Eighth Quarterly Progress Report

X
6

May 1 through July 31, 1994

NIH Contract N01-DC-2-2401

Speech Processors for Auditory Prostheses

Prepared by

Dewey T. Lawson, Blake S. Wilson and Mariangeli Zerbi

Center for Auditory Prosthesis Research
Research Triangle Institute
Research Triangle Park, NC 27709

CONTENTS

I. Introduction 3

II. Further Studies of Complex Tone Perception by Implant Patients 4

III. Glossary of Musical Terms. 43

IV. References 44

V. Appendix I. 45

VI. Appendix II 50

VII. Plans for the Next Quarter 55

VIII. Acknowledgements 56

Appendix A: Summary of Reporting Activity for this Quarter 57

Appendix B: Notice of errata to Fourth Quarterly Progress Report 60

I. Introduction

The purpose of this project is to design and evaluate speech processors for implantable auditory prostheses. Ideally, the processors will extract (or preserve) from speech those parameters that are essential for intelligibility and then appropriately represent those parameters for electrical stimulation of the auditory nerve or central auditory structures. Work in the present quarter included the following:

1. Studies with Ineraid subjects SR2 and SR14. The studies for both subjects included measures of intracochlear evoked potentials for a variety of stimuli. The studies with SR14 also included (a) initial fitting of a *continuous interleaved sampling* (CIS) processor, (b) measures of speech reception with CIS processors using different pulse durations and rates, and (c) measures of speech reception with single-channel processors, for comparison with evoked potential and psychophysical results obtained at the University of Iowa (by Carolyn Brown and Paul Abbas) and in our laboratory for single electrodes with SR14's implant. The studies with SR2 also included additional measures of complex tone perception with a CIS processor.
2. Initial studies with the first patient in the Nucleus percutaneous series, NP-1. Studies included evaluations of CIS and *spectral peak* (SPEAK) processing strategies.
3. Hosting a site visit for Terry Hambrecht and Bill Heetderks, to review and discuss project activities and plans (site visit held at RTI and Duke University Medical Center, July 14 and 15).
4. Presentation of project results in an invited lecture at the *127th Meeting of the Acoustical Society of America* (Cambridge, MA, June 8).
5. Continued preparation of manuscripts for publication, including an invited paper for *Advances in Otolaryngology -- Head & Neck Surgery*, on "Advances in coding strategies for cochlear implants."

In this report we present results from further studies of complex tone perception with cochlear implants. Results from the additional studies indicated in points 1 and 2 above will be presented in future reports.

II. Further Studies of Complex Tone Perception by Implant Patients

Background

Preliminary pilot studies of multichannel processing of complex tones were described in a previous Quarterly Progress Report for this project [Lawson, *et al.*, 1993]. A variety of digitally synthesized complex tone stimuli were input to a continuous interleaved sampling (CIS) processor which in turn stimulated a research subject's intracochlear electrode array. The data collected included anecdotal descriptions of the percepts elicited by individual stimuli, anecdotal descriptions of the perceived differences between members of various pairs of complex stimuli, and surveys of relative overall pitch judgments within such complex stimulus pairs.

Each **stimulus** was approximately 0.5 seconds in duration (22,000 samples at 44.1 kHz), including approximately 11 msec each of linear fade-in and fade-out (500 samples each). Each stimulus was presented from a digitally synthesized file of 16-bit samples, constructed by adding pure tone sinusoidal partials selected from the harmonics either of a single fundamental or of two fundamentals differing by a chosen musical interval. Single fundamentals were chosen from a four octave *equal tempered* scale ascending from 110 Hz. When two fundamentals were to be separated by a consonant pitch interval within a single stimulus, however, the frequency interval was made exact (*i.e. just intonation* was used within stimuli). Consistent with the spectral envelopes of typical musical tones, relative nth harmonic amplitudes proportional to 1/n were chosen to ensure relatively strong beat phenomena. [See Rossing, 1990]. A glossary of musical terms used in this report may be found on page 43.

For simplicity of analysis, we also required that each partial used in a stimulus fulfill an additional criterion with respect to the CIS processor for which it was intended. The criterion ensured that each partial be represented exclusively in a single processing channel -- falling at a frequency that put it within 1 dB of the maximum sensitivity of that channel's input passband, for instance, and at least 10 dB down in any adjacent band. In some cases a minimum of 20 dB of adjacent band rejection was imposed, further reducing the number of available partials.

Within those constraints, we designed combinations of partials to test the efficacy and relative salience of various potential mechanisms for conveying subtleties of perceived pitch and timbre to cochlear implant patients. Examples of such mechanisms included harmonic consistency of partials between and within channels and beat rate patterns between and within channels. Some stimuli were designed with conflicting cues to assess their relative salience.

The research **subject** chosen for the pilot studies was Ineraid patient SR2. He was selected on the basis of (1) excellent performance with existing processor designs, already extensively studied in our laboratory and elsewhere, (2) exceptional analytic and descriptive abilities regarding his auditory percepts, (3) experience as a musician -- both before losing his normal hearing and recently with an analog clinical prosthesis, and (4) familiarity with some basic music theory. Percutaneous access is available to all six of the electrodes implanted in SR2's right cochlea. He is right handed.

The CIS **processors** used in the preliminary studies included a standard 6-channel design [number 163b] that had been used by the subject for a wide range of previous studies in our laboratory, an 11-channel virtual channel interleaved sampling (VCIS) processor [200b] also previously evaluated in our lab, and a six channel CIS variation [284] without the normal preemphasis. The parameters for these processors are included in a table accompanying the discussion below of an additional processor used in more recent complex tone studies. [For a general description of the design of CIS processors see Wilson, *et al.*, 1991. VCIS designs are described in Wilson, *et al.*, 1994.]

The **experimental conditions** explored in the preliminary pilot studies included presentation of the two complex tones of a stimulus pair with and without a one second intervening delay. As the preemphasis filter typically included in CIS and VCIS processors effectively contributes a spectral weighting proportional to harmonic number over part of the represented frequency range (attenuating components below 1.2 kHz at 6 dB/octave), a condition essentially correcting for this effect was included among the pilot studies. The effect of order of presentation within each stimulus pair also was explored. There was no balancing of the overall loudness across stimuli, within or among pairs.

Preliminary **results** from those pilot studies included indications of the importance of channel (place) cues, the importance of intrachannel beat frequencies as (temporal) cues, and the predictability of relative strengths among competing spectral cues in some cases. We noted two distinct patterns of changes in percept that occurred after extended initial comparisons of certain pairs of stimuli: (1) an irreversible change, after which the original percept could not again be found by the subject, and (2) the sudden emergence of an ambiguity, with the subject thereafter able to obtain either percept at will. Examples of complex tone stimuli were found for which the use of preemphasis filtering, the choice of pair presentation order, and the imposition of an interstimulus delay would, individually or in combination, dramatically affect the subject's percepts We observed, on occasion, a surprising ability of the subject to recognize musical intervals and accurately to match his (unmonitored) voice pitch to that of an electrically conveyed complex tone.

Summary of New Studies

The previously reported complex tone studies were carried out in June, 1993. Subject SR2 since has returned to our laboratory for additional speech processor studies on four occasions: in December of 1993 and in March, May, and August of 1994. The present report will describe further complex tone investigations carried out as time permitted during those visits. We shall discuss six distinct new studies, as outlined on the next page. Two of the studies represent systematic explorations of perceptual categorizations that the subject volunteered anecdotally during the original pilot studies. A third was designed to exploit and further explore some of the subtle abilities demonstrated by the subject in the earlier work. The remaining studies were designed to probe the limits of specific previously observed abilities and effects.

Outline of Studies Described in this Report:

- ***Identification of Constituent Tones.*** The subject was asked whether or not a candidate tone was made up entirely of partials contained in a more complex reference tone. The data were analyzed both (1) in terms of how accurately the subject's responses could be predicted using various hypotheses involving potential cues, and (2) in terms of how various structural attributes of stimuli affected the subject's accuracy. A description of this study begins on page 13.
- ***Processor Bandpass Filter Order Effects.*** This study was a test of the sensitivity of our complex tone data to traces of spectral components in processor channels adjacent to the ones for which they were intended. Constituent tones identification tests were repeated with a CIS processor based on 24th order, rather than 12th order, bandpass filters. A description begins on page 18.
- ***Interval Consonance Judgments in a Nontraditional Context.*** A particular nontraditional musical scale supplies many of the same structural cues present in traditional consonant intervals, but sounds quite different to people with normal hearing. This study investigated whether a subject using a CIS processor could detect a difference. It collected anecdotal descriptions of both sequential and simultaneous complex intervals and is described beginning on page 21.
- ***Systematic Examination of Perceptual Category Assignments: Single Stimuli.*** A number of descriptive terms volunteered frequently by the subject during earlier complex tone studies formed the basis for this more structured, automated interview regarding a wide range of stimulus tones. Analysis of these data included association of descriptive categories both with stimulus structures and with the use of other such categories. A description begins on page 25.
- ***Systematic Examination of Perceptual Category Assignments: Stimulus Pairs.*** The same tones described in the single stimulus interviews later were presented in pairs in another automated interview. Comparison judgments along eight different perceptual dimensions were obtained. While their analysis is not yet complete, the study and preliminary findings are described beginning on page 31.
- ***Inconsistency Detection Thresholds.*** Complex tones typically consist of several harmonics of a common fundamental. In such cases there is a high degree of consistency among the partials, each one being an integer multiple of the fundamental frequency and any pairs of adjacent harmonics beating at that same frequency. This study was designed to determine roughly how large an inconsistency must be in order to alter a complex tone percept for a CIS processor user. The stimulus pairs perceptual category interview was repeated for pairs of complex tones with gradually increasing inconsistencies in beat rates and/or absolute frequencies. A description of the study begins on page 35.

Most of the **stimuli** used in all these studies were like those used earlier. One study additionally required the synthesis of partials based on fundamentals separated by an arbitrary frequency interval. In another, stimuli were restricted to odd harmonics only and to fundamentals along a highly unusual musical scale. The basic approaches to stimulus design and synthesis, however, have remained the same throughout all the studies to date.

While our complex tone studies continue to be restricted to work with a single **subject**, knowledge gained from other types of investigations with him have added whole new dimensions to the unique set of advantages he offers such pilot research. In psychophysical experiments we have obtained periodicity pitch saturation data for the same pulse configuration used in SR2's processors (33 μs per phase balanced biphasic). Direct intracochlear evoked potential (EP) measurements for similar stimuli have indicated an accompanying onset of a failure of EP magnitudes to accurately represent each pulse's amplitude within a stimulus pulse train. We have demonstrated that an ensemble model of electrical neural stimulation can accurately predict SR2's EP responses to a wide range of stimulus patterns like those produced by his processors. Finally, SR2 recently has begun use of a six channel, 40 μs/phase CIS strategy on an everyday basis as part of a study by Eddington, *et al.*, at Massachusetts Eye and Ear Infirmary.

Relevant results of the psychophysical study are summarized in Fig. 1. 200 ms trains of 33 μs/phase pulses at various rates were presented to the subject's electrode 3 (numbered from the apical end of the array). Amplitudes were adjusted for most comfortable loudness at each rate and then for constant loudness across all stimuli. The subject was asked to nominate a pitch for each stimulus (on a scale of 0 to 100) and the results are displayed here for 30 randomized presentations at each rate. The dependence of perceived pitch on pulse rate decreases markedly above a rate of 400 pps and disappears somewhere between 800 and 1600 pps. [Wilson, *et al.*, 1994a.]

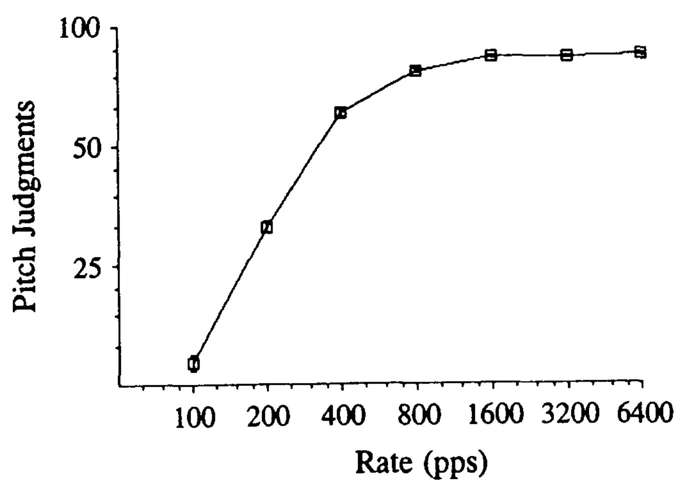

Figure 1. Saturation of Perceived Pitch with Increasing Pulse Rate

Similar 200 ms trains of constant amplitude 16.4 μs/phase pulses were presented to the subject's same electrode while intracochlear EPs were recorded differentially between the unstimulated adjacent electrode 4 and an external electrode on the scalp over the ipsilateral mastoid. We found that during these 200 ms tones the measured EPs for successive pulses would typically share a constant amplitude for rates of up to 400 pps. At higher pulse rates, however, there was an increasing tendency of alternate pulses to elicit EPs with significantly different alternating amplitudes, *i.e.* an apparent decreasing ability of the eighth cranial nerve to convey the relative amplitudes of each succeeding pulse. [See Fig 11 in Wilson, *et al.*, 1994a.]

In a further experiment directly relevant to the representation of our complex tones by CIS processors, SR2's electrode 3 was stimulated with continuous carrier trains of 1000 pps 33 μs/phase pulses whose amplitudes were sinusoidally modulated at various frequencies. The depth of modulation was 100%. Fig. 2 displays 20 ms segments of these recorded EPs for six modulation frequencies ranging from 50 to 400 Hz. For modulation rates at and below 200 Hz, SR2 described his percept as "smooth and tonal", while 300 Hz modulation of the 1000 pps carrier was perceived as "rough and complex" and 400 Hz modulation as having two separate tones. (In order to avoid overlap among EPs, the pulse carrier rate in this experiment was lower than either of those used in our complex tone study processors. Deconvolution techniques may allow measurements for significantly higher pulse carrier rates in the near future.)

The relationships among (1) pulse rate for each processor channel, (2) maximum amplitude modulation rate within each channel, (3) the patterns of stimulation evoked on the eighth nerve, and (4) the nature of the percept produced have obvious relevance to our complex tone studies. An ensemble neural model that could accurately predict EPs for a wide range of such conditions would be of enormous value in the design and interpretation of complex tone perception experiments. Such a model, under development in our laboratory under a separate NIH project, has achieved the necessary level of performance. In Fig. 3 we have plotted EP magnitudes from the data of Fig. 2 (open squares, connected by lines) along with predictions of the model (filled squares). Correlation coefficients over 50 and 200 ms intervals are shown to the right of each 50 ms plot. [Further improvements in the agreement between model and observed EPs can be expected. The benefit of adding membrane noise to the model already has been shown. See Wilson, *et al.*, 1994a.]

