

CODING OF VOICE SOURCE INFORMATION IN THE NUCLEUS COCHLEAR IMPLANT SYSTEM

P. A. JONES, MCOGSC; H. J. MCDERMOTT, PHD; P. M. SELIGMAN, PHD; J. B. MILLAR, PHD

From the Human Communication Research Centre and the Department of Otolaryngology, University of Melbourne, Melbourne (Jones, McDermott), Cochlear Pty Limited, Sydney (Seligman), and Australian National University, Canberra (Millar), Australia.

Two studies are reported in which the effectiveness of explicitly coding voicing and fundamental frequency information for the Nucleus cochlear implant was investigated. In the first study, the voicing perception of a group of three experienced Multipeak users was evaluated when they were using Multipeak and a modified Multipeak in which the explicit fundamental frequency and voicing cues were eliminated and replaced with a 250-Hz constant rate of stimulation. The results of consonant and monosyllabic word tests showed that there was no significant difference in the subjects' ability to discriminate voicing. In the second study, the ability of a group of five experienced users of the constant rate spectral maxima sound processor (SMSP) strategy to discriminate suprasegmental contrasts was evaluated when they were using the SMSP strategy and a modified SMSP strategy that included a rate-encoded representation of the fundamental frequency on the most apical stimulation channel. The results of intonation, roving stress, and question-statement tests showed that there was no significant difference between the scores recorded with these strategies. Since the temporal voicing cue is not a primary cue to voicing discrimination for Multipeak users, and the provision of an additional rate cue to the SMSP strategy does not improve SMSP users' ability to discriminate suprasegmental contrasts, the results of these studies indicate that in the cases investigated, the coding of voice source information by rate of stimulation does not significantly augment the cues present in the spatially distributed constant rate stimulation pattern.

INTRODUCTION

Voice source information is the acoustic information about the motion of the vocal folds during speech. In terms of speech perception, this voice source information contributes to the perception of the phonetic feature of voicing and to the perception of suprasegmental information that helps convey the higher-level rhythm and stress patterns of utterances and provides cues to word and syllable segmentation, speaker identity, question-statement contrasts, and the emotional state of the speaker. Since voice-source information is difficult to discern visually, the speech-coding strategies for the Nucleus implant have traditionally included an explicit encoding of this information. For example, within the Multipeak strategy,¹ variation in stimulation rate is used to code both the

fundamental frequency (F_0), by the rate at which groups of spectral components are presented in voiced speech segments, and the phonetic feature of voicing, by the presence or absence of F_0 periodicity in the stimulation pattern. The more recent spectral maxima sound processor (SMSP)² does not include an explicit coding of this information — rather, the filter bank-derived estimates of the speech signal's spectral components are presented at a constant rate of stimulation. Such a strategy successfully transmits voicing and suprasegmental information because the speech is highly redundant. In this paper the effectiveness of the explicit coding of voice source information by rate of stimulation is investigated by comparing coding strategies that do and do not maintain an explicit rate encoding of voice source information.

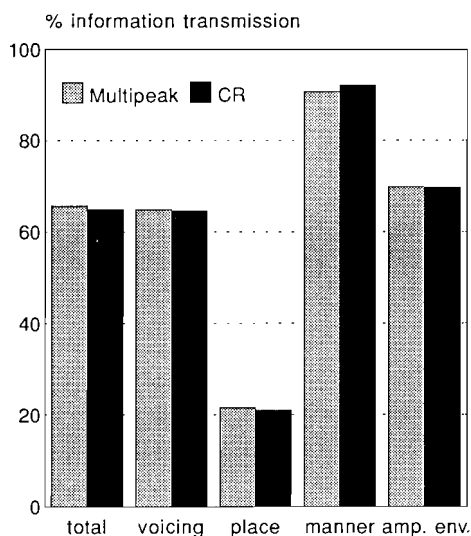


Fig 1. Group mean information transmission scores on medial consonant test for Multipeak and constant rate (CR) speech-coding strategies.

Two speech-coding evaluations were conducted. The first investigated the voicing perception of Multipeak users when the rate-encoded voice source information is removed and replaced with a constant stimulation rate. The second study, based on the finding that the standard 250-Hz carrier rate of the SMSP is insufficient to provide a full representation of F0 through the mechanism of amplitude modulation³ and the psychophysical adequacy of the rate domain to transmit F0 information,⁴ investigated the discrimination of suprasegmental contrasts by SMSP users when a full, rate-encoded representation of F0 is incorporated into this strategy.

