ELECTRODOGRAPHIC ANALYSIS AND FIELD EVALUATION OF THE SPEAK CODING STRATEGY

N. DILLIER, PhD; H. BÖGLI; W. K. LAI, MD

From the ENT Department, University Hospital Zürich, Zürich, Switzerland. This work was partly supported by grant 32-37932.93 from the Swiss National Research Foundation.

The Speak speech-coding strategy for the Nucleus Minisystem-22 cochlear implant continuously analyzes the speech signal using 20 digitally programmable band-pass filters and presents up to 10 spectral maxima to the implanted electrodes. To analyze the performance of this system for a variety of speech sounds in quiet and noise, the stimulation patterns of the implanted electrode array were reconstructed from the transmitted radio frequency signals by software as electrodograms and compared to electrodograms generated by other speech-coding strategies, as well as to the spectrograms of the input signals. The performance with the Speak strategy relative to that with the Multipeak (Mpeak) speech-processing strategy was also evaluated in a field trial study with 20 native German-speaking cochlear implant users from four European implant centers, involving a variety of auditory perceptual tasks in an ABAB paradigm over a 12-week period. Vowel, consonant, and monosyllable word tests, as well as sentence tests in quiet and noise, were conducted. Significant differences in group mean scores for most speech recognition subtests were obtained for the Speak versus the Mpeak strategy, with the largest overall improvements observed for the sentence tests in noisy conditions.

INTRODUCTION

Cochlear implant coding and processing strategies are designed to overcome the limited channel capacity of the artificially stimulated auditory nerve by selecting the necessary and useful portions of the speech information and transforming acoustic signals into electroneural stimulation according to individual topological requirements for loudness and pitch perception, while avoiding electrode interaction and current summation effects. The stimulation pattern generated by the speech processor in the implanted array of electrodes can be visualized and analyzed by special software. The term "electrodogram" is used for the graphic displays of electrode activity over time and place and is analogous with the term "spectrogram" for a display of the distribution of energy in both frequency and time.

Two different coding strategies for the Nucleus 22-channel cochlear implant system were considered in this investigation. The multipeak (Mpeak) strategy, which was introduced in 1989 with the Nucleus miniature speech processor (MSP),[1] determines the frequency and amplitude of the first two formants, as well as the energy in two or three higher-frequency bands, and generates biphasic stimulation pulses on up to 4 out of 22 possible electrodes at a repetition rate dependent on the voice fundamental frequency. The frequency-to-electrode mapping for the second formant overlaps the mapping of the higher-frequency bands.

The Speak strategy, which is based on the spectral maxima sound processor (SMSP) strategy,[2] divides the audio spectrum into 20 frequency bands and continuously selects up to 10 prominent spectral components to produce interleaved stimulation pulses on up to 10 different electrodes within each analysis interval. The rates of stimulation on each channel vary between approximately 180 and 300 Hz, depending on the spectral composition, the sound intensity, and the individual's speech processor program. The frequency-to-electrode mapping follows the normal cochlear tonotopic order.

The Mpeak and Speak coding strategies were evaluated in a multisite field study involving eight cochlear implant centers in English-speaking countries[3] and eight centers in non–English-speaking countries. The results for 20 subjects from four German-speaking cochlear implant centers are summarized below (full details to be published by Dillier et al[4]).

METHODS

The electrodographic analysis was performed with speech processors programmed with two different test maps. The medium-level map contained threshold (T) and comfortable (C) levels at 100 and 130 stimulus level (SL) units, respectively, while the low-level map used T and C levels of 30 and 60 SL, respectively. Digitally recorded or synthesized speech was presented via a loudspeaker to the Nucleus headset microphone and processed by the Mpeak or Speak coding strategy. The digital transmission coil signal (pulse coded at 2.5 MHz) was picked up by the antenna of the Nucleus dual processor interface (DPI), converted to numeric parameters, and captured by the visualization and analysis software, which then generated two- and three-dimensional electrodograms, allowing quantitative measurements of average pulse rates and stimulus level distributions for selected segments of the stimulus patterns. Synthetic syllables /ba/, /da/, and /ga/ were generated with the Klatt synthesizer program. The steady state frequencies of the five formants were 720, 1,240, 2,500, 3,600, and 4,500 Hz. Formants 4 and 5 were kept constant over the whole stimulus duration. All formant transitions occurred during the first 40 milliseconds. The starting frequency of the first formant was 200 Hz. Starting frequencies for the second and third formants were 942 and 1,904 Hz for /ba/, 1,700 and 2,791 Hz for /da/, and 1,650 and 1,904 Hz for /ga/, respectively. The fundamental frequency for all stimuli started at 103 Hz, rose linearly to 125 Hz during the formant transitions, and dropped to 50 Hz at the end of the syllable.[5]