Clearly, for any given carrier pulse rate there will be a maximum modulation rate that can be represented accurately, and a corresponding limitation on the processor channels' temporal envelopes in response to a steady state complex tone. The potential use of intrachannel beat rates as complex tone spectral cues, for instance, will be circumscribed by such limitations. The maximum steady state modulation rate in a processor channel that could result from one of our complex tone stimuli would be beating between two partials at the extreme frequencies satisfying our exclusivity criteria for that channel. For a typical six channel CIS processor with bandpass filters of 12th or higher order, this means approximately the following maximum beat rates:

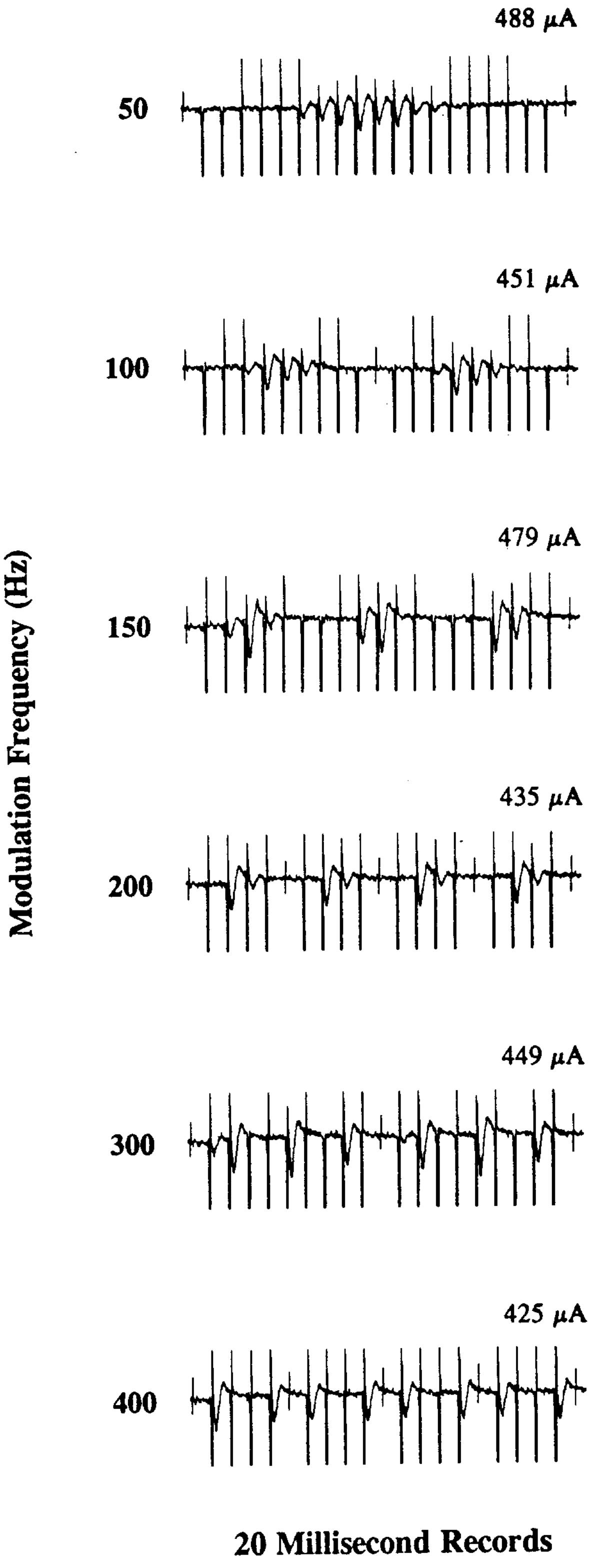

Figure 2. Intracochlear Evoked Potentials: 33 μs/phase, 1000 pps Carrier

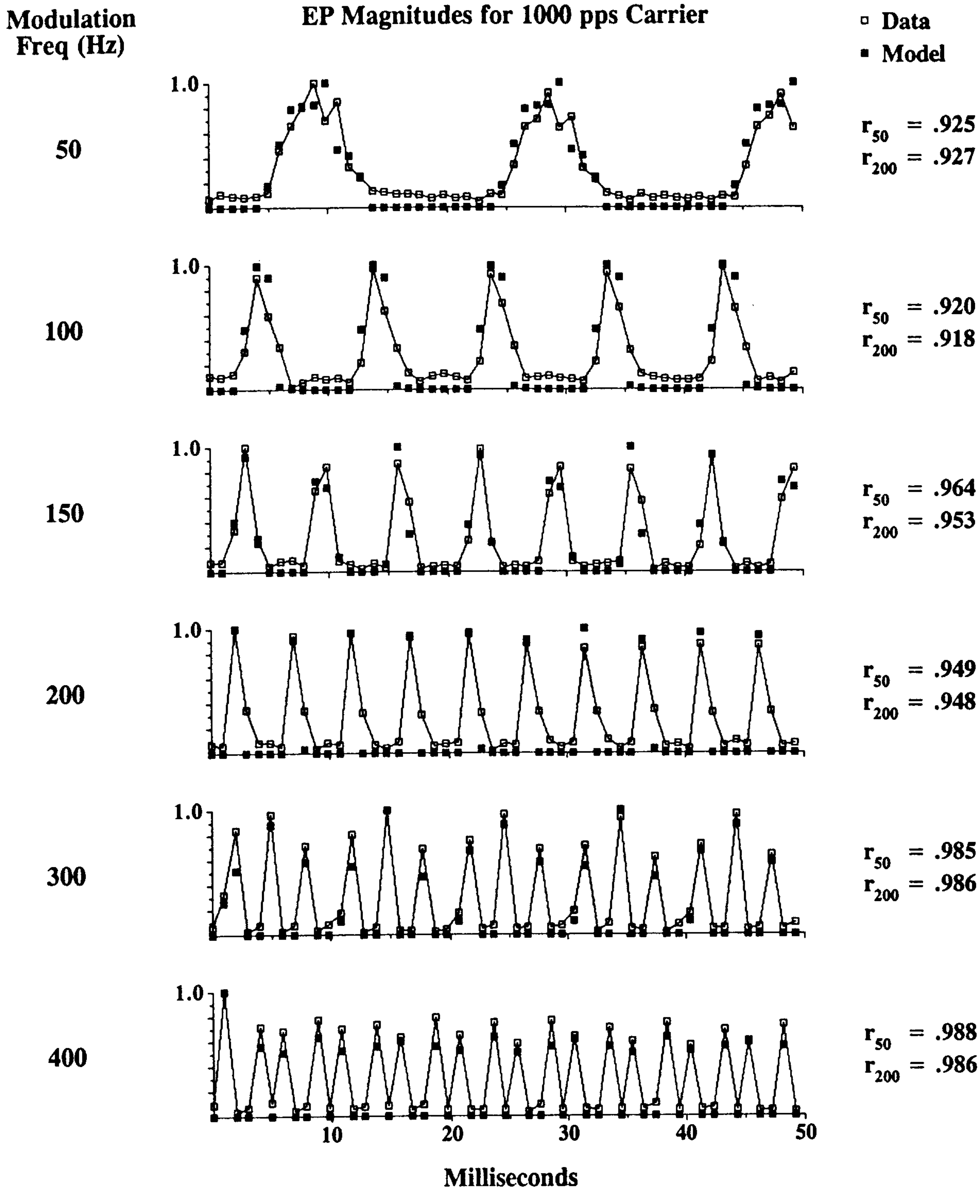

Figure 3. Comparison of Neural Ensemble Model Results to EP Data

Channel	Maximum Beat Rate
1	225 Hz
2	250
3	400
4	640
5	980
6	1800

Note that, in the absence of any low pass filtering of channel envelopes, the highest beat rates in channels 1, 2, and 3 would lie within the linear range of SR2's perceived pitch *vs.* pulse rate data, that the highest beat rates in channel 4 would extend into the reduced sensitivity range, and that channels 5 and 6 could receive beat rates above the subject's pitch saturation rate for such pulses. Note also that the use of a fourth order 400 Hz low pass smoothing filter on the envelope of each channel (as is the case in typical CIS processors, including the ones in our complex tone studies) should substantially reduce the incidence of modulations that are badly distorted on the eighth nerve.

Modeling studies and/or EP measurements for typical CIS rates (2525 pps for SR2's best six channel processors) may lead to a more optimal choice of envelope smoothing filter. In Fig. 4 we show the results of such a modeling study, including the amount of membrane noise that has been found to optimize agreement with measured EPs for lower pulse rate stimuli. Note that these predictions for both 400 and 600 Hz modulation frequencies show strong EP amplitude variations at 100 Hz -- the beat frequency with respect to a 500 Hz subharmonic of the carrier rate in both cases . Use of even higher carrier pulse rates (and correspondingly shorter pulse durations) may substantially improve the representation of useful modulation frequencies on SR2's auditory nerve. Also, reduction in the cutoff frequency of the envelope smoothing filter might reduce or eliminate the distortion shown in the 400 Hz modulation panel of Fig. 4.

Corresponding intracochlear EP data may soon be available for higher pulse rates. We note in passing that the EP data already displayed indicate that the 11 msec onset and offset ramp transients of our complex tone stimuli will not approach any modulation speed limit for our studies' pulse carrier rates.

One additional **processor** was used in the course of our recent complex tone studies. Otherwise identical to processor number 163b, number 355 incorporated 24th rather than 12th order bandpass filters. Parameters varying among the four processors are indicated in the following table; principal parameters common to all four included balanced biphasic pulses 33 μs per phase in duration, full wave rectification and 400 Hz fourth order smoothing on each channel's envelope, and a staggered order for stimulating the channels (*e.g.* 6-3-5-2-4-1).

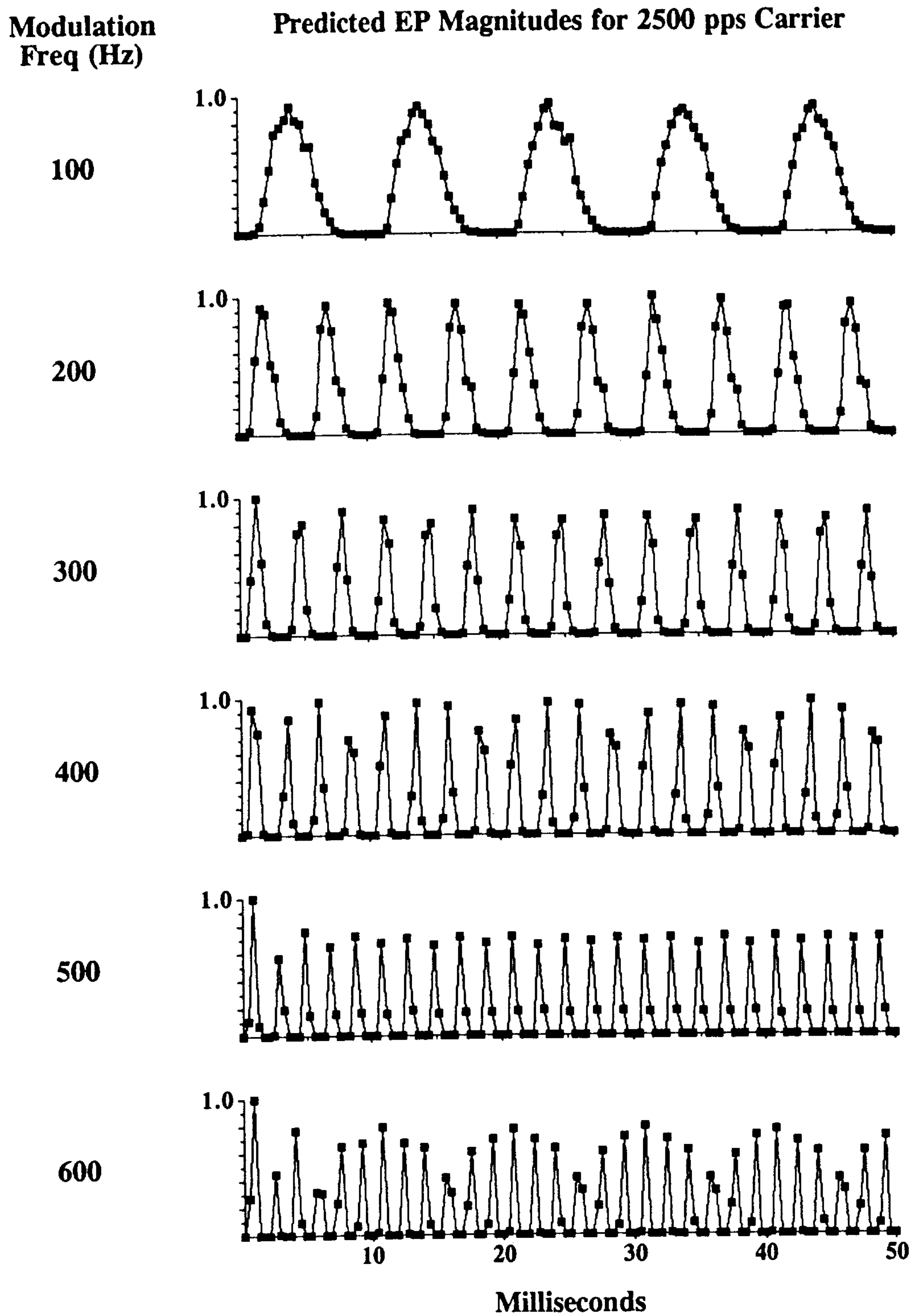

Figure 4. Neural Ensemble Model EP Predictions for a Higher Carrier Rate

Processor	Channels	BP filters	Pulse Rate	Preemphasis?
163b	6	12th order	2525 pps	yes
284	6	12th	2525	**no**
355	6	**24th**	2525	yes
200b	**11**	12th	**1364**	yes

The **experimental conditions** in the more recent studies were a subset of those used earlier. A one second delay was imposed between the two tones of every stimulus pair and a minimum delay of one second was required between any two stimulus pairs. The processors used in the recent work always included the usual preemphasis.

We turn now to a description of the individual new studies, in the order in which they were conducted.

Identification of Constituent Tones

A conclusion from the original pilot studies was that subject SR2 was capable of subtle enough distinctions among complex tone percepts for us to undertake a constituent tone identification study. An interactive computer program was written to administer such a study, organized into panels each of which contained one **reference stimulus** and five **candidate stimuli**. The subject's task was to identify each of the candidates that was a constituent of the more complex reference. Fig 5 shows the display window that, along with a computer mouse, was used by the subject to control the presentation of stimuli and to record his responses.

In the figure, the window is shown just as it appeared when beginning consideration of a new panel of stimuli. Whenever the subject clicked his mouse on the large **Play Reference** "button" a digital recording of the reference stimulus would be played into his 163b processor, subject only to a minimum delay of one second between any two playback operations. The smaller **Play 1** button was also available at this initial stage for playing the first candidate stimulus on demand, again subject to a minimum one second delay between stimuli. The buttons for playing the other four candidate stimuli, their labels dimmed, were not yet available. After listening to the 1st candidate and the reference, as many times as desired and in any order, the subject was asked to click the mouse on one of the two "radio buttons" to the right, indicating whether the first candidate tone sounded as though it was included in the reference tone. The only clarifying instruction available to the subject was *'If there is anything in the candidate tone that is not in the reference tone, answer "No".'* Early in the study SR2 volunteered that the task "was like picking out individual notes in a remembered chord." As soon as a response was given for the first candidate, all labels for that candidate were dimmed and the corresponding buttons disabled, while the labels for the second candidate were intensified and those corresponding buttons made functional. This sequence was repeated, with the subject comparing only a single

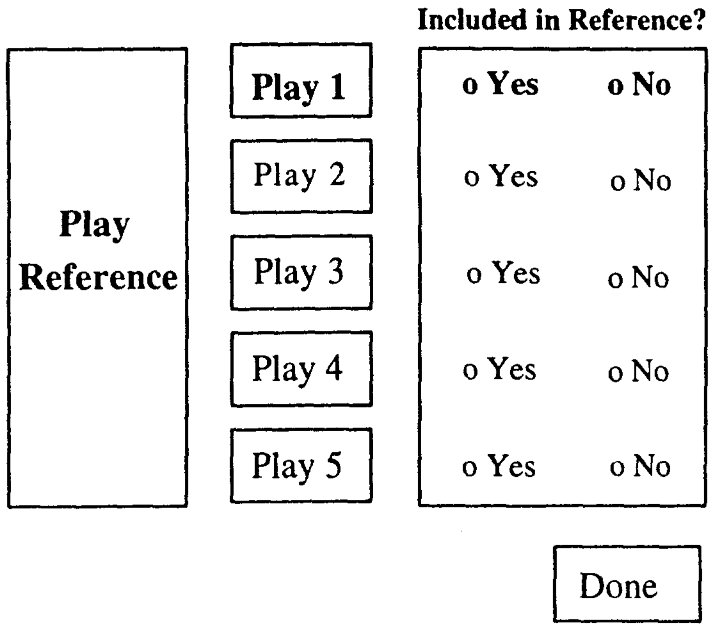

Figure 5. Subject's Control Window for Constituent Tones Identification Test

successive candidate to the reference, until a response was given for the last of the panel, the fifth candidate. At that point all the labels were intensified and all the buttons enabled, along with the **Done** button. The subject was then free to listen to any of the candidates and the reference, any number of times and in any order, and revise any of his previous judgments using the same radio buttons. When he was satisfied with the final result he could click on the Done button, restoring the window to the condition shown in Fig. 5 and bringing up a new panel of stimuli. The supervising program recorded the number of times each stimulus was played at each juncture, and the initial and final choices made for each candidate. An investigator adjusted the playback master gain as necessary to keep each reference tone loudness close to but no more than MCL. A total of 139 panels were administered, divided into four separate sessions over a nine day period.

The reference and candidate stimuli were chosen to evaluate the perceptual effects of a wide range of potential cues of pitch and timbre. A table summarizing the major characteristics of each panel's stimuli is presented in Appendix I. Those characteristics were the fundamental on which each stimulus was based and lists of the included harmonics of each and of the processor channels in which they would be represented.

Using the information contained in Appendix I it is a straightforward task to produce attribute matrices indicating which candidate tones share which characteristics with their respective reference tones. When a given attribute of a particular candidate tone is exactly matched within the corresponding reference tone, the matrix entry for that comparison is set to a binary 1. Examples of such matrices are included as Appendix II.

