slope of the curves) was greater than one, meaning that a change in sensory bottom-up information (isolated phonemes) leads to an even greater change in word scores. This finding is in accordance with the study of Kong et al. (2015) who reported that the measured effect of electric-acoustic stimulation was larger if measured with high-context sentences compared with low-context sentences. So, the use of speech materials with context information is more sensitive to changes in bottom-up information than tests that aim to measure the amount of bottom-up information directly, for example, a non-word repetition test (e.g., Moberly et al., 2017).

Speech Intelligibility and Context Factors in Relation to the Reading Span and Age

The use of contextual information differed between CI users. This individual ability was best reflected by the individual k factor. The k factor was significantly positive correlated to verbal working memory as measured with the Rspan, if the effect of the proportion correctly recognized phonemes was partialled out. This is an indication that lexical-cognitive processing plays a role in the use of contextual information. Furthermore, the Rspan was significantly correlated with the PC words from sentences and the PC of sentences in quiet, but not with scores from CNC words. This suggests that the recognition of CNC words does not rely much on working memory capacity, because these words are short and relatively little processing is required. Understanding of sentences is more likely to depend on working memory. For example, if one of the first words of a sentence was not recognized, the last word of a sentence could make it much easier to predict the missed word. But such a prediction requires that the sentence is kept in the working memory and that some processing is done. This finding is

in accordance with other studies that reported significant positive correlations between a measure of speech perception and a measure of verbal working memory span (Heydebrand et al., 2007; Holden et al., 2013; Tao et al., 2014).

Interestingly, the capacity of using contextual information in sentences was only associated with working memory and not with age. As we found a negative correlation between working memory and age, as expected, we could also expect that older people have more difficulty in using context. This idea is supported by Wingfield, Alexander, and Cavigelli (1994) who found that older adults are less effective in retrospective identification of an unrecognized word that is followed by context words. Other studies reported a greater degree of interference from other words in older adults that may negatively affect the retrospective identification from contextual information (Amichetti et al., 2018; Lash, Rogers, Zoller, & Wingfield, 2013; Sommers, 1996; Sommers & Danielson, 1999). However, there is also an effect of aging on using context in the opposite direction, as older adults have on average a larger vocabulary size than younger adults (Burke & Peters, 1986; Verhaeghen, 2003), which could help with recognition of indistinct words from context. The combined effect of these factors is that in older adults, word recognition is facilitated by sentence context to an equal or greater degree than in young adults (Amichetti et al., 2018; Dubno, Ahlstrom, & Horwitz, 2000; Grant & Seitz, 2000; Nittrouer & Boothroyd, 1990; Pichora-Fuller, Schneider, & Daneman, 1995). This might explain our finding that the k factor was not related to age.

Suitability of an Ecologically Valid Sentence Test for Testing CI users and Recommendations for Clinical Practice

The results of this study suggest that an ecologically valid sentence test is suitable for testing speech intelligibility in CI users if word scoring is used. It appeared that the sentences were not too difficult to recognize for CI users.

The suitability of a test depends on the goal of the test. If the goal is to investigate differences in stimulation strategies or different signal processing options, it is recommended to use speech materials with contextual information within the sentences, word scoring, and a target PC in the mid-range (between 0.3 and 0.7). For CI users having a PC words from sentences in quiet $\geqslant 0.7$, the addition of noise is advised to bring the PC in the responsive mid-range. This recommendation is based on the results in Figure 5, showing that the sensitivity to reliably measure differences between conditions is best if a sentence test with word scoring is used. As explained earlier, the context effect increases the responsiveness to differences in sensory bottom-up information on speech scores.

If the goal is to measure the longitudinal improvement in speech perception due to treatment with CI, the use of the same speech tests pre- and postoperatively is required. From the two speech materials used in this study, the CNC words with phoneme scoring seem to be the best candidate for a longitudinal analysis, because with CNC phoneme scoring there is less risk of a floor or ceiling effect than in a sentence test. The use of phoneme scoring is recommended, because the responsiveness-reliability ratio is better for phoneme scoring than for word scoring (Figure 5, Panel c).

If one wants to combine both goals, we recommend the use of an ecologically valid sentence test with word scoring in combination with a CNC word test with phoneme scoring. The scoring of elements is recommended because it has the best test–retest variability. The combination of a CNC test and an ecologically valid sentence test allows the calculation of the k factor, as a measure of the use of contextual information by the individual patient. This provides a clinical specialist with a measure of the amount of top-down processing in an individual CI user.