VOICING

Method. In the first study, the voicing perception of three experienced postlingually deaf adult Multipeak users was evaluated when they were using Multipeak and when they were using a constant rate strategy (CR) created by forcing Multipeak into its unvoiced mode and including a fixed delay at the end of each stimulation frame such that an average stimulation rate of 250 Hz was achieved. Consequently, CR presents the four spectral components F5, F4/F3, F2, and F1, in which F4/F3 indicates that F4 is selected if F2 is in a high-frequency region and F3 otherwise, at a rate that varied by less than 10% around the average.

The data for this study were collected in four evaluation sessions for each subject, two with each strategy in an alternating sequence. Within each session the implantees responded to lists of /aCa/ consonants, each consisting of four repetitions of 12 consonants, and lists of monosyllabic words, each consisting of 50 unique CNC words.⁵ These lists were recorded by a male talker and presented at a signal level of 65 dB A-weighted sound pressure level (dBA). Considered across sessions and subjects, an unbalanced design was implemented, with a maximum amount of data collected in each fixed-length evaluation session. In total, the number of consonant lists presented in the Multipeak and CR conditions was 16 and 18, respectively, and 28 lists in each condition in the case of the word test.

Results. An analysis of variance of the percent correct

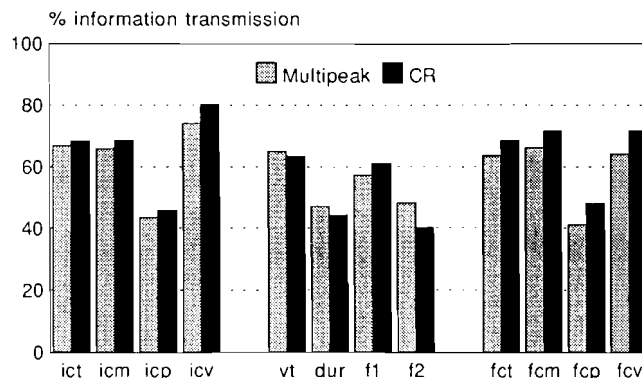


Fig 2. Group mean information transmission scores on monosyllabic word test for Multipeak and constant rate (CR) strategy. Total (t), manner (m), place (p), and voicing (v) categories are shown for initial (ic) and final (fc) position consonants. Categories total (t), duration (dur), F1 frequency (f1), and F2 frequency (f2) are shown for medial vowels.

scores for the medial consonant test ($F[1,27] = 0.34, p = .567$) and the various aspects of the word test (phonemes: $F[1,8] = 1.66, p = .233$; initial consonants: $F[1,8] = 1.66, p = .756$; and final consonants: $F[1,8] = 2.96, p = .124$) with subject and strategy as factors showed that there was no significant difference between the strategies and no significant interaction terms. The group mean information transmission analysis⁶ results for the medial consonant and the word tests are displayed in Figs 1 and 2, respectively. An analysis of variance of the individual information transmission scores for the medial consonant test (voicing: $F[1,2] = 0.3, p = .872$) and word test (initial consonant voicing: $F[1,2] = 5.52, p = .143$; final consonant voicing: $F[1,2] = 0.65, p = .504$) with the factors of strategy and subject showed that there were no significant differences between the strategies. Limiting the discussion to voicing, the results of both the consonant and word tests show that the CR stimulation pattern contains at least as much information about voicing as does the F0-based stimulation pattern. As a consequence of the multiplicity of cues to voicing that exist within the spatially distributed stimulation pattern — voice onset time and spectral weight, for example — the results suggest that explicitly coded temporal periodicity is not an essential cue to voicing discrimination for Multipeak users.

FUNDAMENTAL FREQUENCY

Method. In the second study the ability of a group of five experienced postlingually deaf adult users of the CR SMSP processor to discriminate suprasegmental contrasts was evaluated when they were using the SMSP and a modified SMSP strategy that included a rate-encoded representation of the F0 on the most apical stimulation channel. This dual-rate strategy, SMSP + F0, was implemented by using an MSP programmed with Multipeak as a source of interrupts to an SMSP such that the reception of an F0-based signal caused the immediate delivery of the most apical SMSP filter bank channel.

The data were collected in three evaluation sessions for each subject. Within each session, three speech tests were administered in each strategy condition: a novel randomization of the 50-item roving stress (RST) and question-statement (Q/S) tests from the Speech Pattern and Contrasts

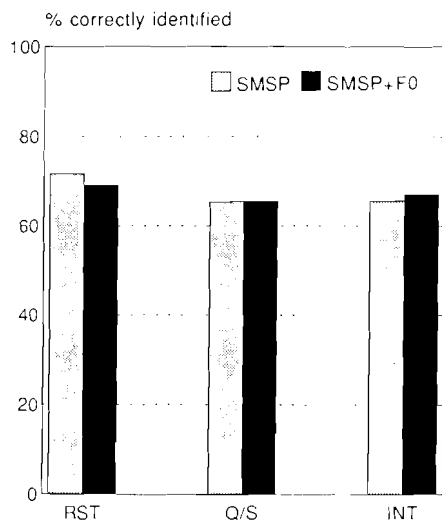


Fig 3. Group mean percent scores for roving stress test (RST), question/statement test (Q/S), and intonation test (INT) for spectral maxima sound pressure (SMSP) and SMSP + F0 speech-coding strategies.