One instructive way to use these matrices is in comparisons of a subject's identifications of constituent tones with predictions based on the assumption of effective use of particular attributes as cues. Given the hypothesis that a particular attribute or set of attributes underlies the subject's identifications, one need only compute the product of the corresponding matrix entries for each candidate tone (*i.e.* perform logical ANDs) and compare the results with the subject's response set. The results can be divided into percent correct predictions (the sum of predicted constituent identifications that were in fact made by the subject and predicted rejections that matched the subject's judgments), percent errors in which predicted constituent identifications were not made by the subject, and percent errors in which predicted rejections were identified as constituents by the subject. Since the subject's responses were binary, prediction accuracies must differ significantly from 50% in order to support any conclusion. Such comparisons can be made for various relevant subsets of test panels, to search for evidence of reliance on different combinations of attributes as cues in different situations. It is also possible to compare the subject's initial responses (based on ordered consideration of one candidate at a time) and final judgments (after opportunity to cross-compare all stimuli within a panel and alter any of the initial responses).

As examples of these techniques, we shall discuss several subsets of our 139 constituent identification test panels. We begin with a set of panels for which each candidate is a single pair of adjacent harmonics, sometimes represented in a common channel and sometimes in separate, adjacent channels. The reference stimuli vary in complexity from two to five harmonic partials. This set includes 41 panels of stimuli [numbers 17-32, 67-82, and 107-115. See Appendices I and II for stimulus attributes in detail], so a total of 205 responses are represented. In nine of the panels the fundamental frequencies on which some of the candidate and reference tones' harmonics are based differ by one semitone ($G^{\#}$-208 Hz); in all other cases the fundamental frequency is a constant G-196 Hz across stimuli and panels. For this number of responses, a 6% deviation from chance (*i.e.* from 50%) is significant at a confidence level of 95%.

Fig. 6 displays statistical prediction results for a wide range of hypotheses about what attributes might influence SR2's responses for these panels. The hypotheses are, from left to right with their labels capitalized: **C**hannel(s) stimulated, **H**armonics used, appearance of adjacent harmonics in **D**ifferent channels, appearance of adjacent harmonics in a common (**S**ame) channel, and absolute **F**undamental frequency [which is also the beat frequency between any two adjacent harmonics]. The remaining hypotheses invoke *pairs* of those five potential cues, requiring that *both* match to predict identification of a candidate as a constituent of the reference tone. Note that the simultaneous invocation of the three attributes channel(s), harmonics, and fundamentals (**C**, **H**, and **F**) will produce predictions of analytically perfect (*i.e.*, all "correct") responses. (Depending on the subset of panels under consideration, other attribute combinations may also produce predictions of perfect responses.) The three rows of bar graphs correspond to (1) the percentage of correct predictions of SR2's responses under each hypothesis (based on the sum of both correctly predicted "Yes" and "No" responses), (2) the percentage of candidates for which "Yes" responses were predicted but the subject responded "No", and (3) the percentage of candidates for which "No" was predicted but a "Yes" was entered by SR2. Each pair of bars indicates any difference between the subject's initial and final judgments.

Clearly the most predictive single attribute hypotheses in this case are channel(s) (**C**) and harmonics (**H**). Combining those two (**CH**) produces only a slightly better set of predictions, as

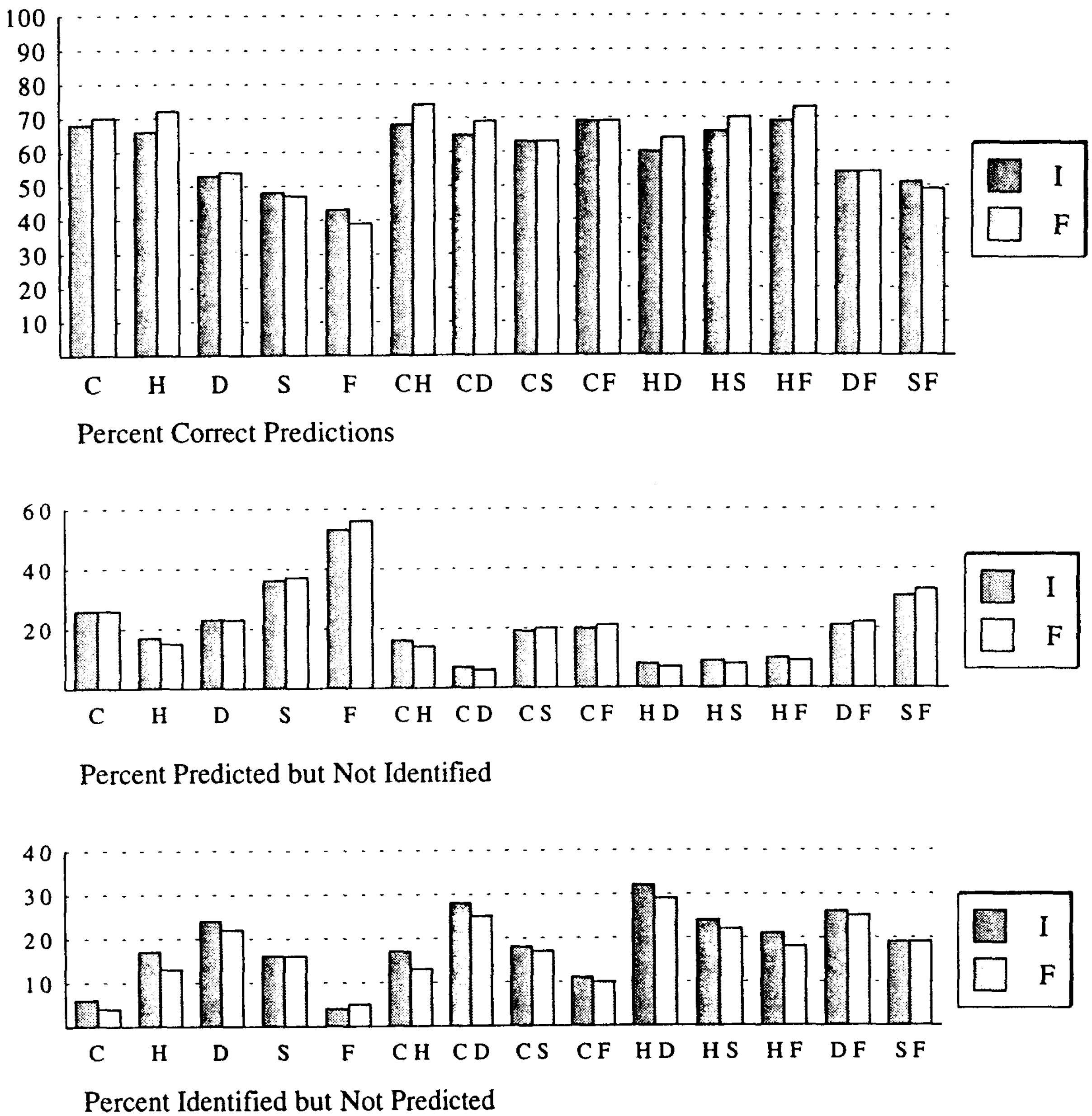

Figure 6. Example of Hypothesis Analysis, Constituent Tones Study

does adding the weakest attribute -- fundamental frequency -- to harmonics (**HF**). In fact, **CHF** (a prediction of analytically perfect responses, not shown in the figure) is a slightly better prediction of the subject's responses (69% of initial, 73% of final responses) than any of these. [Note that, for this subset of panels, **DS** also would predict perfect identifications, and that the fundamental cue (**F**) was relevant in only nine of the 41 panels. Overall, the subject made about twice as many false positive identifications as false negative exclusions.]

From the figure we see that the fundamental attribute (**F**) alone is most likely to predict identifications that the subject will *not* make, while the different channel pair cue (**D**) is the single attribute most often failing to predict his "Yes" responses. Between the best two single predictor attributes, channel errors (**C**) are mostly false positives while harmonics alone (**H**) yields a more balanced set of errors. Between the two single attributes that were the intended focus of this set of panels -- same channel and different channel adjacent pairs (**S** and **D**) -- the latter is a bit more successful as a predictor of the subject's responses. In general the best predictions tended to be slightly better for the final judgments than for the initial ones.

The same full range of hypotheses has been analyzed for each of the other subsets of panels to be discussed, but in the interest of brevity only the general patterns of results will be presented.

Single-partial candidate tones, reference tones composed of 2-5 partials; all from the same single harmonic series: This subset included 36 panels (1-16, 49-66, and 99-100). The common harmonic series was based on a G-196 Hz fundamental. The best single attribute predictions of SR2's responses were obtained for **H** cues (64% initial, 67% final) and **C** cues (62% initial, 68 % final). **CH** [which constituted perfect identification for this subset] yielded the same prediction scores as **H** alone.

Two-partial candidates, 2-5 partial references; same single harmonic series: There were 64 panels satisfying these conditions (17-48, 67-98), all based on the same G-196 Hz fundamental. Again **H** alone (77% initial, 81% final) was a slightly better predictor than **C** alone (76% initial, 78% final), and the [again perfect] **CH** prediction had essentially the same accuracy (77% initial, 82% alone) as **H** alone.

Single-partial candidates, 2-5 partial references; ***mixed use of two fundamentals:*** The two fundamentals were separated by a semitone interval, G-196 Hz and G#-208 Hz. This comparison subset included only 6 panels (101-106) comprising 30 responses. The best predictions using a single attribute were for **H**, with a difference between initial and final subject judgments that was unusual both in size and direction (70% initial, 60% final). **CH** predictions achieved exactly the same scores, but in this case analytically perfect criteria (**CHF**) eliminated one error that the subject also made (67% initial, 57% final). [Given the small number of responses a 10% deviation from chance score is significant at a confidence level of about 80%, a 20% deviation, however, is significant at a 98% confidence level.]

Two-partial candidates, 2-5 partial references; mixed use of two fundamentals: This was another comparison subset of only 8 panels (107-114), using the same pair of fundamentals. All single-cue predictions produced scores in the 48-53% range, with the highest being for **F** and **D** (53% initial, 50% final in each case). In this case, however, combining the two best single predictors (**FD**) produced a poorer prediction (50% initial, 43% final) and the best double cue prediction was **FS** at 55% initial and 58% final. A prediction of perfect identification (**CHF**) was less faithful to the subject's performance (53% initial, 50% final). Here the [less correct] **FS** criteria not only got higher absolute prediction scores, but also were consistent with the identification errors introduced by the subject between his initial and final judgments. Note, however that each of these differences involved only a very few responses out of the 40 in this small subset and cannot support any conclusions at this point. [With this few responses there can be only 60% confidence in a 5% deviation from chance in the prediction accuracy. The confidence level rises to 90% for a 10% deviation from a chance score (*i.e.* from 50%).]

Two-partial candidates, 3-8 partial references; various different fundamentals: A final small comparison subset of 7 panels (115-121) used stimuli based on a wider variety of fundamental frequencies. The best single predictor was **C** (74% initial, 83% final). All other single predictors were substantially less successful. In fact the only other predictor that could be combined with **C** without a decrease in predictive accuracy was **S**: **CD**, for instance, (51% initial, 43% final) was little better than **D** alone. Again, the limited number of responses in the subset leaves large uncertainties, but the best hypothesis predictions were notably better than a prediction of analytically perfect responses (**CHF**; 60% initial, 51% final), and again both in terms of absolute prediction score and consistency with the changes made by SR2 between initial and final responses. [A 20% deviation from chance in this case is significant with 98% confidence.]

Two-partial candidates, 4-partial references; various fundamentals: A larger subset of 18 panels (122-139) shared these conditions. Here **F** (73-74%) was the best single attribute for prediction, and explained the bulk of the 74-76% performance of both **HF** and perfect identification (**CHF**). **D** alone was also a good predictor (72-71%) but its performance was not improved by combination with any other cue. [With this many responses, a 10% deviation from chance is significant at a 95% confidence level.]

The overall picture that emerges is that harmonics is at least as good a predictor of SR2's performance on these tasks as channel(s) cues for the simpler (less complex) panels, and the combination (**CH**) somewhat better than either alone. Where the task is the recognition of pairs of adjacent harmonic partials there are indications that, depending on the context, both intrachannel (**S**) and interchannel (**D**) adjacent harmonic pair cues may be utilized.

Processor Bandpass Filter Order Effects

All the constituent tones identification studies above were conducted during SR2's visit in December of 1993. In the course of subsequent visits to our laboratory in March and May of 1994 we found time to repeat those studies with a modified processor, number 355, with 24th order bandpass filters defining the channels instead of the 12th order filters of number 163b. In all other respects the two processors were identical. We were interested in any effects an even greater degree of adjacent channel rejection might have on complex tone percepts. [In the present comparison studies, all the stimuli were composed of partials satisfying the exclusivity criteria for the 12th order filters with a minimum of 10 dB adjacent channel rejection. In this case the 24th order filter processor provides better adjacent channel rejection for partials near boundaries between channels. Alternatively, the 24th order filters could allow the qualification of additional partials, closer to those boundaries, under the same exclusivity criteria.]

Unfortunate circumstances raised questions about some of the complex tone data obtained during both these visits. In March, fluid seepage at the margin of the subject's percutaneous connector was copious enough to provide a stimulation current shunt between electrode leads that produced noticeable changes in percepts on several occasions during the complex tone tests. Those tests were interrupted frequently to let the subject listen to live speech in an effort to detect and correct such shunts as early as possible. Only the first 100 panels were administered before the frequency of delays for connector cleanings caused us to abandon the attempt. The stimulation on some channels for some of those panels likely was reduced by current shunts at the connector. In May the full 139 panels were run, but subsequent evidence indicated the likelihood that processor channel number 5 was delivering no stimulation at the time. Since channel 5 was involved only in panels 116-121 and 132-139, though, the first 100 panels of the May data can be used as a standard against which to assess defects in the March results. In doing hypothesis prediction analysis of the affected portions of the May data, both the normal matrices (Appendix II) and a special set based on the assumption of no channel 5 stimulation were used for comparison.

We will discuss the 24th order bandpass studies, then, in two sections: (1) the four subsets for which straightforward comparisons are possible between the December and May data, two of which replace the questionable data from March, and (2) the subsets whose analysis is complicated by the possibility of a nonfunctioning channel 5 in the May data. The first section includes the subsets with 2-5 partial reference tones that span single and two-partial candidate conditions based on either the same single fundamental or a pair of fundamentals separated by a semitone. The second section will include the two subsets with stimuli based on a range of various fundamentals.

Within the first section, the two large subsets with all stimuli based on a single fundamental yielded the same patterns of hypothesis prediction scores as before. For the subset with single-partial candidate tones, the highest absolute scores from December, March, and May were all about the same, the May scores being slightly lower than the others but not significantly so. Among the two-partial candidate tone subset scores, those for 24th order filters were higher than for 12th order and the May scores higher than the March ones. The **HC** hypothesis score was 83% initial, 86% final for the 24th order May data, for instance, compared to 80% initial, 85% final for the questionable March studies and 77% initial, 82% final for the December 12th order analysis.

The two smaller comparison subsets in the first section -- based on two fundamentals separated by a semitone -- showed a similar distinction between single and two-partial candidate results. For the single-partial subset, the large drop in the **H** prediction score from initial to final judgments was not seen in the 24th order data; the initial **H** score dropping and the final **C** score rising, making the latter hypothesis the most predictive (63% initial, 67% final). While the best hypothesis scores thus remained comparable between the two filter orders, however, the 24th order scores were higher *with respect to an analytically perfect prediction* (which was 66% initial, 57% final for the 12th order case, and only 53% initial, 50% final for the 24th order data). That is, the subject made more errors and the best hypothesis predicted them equally as well.

Again in the subset with two-partial candidates, the 24th order prediction scores were definitely higher with, however, the **D** hypothesis predictions improving *vis a vis* the **S**, both alone and in conjunction with **F** [a 10% deviation from chance is significant with 90% confidence].

	----------12th order filters----------					-----------24th order filters---------				
hypothesis	**F**	**S**	**D**	**FS**	**FD**	**F**	**S**	**D**	**FS**	**FD**
initial %	53	48	53	55	50	48	38	63	50	60
final %	50	50	50	58	43	58	33	68	60	70

For the 24th order filter data in the second section of subsets, the principal difference between results assuming that channel 5 was present and those assuming it inoperative was substantially improved **C** and **S** predictions in the latter case. In the smaller subset of this section -- the one whose reference tones were made up of from 3 to 8 partials -- **D** was a 89% predictor alone and both **FC** and **FD** achieved 94% scores, equal to a prediction of analytically perfect responses. This represented a substantial improvement over the most accurate prediction for the 12th order case (**C**, 74% initial, 83% final). The larger subset of the second section -- with 4 partials in each reference stimulus -- yielded essentially the same pattern for predictions based on single attributes for the two filter orders, with **F** rising from 73-74% to 74-78% for the latter case. Among multiple attribute predictions, however, **FC** and **FS** (79-82%) for the 24th order case replaced the best combination for the 12th order data **FH** (74-76%). Although **S** remained a much weaker predictor than **D** when used alone (47-46% *vs.* 73-74%) It was the only one of the two that could usefully supplement **F**.

