Sixth Quarterly Progress Report

July 1 through September 30, 2003
NIH Project N01-DC-2-1002

Speech Processors for Auditory Prostheses

Prepared by

Reinhold Schatzer, Blake Wilson,

Robert Wolford and Dewey Lawson

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

CONTENTS

I. Introduction

The main objective of this project is to design, develop, and evaluate speech processors for implantable auditory prostheses. Ideally, such processors will represent the information content of speech in a way that can be perceived and utilized by implant patients. An additional objective is to record responses of the auditory nerve to a variety of electrical stimuli in studies with patients. Results from such recordings can provide important information on the physiological function of the nerve, on an electrode-by-electrode basis, and can be used to evaluate the ability of speech processing strategies to produce desired spatial or temporal patterns of neural activity.

Work and activities in this quarter included:

- Continuing studies with local subjects ME16 and ME22, implanted bilaterally with Med-El Tempo+ devices.
- Three weeks of studies with return subject ME-15, implanted bilaterally with Med-El Tempo+ devices, July 14 – August 1.
- A return visit by consultant Enrique Lopez-Poveda from Salamanca, Spain, July 17-30, coinciding with the visit of subject ME-15.
- A visit by Craig Buchman, MD, Chief, Division of Otology/Neurotology and Skull Base Surgery, UNC-CH, Chapel Hill, NC, August 1.
- Two weeks of studies with return subject ME-3, implanted bilaterally with Med-El Tempo+ devices, August 4-15.
- Attendance by Xiaoan Sun at a training course on “Designing for Performance with Xilinx FPGA” in Raleigh NC, August 13-14.
- Attendance by Blake Wilson, Dewey Lawson, Lianne Cartee, Reinhold Schatzer, Xiaoan Sun and Robert Wolford at the 2003 Conference on Implantable Auditory Prostheses (CIAP 2003), Asilomar CA, August 17-22.
- A presentation by Blake Wilson on “Evaluation of combined EAS in studies at the Research Triangle Institute” at CIAP 2003.
- A poster presentation by Reinhold Schatzer on “A Novel CI speech processing structure for closer mimicking of normal auditory functions” at CIAP 2003.
- Two weeks of studies with return subject ME-10, implanted bilaterally with Med-El Tempo+ devices, August 25-September 5.
- One week of studies with percutaneous return subject SR-9, September 15-19.
- A visit by consultant Marian Zerbi, September 13-15.
- Initial surgeries for two of the Nucleus Contour Electrode percutaneous study patients at Duke University Medical Center, Durham, NC.

- Two weeks of combined electric and acoustic studies with return subject ME-19, September 22-October 3.

In addition to the above-mentioned activities, work continued on analyses of previously collected data and on the preparation of manuscripts for publication. In addition, initial experiments were conducted during the quarter to evaluate several possibilities for increasing the dynamic range of stimulation with cochlear implants. The data from these and planned experiments will be presented in a future report.

In the present report we provide a discussion of preliminary studies with a signal processing strategy that is aimed at better mimicking the functions observed in the inner ear in normal hearing. Thus far, the strategy is based on a model of a nonlinear human cochlear filter bank, which replaces the linear filter bank found in clinical cochlear implant speech processors. The cochlear filter bank closely mimics the properties of the human basilar membrane in terms of frequency selectivity and nonlinearities observed in the membrane response.

Future activities will be focused on complementing the current processing strategy implementation with a model for mimicking the nonlinear compression and adaptation properties of the Inner Hair Cell (IHC) and auditory nerve synapse.

II. Signal processing strategies for a closer mimicking of normal auditory functions

Background

In the normal hearing process, the peripheral auditory system includes several highly nonlinear stages, as illustrated in the simplified block diagram in Figure 1. The input from the stapes is filtered in a nonlinear way by the basilar membrane and its associated structures. The nonlinearity is produced by an active feedback loop involving electromotile contractions of the outer hair cells (e.g., Dallos, 1992). It is absent in many cases of sensorineural hearing loss, as a consequence of damage to or destruction of the outer hair cells.

Movements of the basilar membrane are sensed by the inner hair cells (IHCs), which excite adjacent Type I fibers of the auditory nerve through release of chemical transmitter substance in the synaptic cleft between an IHC and 10-20 Type I fibers. The IHC membrane response rectifies and compresses the signal from the basilar membrane movements, and the IHC membrane response also attenuates strongly frequencies above about 1 kHz. A further, noninstantaneous compression occurs at the synapse. The compression is strong and has at least two time constants, one on the order of 5 ms and the other on the order 40 ms or more.

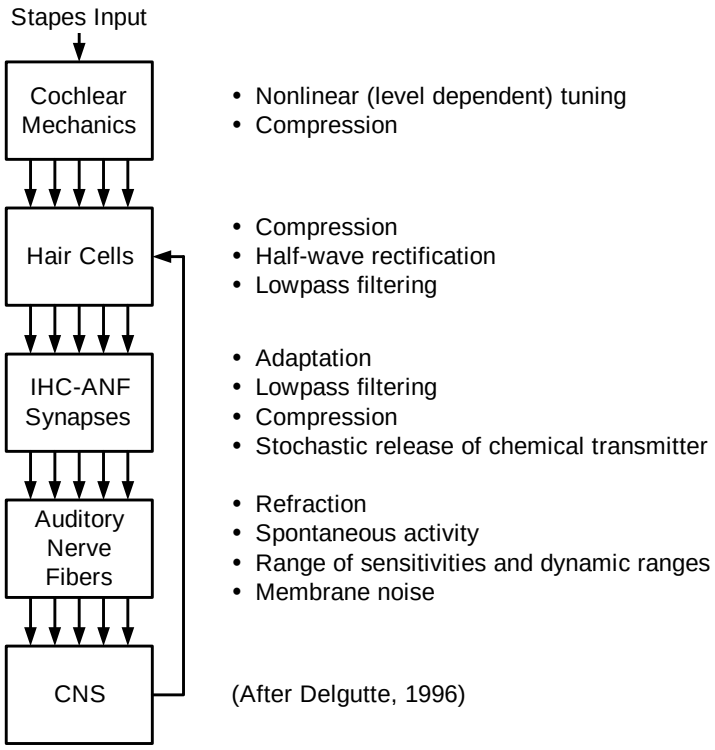

Figure 1: Simplified block diagram of the normal auditory periphery. Abbreviations: IHC-ANF – Inner Hair Cell-Auditory Nerve Fiber, CNS – Central Nervous System. Adapted from Delgutte, 1996.

Processes and properties of the auditory periphery at the level of the nerve fibers include the stochastic release of chemical transmitter substance into the synaptic cleft (giving rise to the spontaneous activity of the neurons), the differences in sensitivities and dynamic ranges of auditory nerve fibers, as well as their refraction properties and membrane noise originating at the nodes of Ranvier. The spontaneous discharge rate of neurons correlates to their sensitivity and dynamic range: Fibers with low rates of spontaneous discharge have high thresholds and relatively wide dynamic ranges, high spontaneous rate fibers have low thresholds and relatively narrow dynamic ranges, and medium spontaneous rate fibers have intermediate thresholds and dynamic ranges.

A simple model of processing in the peripheral auditory system is presented in Figure 2. The block for modeling the responses of the basilar membrane at a particular point along the cochlear partition includes a feedback loop whose output controls the sharpness (Q factor) and gain of a bandpass filter. The time-varying tuning and gain approximates the time-varying (and amplitude dependent) tuning and gain of the basilar membrane.

The output of the time-varying bandpass filter is rectified and lowpass filtered, to reflect processing at the IHC membrane. The corner frequency of the lowpass filter is 1 kHz, matching that of the membrane response. Signal processing at the IHC/neuron synapse is modeled with three feedback loops. The time constants of the loops are selected to approximate the properties of adaptation in normal hearing.

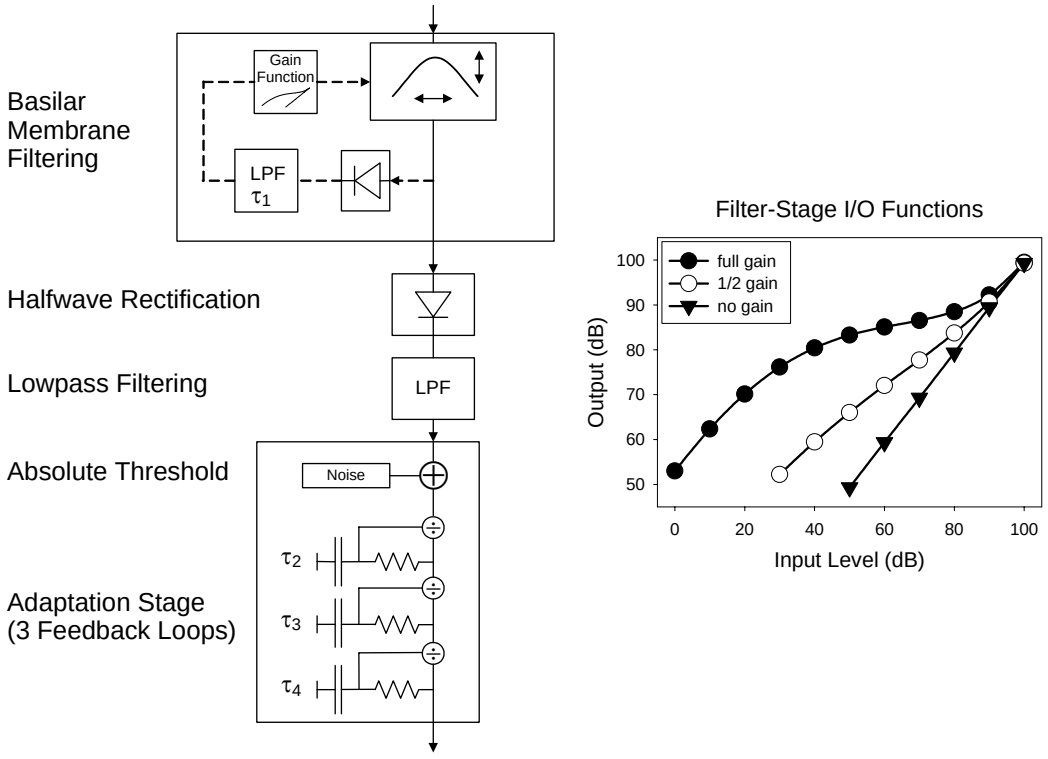

Figure 2: Model of the "effective" signal processing in the auditory periphery (left panel) and input/output (I/O) functions (right panel) for the nonlinear response of the basilar membrane filter stage in that model. Typical values for the time constants in the model are 15 ms for τ_1, 129 ms for τ_2, 253 ms for τ_3 and 500 ms for τ_4. The cutoff frequency for the lowpass filter in the forward path usually is set at 1 kHz, corresponding to the corner frequency of the IHC membrane. Effects of different gains in the feedback loop for the basilar-membrane filtering stage are illustrated in the right panel. A highly nonlinear response is observed with full gain, whereas a linear response is observed with no gain (effectively removing the feedback loop from the model). Adapted from Kollmeier et al., 1998.