Limitations

This study had several limitations. First, the test–retest reliability was derived from Equation 5 and was not actually measured. However, the test–retest reliability may not only originate from variance due to the binomial distribution but may be also influenced by variability between sentence lists. List equivalency is only known for NH listeners, not for CI users. But since lists were randomized over participants and the number of sentences was relatively large ($n = 26$), it is reasonable to assume that differences between sentence lists were small and averaged out. Second, no data for performance below 50% correct phonemes and sentences were included, because participants must be able to perform an adaptive measurement of the SRTn at 50% correct. A third limitation is that the mean age of the CI group and the NH group was different. An analysis of the effect of age in the CI group showed that the ability to use context was not associated with age, but a comparison of age-matched groups would have been even better, because this would have given the opportunity to compare both groups directly. The ability to use contextual information appeared to be an important factor in explaining individual differences in speech intelligibility. In this study, the contextual information came from context information within words and within sentences. In many daily situations, there is even more contextual information: supra sentence information from the topic of a discussion and visual information from speech reading and more general nonverbal communication cues.

These types of context information make even greater demands on the cognitive processing. We believe that the k factor is indicative for these types of context information as well, because the k factor reflects the capability of an individual to make use of context information and is also related to working memory capacity.

Conclusions

1. CI users rely significantly more on contextual information in speech perception than normal-hearing listeners. This was true for both isolated words and sentences.
2. The ability to use contextual information differs between CI recipients, and this ability is related to verbal working memory capacity regardless of age, indicating that postprocessing of the scarce sensory information is dependent on cognitive abilities.
3. The k factor is a good overall measure of the use of contextual information within speech.
4. Presence of contextual information in the speech of a test improves the responsiveness of the test to differences in sensory bottom-up information between conditions.
5. Contextual information increases the risk of a ceiling effect in the speech test, at least for high-performing CI listeners, but this potential problem can be mitigated by adding noise to bring the scores back into the responsive range.

Author Contributions

J. G. D. designed the study, did the analyses, and drafted the manuscript. A. G. revised the manuscript. Both authors approved the final version of the manuscript for submission.

ORCID iD

J. Gertjan Dingemanse (iD) http://orcid.org/0000-0001-8837-3474

References

Amichetti, N. M., Atagi, E., Kong, Y. Y., & Wingfield, A. (2018). Linguistic context versus semantic competition in word recognition by younger and older adults with cochlear implants. *Ear and Hearing*, *39*(1), 101–109. doi:10.1097/AUD.0000000000000469

Başkent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J., ... Gaudrain, E. (2016). Cognitive compensation of speech perception with hearing impairment, cochlear implants, and aging: How and to what degree can it be achieved? *Trends in Hearing*, *20*. doi:10.1177/2331216516670279

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, *13*(3), 108–112. doi:10.3109/03005367909078884

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Boothroyd, A., Hanin, L., & Hnath, T. (1985). *A sentence test of speech perception: Reliability, set equivalence, and short term leaning (Internal Report RCI10)*. New York, NY: University of New York.

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of America*, *84*(1), 101–114. doi:10.1121/1.396976

Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology*, *34*(5), 260–284.

Bronkhorst, A. W., Bosman, A. J., & Smoorenburg, G. F. (1993). A model for context effects in speech recognition. *The Journal of the Acoustical Society of America*, *93*(1), 499–509. doi:10.1121/1.406844

Bronkhorst, A. W., Brand, T., & Wagener, K. (2002). Evaluation of context effects in sentence recognition. *The Journal of the Acoustical Society of America*, *111*(6), 2874–2886. doi:10.1121/1.1458025

Burke, D. M., & Peters, L. (1986). Word associations in old age: Evidence for consistency in semantic encoding during adulthood. *Psychology and Aging*, *1*(4), 283–292. doi:10.1037/0882-7974.1.4.283

Dingemanse, J. G., & Goedegebure, A. (2018). Optimising the effect of noise reduction algorithm ClearVoice in cochlear implant users by increasing the maximum comfort levels. *International Journal of Audiology*, *57*(3), 230–235. doi:10.1080/14992027.2017.1390267

Dingemanse, J. G., Vroegop, J. L., & Goedegebure, A. (2018). Effects of a transient noise reduction algorithm on speech intelligibility in noise, noise tolerance and perceived annoyance in cochlear implant users. *International Journal of Audiology*, *57*(5), 360–369. doi:10.1080/14992027.2018.1425004

