# Whispered speech as input for cochlear implants

Hans Rutger Bosker[1], Jeroen Briaire[2], Willemijn Heeren[1],
Vincent J. van Heuven[1] and Suzanne Jongman[1]


*Leiden Institute for Brain and Cognition (LIBC)*
*[1]Phonetics Laboratory, Leiden University Centre for Linguistics*
*(LUCL) / [2]Leiden University Medical Centre (LUMC)*

## 1.    Introduction

Cochlear implants (CIs) enable the (near-)deaf to perceive sounds and speech. A body-external microphone sends signals to a speech processor, which transforms the speech signal into electrical pulses (Loizou 2006). These pulses are transferred to an array of electrodes in the inner ear, which stimulate the auditory nerve. These signals can be interpreted by the brain as rather crudely defined time-variant sound spectra. The quality of the sound transmission is different from that in normal-hearing people, because, among other things, the number of electrodes (currently a maximum of 22) and the frequency resolution is limited. One of the crucial disadvantages of CIs is that they hardly transmit low frequencies, so that information on periodicity (fundamental frequency or F0) is minimally available to CI-users. Cues on, for example, voicing, word tone, sentence melody, stress and speaker sex, are therefore reduced for CI-users.

In whispered speech the vocal folds do not vibrate, so that F0 is absent in this type of speech as well. There are, however, indications a whisperer compensates for this lack of phonation. Tartter (1989) has demonstrated that listeners can correctly perceive the voiced-unvoiced contrast in whispered plosives. Tartter suggests that other cues, such as duration of aspiration, F1 frequency and intensity of the burst, are utilised by the listener to decide on

voicing in consonants in whispered speech. Also, Tartter & Braun (1994) have shown that the distinction between sad and neutral, or happy and neutral emotion are both correctly perceived in whispered speech. Not only do listeners extract information on voicing and emotion from whispered speech (despite the absence of F0), it has also been shown that listeners can correctly identify the pitch of a whispered vowel /a/ as normal, high or low (Higashikawa et al. 1996).

From Tartter (1989), Tartter & Braun (1994) and Higashikawa et al. (1996) (but see also McGlone & Manning 1979), it is understood that whisper can still, despite the absence of F0, cue correct perception of tone, voicing, and emotion. In line with Lindblom's (1990) Hyper & Hypo theory of phonetic variation we would suggest that whisperers in some way compensate in the higher frequencies of the speech signal for the lack of F0 so as to accommodate the increased listener needs. Such compensations might be especially beneficial for CI-users, for whom the low frequencies are also to a large extent absent.

In addition, vowels in normal speech have much more overall intensity than consonants. Yet, consonants contribute more to the identity of words than vowels do (Van Heuven 2008, and references therein). In whisper, vowels and consonants have roughly equal intensity, so that the important information in the consonants is relatively louder. This might facilitate word recognition and improve sentence intelligibility.

From the above we derive the following hypotheses with respect to the suitability of phonated versus whispered speech to CI-users.

1. Whispered speech may contain enhanced secondary cues to vocal pitch (to compensate for the lack of F0) so that sentence melody including accentuation of specific, typically lexically stressed, syllables, speaker sex and emotion are better perceived in whispered than normally phonated CI-input.

2. In comparison with vowels, consonants are relatively louder in whispered speech than in phonated speech, arguably yielding better intelligibility of whispered CI-input.

In this study we investigated which type of speech, whispered or normal, provides better input to CI-users. Our hypothesis is that whispered speech is more beneficial to CI-users than phonated speech, because there is compensation in whispered speech for the lack of F0 and because consonants are relatively louder in whisper than in normal speech.

## 2.    Method

We have tested our hypotheses by means of three experiments, each addressing one or two aspects of speech recognition:

1. the perception of *lexical stress* and *speaker sex identification,*
2. the perception of *sentence melody,*
3. overall *segmental intelligibility*.

