AUDITORY INFORMATION PROCESSING WITH NERVE-ACTION POTENTIALS

Marcus Holmberg and Werner Hemmert

Infineon Technologies Inc, Corporate Research ST, Otto-Hahn-Ring 6, 81730 Munich, Germany

ABSTRACT

Neural sound processing requires compression of the huge dynamic range of acoustic signals to the limited dynamic range which can be coded by discrete action potentials. We present an inner ear model followed by a compression stage and a realistic sensory cell model. Frequency analysis was achieved using a wave-digital filter model of the inner ear's hydrodynamics. The dynamic compression stage entailed nonlinear and reasonably broad response curves. The model coded speech signals into trains of nerve-action potentials, which were fed into a network of neurons featuring spectral receptive fields with lateral inhibition. When their receptive field width was changed, these neurons coded spectral aspects of sounds and they detected the formants in speech signals.

1. INTRODUCTION

Information coded within the spike-trains of auditory nerve fibers provides the basis for spectral, temporal and spatial information processing at higher levels in the auditory pathway. Extracting this information is crucial for robust speech recognition, especially in noisy environments. Neural processing requires compression of the huge dynamic range of natural acoustic signals to the limited range coded by neurons. This is primarily achieved by an active, nonlinear amplification process in the inner ear. The inner ear is also responsible for the spectral decomposition of sound signals. High frequency sounds stimulate only the basal part of the inner ear, whereas low frequency sounds predominately excite its apical part. This so-called tonotopic frequency separation principle is conserved or even enhanced at all higher levels of the auditory pathway. In this paper, we present a model for realistic coding of acoustic signals into spike trains of the auditory nerve. These spikes are the inputs to a model of spectral processing, which extracts formants of speech sounds.

2. MODEL OF SOUND PROCESSING

The model of the peripheral hearing system consists of a simplified middle ear, a model of inner ear hydrodynamics followed by a compression stage, and an inner hair cell.

2.1. Inner ear model with large dynamic compression

The middle ear transforms sound signals into vibrations which elicit a traveling-wave along the basilar membrane (BM) in the inner ear. The traveling-wave is responsible for the spectral decomposition of sound on the BM. High frequency signals reach their maximum close to the basal end of the inner ear, while low frequency stimuli travel more apically. BM vibrations were calculated with a computationally efficient wave-digital filter model comprising of 100 sections [1]. The wave-digital filter model provides a solution for the passive inner ear hydrodynamics (one-dimensional case), which describes the vibration of the BM at high sound levels (compare Fig. 1). The frequency map of the resonators was adjusted according to Greenwood's [2] map for the human inner ear.

In the living hearing organ, low-level sounds are mechanically amplified, probably by the outer hair cells (OHCs). The OHCs are thought to sense the hearing organ's motion and feed back mechanical energy into its vibration. This active amplification stage both boosts the vibration amplitudes and significantly sharpens the traveling-wave at low levels. The amplification saturates at high levels, which is the basis for compression. Measurements have shown that the amplification is higher than thousand-fold ($> 60\,\mathrm{dB}$, [3]). As the details of the amplification process are still unknown and as the implementation of a feedback amplifier with such a high gain poses severe stability issues, we have chosen a different realization. Instead of amplifying the BM response or altering the quality factors of the transmission-line model [4], we added second-order resonators at the outputs of the cochlear filter bank and modulated their quality factors. Note that these passive resonators amplify only the vibration *amplitude*, not the *energy*. Quality factors were altered depending on the instantaneous displacement of each resonator, following a Boltzmann-function similar to the sensitivity function of OHCs' transduction channels. By cascading four resonator stages and modulating their quality factors from ten to one, we realized large amplification (up to 80 dB in the high frequency range, less at frequencies below 1 kHz) together with reasonably broad filter shapes. Fig. 1 shows the BM response along the cochlea for a 2.5 kHz tone presented at various sound levels. At high levels, responses share the