Thus we are left with the indication that increasing adjacent channel rejection for partials near the spectral boundaries between channels can lead to more simply predictable response patterns on a constituent tones identification task -- at least in many cases where the candidate tones are composed of two partials. During the March visit, limited data were also taken using the same 24th order processor while repeating some of our earlier pilot study surveys exploring perceived pitch differences between pairs of complex tones [Lawson, *et al.*, 1993]. Those data, obtained both with and without a one second delay between tones, have not yet been analyzed. Nor have we yet undertaken analysis of patterns within the prediction errors for various hypotheses and subsets, or a search for any systematic variation in prediction scores with the degree of reference tone complexity.

As a final overview of the constituent tones identification results, we combine all the responses from all three visits -- both 12th order and 24th order processors -- into a single table (next page) showing the percent of the subject's responses that were "correct" in the sense of describing the analytic acoustic structure of the stimuli. The stimuli are sorted according to the same five attributes we have used in the predictive analysis above. Each of the first eight lines corresponds to a possible configuration of the cues **F**, **H**, and **C** between a candidate a reference tone. [The lines are numbered, as well as labeled, to facilitate the discussion that follows.] The first three columns of numbers represent the responses for each configuration sorted into three cases: (1) where there are no pairs of adjacent harmonics in the candidate, (2) where the candidate contains a pair of adjacent harmonics presented within a single channel ["intrachannel", S], and (3) where the candidate contains such a pair presented in different channels ["interchannel", **D**]. The final column summarizes all the responses for each **FHC** configuration and a final row summarizes all the responses in each column. Each numerical entry gives the number of responses and indicates the percentage of those responses that were analytically correct. For the row corresponding to **F**, **H**, and **C** all being the same between reference and candidate [the 8th row] a "correct" response means an identification -- a "Yes" response by the subject. For all other rows, "No" responses have been scored as correct.

----attributes----			------------adjacent harmonic pairs-----------			
same	different		none	intrachannel	interchannel	TOTALS
	FHC	1			89% of 36	89% of 36
C	**FH**	2	50% of 8		71% of 132	70% of 140
H	**FC**	3		100% of 8	100% of 4	100% of 12
F	**HC**	4	95% of 864	88% of 240	96% of 396	94% of 1500
HC	**F**	5	44% of 52	43% of 28	78% of 276	70% of 356
FC	**H**	6	51% of 290		71% of 242	55% of 580
FH	**C**	7	83% of 6			83% of 6
FHC		8	54% of 766	70% of 124	63% of 260	58% of 1150
TOTALS			71% of 1986	79% of 400	79% of 1346	74% of 3780

On line 8 we have the only cases in which the candidates were, physically, true constituents of their reference tones. Note that recognition of those candidates was not all that frequent, except when a pair of adjacent harmonics was present within a single channel.

Turning now to the rest of the table, we note that the subject correctly detected differences between candidate and reference tones with accuracies of 70% or better whenever they differed in at least two attributes [lines 1-4]. The weakest of those performances was in line 2, where there were common channels of stimulation. In line 5 we see that the presence of interchannel pairs was very helpful in conveying differences in fundamental frequency, while intrachannel pairs were not. Similarly, in line 6 interchannel pairs were quite helpful in recognizing differences in harmonics alone. When fundamental frequency was the only attribute in common between stimuli [line 4], the presence of intrachannel pairs [producing beat rates consistent with that fundamental] seems to have masked the other differences on occasion. In the absence of adjacent harmonic pairs in the candidates [first column], channel differences were more easily recognized [line 7] than purely harmonic differences [line 6], with the combination of differences in both attributes [line 4] leaving little doubt. Differences of fundamental frequency alone [line 5] were difficult to recognize in the absence of an interchannel pair.

Interval Consonance Judgments in a Nontraditional Context

As part of our effort to understand how information supporting SR2's judgments about timbre and consonance is conveyed [see Lawson, *et al.*, 1993], we decided to isolate some of the cues known to underlie such judgments in normal listeners through the use of a decidedly nontraditional system of musical scales and intervals. The system we chose was the BP or Bohlen-Pierce scale [Matthews and Pierce, 1989].

The most common musical scale in the West today is an equal tempered (ET) scale that spans an octave in twelve successive semitone intervals. Since the octave corresponds to a ratio of 2:1 between fundamental frequencies, each equal tempered semitone involves a frequency ratio of the 12th root of two (approximately 1.059). The BP scale is mathematically similar in that it spans a frequency factor of three in 13 equal intervals, each a ratio of the 13th root of three (approximately 1.088). The two scales are perceived as quite different by listeners with normal hearing.

The BP scale was chosen for our purposes, not merely because it is different, but because it also allows fulfillment of many of the characteristics that contribute to the perception of consonance for more traditional intervals. This is true so long as all upper partials of the BP complex tones are restricted to frequencies that are *odd* integer multiples of the corresponding BP fundamental.

The traditional ET scale provides acceptably close approximations to exact (just) consonant intervals such as the major fifth [frequency ratio 3:2], perfect fourth [4:3], major third [5:4] and minor third [6:5]. Such intervals between complex tones are characterized by the absence of low rate inter-partial beating and the presence of harmonically-related higher rate beats. Tones with only odd-harmonic partials based on the BP scale offer close approximations to such exact interval ratios as 5:3, 7:5, and 9:7, with quite similar patterns among partials available to support judgments of consonance and dissonance. [*Music theory note:* The BP scale supports 3:5:7 major triads and 5:7:9 minor triads that are highly analogous to the traditional 4:5:6 major chords. The BP equivalent of the major diatonic subset of the traditional chromatic scale (0, 2, 4, 5, 7, 9, and 11 ET semitones above the tonic; *e.g.* C, D, E, F, G, A, and B for a tonic C) involves nine rather than seven tones (0, 1, 3, 4, 6, 7, 9, 10, and 12 BP semitones above the tonic), successive tonics being separated by a 3:1 "tritave" rather than the traditional octave.] [*Physics note:* The design of acoustical (as opposed to electronic) musical instruments for performance of BP scale music would be complicated by the necessity of eliminating all spectral components at even-integer multiples of the fundamental frequency. Regardless of the type of instrument, there is the issue of even-harmonic components being contributed by the mechanical parts of the listener's ear, a concern neatly avoided in the present studies!] By having SR2 compare perceived consonance among complex intervals based on BP and ET scales we hoped to gain some insights into the roles of various consonance cues.

A set of 39 fundamentals along a BP scale ascending from 110 Hz (spanning three "tritaves") was examined, and all partials satisfying our 10 dB exclusivity criteria for a 12th order six channel CIS processor identified. Several sets of stimuli were synthesized to support a series of anecdotal response studies: (1) sequentially presented ET intervals (9 stimuli, each containing harmonics 2, 4, 6, and 8); (2) sequentially presented BP intervals (11 stimuli containing harmonics 3, 5, 7, and 9 and 6 stimuli containing harmonics 1, 3, 5, and 7); (3) simultaneously presented BP complex intervals (16 stimuli using harmonics 3, 5, 7, and 9 and three stimuli using 1, 3, 5, and 7); (4) simultaneously presented ET complex intervals (eight stimuli using harmonics 2, 4, 6, and 8). Each of these sets served in succession as the basis for a session with SR2, with an experimenter presenting the stimuli in various combinations and recording the subject's anecdotal responses. In a final two sessions simultaneous BP and ET complex intervals were intermixed sequentially for direct percept comparisons. The subject was not informed about the nature of the stimuli until after the entire series of sessions. These studies were done in May of 1994 using processor 163b. The playback level was adjusted at the beginning of each session so as not to exceed MCL for the stimuli of that session.

The principal complex intervals involved in these investigations are summarized in the following table. Only even harmonics were used in synthesizing the ET stimuli, only odd in the BP stimuli.

Ratio of Fundamental Frequencies		Interval in Semitones BP	ET	Interval Name	
5:4	1.250		4	Major Third	1
9:7	1.285	3			2
4:3	1.333		5	Perfect Fourth	3
7:5	1.400	4			4
3:2	1.500		7	Perfect Fifth	5
5:3	1.667	6			6
2:1	2.000		12	Octave	7
7:3	2.333	10			8

In some respects the sequential intervals sessions produced quite similar results with both scale systems. Relative pitch was reported to be more ambiguous as intervals decreased for frequency ratios below 1.4, but less so at higher absolute frequencies. Those patterns were consistent across substantial variations in the pattern of channel stimulation from one tone to another. More of the BP stimuli contained two partials represented in a single processor channel than was the case for the corresponding ET stimuli. On every occasion that a BP stimulus *did not* include such a pair of partials the subject volunteered that it "was hard to pin down". Other spontaneous comments during the BP session included: "It's not falling on the note I want it to fall on", "This is a difficult relationship", "The second is hard to relate to the first", and "It's hard to identify the interval -- it doesn't fit into the socket of any identifiable interval." No such comments were offered during the ET session. Both sessions were organized so that several pairs separated by the same interval but occurring at different absolute pitches would be considered sequentially. SR2 frequently sang intervals (unmonitored) as he considered these stimulus tones, his vocalizations tending to become more accurate both in interval and in absolute frequency after he had heard the same interval at different absolute pitches. Among the ambiguous pairs, there were examples of intervals that initially were perceived as ambiguous and remained so, ones that became ambiguous only after repeated listening, and cases in which an unambiguous initial perception changed irreversibly after repeated listening.

When it came to simultaneous complex intervals (single tones synthesized by summing partials from both fundamentals), the BP session was conducted before the ET. The subject was told nothing about the nature of the stimuli except that he was to describe each one rather than, as before, considering them in pairs. The first stimulus considered was a low frequency 6-semitone BP interval (ratio 1.667) using harmonics 3, 5, 7, and 9. SR2 found it profoundly puzzling at first, describing it as "several tones -- almost an organ note in character -- synthetic but pleasant". The second stimulus was the same interval, played one BP semitone [ratio 1.088] higher, and SR2's first comment was "very similar in texture and character -- higher than the first one by a little more than a full step". When the sequence of two such intervals was repeated at substantially higher absolute pitch, however, [fundamentals near 330 Hz, compared to the previous 130 Hz; the lowest partial frequencies were 990 and 390 Hz, respectively] he initially detected no difference between them. After repeated listening he came to hear "an additional high frequency component" in the upper one. The next stimulus was based on exactly the same fundamentals, but used harmonics 1, 3, 5, and 7 instead of 3, 5, 7, and 9. SR2 immediately described the overall pitch as lower and described the tone as "more compact" and the pitch "more stable". After repeated listening he compared this stimulus to the preceding one in the following terms: "It sometimes seems like there is a perfect fifth between the last two; otherwise this one is the same note in a different voice, or with lower harmonics." [The interval between the lowest partials -- 3 and 1 -- of these same fundamentals was, in fact, a ratio of 3, *i.e.* an octave plus a perfect fifth.] When the next stimulus involved the same interval and partials a BP semitone higher, the subject struggled to interpret it in comparison to the previous one, mentioning "two different organ chords with one note in common" and a "minor to major chord" transition before settling on the same interval "up by a major third". Such insightful comments continued as successively narrower simultaneous BP intervals were presented. SR2 began to hear most of the stimuli as combinations of two different pitches. At one point near the end of a sequence of 4-semitone intervals [ratio 1.4] he was struggling to decide whether the interval between the two pitches he heard was a perfect fourth [1.33] or a perfect fifth [1.5]. Frequently he would return to finer grained analyses of spectral components rather than perceive a stimulus as a complex tone pair. Often his remarks were consistent with struggles to resolve competing consonance cues in terms of a traditional experience and vocabulary that allowed no such inconsistencies. Overall, he seemed both to accept these nontraditional constructs as analogous to familiar musical tones and to find them profoundly difficult to describe in those terms.

In the corresponding ET session, considered in order of increasing interval width, SR2 was perceiving the stimuli as two distinct notes by the time perfect fourths [ratio 1.333] were being played and confidently identified the second perfect fifth as such. Octave intervals sounded to him "like more than one note -- at least two", but he did not volunteer the interval. Two octave intervals at different absolute pitches were heard as "having a lot in common, but I can't pin it down". That session ended with one of the most entertaining, if not enlightening, percepts of the study: a rapidly played octave-fifth-fourth-third sequence, all based on the same lower fundamental, evoked memories of tuning a ukulele to the words "my dog has fleas".

In the final two sessions of these studies the very sequence of complex intervals shown in the table on the preceding page was explored, in two significantly different absolute frequency ranges [lower fundamentals ranging from 141 to 235 Hz and from 256 to 330 Hz, respectively].

The subject was asked to consider three stimuli at a time, always corresponding to three successive lines from the above table. Thus, half the time he was considering one BP interval bracketed [in sequence and in width] by two ET ones and the rest of the time *vice versa.* For economy, we will refer to the stimuli using the line numbers at the right edge of the table. In the lower frequency session the stimuli were presented in increasing number order, and in the reverse order during the higher frequency session. The first session -- the lower absolute frequency one -- began with the first three stimuli from the table, a complex major third and perfect fourth bracketing an intermediate complex BP interval. SR2 commented that the first and third were "very similar in feature -- a two-note sound" with the third higher in pitch than the first "by a major third or perfect fourth". The middle stimulus, on the other hand, was described as "unrelated to one and three: the character of it is more . . . discordant". The next two sets of three stimuli, 2-3-4 and 3-4-5, were both accepted as having "pretty much the same character for all three", except that BP stimulus 4 was described as "a bit flat in the former set and "a bit tighter" in the latter. For sets containing two ET and one BP stimulus, the subject frequently tried to characterize the overall pitch of a stimulus *vis a vis* the others in its set of three, with variable accuracy and consistency. While in general SR2 seemed to become less aware of differences between stimuli from the two scale systems as the sessions proceeded, BP stimulus 6 -- the 5:3 interval -- was a notable exception. In the 4-5-6 context, stimulus 6 was judged "different from [4 and 5] -- made of two notes that don't belong together", while in the 5-6-7 set it was described as "weird -- hard to assign it to a pitch with respect to [5 or 7] -- it wants to be down a perfect fifth but it's not." [The lower fundamental was 141 Hz for stimulus 6 and 196 Hz for both the other two.] In the 6-7-8 context stimulus 6 was "flat or sour". In the subsequent higher frequency session, stimulus 6 was again singled out, but only in the initial 8-7-6 context, as "somehow off from where you'd want it to be". With that single exception, all the stimuli in all the sets of the final session were accepted as "very similar in feature" or "close in character -- the same instrument".

Clearly SR2 was able immediately to detect inconsistencies between complex intervals based on the two scale systems, an impressive feat given the known crudeness of stimulation with cochlear implants. Apparently, with very limited experience, he also came to accept mixtures of the two on the basis of the structural patterns that were kept consistent among the stimuli. Despite his abilities with these subtle distinctions, of course, he often made substantial errors in (supposedly, much less subtle) relative pitch judgments. Detailed follow-up studies are planned. One particular item for further investigation is a non-analogous feature of complex BP tones that may be more significant for CIS processor users than for listeners with normal hearing: when adjacent BP overtones are represented in a single processor channel the resulting beat rates always correspond to the *octave* of the fundamental (the difference between successive *odd harmonics).* The octave is, of course, an even harmonic.

Systematic Examination of Perceptual Category Assignments: Single Stimuli

While such anecdotal studies certainly have their rewards, they are always conducted in the hope of their leading to subsequent, more highly controlled experiments. In our earliest complex tone studies with SR2 [Lawson, *et al.*, 1993] the subject volunteered a number of descriptive terms quite frequently. In an effort to gauge the stability and precision of these categories and the degree of independence among them we included in SR2's August 1994 visit systematic interviews regarding single stimulus percepts and percept pair comparisons. The interviews were computer controlled, strictly limiting the descriptive terminology and requiring every category to be considered on each occasion, thus avoiding some of the more obvious potential pitfalls of the anecdotal reports from the pilot studies.

Descriptive categories were assigned to each stimulus by the subject, using a mouse and a sequence of two control display windows. The first of those windows is shown here as Fig. 7.

Here is a new sound:

Play

Choose the one best description for the sound:

o Single pure tone

o Single complex tone

o Pleasant combination of tones

o Dissonance

o Noise

Done

Figure 7. First Control Window, Single Stimulus Category Study

The **Play** button could be clicked as many times as desired, to listen to the current stimulus tone. A minimum delay of one second was imposed between successive presentations. In this window the five options were mutually canceling "radio buttons". Once the subject had highlighted the best one by clicking on it, and was satisfied with the choice, he moved on to the next window by clicking on the **Done** button.