The panel to the right in Figure 2 shows one result of the nonlinearities in the responses of the basilar membrane and associated structures – a nonlinear growth of response at the characteristic frequency (CF) or most-responsive frequency of the equivalent basilar-membrane filter for a sinusoidal input. The response at CF is approximately linear up to sound pressure levels (SPLs) of about 30 dB, but becomes highly nonlinear for SPLs between about 30 and 80 dB SPL.

Also shown in the panel are growth functions associated with reduced amounts of feedback and nonlinearity in the model. Those latter conditions approximate the situation in many cases of sensorineural hearing loss, in which the outer hair cells are damaged or missing. The response with the feedback set at zero is perfectly linear, as would be expected. The higher threshold of response, along with the relatively rapid growth of response for higher input levels, is consistent with the high threshold of response and rapid growth of loudness found in cases of sensorineural hearing loss, the phenomenon of "loudness recruitment."

The nonlinear responses at the basilar membrane and associated structures produce sharp tuning at low input levels, and broader tuning and a downward shift in the most-responsive frequency at higher input levels. The response of the membrane complex is highly nonlinear at and near the spatial position of the CF and roughly linear at other places. In normal hearing, responses at one place along the basilar membrane influence responses at other places. Such lateral or longitudinal interactions among equivalent filters are not included in the model of Figure 2.

Present processing strategies for cochlear implants, such as the Continuous Interleaved Sampling (CIS) strategy shown in Figure 3, provide only a very crude approximation to processing in the normal cochlea. For example, a bank of linear bandpass filters is used instead of the nonlinear and coupled filters that would model normal auditory function. Also, a single nonlinear map is used to produce the overall compression that the normal system achieves in multiple steps. The compression in the standard CIS processor is instantaneous, whereas compression at the IHC/neuron synapse in normal hearing is noninstantaneous, with large adaptation effects.

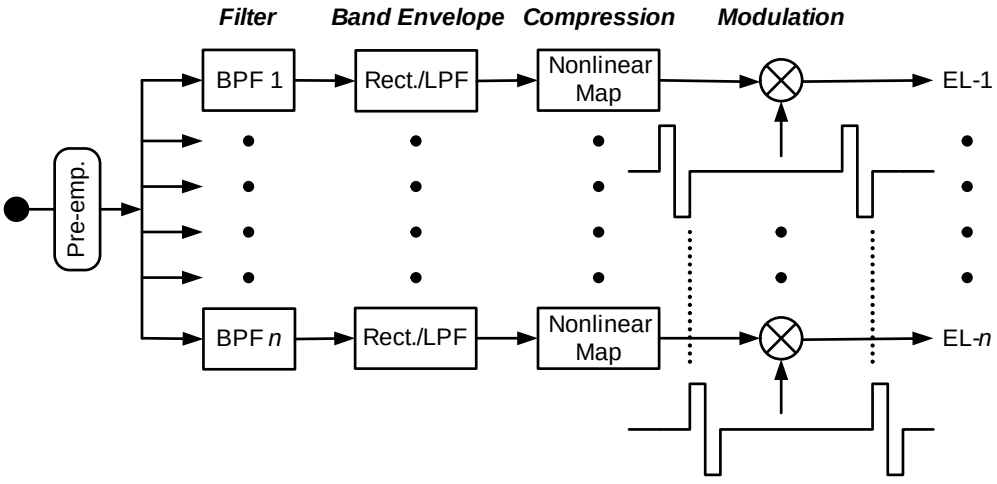

Figure 3: Standard CIS processor structure. Abbreviations: Pre-emp. – Pre-emphasis, BPF – Bandpass Filter, Rect. – Rectifier, LPF – Lowpass Filter, EL – Electrode. Adapted from Wilson et al., 1991.

Deng and Geisler (1987), among others, have shown that nonlinearities in basilar membrane (BM) filtering greatly enhance the neural representation of speech sounds presented in competition with noise. Similarly, results presented by Tchorz and Kollmeier (1999) have indicated the importance of adaptation

at the IHC/neuron synapse in representing temporal events in speech, especially for speech presented in competition with noise. Based on these findings, a new processor structure, designed to provide a closer mimicking of normal auditory functions, is suggested in Figure 4.

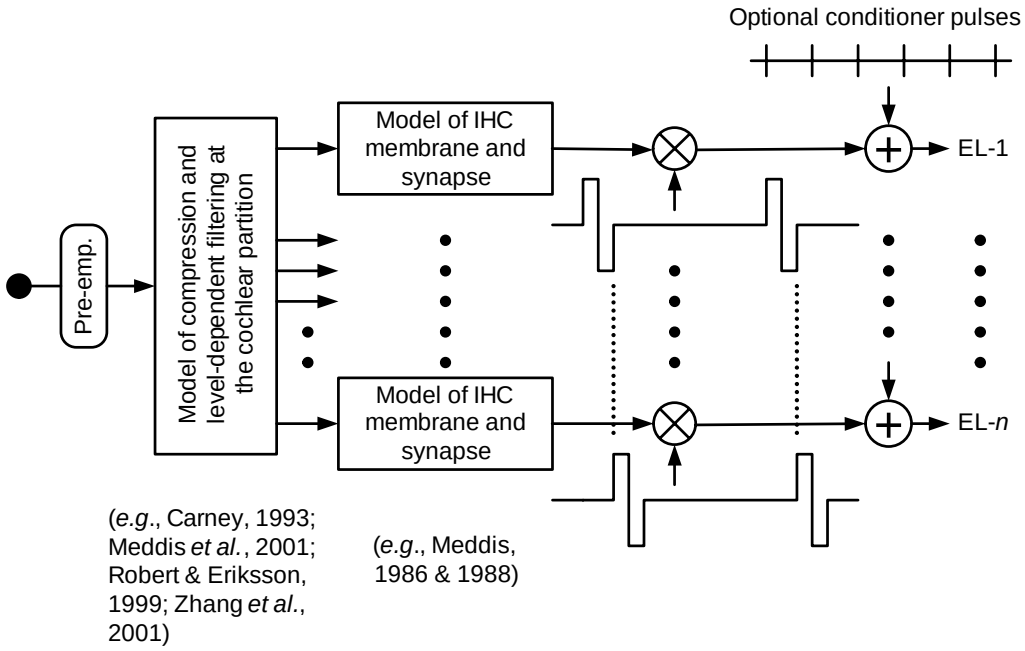

Figure 4: Processor structure for a closer mimicking of normal auditory functions. Possible models to be incorporated into a speech processor design are listed beneath the corresponding blocks.

The structure incorporates the nonlinearities observed in the healthy inner ear, and it does so by utilizing models that have been developed to describe and understand the normal functions. Examples of such models are listed beneath the corresponding blocks in the diagram.

Note that a compression table (or nonlinear map, as in Figure 3) is not included in this processor design. The multiple stages of compression implemented in the auditory models should provide the overall compression needed to map the wide dynamic range of processor inputs onto stimulus levels appropriate for neural activation (some scaling may be needed, but the compression functions should be at least approximately correct). The compression achieved in this way would be much more analogous to the way it is achieved in normal hearing.

Conditioner pulses or high carrier rates may be applied if desired, to impart spontaneous-like activity in the nerve and stochastic independence among neurons (Wilson et al., 1997; Rubinstein et al., 1999).

Implementation of the new processor structure

The development of a speech processor structure as illustrated in Figure 4 involves several steps, in which the standard CIS processor components shown in Figure 3 are gradually replaced with processing blocks that are based on models for a better mimicking of normal auditory functions:

1. Substitution of the Dual Resonance Non-Linear (DRNL) filter bank (Meddis et al., 2001; Lopez-Poveda and Meddis, 2001) for the bank of linear bandpass filters in a standard CIS processor. Such a processor structure will still use an envelope detector that consists of a half- or full-wave rectifier and a lowpass filter, as in the standard CIS strategy.
2. Substitution of the Meddis inner hair-cell (IHC) model (Meddis, 1986 & 1988) for the envelope detector and a portion of the compression table in a standard CIS processor. The cleft contents [c(t)] signal will be used as the output of the Meddis model, rather than the spike timing signal, which is not relevant for the present application.
   As an alternative to the Meddis IHC model, simpler models for reproducing adaptation and compression properties of the IHC/auditory nerve fiber complex (e.g. Kollmeier et al., 1998) might be appropriate as well.
3. Combinations of 1 and 2 and fine-tuning of inter-stage gains and compression in the various stages.

The primary focus thus far has been in the implementation and evaluation of a DRNL filter bank. In this report, we describe the implementation details and discuss our preliminary findings from comparison measures between a standard CIS processor and speech processing structures based on variations of the nonlinear DRNL filter bank.

The DRNL filter

Lopez-Poveda and Meddis (2001) developed a human cochlear filter bank by adapting a computational model of animal basilar membrane (BM) physiology (Meddis et al., 2001) to simulate human BM nonlinearity as measured by psychophysical pulsation-threshold experiments. The approach was based on the dual-resonance nonlinear filter (DRNL filter) illustrated in Figure 5.