characteristics of a passive traveling-wave with a gradual build-up starting from the cochlear base and a sharp roll-off after the relatively broad maximum is reached (Fig. 1, 100 dB line). At low levels (0 – 20 dB), filter shapes are narrow and almost symmetrical. The amplitudes at the most sensitive location are greatly amplified, so that even faint sounds cause excitation above threshold (rate-threshold was estimated between 1.5 nm and 100 μm/s [3]). At the most sensitive place, the growth function of BM displacement is compressed and follows approximately a cube root law at medium levels. At more basal locations, growth functions are almost linear. For increasing intensities, the excitation pattern grows highly asymmetrical, which is well-known as the upper spread of masking.

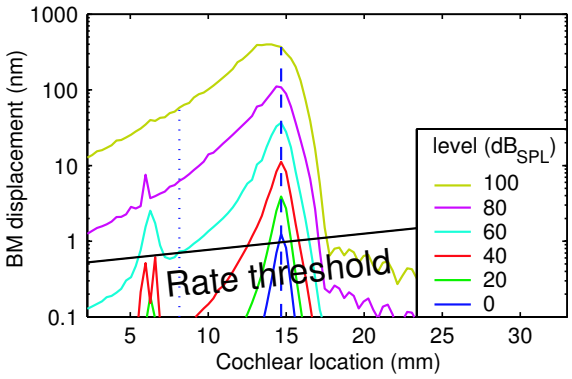

Fig. 1. Model-excitation pattern (RMS) for a 2.5 kHz pure tone at various levels (human cochlea). Note that responses are greatly compressed at the characteristic location (dashed line, 14.5 mm) and almost linear at more basal locations (dotted line). The black line indicates the estimated rate threshold of the auditory nerve.

2.2. Inner hair cell model

Basilar membrane vibrations drive the stereociliary bundles of the sensory cells, the inner hair cells (IHC, Fig. 2), by fluid motion. As a first approximation, fluid friction and hair bundle stiffness form a high-pass filter. The corner frequency is taken to decrease (2000 – 200 Hz) along the length of the cochlea as a result of increasing length and decreasing stiffness of the bundles.

When mechanical stimuli deflect the hair bundle, ion channels at its tip open and a transduction current, mainly carried by K^+-ions, depolarizes the cell. Driving source is the difference between the endocochlear potential (EP, +90 mV) and the membrane potential of the IHC (U_M, -45 mV). The opening probability of the mechano-electrical transduction channels follow a second-order Boltzmann function [5]. This function shows saturation for displacements larger than about 100 nm, limiting

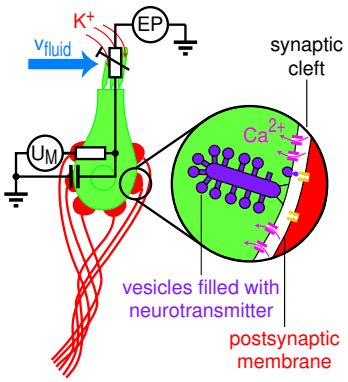

Fig. 2. Schematic view of the inner hair cell with its synaptic region magnified. Potassium ions (K^+) are driven into the cell by the difference between its resting potential (U_M) and the endocochlear potential (EP).

the dynamic range of IHCs to about 40 dB. The transduction current depolarizes the IHC membrane (capacity: 10 pF, conductivity: 60 nS). Depolarization of the membrane activates voltage dependent calcium channels located close to the cell's synaptic terminals and Ca^{2+}-ions enter the cell. These Ca^{2+}-channels were modeled as L-type calcium channels. Elevated Ca^{2+}-levels cause fusion of synaptic vesicles with the cell membrane. The neurotransmitter contained in the vesicles diffuses across the synaptic cleft (not modeled), binds to the receptors in the postsynaptic membrane, causes the membrane to depolarize, and a nerve action potential is propagated along the auditory nerve fiber (ANF) towards the brain.