The second primary window is shown in Fig. 8. Here the function of the **Play** button is the same as before, but as many of the categories as appropriate (or none) could be checked before moving on by clicking on **Done**.

For the same sound:

Play

Check as many as apply:

o **Modulated**

o **Rough**

o **Buzzy**

o **Synthetic**

o **Vowel-like**

o **Chord-like**

o **Smooth**

Done

Figure 8. Second Control Window, Single Stimulus Category Study

An auxiliary message then would appear on the screen, asking the subject *"Did the sound seem to change after repeated listening?"*. Clicking on the **No** button provided would cause the return of Fig. 7's window with a new stimulus tone loaded. Clicking **Yes**, that there had been a change, would summon a sequence of two additional queries, with each indented line available for clicking as a response:

After it changed, could you then
- o *hear it "BOTH ways"?*
- o *only hear it the "NEW way"?*

and then

Have your answers described
- o *only the ORIGINAL way it sounded?*
- o *only the NEW way it sounded?*
- o *BOTH?*

Depending on the answers to those questions, the two primary control windows would appear again for the same stimulus with one of the following three pairs of labels replacing the *"Here is a new sound:"* and *"For the same sound:"* appearing in Figs. 7 and 8: (1) *"Listen to the same sound again and describe the NEW way it sounds:"* and *"Again, describing the new way it* *sounds:", (2) "Listen to the same sound again and describe the ORIGINAL way it sounded:"* and *"Again, describing the original way it sounded:"*, and *(3) "Without listening to the sound again, try to describe the ORIGINAL way it sounded:"*, and *"Again, describing the original way it sounded:"*. In case (3), of course, the **Play** button would not appear in either window.

The supervising computer program recorded the number of times each stimulus was played from each window and the categories selected in that window, as well as the history of any noted changes in percept.

In studies over two days in August 1994, with SR2 using the 163b processor, this interview was administered for a total of 383 stimuli satisfying our 10 dB channel exclusivity criteria for a six channel 12th order bandpass CIS processor. The stimuli included single complex tones with from one to eight harmonic partials, and a few instances (12) of simultaneous combinations of single harmonics or pairs of adjacent harmonics based on two different fundamentals. Most of the stimuli used in previous complex tone studies (except Bohlen-Pierce scale tones [see above]) were included. Some stimuli appeared more than once, to allow assessment of the repeatability of the subject's judgments. The playback gain was set initially to ensure that no stimulus would produce a percept above MCL in loudness, and not changed during the studies. All stimuli were audible, but a few were reported to be quite soft.

The number of stimuli assigned each of the five mutually exclusive categories in the first control window is shown below:

Single Pure Tone	110
Single Complex Tone	248
Pleasant Combination of tones	9
Dissonance	16
Noise	0

Lists were prepared of stimuli in each of these categories except the largest, and each such list was examined for patterns in the structures of its members. These lists also may prove useful in constructing stimulus sets for future studies. The stimulus tones categorized as **dissonances** all involved intrachannel pairs of adjacent harmonics with beat frequencies close to 200 Hz, most occurring in channel 3. All of the stimuli so identified were made up of harmonics of single fundamentals. Only about half of the stimuli categorized as **pleasant combinations of tones** in fact involved harmonics of more than one fundamental. All but one included four or more harmonics and both intrachannel and interchannel occurrences of adjacent partials. In most cases there were intrachannel beats in two different channels. The one exception had two adjacent harmonics of a high fundamental in channel 6, with a beat rate (1245 Hz) greatly attenuated or completely eliminated by the channel envelope's 400 Hz low-pass smoothing filter. Among the 110 stimuli described as sounding like **single pure tones**, 23 were just that, while 59 had adjacent harmonic pairs but none in common channels, 22 had *only* intrachannel pairs, and 6 had pairs of both types. Only 3 of the 22 stimuli with purely intrachannel pairs had beat rates near 200 Hz (*c.f.* the pattern for dissonances). All but one of the 6 with pairs of both types had intrachannel beat rates at or above 400 Hz.

A change in percept while considering these single stimuli was reported in only five cases, none involving any change in categorization.

The following table shows the percentages of each of these four initial categories that then received each of the seven "second window" labels. Percentages exceeding 65% are shown in boldface type.

	-----------Percentage of Initial Category receiving Label-----------			
Label	Pure	Complex	Pleasant Comb.	Dissonance
Modulated	1	5	**67**	12
Rough	2	17	44	**88**
Buzzy	12	42	**67**	38
Synthetic	25	**93**	**100**	**100**
Vowel-like	6	6	0	0
Chord-like	0	19	56	**69**
Smooth	**86**	51	22	6

All stimuli heard as pleasant combinations of tones, all categorized as dissonances, and virtually all called single complex tones also were described as "synthetic". Interestingly, only stimuli described as single pure tones largely escaped that label, suggesting that it was intended to convey complexity rather than imply an artificial source. The label "modulated" was attached primarily to pleasant combinations of tones and "rough" was most strongly linked to dissonances. "Buzzy" was used to modify a wide range of categories and was not especially associated with dissonance. The labels "chord-like" and "rough" were largely applied to dissonances and pleasant combinations, and "smooth" to perceived single tones -- pure and complex.

The wide variation in number of stimuli assigned to each of the four initial categories should be kept in mind when interpreting these results. To that end, we also show the same data expressed in terms of the percentage of all uses of each of the seven labels that occurred within each of the four initial categories.

	----------Percentage of Label Use by Category--------						
Category	Modul	Rough	Buzzy	Synth	Vowel	Chord	Smooth
Pure Tone	5	3	10	10	32	0	42
Complex Tone	57	69	83	81	68	73	56
Pleasant Comb.	29	7	5	3	0	8	1
Dissonance	10	23	5	6	0	17	<1

In the following table we indicate the total number of times each of the labels from the second window was used, and the percentage of uses of each label that were accompanied by the use of each other label. The values above 90% are shown in boldface type while those in the next cluster of values, near 50%, are underlined.

	Total Count	Percent Accompanied By Modul	Rough	Buzzy	Synth	Vowel	Chord	Smooth
Modulated	22		45	36	**91**	23	18	32
Rough	61	16		52	**98**	0	28	2
Buzzy	130	6	25		**99**	0	13	8
Synthetic	284	7	21	45		4	21	46
Vowel-like	22	23	0	0	50		0	**91**
Chord-like	64	6	27	27	**94**	0		48
Smooth	225	3	0	5	58	9	14	

Virtually all occurrences of the labels *modulated, rough, buzzy,* and *chord-like* were accompanied by the label *synthetic*, but roughly half the uses of *synthetic* were in association with about half the occurrences each of *smooth* and *vowel-like*. Virtually all the stimuli deemed *vowel-like* were heard as *smooth*, with a quarter of them also described as *modulated.* A third of all *modulated* tones were considered *smooth*, a third *buzzy*, and almost half *rough.* For *chord-like* tones the proportions were one half *smooth*, and a third each *rough* and *buzzy.* Half of all *rough* stimuli were also *buzzy* but only one quarter of all occurrences of *buzzy* were accompanied by the label *rough. Smooth* occurred almost as frequently as *synthetic*, with *buzzy* being the next most used label. The most seldom used labels in this set were *modulated* and *vowel-like*.

Lists were produced of all stimuli identified as *modulated, rough,* and *chord-like*, in the hope of identifying common structural elements associated with those labels. There were 21 different stimuli in the *modulated* list, 42 in *rough*, and 41 in *chord-like*. There were 9 different stimuli in common between the *modulated* and *rough* lists, 13 between *rough* and *chord-like* lists, and three between *chord-like* and *modulated.* Only one stimulus received all three labels. In the table that follows we show the percentages of stimuli in each of those six groups with (1) **no** intrachannel beats between adjacent partials, (2) such beats in **all** channels stimulated, and (3) a **mixture** of channels with and without such beats. The single stimulus common to all three lists was of the mixed type. Also shown is the range of such beat frequencies in each list.

List	percentage of stimuli: No beats	Beats in all	Mixture	Beat Freq. Range
Modulated	5	76	19	49-1295 Hz
Rough	5	76	19	49-622
Chord-like	22	32	46	123-1245
Mod + Rough	11	67	22	49-233
Rough + Chd	0	54	46	123-233
Chd + Mod	0	67	33	208-1245

In terms of these three structural classes there seems to be a clear statistical distinction between *chord-like* and the other two single-label lists. Furthermore, while the single-label *modulated* list and *rough* list have indistinguishable structural statistics in terms of these classes, a distinction emerges when we compare the paired-label " *chord-like and modulated*" and "*rough* and *chord-like*" lists.

The full range of channels was involved in the intrachannel beats in all six lists. Beat frequencies between adjacent partials ranged from 49 to 1245 Hz among the stimuli in these lists, with values below 123 Hz and above 622 Hz occurring in relatively few stimuli. Thus differences in the beat frequency ranges for the three single-label lists are probably not significant. The lack of beat frequencies above 233 Hz in the "*modulated* and *rough*" and "*rough* and *chord-like*" paired-label lists, on the other hand, may be significant, as may be the absence of beats below 208 Hz for the "*chord-like* and *modulated*" combination.

Systematic Examination of Perceptual Category Assignments: Stimulus Pair Comparisons

The same stimuli that had been considered individually in the category assignment studies discussed above also were employed later in systematic computer controlled interviews regarding comparisons of complex tones presented in pairs. This study sought, again in a more controlled context, to explore perceived differences along several dimensions volunteered anecdotally by SR2 in the course of our earliest studies with complex tones. The initial control window for a new pair of stimuli is shown in Fig. 9.

Clicking a computer mouse on the **Play** button caused playback of the two stimuli, in a fixed order and separated by a one second delay. This button could be used as many times as desired, subject only to a minimum delay of one second between the presentations of any two stimuli. Any of the four mutually exclusive options could be selected by clicking on the corresponding mutually canceling "radio button". Once the subject was satisfied with his choice, he could proceed to the next window by clicking on the **Done** button. In this study the **Different** button was used to signal that change in percept after repeated listening to a pair. Clicking on it caused an instruction to appear in the window: "*Continue to answer for the ORIGINAL way the tones sounded.*"

As indicated in Fig. 9, the first window asked for a judgment of the relative overall pitch of the two complex tones, a judgment sought for many such pairs in the earliest pilot studies [*cf.* Fig. 1 of Lawson, *et al.*, 1993]. In this case, if the first response indicated a difference in overall pitch between the two tones, a second window proceeded to ask for an assessment of the degree of difference.

New Tones: Overall Pitch

- **o 1 Higher**
- **o Same**
- **o 2 Higher**
- **o 2 Both higher and lower in pitch than 1**

Play **Done** **Different**

Figure 9. First Control Window, Stimulus Pair Category Comparisons

The **Play** and **Done** buttons continued to be available, with the options changing to:

The difference is:

- o *A musical second or less*
- o *A musical third*
- o *A musical fourth*
- o *A musical fifth*
- o *A musical sixth or seventh*
- o *An octave*
- o *More than an octave*

This categorization of relative overall pitch was followed by similar questions along a variety of perceptual dimensions, all for the same pair of stimulus tones. In each case an initial window determined whether or not a difference was perceived and, if so, a secondary window asked for an indication of the degree of difference. Each window included the **Play** button, a set of "radio buttons" for the options and the **Done** button. The secondary window text in each remaining case was:

The difference is:

- o *Tiny*
- o *Small*
- o *Moderate*
- o *Large*

The primary dialog texts for the remaining perceptual dimensions were as follows:

Loudness:

- o *1 Greater*
- o *Same*
- o *2 Greater*

Pureness:

- o *1 Greater*
- o *Same*
- o *2 Greater*
- o *Doesn't apply to either*

Roughness:

- o *1 Greater*
- o *Same*
- o *2 Greater*
- o *Doesn't apply to either*

Strength of Modulation:

- o *1 Greater*
- o *Same*
- o *2 Greater*
- o *Doesn't apply to either*

Rate of Modulation:

- o *1 Higher*
- o *Same*
- o *2 Higher*
- o *2 Has both Higher and Lower Rates than 1*
- o *Doesn't apply to either*

Tone Complexity:

- o *1 Greater*
- o *Same*
- o *2 Greater*

Consonance / Pleasantness:

- o *1 Greater*
- o *Same*
- o *2 Greater*

If the subject had clicked on the **Different** button early in the interview for this pair of tones the program now would return to the window of Fig. 9, but with the message "*Now answer for the NEW way these tones began to sound*". Otherwise the next window would look exactly like Fig. 9 and control a new pair of complex tones. In either case the full sequence of windows would follow.

The supervising computer program recorded each response, as well as the number of times each pair was played at each juncture.

As mentioned above, the categories chosen for this interview had been volunteered frequently by the subject in earlier anecdotal responses. Clearly this set of categories would not be appropriate for many cochlear implant subjects!

A total of 118 different complex tone pairs were used as stimuli for this study. Each appeared twice, once in each order of presentation. Playback levels were adjusted as necessary for each pair, to be certain that no amplitude clipping occurred at the input to the subject's 163b processor and that all stimulus pairs were presented near his MCL. The studies were done during SR2's August 1994 visit to our laboratory.

Later in the same visit, the 32 stimulus pairs that had received ambiguous relative overall pitch reports were presented again, along with 12 new stimulus pairs, in a different procedural context. The only change in the interview program for this case was that the single **Play** button in Fig. 9 and related windows was everywhere replaced by separate **Play 1** and **Play 2** buttons, allowing the subject to listen to the two stimuli of each pair in any order. The minimum one second delay between any two stimuli was maintained.

The data from these studies are contained in a pair of matrices, one outlining SR2's judgments along the nine perceptual dimensions and the other describing each stimulus pair in terms of acoustic and processed structure (*i.e.* in terms of fundamental frequencies, harmonic partials, channel assignments, and beat frequencies). These matrices will immediately support better choices of stimuli for future complex tone studies, enabling us to design tests that are more precisely focused. Full analysis of these data, on the other hand, is just beginning and will be reported in a future progress report. We will be searching for correlations between combinations of stimulus structure features and perceived characteristics.

Pending a more complete analysis, some preliminary glimpses of patterns in the data may be of interest. In addition to the 32 instances of ambiguous relative overall pitch among the initial 118 complex tone pairs, for instance, there were four cases of firm reversal of relative pitch percept with changing presentation order. There were no spectral content distinctions that were unanimous across presentation order reversal. There were five instances in which same/different overall pitch judgments for normal/reversed order of presentation were paired with different/same spectral content judgments. As a final glimpse for now, the following table shows the distribution of pitch difference magnitude judgments for two groups of stimulus pairs: those judged to have different overall pitches regardless of order of presentation, and those judged to have different overall pitches when presented in one order but to share the same overall pitch when presented in the opposite order.

Pitch Difference	-----Percentage of Responses----- Difft/Difft	Difft/Same
Second or less	2	0
Third	33	76
Fourth	30	24
Fifth	13	0
Sixth or Seventh	17	0
Octave	2	0
More than Octave	3	0

Inconsistency Detection Thresholds

The final study to be discussed in this progress report utilized the same interview program just described -- the version allowing separate access to the **Play 1** and **Play 2** functions. Also conducted during SR2's August 1994 visit, this study was designed to gauge roughly how large certain internal inconsistencies in complex tones would have to be in order to be noticed by the subject. The 21 stimulus pairs were delivered to processor 163b at MCL. Each stimulus provided spectral components to both processor channels 2 and 3, and to no other channel. Each involved harmonics 3 through 6. The following table indicates the frequencies of the pure tone partial(s) in each channel and any beat frequency present in each channel. All values are frequencies in Hz and the variables in each case are shown in bold face. H is an abbreviation for harmonic number, Ch and Chan for channel number.