The DRNL filter represents each of a limited number of individual sites along the cochlear partition as a tuned system consisting of two parallel processes, one linear and the other nonlinear. The linear path consists of a bandpass function, a lowpass function and a gain/attenuation factor, *g*, in a cascade. The nonlinear path is also a cascade consisting of a bandpass function, a compression function, a second bandpass function, and a lowpass function, in that order. The output of the system is the sum of the outputs of the linear and nonlinear paths.

The compression function in the nonlinear path is linear at low signal levels,

$$y_a(t) = a \cdot x(t),$$

where *x(t)* is the output of the first bandpass filter in the nonlinear path and *a* is a parameter. At higher signal levels, the response is compressed according to the power function

$$y_b(t) = b \cdot |x(t)|^c \operatorname{sgn}(x(t)),$$

where *b* and $c < 1$ are parameters, and sgn(*x*) is the signum function (-1 for $x < 0$, +1 for $x > 0$, 0 for $x = 0$). At all signal levels, the smaller result of the two functions $y_a(t)$ and $y_b(t)$ is chosen, i.e.

$$y(t) = \operatorname{sgn}(x(t)) \cdot \min(a|x(t)|, b|x(t)|^c).$$

Note that *a* and *b* are used to control the gain of the nonlinear path. Their relative value determines the compression threshold.

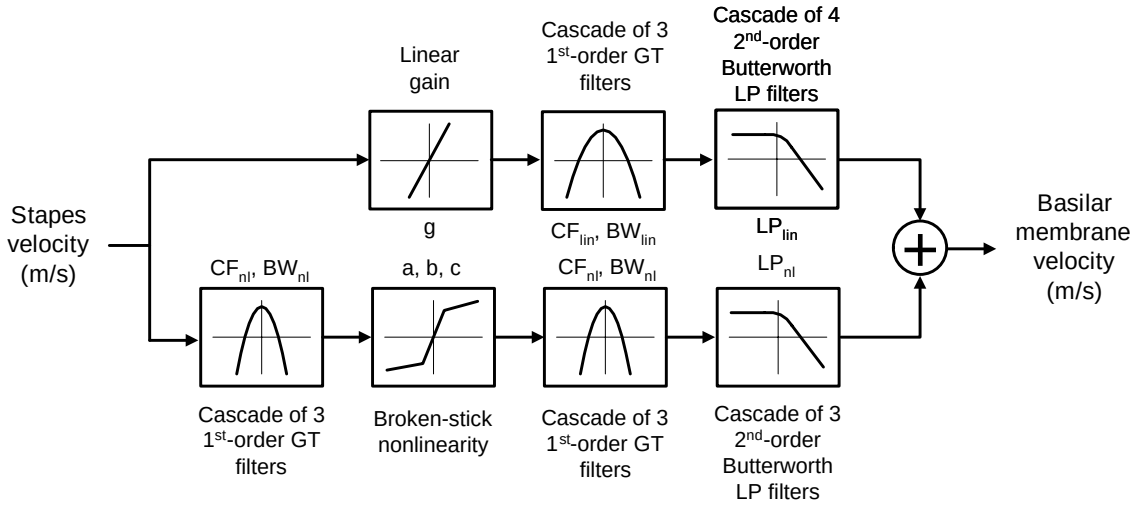

Figure 5: The DRNL filter structure. The input is in the form of stapes velocity and is applied to two signal-processing paths: one linear (top) and one nonlinear (bottom). The outputs of the two paths are summed to form the overall filter output, which corresponds to basilar membrane velocity. The individual bandpass and lowpass functions consist of cascades of bandpass and lowpass functions. The parameters of each block are shown in the space between the two processing paths.
Abbreviations: GT – gammatone, CF – center frequency, BW – bandwidth, LP – lowpass.

The output of the nonlinear path is linear at very low signal levels but is otherwise compressive. The output of the linear path is, of course, linear at all signal levels. At low signal levels, when both outputs are linear, the response of the linear path to tones near CF is typically weak and the response of the nonlinear path dominates the output of the system as a whole. At higher signal levels, the output of the nonlinear path will eventually become less than the output of the linear path, because the former is subject to compression. As a consequence, the output of the system is approximately linear at very high signal levels.

Because of the multiple bandpass filter stages in the nonlinear path, and the nonlinear output dominating the overall filter output at low input levels, the DRNL filter shows a sharply tuned response at low levels. Parameters for the linear path are chosen such that it has a broader response and a lower resonance frequency than the nonlinear path. Thus, at higher input levels, where the linear path dominates the aggregate output, the DRNL filter shows a broader response and a shift in resonance frequency towards lower frequencies.

These properties of the DRNL filter are illustrated in Figure 6, which compares the input/output (I/O) function of a standard CIS processor filter channel (left panel) to the I/O function of a DRNL filter channel (right panel). The functions show the peak amplitude of the channel output envelope signal for a sinusoidal input at different levels and frequencies. Input sinusoids were 200 ms long with 5 ms onset and offset ramps. The peak output amplitude was measured during the last 50 ms of the signal duration in order to avoid any effects that may occur at the onset. For these measures, the filter bank was not preceded by a pre-emphasis filter. A nonlinear maplaw compression was applied only to the standard CIS channel output, whereas the DRNL filter's response is inherently compressed.

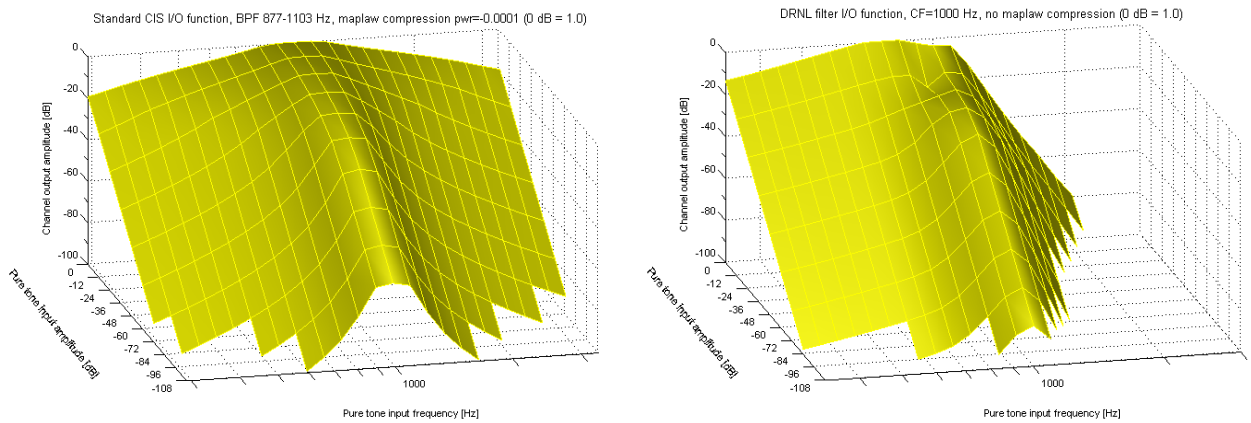

Figure 6: I/O functions of a standard CIS vs. DRNL filter channel at the 1 kHz site. The functions illustrate the peak amplitude of the channel output envelope for a sinusoidal input at different levels and frequencies. The channel output does not include the pre-emphasis filter in either case, while the nonlinear compression table is present only for the standard CIS channel.

The most prominent features of the DRNL filter – which mimic the response of the basilar membrane in the mammalian cochlea to pure tone inputs (Oxenham and Bacon, 2003) – can be summarized as:

- Asymmetric filter shape, with a gradual roll-off at frequencies below the characteristic frequency and a steep roll-off above the most-responsive frequency
- Compressive growth function at and near CF with return to linearity for very high input levels, and with linear growth across all levels for frequencies not near the CF.
- Downward shift of the most-responsive frequency for high input levels
- Level-dependent tuning: Sharp response at low levels, with the tuning getting broader as the input level increases. Note that a constraint applied by Lopez-Poveda and Meddis (2001) in fitting the DRNL parameters to human psychophysical data was for the DRNL filter functions near threshold to have 3 dB-down bandwidths that are consistent with the human auditory filter bandwidth specified in the formula given by Glasberg and Moore (1990). We use the same parameter values as a basis for our implementation of a DRNL filter bank in a speech processor.

An emergent property of the DRNL filter is that the effects of signal level on filter width, most-responsive frequency and growth function are achieved with a level-independent set of model parameters. This fact facilitates the implementation of such a filter bank in the context of a speech processor and was one of the reasons why we selected this model for trying to better mimic the properties of the basilar membrane.

The DRNL filter I/O function in Figure 6 shows a distinct notch below CF for an input level of –12 dB. Such notches are the result of phase cancellation between the outputs of the linear and nonlinear paths (Lopez-Poveda and Meddis, 2001).

Implementation of a DRNL filter bank

For the implementation in a speech processor, the original DRNL filter structure in Figure 5 was modified as illustrated in Figure 7. The original DRNL filter assumes an input in stapes velocity (m/s) and produces an output in basilar membrane velocity (m/s). In order to map the filter's input/output and compression range to the numeric range covered by the digitized speech samples in a speech processor,

and in order to allow for an implementation of the DRNL filter in a real-time Digital Signal Processor (DSP) at a later point, the DRNL structure and parameters were modified as follows:

- Scaling of linear (multiplicative) DRNL parameters (*a*, *b*, *g*) with respect to *a* (*a* becomes 1)
- Down-scaling of signal with linear gain *g* is done after filter stages to preserve numeric precision in filter calculations
- Addition of output scaling factors *H* and *qb*

The effective bandwidth, the shift in most-responsive frequency and the compressive properties of the DRNL filter are not affected by the input/output range remapping.