Dubno, J. R., Ahlstrom, J. B., & Horwitz, A. R. (2000). Use of context by young and aged adults with normal hearing. *The Journal of the Acoustical Society of America*, *107*(1), 538–546. doi:10.1121/1.428322

Ebrahimi-Madiseh, A., Eikelboom, R. H., Jayakody, D. M., & Atlas, M. D. (2016). Speech perception scores in cochlear implant recipients: An analysis of ceiling effects in the CUNY sentence test (Quiet) in post-lingually deafened cochlear implant recipients. *Cochlear Implants International*, *17*(2), 75–80. doi:10.1080/14670100.2015.1114220

Firszt, J. B., Holden, L. K., Skinner, M. W., Tobey, E. A., Peterson, A., Gaggl, W., . . . Wackym, P. A. (2004). Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems. *Ear and Hearing*, *25*(4), 375–387. doi:10.1097/01.AUD.0000134552.22205.EE

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America*, *110*(2), 1150–1163. doi:10.1121/1.1381538

Gaylor, J. M., Raman, G., Chung, M., Lee, J., Rao, M., Lau, J., & Poe, D. S. (2013). Cochlear implantation in adults: A systematic review and meta-analysis. *JAMA Otolaryngology— Head and Neck Surgery*, *139*(3), 265–272. doi:10.1001/jamaoto.2013.1744

Gifford, R. H., Shallop, J. K., & Peterson, A. M. (2008). Speech recognition materials and ceiling effects: Considerations for cochlear implant programs. *Audiology and Neuro-Otology*, *13*(3), 193–205. doi:10.1159/000113510

Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, reliability, and validity of PRESTO: A new high-variability sentence recognition test. *Journal of the American Academy of Audiology*, *24*(1), 26–36. doi:10.3766/jaaa.24.1.4

Grant, K. W., & Seitz, P. F. (2000). The recognition of isolated words and words in sentences: Individual variability in the use of sentence context. *The Journal of the Acoustical Society of America*, *107*(2), 1000–1011. doi:10.1121/1.428280

Henry, B. A., & Turner, C. W. (2003). The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *The Journal of the Acoustical Society of America*, *113*(5), 2861–2873. doi:10.1121/1.1561900

Heydebrand, G., Hale, S., Potts, L., Gotter, B., & Skinner, M. (2007). Cognitive predictors of improvements in adults' spoken word recognition six months after cochlear implant activation. *Audiology and Neurotology*, *12*(4), 254–264. doi:10.1159/000101473

Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., . . . Skinner, M. W. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear and Hearing*, *34*(3), 342–360. doi:10.1097/AUD.0b013e3182741aa7

Kong, Y. Y., Donaldson, G., & Somarowthu, A. (2015). Effects of contextual cues on speech recognition in simulated electric-acoustic stimulation. *The Journal of the Acoustical Society of America*, *137*(5), 2846–2857. doi:10.1121/1.4919337

Kushner, H. J., & Yin, G. (2003). *Stochastic approximation and recursive algorithms and applications* (vol. 35). New York, NY: Springer.

Lash, A., Rogers, C. S., Zoller, A., & Wingfield, A. (2013). Expectation and entropy in spoken word recognition: Effects of age and hearing acuity. *Experimental Aging Research*, *39*(3), 235–253. doi:10.1080/0361073X.2013.779175

Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants. *Advances in Oto-Rhino-Laryngology*, *64*, 109–143. doi:10.1159/000094648

Luxford, W. M., & Ad Hoc Subcommittee of the Committee on Hearing and Equilibrium of the American Academy of Otolaryngology-Head and Neck Surgery. (2001). Minimum speech test battery for postlingually deafened adult cochlear implant patients. *Otolaryngology—Head and Neck Surgery*, *124*(2), 125–126. doi:10.1067/mhn.2001.113035

McRackan, T. R., Bauschard, M., Hatch, J. L., Franko-Tobin, E., Droghini, H. R., Nguyen, S. A., & Dubno, J. R. (201a). Meta-analysis of quality-of-life improvement after cochlear implantation and associations with speech recognition abilities. *Laryngoscope*, *128*(4), 982–990. doi:10.1002/lary.26738

Moberly, A. C., Harris, M. S., Boyce, L., & Nittrouer, S. (2017). Speech recognition in adults with cochlear implants: The effects of working memory, phonological sensitivity, and aging. *Journal of Speech, Language, and Hearing Research*, *60*(4), 1046–1061. doi:10.1044/2016_JSLHR-H-16-0119

Minimum Speech Test Battery. (2011). *New Minimum Speech Test Battery (MSTB) for adult cochlear implant users*. Retrieved from: http://www.auditorypotential.com/MSTBfiles/MSTBManual2011-06-20%20.pdf (accessed March 2018).