	----------Stimulus 1---------				-----------Stimulus 2--------				-----Beat Frequencies------			
	--Chan 2--		--Chan 3--		--Chan 2--		--Chan 3--		---Stim 1---		---Stim 2---	
Pair	H3	H4	H5	H6	H3	H4	H5	H6	Ch2	Ch3	Ch2	Ch3
1	624	832	1040	1248	624	832	**1065**	**1278**	208	208	208	**213**
2	624	832	1040	1248	624	832	**1090**	**1308**	208	208	208	**218**
3	624	832	1040	1248	624	832	**1115**	**1338**	208	208	208	**223**
4	624	832	1040		624	832	**1065**		208		208	
5	624	832	1040		624	832	**1090**		208		208	
6	624	832	1040		624	832	**1115**		208		208	
7	624	832	1040		624	832	**1140**		208		208	
8	624	832	1040		624	832	**1165**		208		208	
9	624	832	1040		624	832	**1190**		208		208	
10	624	832	1040		624	832	**1215**		208		208	
11	624	832	1040		624	832	**1240**		208		208	

	----------Stimulus 1---------				-----------Stimulus 2--------				-----Beat Frequencies-----			
	--Chan 2--		--Chan 3--		--Chan 2--		--Chan 3--		---Stim 1---		---Stim 2---	
Pair	H3	H4	H5	H6	H3	H4	H5	H6	Ch2	Ch3	Ch2	Ch3
12	624		1040	1248	**639**		1040	1248		208		208
13	624		1040	1248	**654**		1040	1248		208		208
14	624		1040	1248	**669**		1040	1248		208		208
15	624		1040	1248	**684**		1040	1248		208		208
16	624		1040	1248	**699**		1040	1248		208		208
17	624		1040	1248	**714**		1040	1248		208		208
18	624		1040	1248	**729**		1040	1248		208		208
19	624		1040	1248	624		**1065**	**1278**		208		**213**
20	624		1040	1248	624		**1090**	**1308**		208		**218**
21	624		1040	1248	624		**1115**	**1338**		208		**223**

Each stimulus again fulfilled our 10 dB exclusivity criteria for 6 channel CIS processors with 12th order bandpass filters. The stimulus pairs were presented in the numerical order shown. Four distinct tests were included in this sequence of pairs, and their stimuli are separated by blank lines in the table above. The subject was given no prior information about the nature of these tests and had no indication of the division of the stimuli into separate tests.

In the first test there were beats in both channels for both stimuli, with the second stimulus' channel 3 partials based on fundamentals that were 5, 10, and 15 Hz higher than the fundamental of all the other partials. This made the beat frequency in that channel increasingly inconsistent, in 5 Hz steps.

The second test supplied unchanging reference beats in channel 2 for both stimuli. While the first stimulus also contained a single, consistent partial in channel 3, however, the second stimulus substituted increasingly inconsistent single partials, again based on a fundamental increasing in 5 Hz steps. Since the partial was a fifth harmonic, it changed in steps of 25 Hz, from 1065 to 1240 Hz.

The third test was quite similar to the second, but with the reference beats in channel 3 for both stimuli and, in channel 2, a consistent single partial in the first stimulus and increasingly inconsistent single partials in the second. Since the third harmonic of the variable fundamental was involved in this case, the frequencies were lower, ranging from 639 to 729 Hz in 15 Hz steps. The relative salience of the inconsistencies in the second and third tests was of particular interest in light of our emerging neural modeling work and intracochlear evoked potential measurements.

In the final test of this pilot series, a fixed single partial was the reference present in channel 2 of both stimuli. Channel 3 received a consistent, constant beat frequency in the first stimulus and an increasingly inconsistent one in the second.

[For purposes of reference, an inconsistency of an equal tempered semitone (a musical minor second or "half step") would lie between the values of pairs 2 and 3, 5 and 6, 13 and 14, and 20 and 21, while an equal tempered whole tone (a major second) is closely approximated by the inconsistencies in pairs 8 and 16.]

There was not time during the August visit to explore additional conditions involving holding a beat rate constant (and consistent with respect to the other components of the stimuli) while increasing the inconsistency of the absolute frequencies producing the beat. That study, among others, is awaiting SR2's next visit.

The following table summarizes selected relevant responses of the subject during these tests. The numbering and grouping of stimulus pairs is the same as in the previous table.

	------Overall Pitch------		----Spectral Content----		Pureness	Complexity
	Difference	Amount	Difference	Amount		
1	2 Higher	Tiny	Same		Same	Same
2	2 Higher	Tiny	Same		Same	Same
3	2 Higher	Tiny	Same		Same	Same
4	2 Higher	Tiny	Same		Same	Same
5	2 Higher	Tiny	Same		Same	Same
6	2 Higher	Small	Same		Same	Same
7	2 Higher	Small	Same		Same	Same
8	2 Higher	Small	Same		Same	Same
9	2 Higher	Moderate	Same		Same	Same
10	2 Higher	Tiny	2 Higher	Small	1 Purer(Sm)	2 More(Sm)
11	2 Higher	Moderate	2 Higher	Small	1 Purer(Sm)	2 More(Sm)
12	1 Higher	[NR]	Same		Same	Same
13	Same		Same		Same	Same
14	2 Higher	Tiny	Same		Same	Same
15	2 Higher	Tiny	Same		Same	Same
16	2 Higher	Tiny	Same		Same	Same
17	2 Higher	Small	1 Higher	Small	2 Purer(Sm)	1 More(Sm)
18	2 Higher	Tiny	Same		Same	Same
19	2 Higher	Small	Same		Same	Same
20	2 Higher	Small	Same		Same	Same
21	2 Higher	Moderate	2 Higher	Small	1 Purer(Sm)	2 More(Sm)

In the first test, the difference in the channel 3 beat rate was salient over the whole tested range (inconsistencies of 5 to 15 Hz) with no detected trend in the percept.

The inconsistencies were again salient over the full range of variation in test two, but this time with the overall pitch difference magnitude showing a clear trend. For the largest inconsistencies, a spectral content difference was reported and the second stimulus was described as less pure and more complex. The marked drop in the magnitude of perceived pitch difference for pair 10 with respect to the pairs on either side of it will be discussed in connection with a similar observation in the next test.

The saliency of the absolute frequency inconsistency in test three was not well established until (in pair 14) it exceeded 30 Hz, while the analogous inconsistency at higher absolute frequency in the second test apparently was salient at 25 Hz or less. The overall pitch difference was described as small when the inconsistency had reached 90 Hz (in pair 17, compared to 75 Hz in the case of the second test), and differences in spectral content, pureness, and complexity were reported at the same point. The direction of all three of those differences, however, was reversed from those reported late in test two. As seen in test two, a sudden drop in the magnitude of perceived overall pitch difference occurred at pair 18, this time accompanied by the disappearance of perceived differences along the other three dimensions.

Some of the perceptual structure detected in tests two and three may be interpreted in terms of an ambiguity in implied fundamental. Let us consider the increasing deviations from a consistent set of harmonics in terms of the frequency *ratios* of the three partials in the second stimuli of both tests, with respect to the reference fundamental. In the first pair of test two (pair 4) the ratios are 3.00, 4.00, and 5.12, reflecting the small departure of the highest partial from a perfectly consistent 5th harmonic frequency. Notice that by the time pairs 10 and 11 are reached, the highest partial's ratio to a consistent fundamental is 5.84 and 5.96 respectively. This may have allowed the subject to interpret the stimulus either as another, even larger departure from harmonics 3, 4, and 5, or a very close approximation to harmonics 3, 4, and **6** of the same fundamental. In test three we began with the 3rd, 5th, and 6th harmonics and ratios of 3.07, 5.00, and 6.00. By the time pairs 17 and 18 had been reached, the ratio of the lowest partial had risen to 3.43 and 3.50 times the consistent fundamental, supporting an alternative interpretation of the three partials as the **7th**, **10th**, and **12th** harmonics of a new fundamental, shifted down by an octave. Hence, perhaps, the contrast in perceived spectral content, pureness, and complexity differences that appeared late in the two tests. Such sudden transitions to alternative interpretations of complex tones in the course of slowly increasing inconsistencies are well known for subjects with normal hearing. [The effect sometimes is called *virtual pitch* or *residue pitch*; see Moore (1989), pp. 167ff., and Houtsma *et al.* (1987).] Anecdotal reports from our earliest pilot studies with SR2 had suggested that he could, on occasion, report perceptual changes consistent with the salience of an implied fundamental. That interpretation now rests on firmer ground.

In the final of the four tests an inconsistent beat rate, in the absence of a simultaneous reference beat rate in another channel, was salient over the whole range explored and showed evidence of a perceptual trend as the inconsistency was increased. The spectral, pureness, and complexity judgments for the largest inconsistency matched those seen in the second test.

Discussion

Summary of principal findings

- The presence of differences in more than one structural attribute of two stimuli (among fundamental frequency, harmonics, and channels involved) substantially increased the subject's accuracy in detecting a difference. This was especially true if one of the differing attributes was channels. In the absence of adjacent harmonic pairs, channel differences were more easily recognized by the subject than purely harmonic differences. For complex tones with relatively few partials, however, patterns in the included harmonics were at least as good a predictor of the subject's constituent tones identifications as patterns in the channels involved. The combination of both patterns was a better predictor than either alone.
- Among complex tones involving adjacent harmonic pairs in various contexts, both intrachannel and interchannel pairs appeared useful to the subject in recognizing differences in (implied) fundamental frequencies. The presence of intrachannel pairs generally increased the subject's accuracy identifying constituent tones but, in comparisons of complex tones based on the same fundamental, the presence of such a pair (beating at a frequency consistent with the fundamental) reduced the likelihood of perceiving simultaneous differences in both harmonic and channel patterns. The presence of interchannel pairs of adjacent harmonics was particularly helpful when the only structural attribute difference was fundamental frequency or harmonic pattern.
- The use of 24th order bandpass filters rather than 12th order made a significant difference in constituent tones identification tests when the candidate tones involved more than a single partial. Simple models predicting the subject's responses on the basis of stimulus structure generally were more accurate for the higher order filters. In some of the most complex comparisons investigated, the higher order filters also supported better accuracy in the subject's judgments themselves.
- Despite many similarities to the stimulation patterns produced by traditional consonant intervals between complex tones, the subject immediately recognized that consonant Bohlen-Pierce intervals were fundamentally different. With experience, he became willing to accept them as analogous.
- The subject reliably reported increases in overall pitch in complex tones with small harmonic inconsistencies in a beat frequency of 208 Hz, whether the simultaneous reference in another channel was a fixed 208 Hz beat or a single partial at 624 Hz. With a fixed 208 Hz reference beat in another channel, he reliably detected small inconsistencies in 624 Hz and 1040 Hz single partials.
- Complex tones described by the subject as "dissonant" were likely to exhibit intrachannel beating between adjacent partials separated by roughly 200 Hz. Complex stimuli described as "single pure tones" included intrachannel beating at any of a wide range of beat frequencies, but seldom around 200 Hz. Stimuli described as "pleasant combinations of tones" usually had four or more harmonics -- whether of a single or different fundamentals -- and included both interchannel and intrachannel pairs.

- Well defined statistical patterns relate several of the subject's perceptual categorizations to structural attributes of the complex tone stimuli they describe.

Implications for speech processor design

Our findings indicate that envelope fluctuations caused both by single harmonic modulation of a channel and by beating between intrachannel pairs of adjacent partials can usefully influence judgments as to the pitch and timbre of musical tones. We have observed this to be true over a range of channel modulation frequencies wider than that fully supported by the CIS processors SR2 has used to date. Combined with a knowledge of the relationship between rate and perceived pitch for this subject and with the results of modeling calculations validated by intracochlear evoked potential measurements, these findings strongly suggest that SR2's perception of speech might be significantly improved by CIS processors utilizing still higher pulse rates and appropriate envelope filtering. More subtle differences in the inputs to a CIS processor can be heard and interpreted if ways can be found to convey them unambiguously via processor channel signals to the eighth nerve.

The present subject is limited to a maximum of six CIS channels. While the pattern of channels stimulated in a given instance clearly is a powerful source of information as to the quality of a sound, intrachannel beats also have proven important. Indications are that increasing the number of perceptually distinct CIS channels would provide more detailed information through patterns of channels stimulated. Beyond some point, however, increasing the number of channels might reduce a user's access to other spectral information. An increased number of increasingly narrow channel passbands would inevitably reduce the number and frequency range of intrachannel beats, and increase the fraction of partials conveyed, perhaps less helpfully, in more than one channel. There may be a very fine balance to be struck between spatial and temporal modes of conveying information via cochlear implants in order to optimize perception of spectral information, and this may be true quite aside from questions of electrode design, electrode position, and the limits of neural spatial resolution.

We have identified some significant abilities and determined some limits imposed on those abilities by present CIS processors. Among other things, our complex tone studies will serve as benchmarks for assessing future improvements in processing strategies. Our now extensive database of stimuli and percepts will make such assessments much more efficient and precise than has been possible to date.

Implications for perception of musical sounds by implant patients

In carefully controlled isolation, a wide range of potential spectral cues contained in the structure of complex musical tones can be conveyed to at least one user of a CIS processor. While that fact alone does not indicate that such cues could be utilized in less controlled circumstances, we note that the abilities demonstrated in these studies begin to offer an explanation for the richness and subtlety of SR2's descriptions of what he hears when highly complex recordings of musical performances are input to various CIS processors.

A musical instrument for the profoundly deaf

In view of the growing number of people relying on CIS processors for hearing and understanding speech, research like that described in this report may find applications not only in improved speech processing strategies but also in musical instruments. We already know enough to contemplate the design of music synthesizers optimized for producing inputs to specific cochlear implant processors. Thus far, of course, our work on complex tone representation and perception via cochlear implants has been limited to a single research subject and we cannot be certain of the generality of our findings. One thing we have seen in that subject, however, we know to be true of others: a strong desire to experience music again with something of its remembered consonance, detail, and subtlety.

Such a synthesizer, generating an analog signal expressly for direct electrical input to a CIS processor, would produce music that need never exist in the form of mechanical vibrations or pressure waves. In the hands of an otherwise profoundly deaf musician on an everyday basis, the musical potential of a reasonably flexible synthesizer design could be explored much more fully and quickly than in a laboratory. With little effort, researchers could look over the shoulder of the musician as he or she synthesized auditory percepts ever closer to remembered musical instrument sounds. Analysis by synthesis is a well established technique, and the knowledge to be gained from such an exercise in this context is obvious. Most musicians would likely find more satisfaction, however, in creating satisfying new timbres without acoustical precedent, and compositions that exploit them. The opportunity of listening to the signals produced by such synthesis notwithstanding, the hearing researcher inevitably will feel handicapped by the inability ever to really know what the musician heard or intended. It is important to remember that such a difficulty is not at all unusual in music, especially in the case of compositions and performance practice from days before the existence of acoustic recording technology. It will be an interesting challenge to see both how well "acoustic" music can be conveyed to the user of a cochlear implant and how well the subtleties of her/his musical creations can be conveyed to people with "normal hearing" and other users of implants.

At this point in our evolving understanding, a musical synthesizer for users of CIS processors would be designed to sum separate analog signals, each intended exclusively for one of the user's processor channels. The simplest approach would be to use a general purpose personal computer to synthesize a set of envelopes, each of which would modulate a fixed-spectrum signal destined exclusively for one channel's passband. The bandwidth of each envelope should be constrained, of course, to conform to the smoothing filters in the processor and, beyond that, to avoid aliasing and to reflect the limits of the user's rate pitch saturation. For the lowest frequency bands, of course, the envelope becomes a waveform itself within the band being conveyed, and the precise nature of the modulated signal would become crucial. The personal computer would be equipped with a commercial sound card to provide the necessary analog output, ideally a card with an onboard digital signal processing capability of its own.

There are several interesting options for the interface by which the processor user would control such a music synthesizer. One possibility would be to provide an interface analogous to that of a traditional analog synthesizer, allowing specification of a combination of pure tone

Fourier components and noise bands, each with its attack-sustain-decay-release envelope generator (generally operating on much coarser time scales than those of the channel envelopes). This approach has the advantages of historical precedent and conceptual immediacy. Even a modest budget of four Fourier components and two noise bands per channel for a six channel processor, however, would amount to 36 operators, each with its own envelope generator. [As demonstrated in this report, we know that SR2, at least, can make use of at least that much complexity per channel.] Another approach would be to provide the user with the capability of graphically drawing envelope waveforms for each channel and/or free-hand editing of such waveforms from a library of computer files. The computer immediately should subject such hand-drawn or modified envelope waveforms to appropriate filtering and then redisplay them for further editing, so that the range of adjustment apparent in the user interface would correspond to what actually could be conveyed via the processor channel. Once a channel waveform was decided upon, the synthesizer could provide the new sound at a variety of pitches and in various combinations. The resulting signal amplitude range would be remapped to appropriately access the user's electrically evoked dynamic range via his/her processor. Initially, more complicated sounds would require non-real-time synthesis, but the delays could be brief enough on a phrase-by-phrase basis to support a satisfying interactive process. Commercial digital sampler hardware eventually could be provided with sample sounds tailored to CIS processor channel exclusivity criteria. At least one reasonably priced personal computer sound card already combines digital to analog conversion, digital file playback, considerable digital signal processing capability, and an onboard synthesizer using a ROM library of digitally sampled sounds.