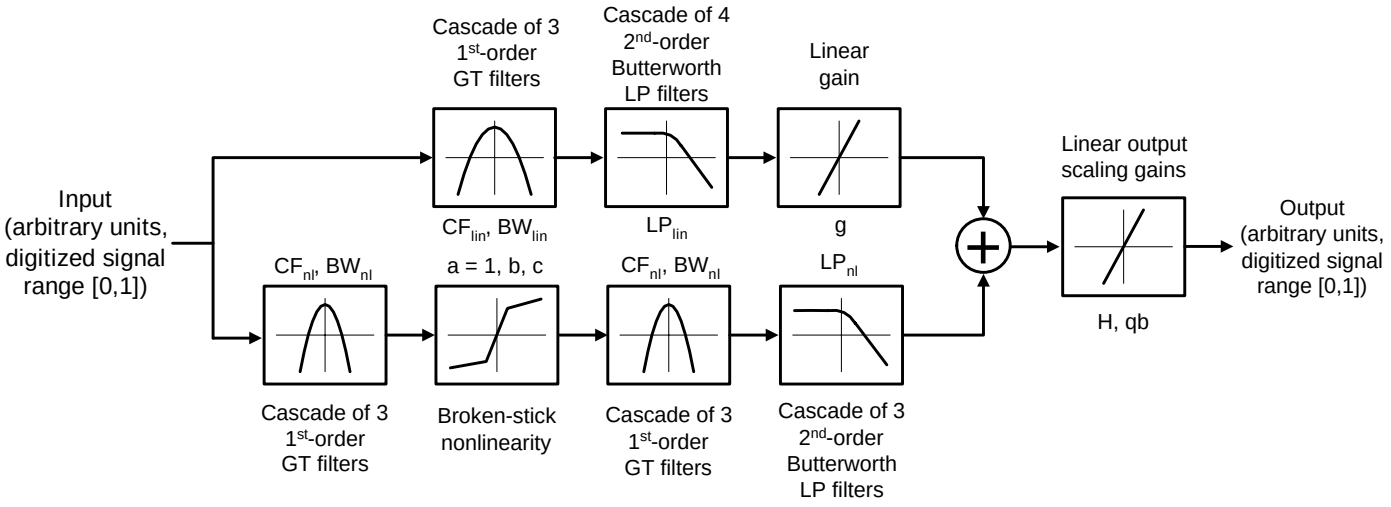

Figure 7: DRNL filter structure as implemented for processing digitized speech samples in the normalized amplitude range [0,1]. Modifications with respect to the original DRNL filter structure, which assumes a stapes velocity input in m/s and produces a basilar membrane velocity output in m/s, include the scaling of gain factors *g*, *a*, *b* and the addition of the linear output scaling gains *H* and *qb*. These modifications map the original DRNL filter's input/output and compression range to the numeric range covered by the digitized speech samples. Note that the effective bandwidth and the compressive properties of the DRNL filter are not affected by this input/output range remapping.

The DRNL parameter set described in Lopez-Poveda and Meddis (2001) for modeling basilar membrane nonlinearity as determined in psychophysical masking experiments for an average of six normal hearing listeners was used as a basis for the implementation of a DRNL filter bank in a speech processor. This parameter set fits the DRNL filter to correspond to the average BM response at six different characteristic frequencies (250, 500, 1000, 2000, 4000 and 8000 Hz). Lopez-Poveda and Meddis also show that the fitted DRNL parameters can be approximated by a linear relationship between the logarithm of each parameter and the logarithm of CF as follows:

$$\log_{10}(parameter) = p_0 + m\log_{10}(CF),$$

where p_0 and m are the regression coefficients. This result suggests a way of creating a filter bank by linear regression of each parameter at intermediate CFs, along a logarithmic scale of frequencies.

Table 1 contains an example of a DRNL parameter set that was the basis for various implementations of a DRNL filter bank in a speech processor. The parameter set corresponds to the average human parameters

in Lopez-Poveda and Meddis, except for *a*, *b* and *g*, which here are constant and do no vary with CF, as they do in the average human set. In Lopez-Poveda and Meddis, *a* in particular varies with CF to compensate in part for the middle ear transfer function that is included in their overall model. Our speech processor design however does not include such a bandpass function to model the middle ear resonance.

Characteristic frequency (Hz)	250	500	1000	2000	4000	8000
DRNL linear path						
No. cascaded GT filters	3	3	3	3	3	3
CF_{lin} (Hz)	244	480	965	1925	3900	7750
BW_{lin} (Hz)	100	130	240	400	660	1450
g	410	410	410	410	410	410
LP_{lin} cutoff (Hz)	CF_{lin}	CF_{lin}	CF_{lin}	CF_{lin}	CF_{lin}	CF_{lin}
No. cascaded LP filters	4	4	4	4	4	4
DRNL nonlinear path						
No. cascaded GT filters	3	3	3	3	3	3
CF_{nl} (Hz)	250	500	1000	2000	4000	8000
BW_{nl} (Hz)	84	103	175	300	560	1100
a	4800	4800	4800	4800	4800	4800
b	0.28	0.28	0.28	0.28	0.28	0.28
c	0.25	0.25	0.25	0.25	0.25	0.25
LP_{nl} cutoff (Hz)	CF_{nl}	CF_{nl}	CF_{nl}	CF_{nl}	CF_{nl}	CF_{nl}
No. cascaded LP filters	3	3	3	3	3	3

Table 1: Example of a DRNL parameter set used to derive parameters for the single DRNL filters in various speech processor implementations. CF_{lin} and CF_{nl} refer to the center frequency of the 1^{st}-order gammatone filters in the linear and nonlinear paths of the DRNL model, respectively.

The parameters for the single DRNL units in a filter bank are created through linear regression from a parameter set as in Table 1. The characteristic frequencies of the single DRNL filters are equally spaced on a logarithmic scale to cover the processor's analysis frequency range. The linear gain factors *a*, *b* and *g* are normalized with respect to *a*, and output gains *H* and *qb* are applied, as appropriate for the normalized DRNL filter implementation in Figure 7. As an example, the resulting DRNL parameters of an 8-channel processor with 8 filter units arranged from 429 Hz to 5907 Hz are shown in Table 2.

Characteristic frequency (Hz)	429	625	908	1321	1921	2793	4062	5907
DRNL linear path								
No. cascaded GT filters	3	3	3	3	3	3	3	3
CF_{lin} (Hz)	416	604	879	1278	1859	2704	3932	5719
BW_{lin} (Hz)	130	173	231	309	413	551	736	984
g	0.085	0.085	0.085	0.085	0.085	0.085	0.085	0.085
LP_{lin} cutoff (Hz)	CF_{lin}	CF_{lin}	CF_{lin}	CF_{lin}	CF_{lin}	CF_{lin}	CF_{lin}	CF_{lin}
No. cascaded LP filters	4	4	4	4	4	4	4	4
DRNL nonlinear path								
No. cascaded GT filters	3	3	3	3	3	3	3	3
CF_{nl} (Hz)	429	625	908	1321	1921	2793	4062	5907
BW_{nl} (Hz)	103	137	183	243	323	430	571	760
a	1	1	1	1	1	1	1	1
b	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
c	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
LP_{nl} cutoff (Hz)	CF_{nl}	CF_{nl}	CF_{nl}	CF_{nl}	CF_{nl}	CF_{nl}	CF_{nl}	CF_{nl}
No. cascaded LP filters	3	3	3	3	3	3	3	3
H	31.7	31.7	31.7	31.7	31.7	31.7	31.7	31.7
qb	4.74	4.74	4.74	4.74	4.74	4.74	4.74	4.74

Table 2: Example of parameters for eight single DRNL filters as implemented in an 8-channel speech processor. Characteristic frequencies (for low input levels) of the single filter units are equally spaced on a logarithmic scale from 429 to 5907 Hz. Gains *a*, *b* and *g* are normalized, and output scaling factors *H* and *qb* are also shown. CF_{lin} and CF_{nl} refer to the center frequency of the 1st-order gammatone filters in the linear and nonlinear paths of the DRNL model, respectively.

At low levels, the 3 dB bandwidth of the modeled DRNL filters closely matches the human auditory filter bandwidth after Glasberg and Moore (1990). In normal hearing, about 16 to 25 "critical band" or "equivalent rectangular bandwidth" channels span the range of speech frequencies, with the exact number depending on the specification of endpoints for that range. A similar number of filter channels may be required to implement an effective DRNL filter bank in a speech processor. If only a small number of DRNL filter channels is used, e.g., 6 to 12, then large portions of the spectrum would not be represented, due to the relatively sharp tuning of the filters, especially at low input levels. However, current implant systems either do not provide such a high number of stimulation electrodes, or channel interactions limit the number of independently-usable sites. As a consequence, we have investigated various strategies for trying to preserve spectral information with a DRNL processor with a lower number of stimulation electrodes, including:

- Increasing the effective bandwidth of the DRNL filter at all input levels by multiplying the bandwidth parameters BW_{lin} and BW_{nl} by a factor greater than 1.
- Taking advantage of the increased number of pitch-distinct sites available with bilaterally implanted subjects by increasing the number of processing channels to equal the number of pitch distinct electrodes and assigning the channel outputs in an alternating, interleaved fashion to electrodes across the two sides ("zipper" approach).
- Applying the "*n*-to-*m*" approach illustrated in Figure 8. Assuming a number of *m* available stimulation electrodes, a filter bank with $n = k*m$ DRNL filters is constructed, where *k* is an integer number greater than 1. The single DRNL filters are equally spaced on a logarithmic scale

within the speech processor's analysis frequency range. The compressed envelope output signals from *k* adjacent filter channels are then combined, either using the average or selecting the maximum, to modulate the carrier pulse trains on each of the *m* electrodes. Figure 8 shows the magnitude and phase response for 22 DRNL filter units at a low input level, where the filters are narrowest. The effect of combining adjacent band outputs using the average or maximum is also shown.