Speech from six native speakers of Dutch (three males, three females, aged 19-23) was recorded for experiment 1. One male and one female speaker of this set also recorded the stimuli for experiments 2 and 3. Recordings were made of the speakers' normal and whispered speech in a sound-attenuating booth in the Phonetics Laboratory of Leiden University. A Sennheiser MKH 416T directional condenser microphone was used for recording at a sampling frequency of 44.1 kHz. Using the Praat speech processing software (Boersma 2001) the intensity of the normally phonated and the whispered stimuli was normalised (whispered speech naturally has a lower intensity than normally phonated speech) by amplifying the digital representation of each single stimulus sentence to the maximum level afforded by the analog-to-digital converter. The mean intensity (in decibels) was then computed for the whispered and phonated version of each sentence as produced by each speaker. Either the whispered or the phonated member, whichever had the higher mean intensity, was then uniformly attenuated such that the mean intensities of the counterpart utterances were equalized.

To simulate speech as Dutch CI-users would perceive it, all stimuli were processed at the ENT department of the Leiden University Medical Centre (LUMC) through an eight-channel noise vocoder (e.g., Litvak et al. 2007, Spahr et al. 2008) using the CI simulator TIGERCIS (http://www.tigerspeech.com/; last visited 19/05/2010). Stimulus processing involved two phases: an analysis phase and a synthesis phase. In the analysis phase, band-pass filters were used to divide the signal into eight non-linearly spaced channels between 351 and 8333 Hz (24 dB/octave roll-off) and a low-pass filter to derive the amplitude envelope from each channel (160 Hz cut-off, 24 dB/octave roll-off). The following settings were used for the eight non-linearly spaced channels: (1) 351-496 Hz, (2) 496-699 Hz, (3) 699-987 Hz, (4) 987-1393 Hz, (5) 1393-1965 Hz, (6) 1965-2773 Hz, (7) 2773-3913 Hz, (8) 3913-8333 Hz.

In the synthesis phase, narrow-band noise was generated around each of the centre frequencies of the eight analysis filters, and modulated with their respective amplitude envelopes.

The eight-channel simulation was selected based on research that has demonstrated that on average normal-hearing listeners perform similar to CI users when listening to eight-channel simulations compared to greater or fewer numbers of channels (e.g. Dorman et al. 1997).

Each of the three experiments was administered to the same 24 listeners (12m, 12f; mean age 20), all of whom were native listeners of Dutch and students at Leiden University. They were paid for their participation. Experiments were run in a quiet lecture room. Stimuli were presented over Quad Electrostatic ESL-63 loudspeakers at a comfortable listening level. Fourteen listeners heard the stimuli in one pseudo-random order, whereas the other ten heard the reverse stimulus order, to counterbalance effects of fatigue or learning.

A session consisted of five practice sentences prior to each experiment proper. For experiments 1 and 2 the inter-stimuli interval (ISI) was 5 seconds, whereas for experiment 3 the ISI was 15 seconds. After every ten stimuli a 1000-Hz tone was played as a reference. Each experiment lasted 15 minutes, making the entire session last for about an hour.

## 2.1. *Perception of lexical stress and male-female identification*

In experiment 1 the perception of lexical stress and the identification of male-female voices were investigated. Presenting listeners with both normally phonated and whispered stimuli that had been processed using a vocoder and consequently testing in which type of speech participants perform better, should provide insight into which type of stimulus can be seen as better input for CIs.

A neutral sentence containing a target word was used: *Ik ga nu het woord …. zeggen*, 'I am going to say the word …. now'. This context was chosen for its non-restrictive semantics and the non-final position of the target word. The target words carried a sentence accent and consisted of di-syllabic non-words limiting the choice to two possible answers: either the first of the second syllable (a 2AFC task). The selected target words consisted of sonorant consonants and two instances of the same vowel: /nunu/, /nini/, /lulu/ and /lili/. The total number of stimuli was 96 (6 speakers × 2 phonation types × 4 non-words × 2 stress positions). Participants were asked to report the position of the stress in the target word by ticking one of two boxes on their answer sheets. On the answer sheet, the words of each sentence were printed, so that participants knew which word was coming and could therefore focus on the stress judgment task.