Having such a synthesizer produce an analog signal for input to the user's processor rather than conveying a digital signal directly would have several advantages: The processor would protect the user from overstimulation, regardless of any error that might occur in the (creatively user-controlled) synthesizer. Typical adjustments to the processor's fitting parameters would not require any adjustment to the synthesizer. Essentially the same synthesizer could be used with a wide range of individual CIS processors, even ones running on different hardware implementations.

Realization of such a music synthesizer on a general purpose personal computer also would offer several advantages: A user could, upon achieving a closer match to a remembered instrument sound or a particularly interesting timbre, not only store it for future use and/or further modification, but also document it for later consideration by researchers, along with remarks and a verbal description of the sound. Modifications to the user interface, the synthesizer algorithms, and the library of envelope waveforms would be facilitated by use of a general purpose computer, as would exchange of favorite sounds among users.

Given the relatively modest cost of the additional hardware needed for such a synthesizer and the significant potential impact on the lives of many cochlear implant users, we are approaching the point in our research at which development and field testing of a prototype may be appropriate, if a funding source can be identified. We already have identified a potential volunteer user.

III. Glossary of Musical Terms

Bohlen-Pierce scale: a nontraditional musical scale that divides each factor of three in frequency (called a "tritave") into thirteen equal intervals, each corresponding to a frequency ratio equal to the thirteenth root of three. Used with complex tones containing only odd harmonics, this scale approximates a different set of small integer ratio intervals than traditional equal temperament [*e.g.* 5:3, 7:5, and 9:7].

equal temperament: the most common musical scale in use in our society; it divides each octave into twelve equal intervals (semitones), each corresponding to a frequency ratio equal to the twelfth root of two.

fundamental: the lowest frequency component of a harmonic series.

harmonics: components whose frequencies are integer multiples of some fundamental; *i.e.*, members of the harmonic series of that fundamental. The frequency of the nth harmonic is n times the fundamental frequency.

just intonation: requiring consonant musical intervals to minimize beating among upper partials by having frequency ratios that are exact ratios of small integers rather than the approximations that result from, say, an equal tempered scale. [The ratios 3:2, 4:3, 5:4, and 6:5 are, respectively, the just intonation perfect fifth, perfect fourth, major third, and minor third. The approximately equivalent intervals in equal temperament are 1.498, 1.335, 1.260, and 1.189.]

octave: the musical interval corresponding to a factor of two in frequency.

partial: any single frequency component of a complex sound, whether or not harmonically related to other components.

pure tone: a single sine wave.

semitone: a musical "half step"; the smallest pitch difference available, for instance, on a piano.

IV. References

D. K. Eddington (1994), private communication.

A. J. M. Houtsma, T. D. Rossing, and W. M. Wagenaars (1987), *Auditory Demonstrations* (Phillips Compact Disc #1126-061 and text), Acoustical Society of America, Woodbury, N. Y.

D. T. Lawson, M. Zerbi, and B. S. Wilson (1993) Speech Processors for Auditory Prostheses, *Fourth Quarterly Progress Report*, NIH project N01-DC-2-2401.

M. V. Mathews and J. R. Pierce (1989), The Bohlen-Pierce Scale, in *Current Directions in Computer Music Research*, M. V. Mathews and J. R. Pierce Eds., MIT Press, Cambridge, pp. 165-173.

B. C. J. Moore (1989), *An Introduction to the Psychology of Hearing, 2nd Edition,* Academic Press, London.

T. D. Rossing (1990), *The Science of Sound, 2nd Edition,* Addison-Wesley, New York.

B. S. Wilson, C. C. Finley, D. T. Lawson, R. O. Wolford, D. K. Eddington, and W. M. Rabinowitz (1991), Better speech recognition with cochlear implants, *Nature* **352**: 236-238.

B. S. Wilson, D. T. Lawson, and M. Zerbi (1994), Speech Processors for Auditory Prostheses, *Sixth Quarterly Progress Report*, NIH project N01-DC-2-2401.

B. S. Wilson, C. C. Finley, M. Zerbi, and D. T. Lawson (1994a), Speech Processors for Auditory Prostheses, *Seventh Quarterly Progress Report*, NIH project N01-DC-2-2401.

V. Appendix I

Attributes of Stimuli
Constituent Tones Identification Study

The 139 panels are numbered in the order in which they were presented and are summarized here one panel per line with the sequence number in the leftmost column.

The next six columns, under the heading "fundamentals" indicate the fundamental pitches on which the reference tone and each of the five candidate tones were based. The number shown in each case is the number of equal tempered semitones above A-110 Hz.

The next group of six columns, "harmonics", lists the harmonics of the respective fundamentals included in synthesizing the reference and each of the candidate tones. (The notation "23", for instance, for the reference tone in panel 1, indicates that the second and third harmonics of the fundamental are included. The frequency of that fundamental can be found from the reference tone column to the left; it is 10 semitones above A-110 Hz, *i.e.* G-196 Hz.)

In the final six columns, under "channels", are listed the processor channels in which each of those harmonics will be represented. (For the reference tone in panel 1, the notation "12" indicates that the lower partial [harmonic 2 of G-196] will be represented in channel 1 and the upper partial [harmonic 3 of the same fundamental] will appear in channel 2.)

This table makes it easy to identify the occurrence of various potential cues that might be expected to influence a subject's judgments, *e.g.* pairs of adjacent harmonics that influence a common channel *vs.* similar pairs that fall in separate channels.

Constituent Complex Tone Tests, December 1993 139 panels

	fundamentals------						harmonics------------------						channels---------------				
seq	ref	c1	c2	c3	c4	c5	ref	c1	c2	c3	c4	c5	ref	c1	c2	c3	c4	c5
1	10	10	10	10	10	10	23	2	3	4	5	6	12	1	2	2	3	3
2	10	10	10	10	10	10	46	2	3	4	5	6	23	1	2	2	3	3
3	10	10	10	10	10	10	45	2	3	4	5	6	23	1	2	2	3	3
4	10	10	10	10	10	10	35	2	3	4	5	6	23	1	2	2	3	3
5	10	10	10	10	10	10	34	2	3	4	5	6	22	1	2	2	3	3
6	10	10	10	10	10	10	48	2	3	4	5	6	24	1	2	2	3	3
7	10	10	10	10	10	10	56	2	3	4	5	6	33	1	2	2	3	3
8	10	10	10	10	10	10	24	2	3	4	5	6	12	1	2	2	3	3
9	10	10	10	10	10	10	234	2	3	4	5	6	122	1	2	2	3	3
10	10	10	10	10	10	10	468	2	3	4	5	6	234	1	2	2	3	3
11	10	10	10	10	10	10	456	2	3	4	5	6	233	1	2	2	3	3
12	10	10	10	10	10	10	568	2	3	4	5	6	334	1	2	2	3	3
13	10	10	10	10	10	10	256	2	3	4	5	6	133	1	2	2	3	3
14	10	10	10	10	10	10	4568	2	3	4	5	6	2334	1	2	2	3	3
15	10	10	10	10	10	10	3456	2	3	4	5	6	2233	1	2	2	3	3
16	10	10	10	10	10	10	23456	2	3	4	5	6	12233	1	2	2	3	3
17	10	10	10	10	10	10	23	23	34	45	56	78	12	12	22	23	33	44
18	10	10	10	10	10	10	46	23	34	45	56	78	23	12	22	23	33	44
19	10	10	10	10	10	10	45	23	34	45	56	78	23	12	22	23	33	44
20	10	10	10	10	10	10	35	23	34	45	56	78	23	12	22	23	33	44
21	10	10	10	10	10	10	34	23	34	45	56	78	22	12	22	23	33	44
22	10	10	10	10	10	10	48	23	34	45	56	78	24	12	22	23	33	44
23	10	10	10	10	10	10	56	23	34	45	56	78	33	12	22	23	33	44
24	10	10	10	10	10	10	24	23	34	45	56	78	12	12	22	23	33	44
25	10	10	10	10	10	10	234	23	34	45	56	78	122	12	22	23	33	44
26	10	10	10	10	10	10	468	23	34	45	56	78	234	12	22	23	33	44
27	10	10	10	10	10	10	456	23	34	45	56	78	233	12	22	23	33	44
28	10	10	10	10	10	10	568	23	34	45	56	78	334	12	22	23	33	44
29	10	10	10	10	10	10	256	23	34	45	56	78	133	12	22	23	33	44
30	10	10	10	10	10	10	4568	23	34	45	56	78	2334	12	22	23	33	44
31	10	10	10	10	10	10	3456	23	34	45	56	78	2233	12	22	23	33	44
32	10	10	10	10	10	10	23456	23	34	45	56	78	12233	12	22	23	33	44
33	10	10	10	10	10	10	23	24	35	46	48	78	12	12	23	23	24	44
34	10	10	10	10	10	10	46	24	35	46	48	78	23	12	23	23	24	44
35	10	10	10	10	10	10	45	24	35	46	48	78	23	12	23	23	24	44
36	10	10	10	10	10	10	35	24	35	46	48	78	23	12	23	23	24	44
37	10	10	10	10	10	10	34	24	35	46	48	78	22	12	23	23	24	44
38	10	10	10	10	10	10	48	24	35	46	48	78	24	12	23	23	24	44
39	10	10	10	10	10	10	56	24	35	46	48	78	33	12	23	23	24	44
40	10	10	10	10	10	10	24	24	35	46	48	78	12	12	23	23	24	44
41	10	10	10	10	10	10	234	24	35	46	48	78	122	12	23	23	24	44

	fundamentals------						harmonics-----------------						channels---------------					
seq	ref	c1	c2	c3	c4	c5	ref	c1	c2	c3	c4	c5	ref	c1	c2	c3	c4	c5
42	10	10	10	10	10	10	468	24	35	46	48	78	234	12	23	23	24	44
43	10	10	10	10	10	10	456	24	35	46	48	78	233	12	23	23	24	44
44	10	10	10	10	10	10	568	24	35	46	48	78	334	12	23	23	24	44
45	10	10	10	10	10	10	256	24	35	46	48	78	133	12	23	23	24	44
46	10	10	10	10	10	10	4568	24	35	46	48	78	2334	12	23	23	24	44
47	10	10	10	10	10	10	3456	24	35	46	48	78	2233	12	23	23	24	44
48	10	10	10	10	10	10	23456	24	35	46	48	78	12233	12	23	23	24	44
49	10	10	10	10	10	10	256	6	5	4	3	2	133	3	3	2	2	1
50	10	10	10	10	10	10	468	6	5	4	3	2	234	3	3	2	2	1
51	10	10	10	10	10	10	23	6	5	4	3	2	12	3	3	2	2	1
52	10	10	10	10	10	10	46	6	5	4	3	2	23	3	3	2	2	1
53	10	10	10	10	10	10	45	6	5	4	3	2	23	3	3	2	2	1
54	10	10	10	10	10	10	35	6	5	4	3	2	23	3	3	2	2	1
55	10	10	10	10	10	10	34	6	5	4	3	2	22	3	3	2	2	1
56	10	10	10	10	10	10	48	6	5	4	3	2	24	3	3	2	2	1
57	10	10	10	10	10	10	56	6	5	4	3	2	33	3	3	2	2	1
58	10	10	10	10	10	10	24	6	5	4	3	2	12	3	3	2	2	1
59	10	10	10	10	10	10	234	6	5	4	3	2	122	3	3	2	2	1
60	10	10	10	10	10	10	468	6	5	4	3	2	234	3	3	2	2	1
61	10	10	10	10	10	10	456	6	5	4	3	2	233	3	3	2	2	1
62	10	10	10	10	10	10	568	6	5	4	3	2	334	3	3	2	2	1
63	10	10	10	10	10	10	256	6	5	4	3	2	133	3	3	2	2	1
64	10	10	10	10	10	10	4568	6	5	4	3	2	2334	3	3	2	2	1
65	10	10	10	10	10	10	3456	6	5	4	3	2	2233	3	3	2	2	1
66	10	10	10	10	10	10	23456	6	5	4	3	2	12233	3	3	2	2	1
67	10	10	10	10	10	10	23	78	56	45	34	23	12	44	33	23	22	12
68	10	10	10	10	10	10	46	78	56	45	34	23	23	44	33	23	22	12
69	10	10	10	10	10	10	45	78	56	45	34	23	23	44	33	23	22	12
70	10	10	10	10	10	10	35	78	56	45	34	23	23	44	33	23	22	12
71	10	10	10	10	10	10	34	78	56	45	34	23	22	44	33	23	22	12
72	10	10	10	10	10	10	48	78	56	45	34	23	24	44	33	23	22	12
73	10	10	10	10	10	10	56	78	56	45	34	23	33	44	33	23	22	12
74	10	10	10	10	10	10	24	78	56	45	34	23	12	44	33	23	22	12
75	10	10	10	10	10	10	234	78	56	45	34	23	122	44	33	23	22	12
76	10	10	10	10	10	10	468	78	56	45	34	23	234	44	33	23	22	12
77	10	10	10	10	10	10	456	78	56	45	34	23	233	44	33	23	22	12
78	10	10	10	10	10	10	568	78	56	45	34	23	334	44	33	23	22	12
79	10	10	10	10	10	10	256	78	56	45	34	23	133	44	33	23	22	12
80	10	10	10	10	10	10	4568	78	56	45	34	23	2334	44	33	23	22	12
81	10	10	10	10	10	10	3456	78	56	45	34	23	2233	44	33	23	22	12
82	10	10	10	10	10	10	23456	78	56	45	34	23	12233	44	33	23	22	12
83	10	10	10	10	10	10	23	78	48	46	35	24	12	44	24	23	23	12

fundamentals--------							harmonics--------------------						channels--------------------					
seq	ref	c1	c2	c3	c4	c5	ref	c1	c2	c3	c4	c5	ref	c1	c2	c3	c4	c5
84	10	10	10	10	10	10	46	78	48	46	35	24	23	44	24	23	23	12
85	10	10	10	10	10	10	45	78	48	46	35	24	23	44	24	23	23	12
86	10	10	10	10	10	10	35	78	48	46	35	24	23	44	24	23	23	12
87	10	10	10	10	10	10	34	78	48	46	35	24	22	44	24	23	23	12
88	10	10	10	10	10	10	48	78	48	46	35	24	24	44	24	23	23	12
89	10	10	10	10	10	10	56	78	48	46	35	24	33	44	24	23	23	12
90	10	10	10	10	10	10	24	78	48	46	35	24	12	44	24	23	23	12
91	10	10	10	10	10	10	234	78	48	46	35	24	122	44	24	23	23	12
92	10	10	10	10	10	10	468	78	48	46	35	24	234	44	24	23	23	12
93	10	10	10	10	10	10	456	78	48	46	35	24	233	44	24	23	23	12
94	10	10	10	10	10	10	568	78	48	46	35	24	334	44	24	23	23	12
95	10	10	10	10	10	10	256	78	48	46	35	24	133	44	24	23	23	12
96	10	10	10	10	10	10	4568	78	48	46	35	24	2334	44	24	23	23	12
97	10	10	10	10	10	10	3456	78	48	46	35	24	2233	44	24	23	23	12
98	10	10	10	10	10	10	23456	78	48	46	35	24	12233	44	24	23	23	12
99	10	10	10	10	10	10	256	2	3	4	5	6	133	1	2	2	3	3
100	10	10	10	10	10	10	468	2	3	4	5	6	234	1	2	2	3	3
101	10	10	11	10	11	11	23	2	2	3	3	4	12	1	1	2	2	2
102	10	10	11	11	10	11	35	3	3	4	5	5	23	2	2	2	3	3
103	11	10	11	10	11	11	23	3	2	3	3	4	12	2	1	2	2	2
104	11	10	11	11	10	11	35	3	3	4	5	5	23	2	2	2	3	3
105	10	10	11	10	11	11	234	2	2	3	3	4	122	1	1	2	2	2
106	11	10	11	10	11	11	234	2	2	3	3	4	122	1	1	2	2	2
107	11	10	11	10	11	10	3456	34	34	45	45	56	2233	22	22	23	23	33
108	10	10	11	10	11	10	3456	34	34	45	45	56	2233	22	22	23	23	33
109	10	10	11	10	10	11	3456	34	34	45	56	56	2233	22	22	23	33	33
110	11	10	11	10	10	11	3456	34	34	45	56	56	2233	22	22	23	33	33
111	10	10	11	10	11	10	23456	23	23	34	34	45	12233	12	12	22	22	23
112	11	10	11	10	11	10	23456	23	23	34	34	45	12233	12	12	22	22	23
113	10	11	10	11	10	11	23456	23	34	34	56	56	12233	12	22	22	33	33
114	11	11	10	11	10	11	23456	23	34	34	56	56	12233	12	22	22	33	33
115	10	10	12	14	15	18	456	56	56	56	56	56	233	33	33	34	34	44
116	22	16	17	22	23	24	12345678	67	67	67	67	67	12344555	44	44	55	55	55
117	22	16	17	22	23	24	123456	67	67	67	67	67	123445	44	44	55	55	55
118	22	16	17	22	23	24	1234	67	67	67	67	67	1234	44	44	55	55	55
119	30	24	30	31	32	33	12345678	67	67	67	67	67	23455666	55	66	66	66	66
120	30	24	30	31	32	33	12345	67	67	67	67	67	23455	55	66	66	66	66
121	30	24	30	31	32	33	123	67	67	67	67	67	234	55	66	66	66	66
122	03	00	01	03	05	07	3456	67	56	56	45	45	1122	22	22	22	22	22
123	03	01	03	05	07	10	3456	56	56	45	45	34	1122	22	22	22	22	22
124	06	01	03	05	07	10	4567	56	56	45	45	34	2233	22	22	22	22	22
125	06	03	04	06	08	10	4567	78	78	67	67	56	2233	33	33	33	33	33

	fundamentals--------						harmonics------------------						channels-------------------				
seq	ref	c1	c2	c3	c4	c5	ref	c1	c2	c3	c4	c5	ref	c1	c2	c3	c4	c5
126	06	03	06	10	13	18	4567	78	67	56	45	34	2233	33	33	33	33	33
127	11	08	11	13	15	18	3456	67	56	45	45	34	2233	33	33	33	33	33
128	15	08	11	13	15	18	4567	67	56	45	45	34	3344	33	33	33	33	33
129	15	11	13	15	17	19	4567	78	78	67	56	56	3344	44	44	44	44	44
130	15	12	15	18	22	26	4567	78	67	56	45	34	3344	44	44	44	44	44
131	19	15	17	19	22	26	3456	67	56	56	45	34	3344	44	44	44	44	44
132	22	15	17	19	22	26	4567	67	56	56	45	34	4455	44	44	44	44	44
133	22	19	21	22	24	26	4567	78	78	67	67	56	4455	55	55	55	55	55
134	22	19	22	25	29	34	4567	78	67	67	45	34	4455	55	55	55	55	55
135	26	22	24	26	29	34	3456	67	67	56	45	34	4455	55	55	55	55	55
136	30	22	24	26	30	34	4567	67	67	56	45	34	5566	55	55	55	55	55
137	30	27	29	30	32	34	4567	78	78	67	67	56	5566	66	66	66	66	66
138	30	27	30	33	36	39	4567	78	67	56	56	45	5566	66	66	66	66	66
139	34	27	30	34	37	42	3456	78	67	56	45	34	5566	66	66	66	66	66