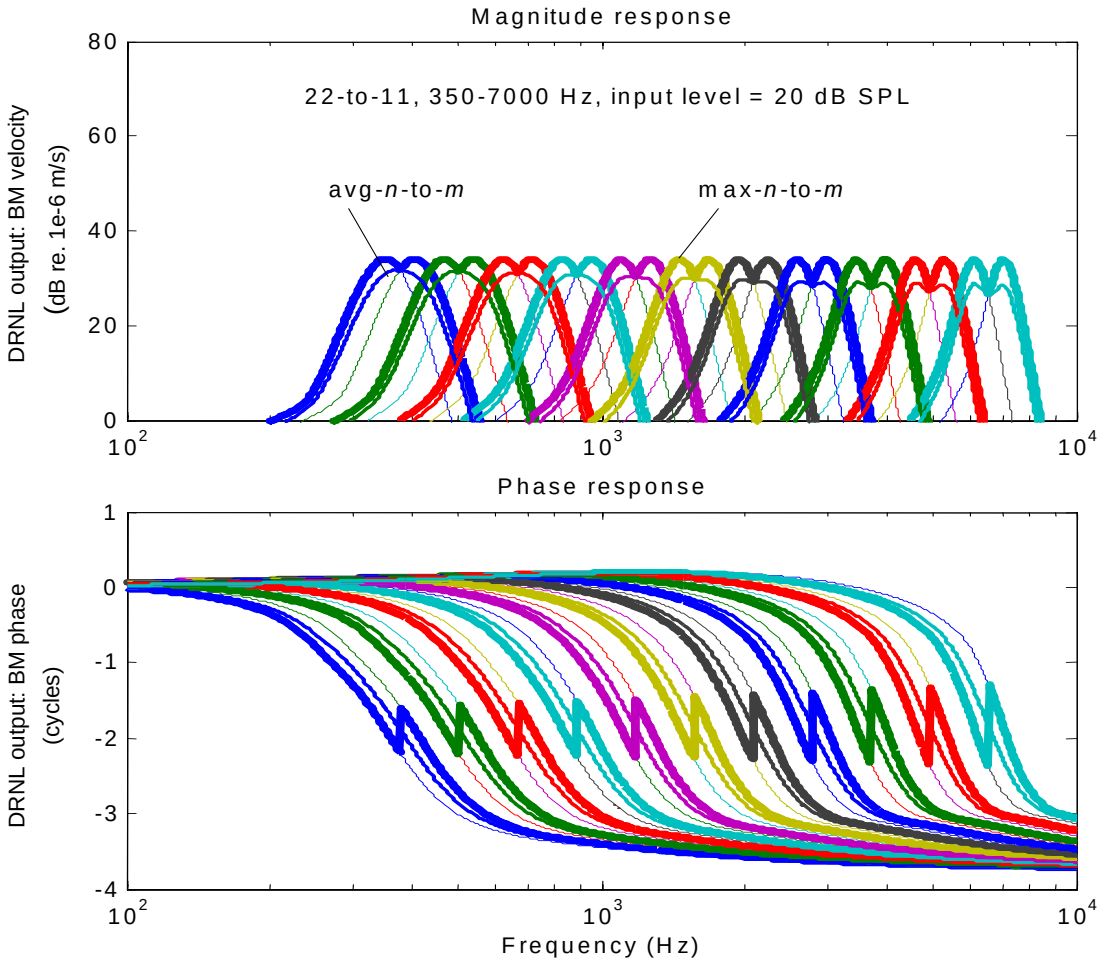

Figure 8: The *n*-to-*m* approach for avoiding spectral gaps in a processor design based on a DRNL filter bank: Let *m* be the number of available electrodes, then $n = k*m$ is the number of DRNL filters in the filter bank, with *k* an integer number greater than 1. The compressed envelope signals from *k* adjacent channels are combined, using either the average or selecting the maximum, to modulate the carrier pulse trains on each of the *m* electrodes. The particular curves shown illustrate the combined magnitude and phase response for a 22-to-11 DRNL filter bank at a low input level, where the filters are narrowest. The thin lines represent the output from the 22 individual DRNL filters. The medium-thick lines represent the 22-to-11 approach, using the average. The thickest lines represent the 22-to-11 approach, using the maximum. Note that the combined magnitude and phase response here is illustrated for DRNL filter band output signals, whereas in a speech processor implementation, the compressed envelope signals of neighboring filter outputs are combined, as opposed to the band outputs directly. As a consequence, the effect of the envelope extraction and maplaw compression stage is not accounted for here.

New electrode arrays, designed to be positioned closer to the inner wall of the scala tympani, may provide better spatial specificity of neural excitation and support a higher number of independent stimulation channels than current electrode designs. Such an electrode array would likely be a better target for the implementation of a speech processor based on a DRNL filter bank, because the better spatial specificity of neural excitation in conjunction with a higher number of stimulation electrodes may support the assignment of a single "human bandwidth" DRNL filter to each electrode.

Initial studies and preliminary results

To date, studies have been conducted with eight subjects. The aim of these initial studies was to explore the parametric space of the DRNL filter bank implementations in a speech processor and to identify configurations that provide a robust performance in terms of speech reception in quiet and noise. Speech reception scores of DRNL processor variations were always compared with a control processor using a standard CIS strategy for the same subject. Results from the studies with each subject have provided direction and guidance for parametric variations to investigate in subsequent experiments. Findings from our studies with DRNL processors so far are discussed in the following section.

Subjects

Table 3 lists the subjects tested to date with speech processors based on a DRNL filter bank. All of them except for SR-3 are recipients of bilateral Med-El implants. The number of usable electrodes for each subject is also shown, along with the subject's native language.

Subject	Implant Left	Implant Right	No. of usable Electrodes Left	No. of usable Electrodes Right	Native Language
SR-3	—	Ineraid	—	6	English
ME-18	C40+	C40+	12	12	English
ME-16	C40+	C40+	12	12	English
ME-22	C40+	C40+	11	11	English
ME-12	C40+	C40+	12	12	English
ME-15	C40+	C40+	11	11	German
ME-3	C40+	C40+	8	8	German
ME-10	C40+	C40	11	8	German

Table 3: Subjects participating in DRNL processor studies.

Speech materials and test procedure

Speech reception was tested using 24 English medial consonants in an /a/-consonant-/a/ context. For the German-speaking subjects, a subset of 16 out of the 24 consonants was chosen as appropriate for German. Prior to their administration in tests with the bilateral subjects, the consonant tokens were processed using a head-related transfer function (HRTF) to simulate a sound coming from the front. The processed consonants were mixed with HRTF-processed CCITT speech-spectrum noise, to simulate a noise source also in front of the subject. For unilateral subject SR-3, the unprocessed consonant tokens were mixed with also unprocessed CCITT noise. Signal-to-noise ratios (SNRs) were selected to avoid ceiling or floor effects for each of the subjects. The consonant test tokens in quiet and noise were then processed off-line

in a MATLAB speech processor implementation (see QPR5, this project) and presented in 10 randomized repetitions for each condition.

Processor designs with different variations of the DRNL filter bank were evaluated in comparison with a standard CIS control processor. Constant parameters across comparisons discussed here were:

- Biphasic carrier pulses at 27 μs/phase, 1515 pulses/s per electrode
- Synchronized, simultaneous carriers on the two sides for bilateral processors
- 1200 Hz/1st-order pre-emphasis highpass filter
- Full-wave rectifier, 200 Hz/4th-order smoothing filter, for the envelope detectors
- Logarithmic compression map (power function exponent pwr = –0.0001) for the standard CIS processors

Results

Figure 9 shows results of a comparison between a standard CIS processor and an early DRNL processor implementation for four bilateral Med-El recipients. Both processors presented eight channel output signals diotically, i.e., the same eight analysis channels were used for each side, with separate mapping functions for each of the 16 electrodes. The subset of electrodes selected for stimulation was kept constant throughout comparison conditions for each subject.

In the standard CIS condition, the analysis frequency range was 350-5500 Hz, measured between the outermost 3 dB-down band edges. The single frequency bands were set to overlap at –3 dB points, and the points were equally spaced on a logarithmic scale of frequencies. In the DRNL filter bank, characteristic frequencies of the single DRNL units were equally spaced on a logarithmic scale from 429 to 5907 Hz. CFs were calculated by splitting a frequency range of 350 to 7000 Hz into eight equally broad bands on a logarithmic scale, and then each CF was set to the center frequency of the corresponding band. The resulting CFs are shifted towards higher frequencies with respect to the center frequencies of the eight standard CIS channels, to qualitatively account for the asymmetric filter shape and CF shift of the DRNL filters.

Because of the inherent compression of the DRNL filter, the power exponent of the maplaw compression function in that processor was set to 0.7, which is almost linear (a power exponent close to zero would result in a nearly-logarithmic function, whereas an exponent of 1 would result in a linear function). This value for the maplaw compression exponent was chosen on an ad-hoc basis, in a comparison of DRNL processors with subject ME-18 using different power exponents. However, this comparison was based on only eight DRNL filter channels spanning two different analysis frequency ranges. We plan to conduct a systematic study of the sensitivity of the maplaw compression parameter in conjunction with a DRNL filter bank that uses a higher number of filter channels and, if required, the *n*-to-*m* approach discussed earlier.

The channel envelope signals in the DRNL processors were extracted from the bandpass filter output signals in the same way as in the standard CIS processor, using a full-wave rectifier, followed by a Butterworth lowpass filter of 4th order at a –3 dB cutoff frequency of 200 Hz.

Speech reception was tested using a 24 medial consonant test in quiet and at a signal-to-noise ratio of +5 dB. Error bars show statistical uncertainties expressed in terms of the standard error of the means.