Twenty adult, postlingually deaf, German-speaking users of the Nucleus cochlear implant volunteered to participate in this study, which was conducted at four implant centers located in the cities of Aachen, Hannover, Kiel, and Zürich according to the same experimental protocol. All participants had at least 16 active electrode channels and 9 months or more

AVERAGE OVERALL STIMULATION RATES (PULSES PER SECOND) WITH MPEAK AND SPEAK

Segment	*Level*	*Mpeak*	*Speak*
Consonant	Medium	909	1,917
	Low	931	2,389
Vowel	Medium	450	1,581
	Low	470	1,956

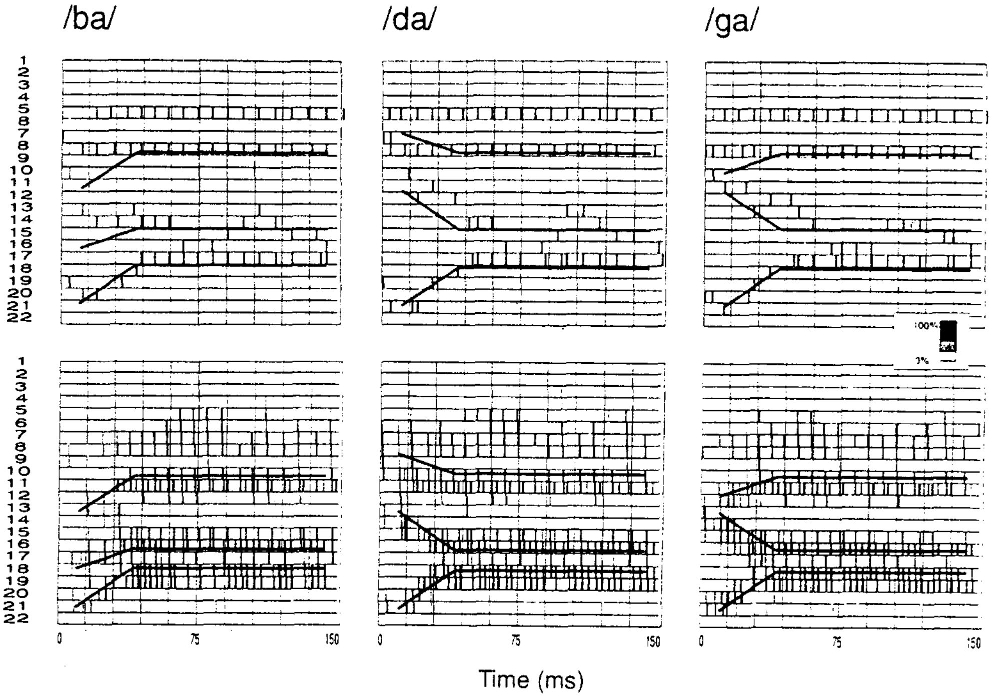

Fig 1. Electrodograms for three synthetic syllables /ba/ (left), /da/ (middle), and /ga/ (right). Upper row, Mpeak; lower row, Speak.

of experience with the Mpeak processing strategy on a regular (12 hours daily) basis, and had reached a monosyllabic word recognition score with their MSP of 10% or above.