Simultaneously with this first part of experiment 1, listeners were asked to provide a judgment on the sex of the speaker. Participants reported their gender judgments by ticking one member of a second pair of boxes on the same answer sheet.

2.2. *Perception of sentence melody*

In experiment 2 the perception of sentence melody was investigated. In Dutch, a sentence with an affirmative word order can be given a question interpretation by placing an interrogative sentence melody on the sentence, even without changing the word order (Van Heuven & Haan 2002, Haan & Van Heuven 2003). Examples of two versions of an affirmative word order are:

(1)  a.    Hij gaat morgen naar Leiden. [affirmative prosody]
           He goes tomorrow to Leiden
           'He goes to Leiden tomorrow.'
     b.    Hij gaat morgen naar Leiden? [interrogative prosody]
           'Is he going to Leiden tomorrow?'

The opposite is also possible: placing an affirmative sentence melody on an interrogative sentence with question word order (Van Heuven & Kirsner 2004). The difference is, however, that the question with affirmative prosody does not receive an affirmative interpretation but rather signals a polite command or suggestion.

(2)  a.    Neemt u de trein naar Ede? [interrogative prosody]
           Take you the train to Ede
           'Do you take a train to Ede?'
     b.    Neemt u de trein naar Ede. [affirmative/suggestive prosody]
           'Please take a train to Ede'

These four types of stimulus sentences were used: sentences with affirmative word order and affirmative sentence melody (1a), sentences with affirmative word order but interrogative sentence melody (1b), sentences with interrogative word order and interrogative sentence melody (2a), or sentences with interrogative word order and affirmative/suggestive sentence melody (2b). For all types of sentences, sentence melody ultimately determines the interpretation. If indeed it is the case that whispered speech is better input for CIs, we expect better performance of participants on whispered-vocoded stimuli in general as compared to normal-vocoded stimuli.

Of each type six different sentences were selected and participants were asked for their judgment whether the sentence had an affirmative/suggestive interpretation versus an interrogative one by ticking the appropriate box on their answer sheet (2AFC task). Again, the sentences (without punctuation) were printed on the answer sheet, so that intelligibility would not be a concern. The

total number of stimuli was 96, i.e. 2 speakers × 2 phonation types × 6 sentences × 2 word orders x 2 sentence melodies.

### 2.3. *Segmental intelligibility*

Speech intelligibility was tested in a dictation task. Listeners heard vocoder-processed whispered and normally phonated sentences from a subset of the Dutch VU-98 sentences (Versfeld et al. 2000) and were asked to write down what they heard as accurately as possible. The VU-98 sentences are specifically selected sets of thirteen sentences for which has been established that the speech reception threshold (SRT) is equal across all sets (Versfeld et al. 2000). This means that all sets yield equal intelligibility scores in principle. Four sets of 13 sentences were selected: two spoken by our male and two by our female speaker, and per speaker one set was whispered and one phonated. The total number of stimuli was 52. i.e. 2 speakers × 2 phonation types × 13 sentences.

### 3.    **Results**

As is known from the literature, long-term exposure to vocoded speech in time leads to improvement of recognition scores (Davis et al. 2005). Therefore, we tested whether we could find any learning effects in our data by comparing the scores obtained from the first and second halves ('blocks') of the experiment separately. Learning trends were absent in all of the experiments. However, when a Repeated Measures Analysis of Variance (RM-ANOVA) was run on the percentage correct scores with Phonation Type and Block as within-subject factors, a main effect of Block, $F(1,23) = 5.9$ ($p = 0.024$), and an interaction between Phonation Type and Block, $F(1,23) = 14.1$ ($p = 0.001$), was found in the first part of experiment 1 investigating lexical stress. These results, however, indicate a fatigue effect, rather than learning.