battery⁷ and a specific test of the detection of rising and falling intonation (INT). An administration of the INT test consisted of nine repetitions of four distinct tokens of a rising, falling, or steady intonation contour imposed on a sustained /m/. All the speech material was spoken by a male speaker and presented via audiotape at a signal level that never exceeded 75 dBA. The first data collection session was preceded by an initial practice session used to familiarize the subjects with the experimental environment and procedures. The speech materials were presented in both strategy conditions, but no explicit training was undertaken. In the evaluation sessions, the subjects' experience with the SMSP + F0 strategy was limited to 5 minutes of conversation, and during evaluation sessions no feedback was given as to the correctness of the subjects' responses.

Results. The group mean percent correct scores for the RST, Q/S, and INT tests are displayed in Fig 3. The chance-level scores for these tests are 33.3%, 50%, and 33.3%, respectively. An analysis of variance of the percent correct scores for the RST test ($F[1,18] = 0.71$, $p = .410$), Q/S test ($F[1,18] = 0.01$, $p = .943$), and INT test ($F[1,73] = 0.12$, $p = .731$) with subject and strategy as factors showed that there was no significant difference between the scores with the differing strategies and no significant interaction terms. There was no learning effect evident in any of the tests. It may

be possible that more experience or a structured program of training might be necessary for the subjects to make use of this additional source of information or that the result represents a ceiling effect in F0 perception with the SMSP due to language knowledge or some more central psychophysical mechanism. However, the failure of the additional rate-encoded F0 information to increase the standard SMSP scores suggests that spectral amplitude and amplitude modulation variations are the primary carriers of stress and intonation information within the SMSP + F0 strategy and indicate that a single channel of rate-encoded F0 information is ineffective in enhancing SMSP performance in cueing stress and intonation.

CONCLUSION

Although these studies investigate voicing and F0 in different coding strategies, the common finding is that the rate domain is shown to be a secondary carrier of the information under investigation in each study. More formally, since the temporal voicing cue is not a primary cue to voicing discrimination for Multipeak users, and the provision of an additional rate cue to the SMSP strategy does not improve SMSP users' ability to discriminate suprasegmental distinctions, the results of these studies indicate that in the cases investigated, the coding of voice source information by rate of stimulation does not significantly augment the cues present within the spatially distributed CR stimulation pattern.

REFERENCES

1. Skinner MW, Holden LK, Holden TA, et al. Performance of postlinguistically deaf adults with the wearable speech processor (WSP III) and the mini speech processor (MSP) of the Nucleus multi-electrode cochlear implant. *Ear Hear* 1991;9:306-12.
2. McKay CM, McDermott HJ, Vandali AE, Clark GM. A comparison of the speech perception of cochlear implantees using the spectral maxima sound processor (SMSP) and the MSP (Multipeak) processor. *Acta Otolaryngol (Stockh)* 1992;112:752-61.
3. McKay CM, McDermott HJ, Vandali AE, Clark GM. The spectral maxima sound processor: recent findings in speech perception and psychophysics. In: Hochmair-Desoyer LJ, Hochmair ES, eds. *Advances in cochlear implants*. Vienna, Austria: Manz, 1994:141-6.
4. Tong YC, Clark GM. Absolute identification of pulse rates and electrode positions by cochlear implant patients. *J Acoust Soc Am* 1985;77:1881-8.
5. Peterson GE, Lehiste I. Revised CNC lists for auditory tests. *J Speech Hear Disord* 1962;27:62-70.
6. Miller GA, Nicely PE. An analysis of perceptual confusions among some English consonants. *J Acoust Soc Am* 1955;27:338-52.
7. Boothroyd A. Perception of speech pattern contrasts from auditory presentation of fundamental frequency. *Ear Hear* 1988;9:313-21.



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Jones, P. A.; McDermott, H. J.; Seligman, P. M.; Millar, J. B.

Title:

Coding of voice source information in the Nucleus cochlear implant system

Date:

1995

Citation:

Jones, P. A., McDermott, H. J., Seligman, P. M., & Millar, J. B. (1995). Coding of voice source information in the Nucleus cochlear implant system. *Annals of Otology, Rhinology & Laryngology*, 104(suppl.166), 363-365.

Persistent Link:

<http://hdl.handle.net/11343/27435>

File Description:

Coding of voice source information in the Nucleus cochlear implant system