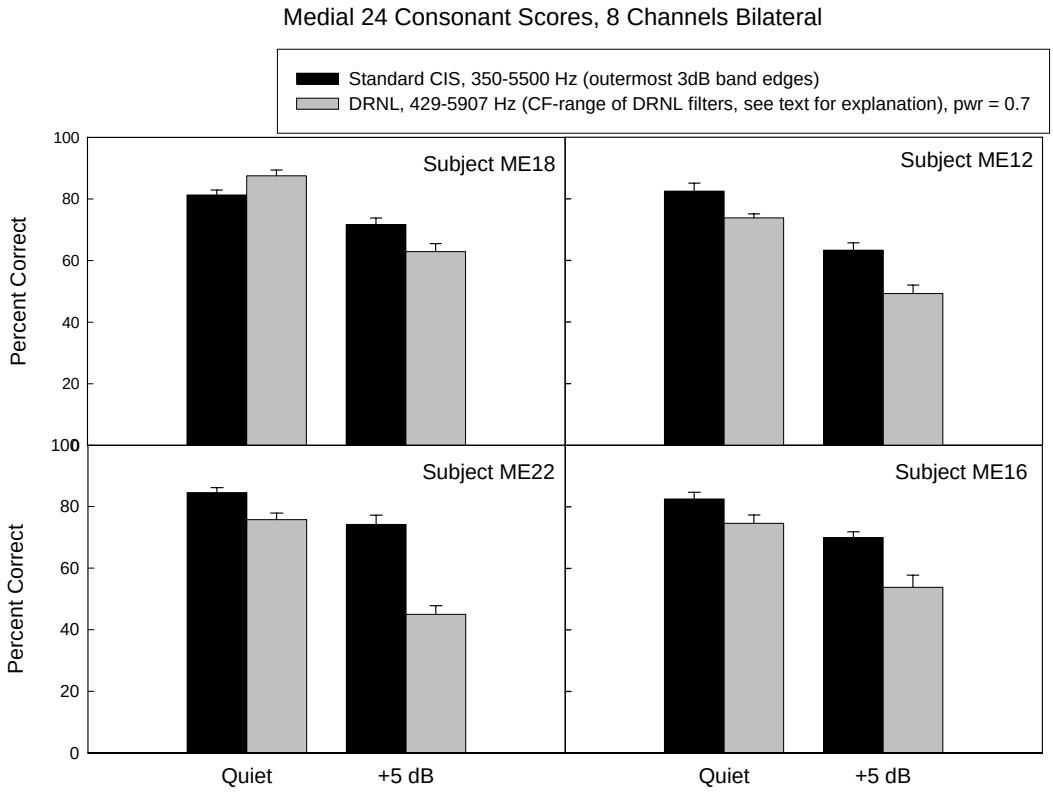

Figure 9: Comparison of consonant identification scores for a standard CIS (black bars) and a DRNL processor (grey bars) with four bilateral Med-El subjects. Both processors presented eight filter channel outputs diotically at a synchronized pulse rate of 1515 pulses per second per electrode. Error bars show the standard error of the means.

The results in Figure 9 show that performance with the 8-channel DRNL processor was comparable with the standard CIS processor for two of the four subjects in the quiet condition only. For the other two subjects in quiet, and for all four subjects in noise, performance with the DRNL processor was poor in comparison to the standard CIS processor. For instance, consonant recognition scores dropped from 74 percent correct with the standard CIS processor to 45 percent with the DRNL processor for ME-22 in the +5 dB noise condition.

With the same four subjects, consonant recognition scores were also measured in quiet and +5 dB SNR for a higher number of channels. The results are presented in Figure 10. In the case of the DRNL processors, 12 bilateral channels were tested, except for ME-22, who only has access to 11 electrodes on each side. Scores are compared to a bilateral 12-channel standard CIS processor for subjects ME-12 and ME-16. For ME-18 and ME-22, DRNL scores are compared to the 8-channel standard CIS results from Figure 9, because no data were collected for a standard CIS processor with 12 channels for ME-18 and 11 channels for ME-22. All stimuli were presented diotically.

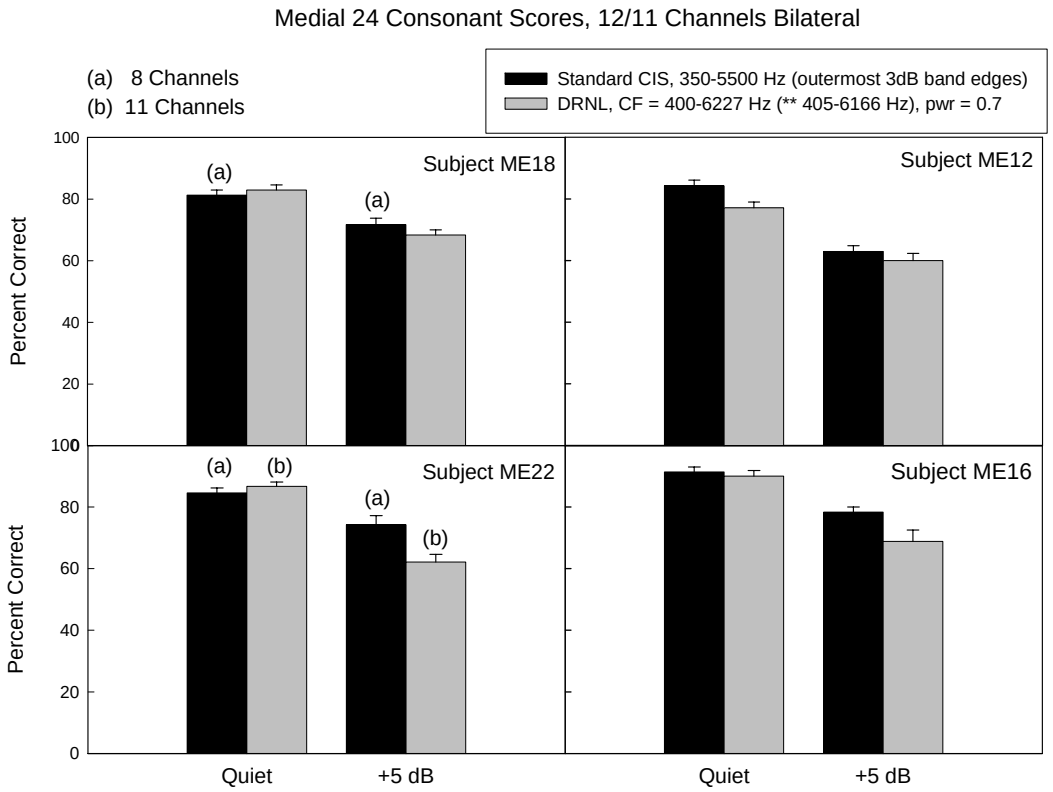

Figure 10: Comparison between a standard CIS processor (black bars) and a DRNL processor (grey bars) using the maximum number of available channels on each side for the same subjects as in Figure 9. The number of DRNL filter channels was 12, except for ME-22, who has only access to 11 electrodes on each side. For subjects ME-18 and ME-22, DRNL scores are compared to the 8-channel standard CIS results from Figure 9 because of the lack of standard CIS comparison data with 12 (ME-18) and 11 channels (ME-22), respectively. Error bars show the standard error of the means.

The higher number of processing channels resulted in increased performance with the DRNL condition in noise for all four subjects (62.9±2.6 to 68.3±1.7 for ME-18, 49.2±2.8 to 60.0±2.3 for ME-12, 45.0±2.8 to 62.1±2.5 for ME-22, 53.8±4.0 to 68.8±3.7 for ME-16). DRNL scores in quiet changed from slightly worse for ME-18 to better for all others (87.5±1.9 to 82.9±1.7 for ME-18, 73.8±1.3 to 77.1±1.9 for ME-12, 75.8±2.1 to 86.7±1.4 for ME-22, 74.6±2.7 to 90.0±1.8 for ME-16). As for the standard CIS processors, the increase of channels resulted in better scores in both quiet and noise for ME-16 and in no significant difference for ME-12.

Generally, performance with the DRNL processors was observed to be better and to come closer to the level of the standard CIS condition for a higher number of DRNL filter channels. Learning effects do not account for the improvement, at least for ME-12, where the 12-channel DRNL processor was tested before the 8-channel condition, and for ME-22, where conditions were tested two weeks apart. The increase in performance may be attributed to the better spectral representation resulting from a higher number of (sharply tuned) DRNL filters spanning the analysis frequency range. This observation led us to investigate the strategies mentioned earlier for trying to preserve spectral information with a DRNL processor and a number of stimulation electrodes that is less than the number of critical bands within the analysis frequency range, using the following approaches:

- Increasing the bandwidth of the single DRNL units in a filter bank
- Taking advantage of the increased number of pitch-distinct sites available with bilaterally implanted subjects by increasing the number of processing channels and assigning the channel outputs in an alternating, interleaved fashion to electrodes across the two sides ("zipper" approach)
- Applying the *n*-to-*m* approach discussed earlier

Figure 11 shows results obtained for subjects ME-12 and ME-16 when increasing both DRNL filter bandwidth parameters BW_{lin} and BW_{nl} by a factor of 1.2. For ME-16, performance was better with the broader DRNL filters, but for ME-12, no consistent increment was observed. With subject ME-16, the data point for wider DRNL bands at CFs that are logarithmically spaced from 400 to 6227 Hz was not measured.

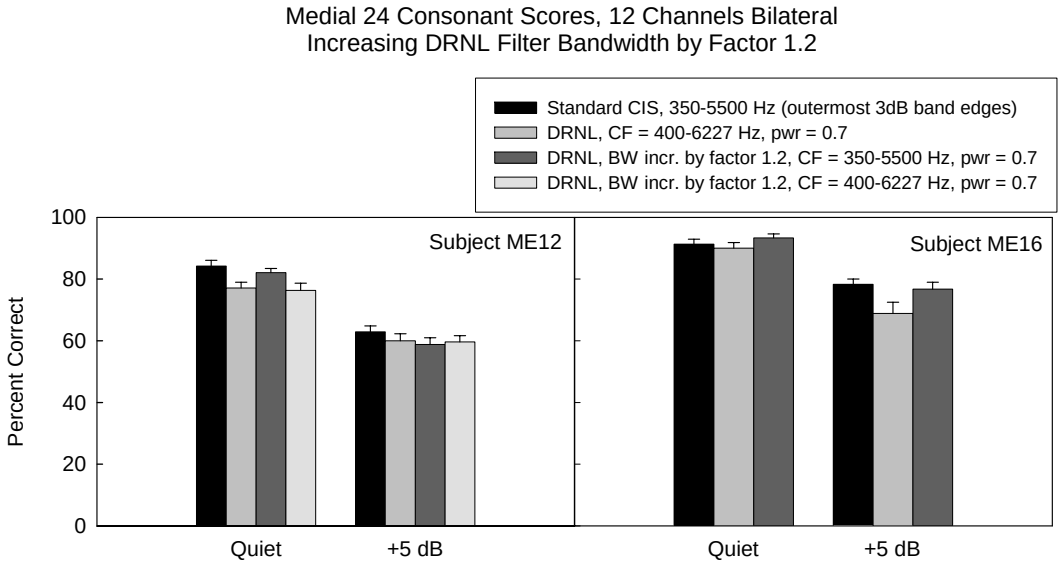

Figure 11: Results of increasing the DRNL filter bandwidth in a 12-channel processor with subjects ME-12 (left panel) and ME-16 (right panel). From the left, bars show 24-consonant identification scores for a standard CIS processor, for a DRNL processor with unmodified bandwidth parameters, and for DRNL processors with increased filter bandwidths. All processors presented stimuli bilaterally using 12 channels on each side at a synchronized pulse rate of 1515 pulses per second.