### 3.1. *Perception of lexical stress and male-female identification*

For lexical stress perception, i.e. the first part of experiment 1, percentages of correct responses were computed and subjected to an RM-ANOVA with Phonation Type and Stress Position as within-subject factors. There was a significant main effect of Phonation Type, $F(1,47) = 8.6$ ($p = 0.005$), as recognition scores were higher for normal speech (90.7%) than for whispered speech (86.9%). The high scores indicate that the task was relatively easy. There was also a main effect of Stress Position, $F(1,47) = 25.3$ ($p < 0.001$). As shown

in figure 1, percentages correct were higher for stimuli with final stress than for stimuli with initial stress.
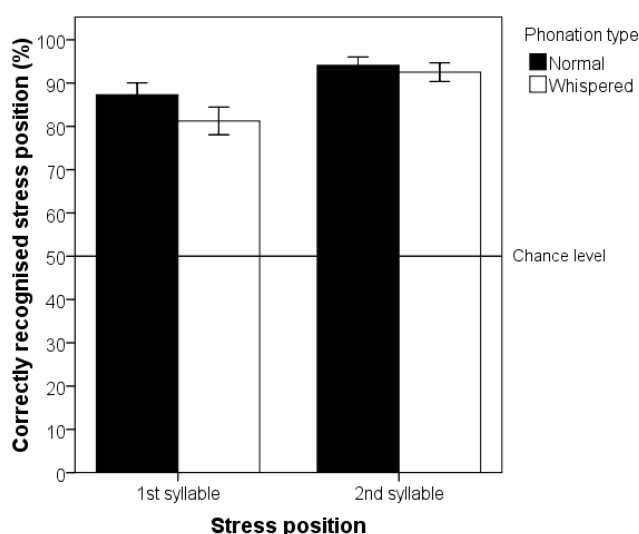


Figure 1: Mean percentage correctly reported stress position broken down by Phonation Type and Stress Position. Error bars: 95% confidence interval.

The second part of experiment 1 concerned the identification of male and female voices. An RM-ANOVA was run with Phonation Type and Speaker Sex as within-subject factors. We expected listeners to perform significantly better on this identification task in whispered than in normal speech. The analysis showed an interaction between Phonation Type and Speaker Sex, $F(1,23) = 26.8$ ($p < 0.001$). Also, main effects of Speaker Sex, $F(1,23) = 12.4$ ($p = 0.002$), and of Phonation Type were found, $F(1,23) = 10.9$ ($p = 0.003$).