Vowel, consonant, and monosyllable word tests, as well as two different sentence tests (Innsbrucker[6] and Göttinger[7] sentence lists) in quiet and noise, were presented via loudspeaker from a digital audiotape or directly from a computer disk in a sound-treated room. The sequence of test lists to be used for each subject was fixed before the start of the experiment. The electronic circuitry for the new encoder was embedded in a body-worn processor case identical to the Nucleus MSP. Baseline levels of auditory performance were established with the Mpeak speech-processing strategy of the MSP. All subjects received both the baseline (Mpeak) and experimental (Speak) conditions twice, in an ABAB paradigm, over a 12-week period. Questionnaires were used to assess musical quality, effects on tinnitus, and subjective performance rating.

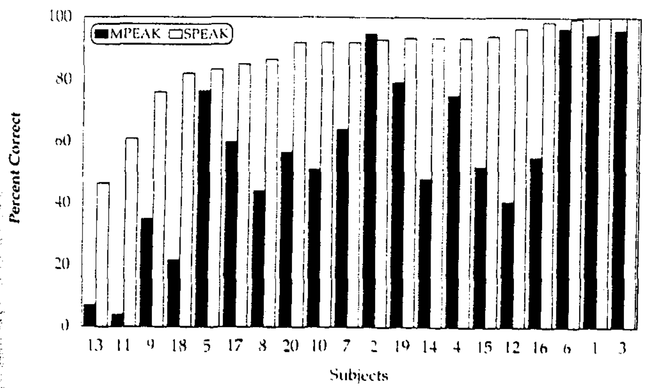

Fig 2. Speech intelligibility scores of 20 subjects with Mpeak and Speak for Innsbrucker sentence test in noise (at 10-dB signal-to-noise ratio). Data are sorted according to results with Speak. Numbers on x-axis represent subject identifications.

RESULTS AND DISCUSSION

The Table shows average overall stimulation rates for the consonant and vowel portions of the naturally spoken German word /scheuen/ (male voice with a fundamental frequency of approximately 120 Hz). The rates for the Mpeak strategy were independent of the T and C levels, but varied by as much as 100% between vowel and consonant portions. The average rates for the Speak strategy, on the other hand, were increased by 20% to 25% for the low-level map in comparison to the medium-level map, but did not differ as much between vowel and consonant portions as the Mpeak rates. The rate difference for the two maps is due to the coding of stimulus level by pulse width in the Nucleus implant system. Note that the measured overall rates include stimuli on all electrodes.

Figure 1 shows the electrodograms for the synthetic syllables /ba/, /da/, and /ga/ as produced by the Mpeak and Speak coding strategies, respectively, overlaid by manually fitted formant trajectories for the first three formants. The electrodograms reveal the difference in average rate for the two strategies, as well as the poor fit of the Mpeak pattern to the transitions of the second and third formants. Because of the overlapping frequency mapping for the second formant and higher-frequency components and the limited number of pulses per pitch period in the Mpeak strategy, separation of the fourth and third formants is not possible. The Speak strategy, on the other hand, is able to reproduce the first three formant transitions quite accurately and also provides information about higher-frequency components.

Figure 2 shows the Innsbrucker sentence test results for the 10-dB signal-to-noise ratio condition for all 20 subjects. Although the intersubject variation is large, statistical analysis of group mean data showed marked improvement of speech recognition scores with the Speak strategy compared to the Mpeak strategy. The largest differences were observed for the Innsbrucker sentence test (36%, 30%, and 28% mean difference for a signal-to-noise ratio of 15, 10, and 5 dB, respectively, and 11% for the quiet condition). Mean differences above 10% were also observed for the Göttinger sentences presented in noise (16%, 16%, and 27% for a signal-to-noise ratio of 15, 10, and 5 dB, respectively), whereas the difference for speech without noise was 8%. All differences between the Mpeak and Speak conditions were significant at a level of .05 or better (paired samples *t* test), with the exception of the vowel test (VO8).

Considerable individual improvements in speech perception were noted for most subjects on at least one measure when using the Speak strategy. Differences between the four centers were observed that were mainly related to geographic speaker and listener variables. Additional tests and the analysis of the performance questionnaires revealed generally more pleasant listening to music, improved identification of musical instruments, and mainly very positive qualitative judgments of sound quality with the Speak coding strategy as compared to the Mpeak strategy.