With subject ME-16 we looked at the "zipper" approach, where 24 DRNL filter channels were arranged at CFs from 350 to 5500 Hz, and filter channels were then assigned in an alternating pattern to the 12 electrodes on each side (the 12 odd filter channels with CFs from 350 to 4879 Hz were assigned to the left implant electrodes, and the 12 even channels from 394 to 5500 Hz were assigned to the electrodes on the right side). Performance using that approach was at the same level as with the wider DRNL filter bands and also at the same level as with the standard CIS processor, as shown in Figure 12. However, note that possible ceiling effects may have obscured a difference for the quiet conditions.

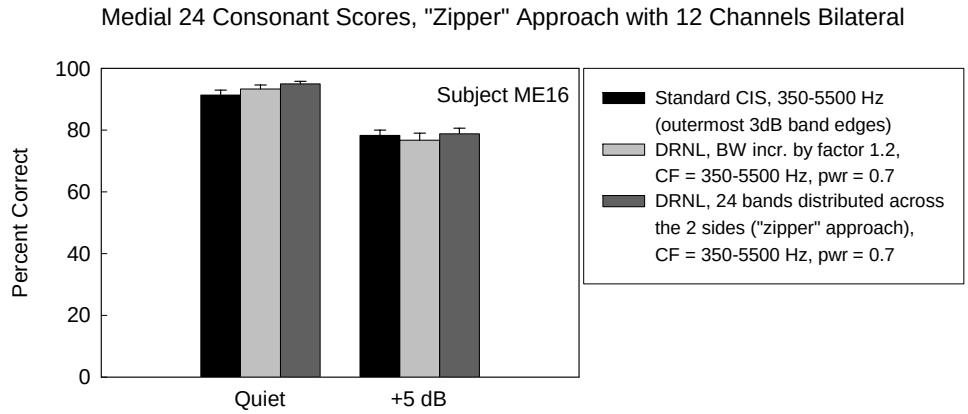

Figure 12: Comparison of 24 consonant scores for subject ME-16 with a bilateral 12-channel standard CIS processor (left bar), a bilateral 12-channel DRNL processor with increased bandwidth (middle bar), and a 24-channel DRNL processor with filter bands distributed between the two sides.

Variations of the *n*-to-*m* approach for combining multiple DRNL filter envelope signals into one processor channel were evaluated with five subjects. Subjects ME-15, ME-3 and ME-10, for which data are shown in Figure 13, are all bilateral Med-El implant recipients and native German speakers. Speech reception for these three subjects was tested using a subset of 16 out of the 24 English consonants, as appropriate for German. Consonant scores were measured in quiet and at a signal-to-noise ratio of 0 dB or +5 dB, as appropriate for avoiding ceiling effects. All processors presented stimuli diotically at a synchronized pulse rate of 1515 pulses per second.

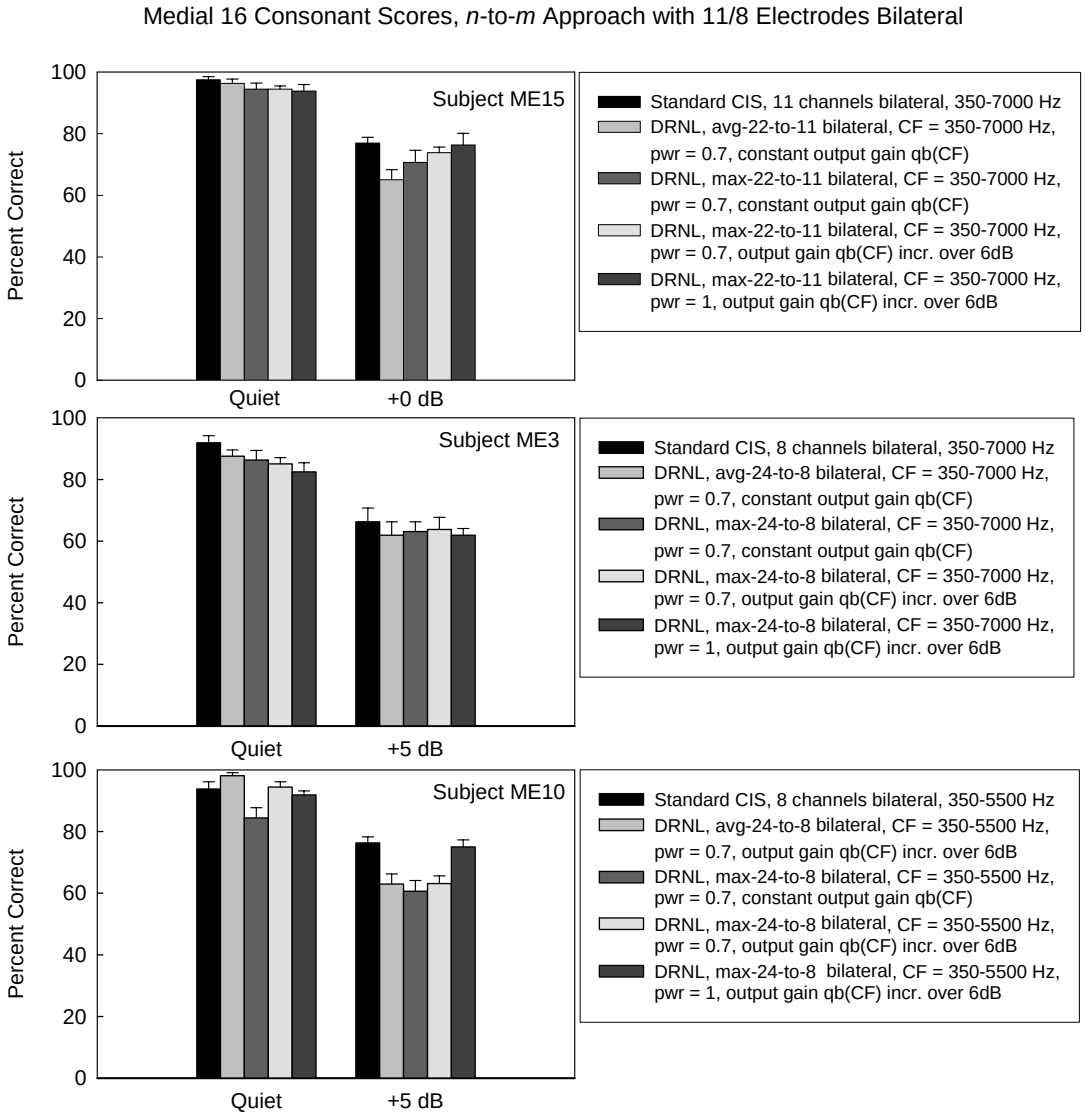

Figure 13: Comparison of a standard CIS processor (leftmost bar) with variations of a DRNL processor using the *n*-to-*m* approach for combining multiple DRNL filter envelope signals into one processor channel (other bars). Data from a 16-consonant identification test in quiet and noise are shown for three German-speaking subjects. The analysis frequency range was 350 to 7000 Hz in processors tested with ME-15 and ME-3, and 350 to 5500 Hz with ME-10. Depending on the number of available electrodes, either two (22-to-11) or three (24-to-8) DRNL filter envelope signals were combined into a channel output. Variations included the combination method of DRNL filter envelopes (average vs. maximum), the gain applied to the DRNL filter output signals (constant *qb* vs. *qb* increasing by 6 dB over the CF range), and the power exponent applied in the maplaw compression stage. Increasing the DRNL filter output gain *qb* by 6 dB over the CF range provides a relatively flat overall frequency response of the DRNL filter bank, if the contributions of all filters are combined, whereas a constant output gain *qb* across filters results in an overall low frequency emphasis due to the asymmetric DRNL filter shape.

For subjects ME-15 and ME-3, performance with the different DRNL *n*-to-*m* variations is similar and comparable to the standard CIS processor results, except for the avg-22-to-11 condition for ME-15 in noise. Subject ME-10 shows similar speech reception scores in both quiet and noise only for the DRNL condition with no back-end compression, i.e. with a linear maplaw compression power of 1.

Results from the evaluation of one *n*-to-*m* DRNL processor condition with our local bilateral subjects ME-16 and ME-22 are shown in Figure 14. These subjects also show a similar performance with the two conditions compared, in both quiet and in noise at +5 dB SNR.