Speaker sex identification was 67% in vocoded normal speech, but 76% in vocoded whispered speech. The interaction indicates that this effect differs between male and female speakers, as illustrated by figure 2. Overall, male voices were better identified in normal speech, whereas female voices were better identified in whispered speech. In addition, we see chance scores for normally phonated stimuli spoken by females, whereas identification of whispered male stimuli remained above chance level in two out of three speakers.
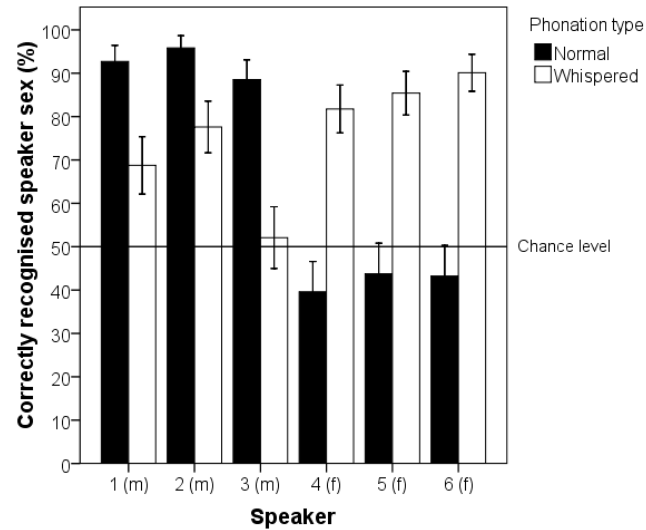
Figure 2. Mean percentage correctly reported sex of speaker broken down by Phonation Type and Speaker (three males and three females). Error bars: 95% confidence interval.
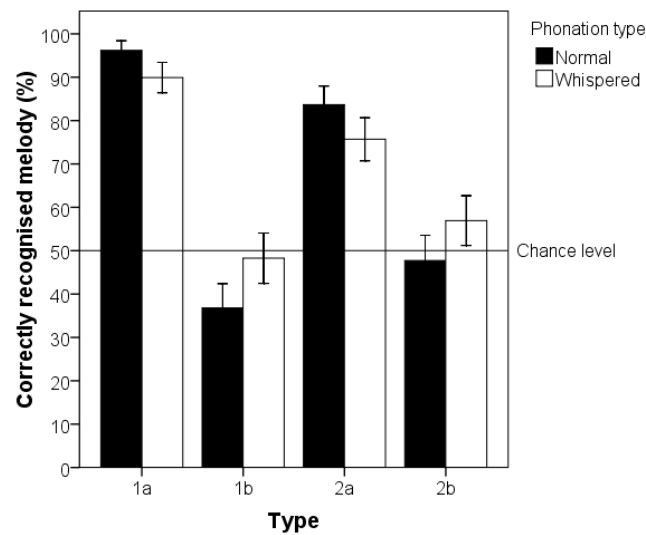


Figure 3: Mean percentage correctly reported sentence melody broken down by Type (1a to 2b) and Phonation Type (clusters). Error bars: 95% confidence interval.

### 3.2. *Perception of sentence melody*

On the percentages of correct sentence melody identification, an RM-ANOVA was run with Phonation Type, Word Order and Sentence Melody as within-subject factors. As for the perception of sentence melody (affirmative/suggestive versus interrogative melody), the score on whispered stimuli was expected to be significantly higher than on normally phonated stimuli over all types of stimuli (1a) to (2b). The results, however, showed no main effect of Phonation Type: $F(1,23) = 1.2$ ($p = 0.289$, ins.). The percentage correct for normal speech was 66.1, and 67.7 for whispered speech. Also, a main effect of Sentence Melody was found, $F(1,23) = 12.6$ ($p = 0.002$), which indicates better scores on stimuli of type (1a) and (2b) as opposed to types (1b) and (2a) (see figure 3). Moreover, a three-way interaction of Phonation Type, Sentence Melody and Word Order was observed, $F(1,23) = 14.9$ ($p = 0.001$).

As is illustrated in figure 3, better scores were found for stimuli of type (1a) and (2a) as opposed to (1b) and (2b). As may be recalled, four stimulus types were used in the experiment:

1a. affirmative word order + affirmative melody
1b. affirmative word order + interrogative melody
2a. interrogative word order + interrogative melody
2b. interrogative word order + affirmative/suggestive melody

The crucial difference between the a and b types concerns the association between word order and prosody. In sentences of type a, word order and prosody are in congruence with each other, whereas in sentences of type b word order and prosody diverge. In other words, in sentences of type a, both word order and sentence melody cue the correct interpretation of the sentence. In type b, however, the word order and sentence melody clash with each other. Word order in this case cues the 'opposite' interpretation from the sentence melody. In (1b) and (2b), therefore, participants had only the prosody at their disposal to report a correct choice and go against the incorrect bias of the word order, thus making the task harder. Therefore, reasoning from our hypothesis that prosody is cued more clearly in whispered speech, we may expect listeners to perform better on whispered stimuli of type 1b and 2b. We tested for this by performing a post-hoc analysis on a subset of all stimuli. On the percentages of correct sentence melody identification of types (1b) and (2b), an RM-ANOVA was run with Phonation Type as a within-subject factor. As expected, a main effect of Phonation Type is found within this subset of the data, $F(1,23) = 9.7$ ($p = 0.005$), even though the scores on stimuli of type b are around chance.