Neurotransmitter release is dominated by a so-called readily releasable pool (RRP) of vesicles, which are located in the synaptic region closely to the cell membrane (see Fig. 2). A large stimulus causes fusion of many vesicles, depleting the RRP. Due to the depletion, the spiking probability of the auditory nerve is reduced in the following few tens of milliseconds, an effect known as adaptation. Vesicle fusion rate dependency on Ca^{2+} concentration and RRP refill was modeled according to recent measurements [6]. Vesicle fusion is a stochastic and quantal process. In addition to depletion of the RRP, the ANF itself has a refraction time, comprising of an absolute refractory period in which no spike can be generated (0.75 ms), followed by a double exponential recovery with time [7].

Although single IHCs exhibit only a limited dynamic range of about 40 dB, a much larger range in sound level has to be coded. This is largely due to the mechanical compression of the inner ear, but additional processes are in place to overcome this so-called dynamic-range-problem. For example, individual IHCs are innervated by low spontaneous rate (LSR) fibers which display a high threshold and large dynamic range, and high spontaneous rate (HSR) fibers

which have low thresholds and limited dynamic range (typically 20 – 30 dB). In our model, different types of ANFs were realized by altering the maximum conductance of the voltage dependent Ca^{2+}-channels, i.e. the number of calcium channels in the vicinity of the synapse, and the Ca^{2+}-concentration in the IHC at rest (compare [8]). The excitation pattern of HSR auditory fibers along the cochlea is plotted in Fig. 3. Input sound was the artificial vowel "∧" as in "*but*". The responses of twenty HSR fibers derived from a single IHC were modeled and plotted. Responses are delayed towards the apex due to propagation delay of the traveling wave. At medium levels, the three formants of the artificial vowel, not individual harmonics, are separated along the cochlea. The fundamental frequency of the vowel was 100 Hz corresponding to a period of 10 ms. Temporal information is coded by phase-locking of the action potentials, that means, auditory nerve fibers fire preferentially at a certain phase of the stimulus. Phase-locking is known to decrease for frequencies above 1 kHz due to the membrane time constant of the IHCs and the limitations of other synaptic processes. However, nerve fibers sensitive to frequencies well above 1 kHz still phase-lock on the envelope of the sound signal. This effect is clearly visible at the location of the second and third formant, where action potentials clearly preserve the periodicity of the phoneme and fire preferentially in 10 ms intervals.

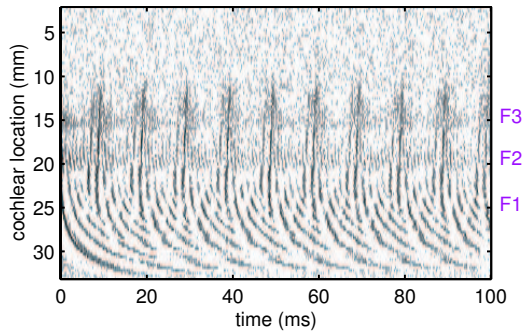

Fig. 3. Firing pattern of auditory nerve fibers along the cochlea to an artificial vowel "∧" (as in "*but*"). The sound consists of 35 sine waves, all harmonics of 100 Hz. Formants are located at F1=500 Hz (81 dB_{SPL}), F2=1.2 kHz (73 dB_{SPL}) and F3=2.3 kHz (69 dB_{SPL}). Note that both spectral and temporal features of the sounds are conserved. Especially periodicity (10 ms) of the phoneme is clearly visible as an amplitude modulation at the formant frequencies.