ACKNOWLEDGMENTS — The field trial study results were contributed by Drs R. D. Battmer, W. H. Döring, and J. Müller-Deile. The support of many staff members of Cochlear AG, Basel, is also greatly acknowledged.

REFERENCES

1. Patrick JF, Seligman PM, Money DK, Kuzma JA. Engineering. In: Clark GM, Tong YT, Patrick JF, eds. Cochlear prostheses. Edinburgh, Scotland: Churchill Livingstone, 1990:99-124.

2. McDermott HJ, McKay C, Vandali AE. A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant. J Acoust Soc Am 1992;91:3367-71.

3. Skinner MW, Clark GM, Whitford LA, et al. Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system. Am J Otol 1994;15(suppl 2):15-27.

4. Dillier N, Battmer RD, Döring WH, Müller-Deile J. Multicentric field evaluation of a new speech coding strategy for cochlear implants. Audiology 1995;34 (in press).

5. Raz I, Noffsinger D. Identification of synthetic, voiced stop-consonants by hearing-impaired listeners. Audiology 1985;24:437-48.

6. Hochmair-Desoyer IJ, Hochmair ES, Fischer RE, Burian K. Cochlear prostheses in use: recent speech comprehension results. Arch Otolaryngol 1980;229:81-98.

7. Wesselkamp M, Kliem K, Kollmeier B. Erstellung eines optimierten Satztests in deutscher Sprache. In: Kollmeier B, ed. Moderne Verfahren der Sprachaudiometrie. Heidelberg, Germany: Median-Verlag von Killisch-Horn GmbH, 1992:330-43.

CHARACTERISTIC FREQUENCY MAPPING IN SUBJECTS USING THE NUCLEUS 22-CHANNEL COCHLEAR IMPLANT SYSTEM WITH PARTIAL AND FULL INSERTION

J. A. EYLES, BSC(HONS), MSC; P. J. BOYLE, BSC, MSC; M. J. BURTON, MA, FACS

From the University College London Hospitals Trust Cochlear Implant Programme, Middlesex Hospital Outpatients Department, London, England.

INTRODUCTION

In this study we compared the performance of the same subjects using two alternative speech-coding strategies: the default frequency boundaries for the multipeak speech-coding strategy, and the characteristic frequency (charf) mapping technique. The latter modifies the frequency allocation on the basis of characteristic frequencies within the cochlea. Whitford et al[1] have shown that a group of 4 patients using the charf mapping showed significant improvement in open-set sentence recognition tests in relatively high levels of background noise. In that study patients alternated between charf and standard mapping on a weekly or fortnightly basis. This short time interval may not have allowed the patients to become fully adjusted to the new map before they had to switch again. In the present study an ABAB design is being used in which the patients are using each strategy for 3 months. The data have been collected for the first charf strategy. The results are presented here.

OUTLINE OF CHARACTERISTIC FREQUENCY MAPPING

Postoperative radiographs were used to determine the insertion depth of the electrode array and the location of the individual electrodes. The modified Stenvers view technique was used.[2] A vertical line drawn from the top of the superior semicircular canal through the center of the vestibule has been shown to transect the round window membrane. Electrodes and stiffening bands posterior to this line were considered extracochlear, and those anterior, intracochlear.

Using the charf mapping program provided by Cochlear Pty Limited, we specified the number of extracochlear stiffening bands and electrodes. The program then determined the frequency allocation for each electrode independent of the insertion depth from the round window, according to the charf of points along the basilar membrane in the normal cochlea.[3,4] The frequency boundary of the electrode that stimulated the

STATISTICAL RESULTS FOR GROUP OF SUBJECTS UNDER FOUR TESTING CONDITIONS

Test	No. of Subjects	Wilcoxon Matched Pairs Signed Ranks Test Result: 1-mo	3-mo
BKB sentences, audition with lipreading	10	.037*	
	11		.04*
BKB sentences, audition only	5	.89	
	6		.46
CNC words, audition with lipreading	9	.36	
	10		.95
CNC words, audition only	6	.23	
	7		.17
CNC phonemes, audition with lipreading	9	.68	
	10		.33
CNC phonemes, audition only	6	.75	
	7		.128

*Significant at 95% level.