3.3. *Segmental intelligibility*

The percentage of correctly perceived words was computed as our measure of segmental intelligibility. An RM-ANOVA was run with Phonation Type as a within-subject factor. The overall segmental intelligibility of the VU-98 sentences was expected to be better for whispered speech than for normal speech. The opposite was found: segmental intelligibility was significantly better for normal stimuli (91.5%,) than for whispered stimuli (77.7%), $F(1,23) = 48.3$ ($p < 0.001$).

## 4.    Discussion

For this study we asked which type of speech, whispered or normal, is more beneficial to CI-users. Based on earlier research we expected (amplified) whispered speech to be more beneficial to CI-users than normally phonated speech for the two reasons listed in the introduction:

1. Whispered speech may contain enhanced secondary cues on vocal pitch (to compensate for the lack of F0) so that sentence melody including accentuation of specific, typically lexically stressed, syllables, speaker sex and emotion are better perceived in whispered than normally phonated CI-input.
2. In comparison with vowels, consonants are relatively louder in whispered speech than in phonated speech, arguably yielding better intelligibility of whispered CI input.

To simulate listeners receiving CI-input, whispered and phonated speech were processed using a vocoder roughly mimicking the stimulus quality of speech as perceived by cochlear implantees.

In the first part of experiment 1, concerning the perception of lexical stress, we found no support for our hypothesis. There was a better score for normal speech than for whispered speech, but both scores were high (i.e. ceiling effect) indicating that the task was (too) easy.

In the results of the second part of experiment 1, concerning the perception of the speaker sex, an interaction of Phonation Type and Speaker Sex was found. Specifically, the interaction suggested a preference for reporting a male voice when hearing normal-vocoded speech and a (somewhat smaller) preference for reporting a female voice when listening to whispered-vocoded speech. On balance, however, the results of the second part of experiment 1 support our

hypothesis: identification of male and female voices was better in whispered-vocoded speech. For non-processed whispered stimuli, Tartter (1989) found high speaker sex identification scores (2 speakers), in line with the findings of Schwartz & Rine (1968) (10 speakers). Lass et al. (1976) reported a more moderate speaker sex identification score of 75% (20 speakers). This seems to suggest that sex identification scores may be dependent on the speakers themselves. The conclusion of Lass et al. (1989), however, that fundamental frequency information is crucial for speaker sex identification is weakened by our high scores on whispered-vocoded compared with normal-vocoded stimuli.

Since the second part of experiment 1, targeting perception of speaker gender, was the only experiment that showed a clear advantage of whispered over phonated input for cochlear implantees, we attempted a preliminary stimulus analysis. We computed long-term average spectra on the vocoded male and female stimuli separately, expecting to find a flatter spectral tilt for the female than for the male speaker (in line with Sluijter 1995, Sluijter & Van Heuven 1996). The analysis revealed a difference in spectral slope, which, however, was entirely caused by differences in the intensity levels of the base band (0-500 Hz, comparable with the lowest filter in the vocoded signals) between the male and female spectra. In the (vocoded) normal speech, the intensity level in the base band of the male speaker was just 1 dB greater than in that of the female speaker. In the (vocoded) whispered speech the intensity difference in the base band between the male and female speaker was much larger (13 dB greater for the male speaker). Although this difference suggest a first explanation of the better discrimination of speaker sex in vocoded whisper than in vocoded phonated speech, a more detailed pairwise by-item comparison will be required to provide a more definitive explanation of the perceptual effect.