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *The Journal of the Acoustical Society of America*, *87*, 2705. doi:10.1121/1.399061

Nittrouer, S., Tarr, E., Bolster, V., Caldwell-Tarr, A., Moberly, A. C., Lowenstein, J. H. (2014). Low-frequency signals support perceptual organization of implant-simulated speech for adults and children. *International Journal of Audiology*, *53*(4), 270–284. doi:10.1121/1.399061.

Oh, S. H., Donaldson, G. S., & Kong, Y. Y. (2016). Top-down processes in simulated electric-acoustic hearing: The effect of linguistic context on bimodal benefit for temporally interrupted speech. *Ear and Hearing*, *37*(5), 582–592. doi:10.1097/AUD.0000000000000298

Pearsons, K. S., Bennett, R. L., & Fidell, S. (1977). *Speech levels in various noise environments*. Washington, DC: U.S. Environmental Protection Agency.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*(1), 593–608. doi:10.1121/1.412282

Rubinstein, J. T. (2004). How cochlear implants encode speech. *Current Opinion in Otolaryngology and Head and Neck Surgery*, *12*(5), 444–448. doi:00020840-200410000-00016 [pii].

Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging*, *11*(2), 333–341. doi:10.1037/0882-7974.11.2.333

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging*, *14*(3), 458–472. doi:10.1037/0882-7974.14.3.458

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., . . . Cook, S. (2012). Development and validation of the AzBio sentence lists.

Ear and Hearing, *33*(1), 112–117. doi:10.1097/AUD.0b013e 31822c2549

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research*, *28*(3), 455–462.

Tao, D., Deng, R., Jiang, Y., Galvin, J. J. 3rd, Fu, Q. J., & Chen, B. (2014). Contribution of auditory working memory to speech understanding in mandarin-speaking cochlear implant users. *PLoS One*, *9*(6), e99096. doi:10.1371/journal.pone. 0099096

Theelen-van den Hoek, F. L., Houben, R., & Dreschler, W. A. (2014). Investigation into the applicability and optimization of the Dutch matrix sentence test for use with cochlear implant users. *International Journal of Audiology*, *53*(11), 817–828. doi:10.3109/14992027.2014.922223

Thornton, A. R., & Raffin, M. J. (1978). Speech-discrimination scores modeled as a binomial variable. *Journal of Speech and Hearing Research*, *21*(3), 507–518. doi:10.1044/jshr.2103.507

van den Noort, M., Bosch, P., Haverkort, M., & Hugdahl, K. (2008). A standard computerized version of the reading span test in different languages. *European Journal of Psychological Assessment*, *24*(1), 35–42. doi:10.1027/1015-5759.24.1.35

van Wieringen, A., & Wouters, J. (2008). LIST and LINT: Sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands. *International Journal of Audiology*, *47*(6), 348–355. doi:10.1080/14992020801895144

Verhaeghen, P. (2003). Aging and vocabulary scores: A meta-analysis. *Psychology and Aging*, *18*(2), 332–339. doi:10.1037/0882-7974.18.2.332

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *The Journal of the Acoustical Society of America*, *107*(3), 1671–1684. doi:10.1121/1.428451

Vroegop, J. L., Dingemanse, J. G., van der Schroeff, M. P., Metselaar, R. M., & Goedegebure, A. (2017). Self-adjustment of upper electrical stimulation levels in CI programming and the effect on auditory functioning. *Ear and Hearing*, *38*(4), e232–e240. doi:10.1097/AUD.000000000 0000404

Wingfield, A., Alexander, A. H., & Cavigelli, S. (1994). Does memory constrain utilization of top-down information in spoken word recognition? Evidence from normal aging. *Language and Speech*, *37*(Pt 3), 221–235. doi:10.1177/002383099403700301

Winn, M. B. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends in Hearing*, *20*. doi:10.1177/2331216516669723

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. *The Journal of the Acoustical Society of America*, *131*(2), 1465–1479. doi:10.1121/1.3672705