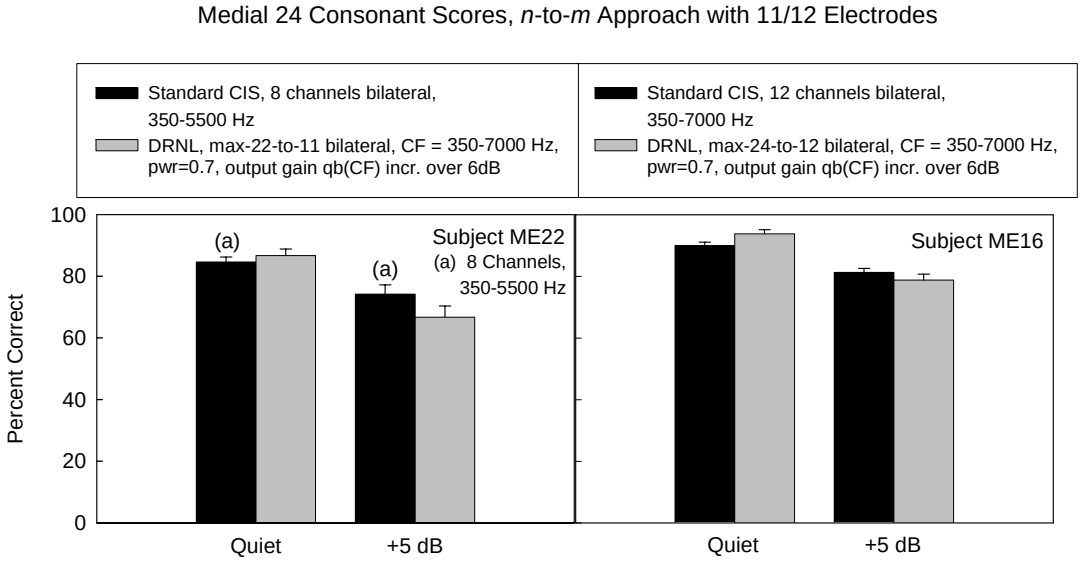

Figure 14: Comparison of a standard CIS processor (black bars) with a DRNL processor using the *n*-to-*m* approach (grey bars) for subjects ME-22 and ME-16. Speech reception was measured using 24 English consonants in quiet and +5 dB SNR. The analysis frequency range was 350 to 7000 Hz, except for the standard CIS condition with ME-22, where only 8-channel data for a frequency range of 350 to 5500 Hz are available.

Conclusion and next steps

The main purpose of our studies thus far has been to explore the large parametric space of the DRNL filter implementation in a speech processor and to determine the effect of a number of variables on speech reception scores. In this exploration we have focused primarily on the properties of the filter bank in the frequency domain. In particular, increasing the number of filter channels using the *n*-to-*m* approach provided a robust performance in consonant identification tests for all subjects tested with such a condition. With this approach, we have been able to identify parameter settings that are capable of supporting speech reception scores comparable to those of a standard CIS processor, in subjects who have had no chronic experience with the new processor design.

However, the results obtained to date have to be considered as preliminary. Future investigations might productively address the following aspects:

- Further exploration of parameter settings for the currently-implemented DRNL filter bank as a model for the basilar membrane frequency selectivity. In particular, a question still to be

addressed is how much additional compression should be applied to the DRNL filter output, or what the optimum amount of compression in the DRNL filter is.

- Comparison of various speech processor designs using test materials in addition to consonants, such as sentences or monosyllabic words.
- Implementation of a full processor structure for mimicking normal auditory functions by combining the DRNL filter bank with models of the IHC membrane and nerve synapse characteristics. Such a processing stage should provide a further compression of the DRNL filter bank output as required, and also model the adaptation properties of the IHC and nerve synapse.
- Application of conditioner pulses or high carrier rates to impart spontaneous-like activity in the nerve and stochastic independence among neurons.
- Evaluation of the new strategies in conjunction with novel electrode designs that may provide a greater spatial specificity of neural excitation and support a higher number of independent stimulation channels. The percutaneous study patients implanted at Duke University Medical Center with the Nucleus Contour Electrode and scheduled for studies in our laboratory in the immediate future are ideal candidates for such an evaluation.

III. References

Carney LH. 1993. A model for the responses of low-frequency auditory-nerve fibers in cat. *J. Acoust. Soc. Am.* 93: 401-17

Dallos P. 1992. The active cochlea. *J. Neurosci.* 12: 4575-85

Delgutte B. 1996. Physiological models for basic auditory percepts. In *Auditory Computation*, ed. Hawkins HL, Mc-Mullen TA, Popper AN, Fay RR, pp. 157-220. New York: Springer.

Deng L, Geisler CD. 1987. A composite auditory model for processing speech sounds. *J. Acoust. Soc. Am.* 82: 2001-12

Glasberg BR, Moore BCJ. 1990. Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* 47: 103-38

Kollmeier B, Derleth R-P, Dau T. 1998. Modeling the "effective" auditory signal processing for hearing-impaired listeners. In *Psychophysical and Physiological Advances in Hearing*, ed. Palmer AR, Rees A, Summerfield AQ, Meddis R, pp. 482-88. London: Whurr.

Lopez-Poveda EA, Meddis R. 2001. A human nonlinear cochlear filterbank. *J. Acoust. Soc. Am.* 110: 3107-18

Lopez-Poveda EA. 2003. An approximate transfer function for the dual-resonance nonlinear filter model of auditory frequency selectivity. *J. Acoust. Soc. Am.* 114: 2112-17

Meddis R, O'Mard LP, Lopez-Poveda EA. 2001. A computational algorithm for computing nonlinear auditory frequency selectivity. *J. Acoust. Soc. Am.* 109: 2852-61

Meddis R. 1986. Simulation of mechanical to neural transduction in the auditory receptor. *J. Acoust. Soc. Am.* 79: 702-11

Meddis R. 1988. Simulation of auditory-neural transduction: Further studies. *J. Acoust. Soc. Am.* 83: 1056-63

Oxenham AJ, Bacon SP. 2003. Cochlear compression: Perceptual measures and implications for normal and impaired hearing. *Ear & Hearing* 24: 352-66

Robert A, Eriksson JL. 1999. A composite model of the auditory periphery for simulating responses to complex tones. *J. Acoust. Soc. Am.* 106: 1852-64

Rubinstein JT, Wilson BS, Finley CC, Abbas PJ. 1999. Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation. *Hear. Res.* 127: 108-18

Tchorz J, Kollmeier B. 1999. A model of auditory perception as a front end for automatic speech recognition. *J. Acoust. Soc. Am.* 106: 2040-50

Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM. 1991. Better speech recognition with cochlear implants. *Nature* 352: 236-38

Wilson BS, Finley CC, Lawson DT, Zerbi M. 1997. Temporal representations with cochlear implants. *Am. J. Otol.* 18: S30-34

Zhang X, Heinz MG, Bruce IC, Carney LH. 2001. A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. *J. Acoust. Soc. Am.* 109: 648-70

IV. Plans for the next quarter

Among the activities planned for the next quarter are:

- Completion of studies with Subject ME-19 through October 3.
- Continuing studies with local subjects ME14, ME16 and ME22, implanted bilaterally with Med-El Tempo+ devices.
- Guest of Honor and invited presentation by Blake Wilson at the Hearing Preservation Workshop, Frankfurt, Germany October 17-19.
- Attendance by Blake Wilson, Dewey Lawson, Reinhold Schatzer, and Xiaoan Sun at the Neural Prosthesis Workshop, NIH, October 21-23.
- Presentation by Dewey Lawson and Reinhold Schatzer, at the Neural Prosthesis Workshop, NIH, October 21-23.
- A visit by Chris Turner, University of Iowa, October 24.
- Two weeks of studies with return subject SR-3, December 8-19.
- Third and fourth surgery for the Nucleus Contour Electrode percutaneous study patients at Duke University Medical Center, Durham, NC.

V. Acknowledgments

We thank volunteer research subjects ME-3, ME-10, ME-15, ME-16, ME-19 and SR-9, who participated in studies conducted during this quarter.

We would also like to acknowledge the exceptional contributions of Enrique Lopez-Poveda and Marian Zerbi, who both serve as a consultant to our group.

Appendix 1: Summary of reporting activity for this quarter

Reporting activity for this quarter, covering the period of July 1 through September 30, 2003, included:

Invited Presentations

Wilson BS, Wolford RD, Lawson DT, Schatzer R, Brill SM: Evaluation of combined EAS in studies at the Research Triangle Institute. *2003 Conference on Implantable Auditory Prostheses*, Pacific Grove, CA, August 17-22, 2003.

Tyler R, Witt S, Dunn C, Kane D, Kenworthy M, Wilson B, Rubinstein J, Gantz B, Preece J, Parkinson A: A framework for cochlear implantation guidelines in the case of monaural and binaural fittings. *2003 Conference on Implantable Auditory Prostheses*, Pacific Grove, CA, August 17-22, 2003.

Additional Presentations

Schatzer R, Wilson BS, Lopez-Poveda EA, Zerbi M, Wolford RD, Lawson DT: A novel CI speech processing structure for closer mimicking of normal auditory functions. *2003 Conference on Implantable Auditory Prostheses*, Pacific Grove, CA, August 17-22, 2003.

Appendix 2: Erratum

In our prior (fifth) Quarterly Progress Report, the case example for a virtual CIS processor in streaming mode on page 11 contains an error in the channel data table and spec file format. The correct example is as provided below.

One use of simultaneous pulses is in virtual CIS (VCIS) strategies, where two or more adjacent electrodes are stimulated simultaneously to obtain a perceived pitch distinct from that of any electrode alone.

As an example of the relationship between the channel data table and the spec file, consider a VCIS processor in which channel 1 stimulates electrodes 1 and 2 simultaneously with respect to return electrode 8, with electrode 2 receiving a stimulus that is one half the amplitude delivered to electrode 1 and of opposite polarity. The corresponding spec file would reflect this situation in the first line of its >VCIS section:

```
>VCIS
8,1:100,2:-50
. . .
```

Later in the spec file, the first line of the psychophysics section would contain the THR and MCL values for VCIS channel 1 (measured with electrodes 1 and 2 simultaneously and with electrode 2's amplitude half that of electrode 1 and of opposite polarity).

```
>psycho
50 100
                . . .
```

The corresponding channel data table would contain amplitude values calculated using the THR and MCL data for each electrode along with the appropriate VCIS weight multipliers. Its first nine entries would contain the following information (with appropriate settings of bit 23 where relevant):

Y:$0	2	[number of simultaneous electrodes for channel 1]
.	1.0 · (100-50) · DACunits/μA	[MCL-THR, weighed for electrode 1]
.	1.0 · (50) · DACunits/μA	[THR, weighed for electrode 1]
.	$E0	[pointer to VCIS channel 1 amplitude multiplier]
.	address for electrode 1	
.	-0.5 · (100-50) · DACunits/μA	[MCL-THR, weighed for electrode 2]
.	-0.5 · (50) · DACunits/μA	[THR, weighed for electrode 2]
.	$E0	[pointer to VCIS channel 1 amplitude multiplier]
.	address for electrode 2	
Y:$9	address for return electrode 8	