Experiment 2 concerned the perception of sentence melody. We found no support for our hypothesis, since the scores on whispered stimuli were equal to the scores on normal stimuli. An explanation for these results might lie in the difficulty of the task (as reported by participants) causing them to rely mostly on the word order and not on the prosody. This hypothesis was tested for by running a post-hoc analysis on only those stimuli which had 'unexpected' prosody, i.e. types (1b) and (2b). In this specific subset of the stimuli, we found support for our hypothesis in a main effect of Phonation Type. This seems to suggest that whispered stimuli are better input than normal stimuli for CIs when the listener is wholly dependent on prosody for the correct interpretation, but as the scores on these types of stimuli are around chance level this suggestion is controversial.

Finally, experiment 3 concerned overall segmental intelligibility. No support for our hypothesis was found, since better scores were observed for normal than for whispered stimuli. This result is in contrast with Ringeling's (1984) study of the intelligibility of whispered, soft and normal speech. He

showed that  Dutch isolated whispered words were 10 percentage points more intelligible (77% correct) than normal speech (67%) while softly spoken words were 35 percentage points less intelligible (32%) than normal speech. There are several methodological differences between the present study and Ringeling (1984). Firstly, Ringeling did not present vocoder-processed stimuli but used naturally produced materials only. Second, Ringeling tested the intelligibility of single words presented in isolation rather than words in sentence contexts. The third, and probably crucial, difference is that Ringeling's materials were presented at an intensity of 20 dB above the individual listener's individual hearing threshold, which was established separately for normal speech, soft speech and for whisper. These levels of presentation seem unrealistically low, since speech in everyday life is typically heard at intensity levels between 60 and 80 dB above the hearing threshold. In our own experiments, we used a rather more straightforward way of intensity normalisation: we amplified all stimuli to the maximum level afforded by the available hardware. As a result all our stimuli, whether originally phonated or whispered, will be much more intelligible than Ringeling's.

All in all, in line with Lindblom's Hyper & Hypo theory of speech production, we expected compensatory measures in whisper and relatively louder consonants in whisper, which should render whispered speech superior input for CIs. Some support for our hypothesis was found in the results on speaker sex identification. These results, however, are compromised by a clear preference for judging normal-vocoded stimuli to be spoken by a male speaker and whispered-vocoded stimuli to be spoken by a female speaker. Further support for the hypothesis is found in an analysis on a subset of the stimuli of experiment 2. Due to scarce support for the hypothesis, we will have to retreat to a more conservative view that states sufficient secondary cues are present in speech to afford some perception of pitch and voicing in whisper. Tartter (1989), for instance, found that duration of aspiration, F1 frequency and intensity of the burst to cue the voicing distinction in whispered consonants. Quite likely these cues are also present in normal speech, causing whispered speech not to have an advantage over normal speech.


## 5.    Conclusion

This study asks which type of speech, whispered or normal, can be seen as better input for CIs. All things considered, we cannot provide a clear, decisive answer to the question which type of speech, whispered or normal, is more beneficial as input to CI-users. Our hypothesis that whispered speech is better input is not

sufficiently supported by the results. The prediction that there should be enhanced secondary cues to vocal pitch in whisper compensating for the lack of F0 is not borne out by the results. But taking into account that there are few studies investigating the acoustic characteristics of whispered speech, new insights into the nature of whispered speech might provide us with new applications of this particular type of speech.

## References

Boersma, Paul 2001. "Praat, a system for doing phonetics by computer". *Glot International* 5.341-345.

Davis, Matt H., Johnsrude, Ingrid S., Hervais-Adelman, Alexis & Taylor, Karen 2005. "Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences". *Journal of Experimental Psychology: General* 134.222–241.

Dorman, Michael F., Loizou, Philipos C. & Rainey, Dawne 1997. "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise band outputs". *Journal of the Acoustical Society of America* 102.2403–2411.