VI. Appendix II

Analysis Matrices
Constituent Tones Identification Study

There is a line for each of the panels, numbered as in Appendix I.

Within each line are five groups, each with five binary digits.

The groups correspond to selected attributes of the stimuli: (from left to right) fundamentals, harmonics, channels, pairs of adjacent harmonics within a common channel (adj same), and pairs of adjacent harmonics in separate channels (adj difft).

The five digits within each group correspond to the five candidate stimuli.

A given digit will be set to 1 if the attribute it represents is shared by the candidate tone it represents and the corresponding reference tone. (If a particular candidate tone contains harmonics 3 and 4 of its fundamental, then the harmonic attribute digit for that candidate will be 1 only if the reference tone also contains harmonics 3 and 4. For purposes of the value of that attribute digit, it does not matter whether there are additional harmonics in the reference. Nor does it matter what the reference's fundamental frequency is.)

The information in these matrices also can be read directly from the corresponding lines of Appendix I.

	funds	harms	chans	adj same	adj difft
1	11111	11000	11100	00000	00000
2	11111	00101	01111	00000	00000
3	11111	00110	01111	00000	00000
4	11111	01010	01111	00000	00000
5	11111	01100	01100	00000	00000
6	11111	00100	01100	00000	00000
7	11111	00011	00011	00000	00000
8	11111	10100	11100	00000	00000
9	11111	11100	11100	00000	00000
10	11111	00101	01111	00000	00000
11	11111	00111	01111	00000	00000
12	11111	00011	00011	00000	00000
13	11111	10011	10011	00000	00000
14	11111	00111	01111	00000	00000
15	11111	01111	01111	00000	00000
16	11111	11111	11111	00000	00000
17	11111	10000	11000	01011	10100
18	11111	00000	01110	01011	10100
19	11111	00100	01110	01011	10100
20	11111	00000	01110	01011	10100
21	11111	01000	01000	01011	10100
22	11111	00000	01001	01011	10100
23	11111	00010	00010	01011	10100
24	11111	00000	11000	01011	10100
25	11111	11000	11000	01011	10100
26	11111	00000	01111	01011	10100
27	11111	00110	01110	01011	10100
28	11111	00010	00011	01011	10100
29	11111	00010	00010	01011	10100
30	11111	00110	01111	01011	10100
31	11111	01110	01110	01011	10100
32	11111	11110	11110	01011	10100
33	11111	00000	10000	00001	00000
34	11111	00100	01100	00001	00000
35	11111	00000	01100	00001	00000
36	11111	01000	01100	00001	00000
37	11111	00000	00000	00001	00000
38	11111	00010	00010	00001	00000
39	11111	00000	00000	00001	00000
40	11111	10000	10000	00001	00000
41	11111	10000	10000	00001	00000
42	11111	00110	01110	00001	00000

	funds	harms	chans	adj same	adj difft
43	11111	00100	01100	00001	00000
44	11111	00000	00001	00001	00000
45	11111	00000	00000	00001	00000
46	11111	00110	01111	00001	00000
47	11111	01100	01100	00001	00000
48	11111	11100	11100	00001	00000
49	11111	11001	11001	00000	00000
50	11111	10100	11110	00000	00000
51	11111	00011	00111	00000	00000
52	11111	10100	11110	00000	00000
53	11111	01100	11110	00000	00000
54	11111	01010	11110	00000	00000
55	11111	00110	00110	00000	00000
56	11111	00100	00110	00000	00000
57	11111	11000	11000	00000	00000
58	11111	00101	00111	00000	00000
59	11111	00111	00111	00000	00000
60	11111	10100	11110	00000	00000
61	11111	11100	11110	00000	00000
62	11111	11000	11000	00000	00000
63	11111	11001	11001	00000	00000
64	11111	11100	11110	00000	00000
65	11111	11110	11110	00000	00000
66	11111	11111	11111	00000	00000
67	11111	00001	00011	00000	00101
68	11111	00000	01110	11010	00101
69	11111	00100	01110	11010	00101
70	11111	00000	01110	11010	00101
71	11111	00010	00010	11010	00101
72	11111	00000	10010	11010	00101
73	11111	01000	01000	11010	00101
74	11111	00000	00011	11010	00101
75	11111	00011	00011	11010	00101
76	11111	00000	11110	11010	00101
77	11111	01100	01110	11010	00101
78	11111	01000	11000	11010	00101
79	11111	01000	01000	11010	00101
80	11111	01100	11110	11010	00101
81	11111	01110	01110	11010	00101
82	11111	01111	01111	11010	00101
83	11111	00000	00001	10000	00000
84	11111	00100	00110	10000	00000

	funds	harms	chans	adj same	adj difft
85	11111	00000	00110	10000	00000
86	11111	00010	00110	10000	00000
87	11111	00000	00000	10000	00000
88	11111	01000	11000	10000	00000
89	11111	00000	00000	10000	00000
90	11111	00001	00001	10000	00000
91	11111	00001	00001	10000	00000
92	11111	01100	11110	10000	00000
93	11111	00100	00110	10000	00000
94	11111	00000	10000	10000	00000
95	11111	00000	00000	10000	00000
96	11111	01100	10110	10000	00000
97	11111	00110	00110	10000	00000
98	11111	00111	00111	10000	00000
99	11111	10011	10011	00000	00000
100	11111	00101	01111	00000	00000
101	10100	11110	11111	00000	00000
102	10010	11011	11111	00000	00000
103	01011	11110	11111	00000	00000
104	01101	11011	11111	00000	00000
105	10100	11111	11111	00000	00000
106	01011	11111	11111	00000	00000
107	01010	11111	11111	11001	00110
108	10101	11111	11111	11001	00110
109	10110	11111	11111	11011	00100
110	01001	11111	11111	11011	00100
111	10101	11111	11111	00110	11001
112	01010	11111	11111	00110	11001
113	01010	11111	11111	01111	10000
114	10101	11111	11111	01111	10000
115	10000	11111	11000	11001	00110
116	00100	11111	11111	11111	00000
117	00100	00000	11111	11111	00000
118	00100	00000	11000	11111	00000
119	01000	11111	11111	11111	00000
120	01000	00000	10000	11111	00000
121	01000	00000	00000	11111	00000
122	00100	01111	11111	11111	00000
123	01000	11111	11111	11111	00000
124	00000	11110	11111	11111	00000
125	00100	00111	11111	11111	00000
126	01000	01110	11111	11111	00000

	funds	harms	chans	adj same	adj difft
127	01000	01111	11111	11111	00000
128	00010	11110	11111	11111	00000
129	00100	00111	11111	11111	00000
130	01000	01110	11111	11111	00000
131	00100	01111	11111	11111	00000
132	00010	11110	11111	11111	00000
133	00100	00111	11111	11111	00000
134	01000	01110	11111	11111	00000
135	00100	00111	11111	11111	00000
136	00010	11110	11111	11111	00000
137	00100	00111	11111	11111	00000
138	01000	00111	11111	11111	00000
139	00100	00111	11111	11111	00000

VII. Plans for the Next Quarter

Our plans for the next quarter include the following:

1. Continued studies with Ineraid subject SR2, primarily to evaluate variations of CIS processors and to extend the range of stimuli for evoked potential recordings.
2. Initial studies with the second patient in the Nucleus percutaneous series (NP-2) and continued studies with the first patient (NP-1). Studies with NP-2 will include evaluations of CIS and *spectral peak* (SPEAK) processing strategies. Studies with NP-1 will include repeated measures with the SPEAK strategy, detailed evaluation of CIS processors using more than six channels, and measures of intracochlear evoked potentials.
3. Presentation of project results at the *25th Annual Neural Prosthesis Workshop*.
4. Continued analysis of EP and speech reception data from prior studies.
5. Continued preparation of manuscripts for publication.

VIII. Acknowledgments

We thank subject SR2 for his enthusiastic participation and insightful reporting.

Appendix A

Summary of Reporting Activity for the Period of

May 1 through July 31, 1994

NIH Project N01-DC-2-2401

Reporting activity for the last quarter included presentation of an invited lecture. The citation is listed below and the published abstract for that talk is reproduced on the next page.

Wilson BS: Progress in the development of speech processors for cochlear prostheses. Presented in the special session on "Electro-Auditory Prostheses," *127th Meeting of the Acoustical Society of America*, Cambridge, MA, June 8, 1994. [Abstract published in J. Acoust. Soc. Am. 95: 2905, 1994.]

9:35

3aSP3. Speech understanding in adult cochlear implant users. Richard S. Tyler, Mary Lowder, George Woodworth, and Aaron Parkinson (Dept. of Otolaryngol., Univ. of Iowa, 200 Hawkins Dr., Iowa City, IA 52242-1078)

This review discusses the speech perception results obtained from postlingually deafened adults with cochlear implants using a variety of speech coding strategies and electrode configurations. The speech feature information obtained from feature and whole-signal processing, and single and multichannel stimulation will be examined. Results will be discussed from implants developed in Innsbruck, London, Melbourne, Los Angeles, Paris, Duren, Utah, and San Francisco. Speech perception tests include simple pattern perception, audiovisual recognition, and the understanding of words without visual or contextual clues. Some patients perform poorly, whereas others are able to converse freely on the telephone. For example, word recognition from unknown lists of 50 words can range from 0% to 78% correct. Factors that contribute to successful implant use will be examined. Older patients and patients who have been deafened for long periods of time tend not to perform as well as recently deafened younger patients. The rate of learning over a five-year period shows large individual differences. Most patients show large improvements over the first 6 to 12 months, but some show continued real gains over 4 years.

10:05

3aSP4. Speech perception and production results in children with multichannel cochlear implants. Mary Joe Osberger (Dept. of Otolaryngol., Indiana Univ. School of Medicine, Riley Res., Rm. 044, Indianapolis, IN 46202)

The results from a number of studies in our laboratory will be presented to demonstrate the benefits that children with profound hearing impairments derive from multichannel cochlear implants. The performance of children with cochlear implants has been evaluated over time and compared to that of children who used conventional hearing aids, grouped according to unaided thresholds. Results showed that speech perception and production skills developed over a time course as long as 5 years in children with prelingual deafness who received multichannel cochlear implants. After roughly 3 years, the scores of prelingually deafened children with multichannel implants were higher than those of children with hearing aids with unaided thresholds of 101–110 dB HL and similar to those of children with unaided thresholds of 90–100 dB HL. New research directions involve examination of lexical effects on word recognition in multichannel implant users. Data suggest that pediatric cochlear implant users are sensitive to the acoustic-phonetic similarities among words, that they organize words into similarity neighborhoods in long-term lexical memory, and that they use this structural information in recognizing isolated words. [Work supported by NIH.]

10:35

3aSP5. Progress in the development of speech processors for cochlear prostheses. Blake S. Wilson (Neurosci. Prog. Res. Triangle Inst., Res. Triangle Park, NC 27709 and Div. of Otolaryngol., Duke Univ. Med. Ctr., Durham, NC 27710)

New strategies for representing acoustic information with multichannel cochlear implants have produced substantial improvements in speech recognition scores for implant users. One of those strategies, *continuous interleaved sampling* (CIS), presents brief pulses in a nonoverlapping sequence across electrode channels, with the pulse amplitudes for each channel reflecting the envelope of a corresponding frequency band of the acoustic input. Recent studies with CIS and related processing strategies will be described, including (a) within-subject comparisons of CIS with the *compressed analog* (CA) processor used in a standard clinical device, (b) parametric and control studies with CIS processors, and (c) a preliminary evaluation of a related strategy, *virtual channel interleaved sampling* (VCIS). VCIS processors add to the single-electrode channels of CIS processors virtual channels, produced by simultaneous stimulation of two or more electrodes and eliciting pitch percepts that are distinct from those of single-electrode channels. In general, the CIS/CA comparisons show higher levels of open-set speech recognition with CIS for each of 11 subjects. Results from the additional studies with CIS processors show how choices of pulse duration, pulse rate, and channel update order affect performance. Initial CIS/VCIS comparisons with three subjects do not show an immediate improvement in speech recognition scores with VCIS. [Work supported by NIH.]

3a WED. AM

11:05–11:15 Break

Contributed Papers

11:15

3aSP6. Mathematical modeling of vowel perception by users of the Ineraid cochlear implant: Temporal and place-of-stimulation cues. Mario A. Svirsky (Speech Commun. Group, Res. Lab. of Electron., MIT, 50 Vassar St., Rm. 36-525, Cambridge, MA 02139)

The Ineraid cochlear implant filters the speech spectrum into four frequency bands and delivers each signal to one of four electrodes in the cochlea. Higher frequency bands stimulate more basally placed electrodes. Ineraid users may recognize vowels using temporal cues, place-of-stimulation cues (such as amount of energy sent to each electrode) or a combination of place and temporal cues. A mathematical model is proposed to account for vowel perception by users of the Ineraid cochlear implant. The model is computationally similar to the one that was proposed for users of pulsatile cochlear implants [M. A. Svirsky and S. H. Svirsky, J. Acoust. Soc. Am. 92(4), 2415–6 (1992)], but the perceptual dimensions involved are different in the two cases. To test the model vowel identification data obtained with "conflicting cue" vowels was used [Dorman *et al.*, J. Acoust. Soc. Am. 92(6), 3428–31 (1992)]. The model was run under three different assumptions: Ineraid users only employ temporal cues for vowel perception, they only employ place cues, or they combine temporal and place cues. Model output fit the data well only under the temporal/place assumption. This suggests that Ineraid users combine information from temporal and place cues in order to recognize vowels. [Work supported by NIH Grant No. R03-DC01721 and Contract No. N01-DC-2-2402.]

11:30

3aSP7. Electrode discrimination measures: Relationship with speech perception and clinical applicability of results. Leslie M. Collins (Dept. of Elec. Eng. and Comput. Sci., 1301 Beal Ave., Univ. of Michigan, Ann Arbor, MI 48109), Teresa A. Zwolan, and Gregory H. Wakefield (Univ. of Michigan, Ann Arbor, MI 48109)

Appendix B

Notice of Errata to Fourth Quarterly Progress Report

page 25, Table: Harmonic 7 should not be listed as meeting the criteria in channel 4 for G-196 Hz, and should be listed in that channel for G#-208 Hz.

pages 32 and 33: The labels and contents for Figures 11 and 12 were interchanged.