2.3. Spectral processing

Spectral processing along the auditory pathway involves integration over multiple frequency channels with both excitatory and inhibitory inputs [9, 10]. Lateral inhibition can thereby enhance frequency tuning at higher processing levels [10]. Neurons with narrow spectral receptive fields were described in the central part of the inferior colliculus [11], surrounded by neurons which integrate over wider frequency areas. In this paper we model layers of neurons for which we vary the width of their synaptic fields systematically and test their reaction to speech sounds. We term these neurons "Q–neurons", as their spectral receptive fields would show varying values of their "quality of frequency tuning" ($Q_{10\,dB}$ is usually used to describe frequency tuning of a single neuron [11]). We assume excitatory inputs in the central response area and lateral inhibition. The shapes of excitatory and inhibitory receptive fields were assumed to be Gaussian, with the width of the excitatory field being half the width of the inhibitory field. The functions we used exhibited zero mean. The effective area of excitation was normalized to 1. Exemplary weight functions are plotted on the right side in Fig. 4. We integrated synaptic inputs using an integrate-and-fire neuron (leaky integration with a time constant of 30 ms). It fired when a value equivalent to five simultaneous action potentials, integrated over the whole receptive field, was reached. As we have no model of the first processing stage in the auditory pathway, the cochlear nuclei, we fed our modeled ANF outputs (20 per frequency channel) directly into the Q–neurons. In this study, we

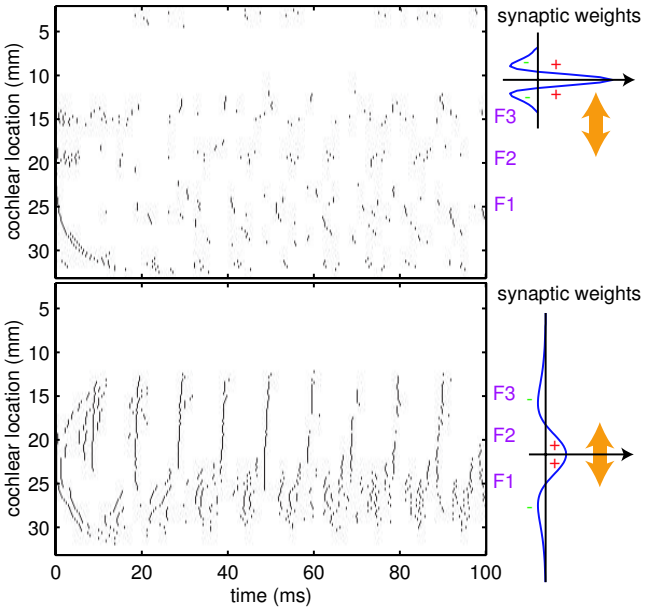

Fig. 4. Response of neurons covering spectral fields with different width. Synaptic weights of the neuron's receptive fields are plotted on the right hand side. Top panel: Receptive field width of 1.8 mm and bottom panel: 6.7 mm (defined as the width of the excitatory regions). Inputs were derived from ANF action potentials as plotted in Fig. 3.

placed one Q–neuron at each frequency channel and labeled it with the corresponding BM position of the ANF

connected to the center of its receptive field. Results for a narrow receptive field (distance between zero-crossings corresponds to ANFs originating from a BM-region spanning 1.8 mm) are shown in Fig. 4 (top panel). This neuron is able to detect and separate the formants F2 and F3. The neuron also fires within the region of the first formant. As the extension of this formant is broader, mostly due to the shape of the frequency-place transformation of the inner ear, Q–neurons innervated from ANFs originating at the location of F1 fire slightly less vivid than at the location of F2 and F3, despite the higher ANF input rate. This changes for Q–neurons with broader receptive fields. They display higher fire rates for F1, while the rates for F2 and F3 are decreasing. Q–neurons with a receptive field of 6.7 mm fire over a region spanning F1 to F3, with the most vivid reaction around F1 (Fig. 4, bottom panel). This figure, once more, highlights the importance of temporal processing: Especially in the region of F3, but also in all other stimulated areas, the 10 ms periodicity of the input sound is clearly visible and enhanced compared to the response of the ANF (compare Fig. 3). This result came as a surprise to us, as we have designed the Q–neurons for spectral processing. Nevertheless, this behavior can be explained by the coincidence-detecting features of the leaky integration performed by the Q–neurons and the large number of input ANFs.