Haan, Judith & van Heuven, Vincent J. 2003. "This is a yes/no question? *Linguistics in the Netherlands 2003* ed. by Paula Fikkert & Leonie Cornips, 59-70. Amsterdam/Philadelphia: John Benjamins.

van Heuven, Vincent J. 2008. "Making sense of strange sounds: (mutual) intelligibility of related language varieties. A review". *International Journal of Humanities and Arts Computing* 2.39-62.

van Heuven, V.J. & Haan, J. 2002. "Temporal development of interrogativity cues in Dutch". *Papers in Laboratory Phonology VII* ed. by Carlos Gussenhoven & Natasha Warner, 61-86. Berlin: Mouton de Gruyter.

van Heuven, Vincent J. & Kirsner, Robert S. 2004. "Phonetic or phonological contrasts in Dutch boundary tones?". *Linguistics in the Netherlands 2004* ed. by L. Cornips & J. Doetjes, 102-113. Amsterdam/Philadelphia: John Benjamins

Higashikawa, Masahiko, Nakai, Ken, Sakakura, Atsushi & Takahashi, Haruhisa 1996. "Perceived pitch of whispered vowels – relationship with formant frequencies: a preliminary study". *Journal of Voice* 10.155–158.

Lass, Norman J., Hughes, Karen R., Bowyer, Melanie D., Waters, Lucille T. & Bourne, Victoria T. 1976. "Speaker sex identification from voiced, whispered, and filtered isolated vowels". *Journal of the Acoustical Society of America* 59.675-678.

Lindblom, Björn E.F. 1990. "Explaining phonetic variation: a sketch of the H&H theory". *Speech production and speech modelling* ed. by W.J. Hardcastle & A. Marchal, 403-439. Dordrecht: Kluwer.

Litvak, Leonid M., Spahr, Anthony J., Saoji, Aniket A. & Fridman, Gene Y. 2007. "Relationship between perception of spectral ripple and speech recognition in

cochlear implant and vocoder listeners". *Journal of the Acoustical Society of America* 122.982-991.

Loizou, Philipos C. 2006. "Speech processing in vocoder-centric cochlear implants". *Cochlear and Brainstem Implants* ed. by Aage R. Møller, 109-143. Basel: Karger.

McGlone, Robert E. & Manning, Walter H. 1979. "Role of second formant in pitch perception of whispered and voiced vowels". *Folia Phoniatrica (Basel)* 31.9-14.

Ringeling, Johannes C.T. 1984. Reducing redundancy in normal, soft and whispered speech: a study on native and near-native perception. Doctoral diss., Utrecht University.

Schwartz, Martin F. & Rine, Helen E. 1968. "Identification of speaker sex from isolated, whispered vowels". *Journal of the Acoustical Society of America* 44.1736-1737.

Sluijter, Agaath M.C. 1995. *Phonetic correlates of stress and accent*. Holland Institute of Generative Linguistics dissertation series No. 15. The Hague: Holland Academic Graphics.

Sluijter, Agaath M.C. & van Heuven, Vincent J. 1996. "Spectral balance as an acoustic correlate of linguistic stress". *Journal of the Acoustical Society of America* 100.2471-2485.

Spahr, Anthony J., Litvak, Leonid M., Dorman, Michael F., Bohanan, Ashley R. & Mishra, Lakshimi N. 2008. "Simulating the effects of spread of electric excitation on musical tuning and melody identification with a cochlear implant". *Journal of Speech, Language, andHearing Research* 51.1599-1606.

Tartter, Vivien 1989. "What's in a whisper?". *Journal of the Acoustical Society of America* 86.1678–1683.

Tartter, Vivien & Braun, David 1994. "Hearing smiles and frowns in normal and whisper registers". *Journal of the Acoustical Society of America* 96.2101-2107.

Versfeld, Niek J., Daalder, Laura, Festen, Joost M. & Houtgast, Tammo 2000. "Method for the selection of sentence materials for efficient measurement of the speech reception threshold". *Journal of the Acoustical Society of America* 107.1671-1684.