3. CONCLUSIONS

We have modeled inner ear hydrodynamics followed by a saturating compression stage. We achieved large dynamic compression, which is required to code real-life sound signals into nerve-action potentials. Compression made the rate-level functions of ANFs grow nearly logarithmically over a dynamic range of up to 60 dB (data not shown). Individual frequency components of voiced sounds are not resolved at higher frequencies, instead, the temporal structure of the vowels, their periodicity, is preserved. Therefore, speech sounds are coded into a spatio-temporal pattern rather than a pure frequency-place code. We modeled neurons with receptive field shapes similar to neurons found in the auditory pathway [11]. Their receptive fields exhibited central excitation and lateral inhibition. When their excitatory receptive field width matched location and bandwidth of a formant, they responded vividly. Balancing excitatory and inhibitory synaptic weights reduced level-dependencies of their responses (results not shown) and suppressed most of the spontaneous activity present at higher frequencies (see Fig. 4). A whole population of these neurons with a range from narrow to wide spectral receptive fields code sound signals in a way equivalent to cepstral analysis, which is successfully applied in conventional automatic speech recognition systems. Like the cepstrum, these neurons code the shape of the logarithmic amplitude spectrum. We noticed also that temporal features are not destroyed by the neurons we designed to perform spectral processing, instead they were even enhanced. We expect a large, yet unused potential exploiting the fine-grained temporal information inherent in spike trains, with the potential to improve speech recognition in noisy and reverberant environments.

4. REFERENCES

[1] H.W. Strube, "A computionally efficient basilar-membrane model," *Acustica*, vol. 58, pp. 207–214, 1985.

[2] D.D. Greenwood, "A cochlear frequency-position function for several species – 29 years later," *J. Acoust. Soc. Am.*, vol. 87, pp. 2592–2605, 1990.

[3] M.A. Ruggero, S.S. Narayan, A.N. Temchin, and A. Recio, "Mechanical bases of frequency tuning and neural excitation at the base of the cochlea," *Proc. Natl. Acad. Sci. USA*, vol. 97, pp. 11744–11750, 2000.

[4] Deng L. and Geisler C.D., "A composite auditory model for processing speech sounds," *J. Acoust. Soc. Am.*, vol. 82, pp. 2001–2012, 1987.

[5] D.C. Mountain and A.R. Cody, "Multiple modes of inner hair cell stimulation," *Hear. Res.*, vol. 132, pp. 1–14, 1999.

[6] T. Moser and D. Beutner, "Kinetics of exocytosis and endocytosis at the cochlear inner hair cell afferent synapse of the mouse," *Proc. Natl. Acad. Sci. USA*, vol. 97, pp. 883–888, 2000.

[7] L.H. Carney, "A model for the responses of low-frequency auditory-nerve fibers in cat," *J. Acoust. Soc. Am.*, vol. 93, pp. 401–417, 1993.

[8] C.J. Sumner, E.A. Lopez-Poveda, L.P. O'Mard, and R. Meddis, "A revised model of the inner-hair cell and auditory-nerve complex," *J. Acoust. Soc. Am.*, vol. 111, pp. 2178–2188, 2002.

[9] A. Klug, E.E. Bauer, and G.D. Pollak, "Multiple components of ipsilaterally evoked inhibition in the inferior colliculus," *J. Neurophysiol.*, vol. 82, pp. 593–610, 1999.

[10] M.L. Sutter, "Shapes and level tolerances of frequency tuning curves in primary auditory cortex: Quantitative measures and population codes," *J. Neurophysiol.*, vol. 84, pp. 1012–1025, 2000.

[11] C.E. Schreiner and G. Langner, "Periodicity coding in the inferior colliculus of the cat. II. Topographical organization," *J. Neurophisiol.*, vol. 60, pp. 1823–1839, 1988.