An auditory model for coding speech into nerve-action potentials

Marcus Holmberg and Werner Hemmert
Infineon Technologies AG, Corporate Research ST, Munich, Germany,
Email: {marcus.holmberg,werner.hemmert}@infineon.com

Introduction

In this paper we present a model for coding arbitrary acoustic stimuli into auditory nerve action potentials. The model consists of an inner ear model with high dynamic compression, and an existing inner hair-cell and auditory nerve-synapse model. For this study, we have tuned the model to recent psychophysical data [3]. As the dynamic range of the sensory cells is limited, the challenge lies in the inner ear providing appropriate compression so that sound signals can be coded into action potentials without loss of information.

Model description

The passive motion of the basilar membrane (BM) is described by the hydromechanics of the cochlea. The solution in the one-dimensional case is described by a transmission-line [5]. Such a transmission-line with 100 sections forms the basis of our BM model. It provides an adequate description of the travelling wave in the living cochlea at high sound levels. At low to medium levels, BM vibration is amplified. The amplification is most probable a result of an active process involving electromotility of the outer hair-cells (OHC), and the mechanism is often referred to as the cochlear amplifier. The cochlear amplifier also significantly increases the frequency selectivity of BM motion. The gradual saturation of the amplifier causes a compression of the dynamic range. Numeral other nonlinear traits of the cochlea, such as two-tone suppression and distortion products, have also been related to the cochlear amplifier.

We model amplification and compression phenomenologically by cascading second-order time-varying resonators after each section (compare [4]). The output of each resonator is fed through a first-order Boltzmann function, mimicking the transducer channels of the OHCs, and thereafter used to modulate the Q-value of the resonator. Each of the resonators compresses a certain stimulus range, determined by the parameter of the Boltzmann function. Modelling results indicate that four resonators in series provide sufficient compression.

The model of the inner hair-cell (IHC) and the synaptic processes is taken from Sumner *et al.* [6]. This model includes low spontaneous rate (LSR) and high spontaneous rate (HSR) fibers innervating the same inner hair cell.

Tuning of a human auditory model

The resonant frequencies of the resonators were adjusted to Greenwood's map of the human cochlea [1]. Characteristic frequencies (CF) below 50 Hz and above 15 kHz were discarded. A measure of frequency selectivity common in psychoacoustics is the equivalent rectangular bandwidth (ERB), often expressed as a quality factor Q_{ERB} (resonant frequency over bandwidth). Forward masking notched-noise experiments [3] have revealed much sharper tuning of the auditory filters than previously believed. In contrast to earlier psychoacoustic estimations, the auditory filter frequency selectivity increases with frequency also at the high-frequency part of the cochlea. The formula

$$Q_{ERB} = 11 f^{0.27}, \qquad (1)$$

(where f is the CF in kHz) fits auditory filters between 1 and 8 kHz. This relationship was extended to the whole frequency range modelled. Maximum Q-values of the resonators in our model were adjusted to attain this bandwidth at low stimulus levels.

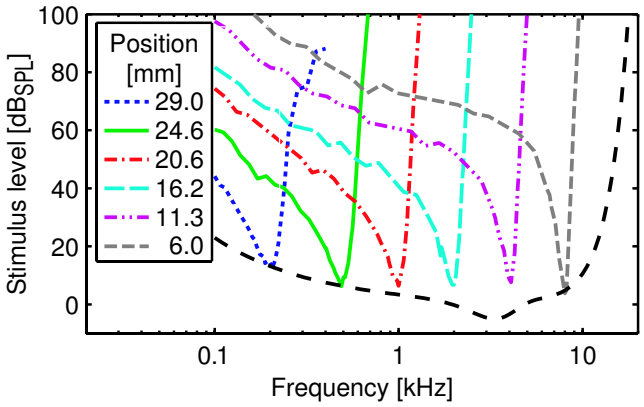

Figure 1: Threshold tuning curves for modelled HSR auditory nerve fibers at cochlear locations as indicated in legend (relative the stapes). The dashed line indicates hearing threshold. The auditory filters show increasing Q-values with CF.

A hallmark of the inner ear is its dynamic compression. The compression enables the inner hair-cells (with a dynamic range of approximately 40 dB) to code the wide range of signal levels occurring in nature. Psychophysical experiments [2] show that for CFs above 0.5 kHz the compression is independent of CF. The growth ratio was estimated to be between 1/3 dB/dB and 1/5 dB/dB (corresponding to a power law with exponent between 0.33 and 0.2). It is however still an open question whether the compression at low CF is a result of cochlear or neural processing. A growth ratio of approximately 0.25 dB/dB was achieved in our model above 1 kHz, whereas compression decreased slightly at lower CFs. The maximum Q-values of the resonators, tuned to match the filter shapes, restrict the possible compression ratio at low frequencies.

Results

Figure 1 shows threshold tuning curves of HSR-fibers at six different locations in the cochlea. Each point denotes the stimulus (pure tone) level necessary to measure a statistically significant increase in firing rate. The curves thus represents the auditory filters at lowest working sound levels. The Q-values clearly increase with increasing CF (more basal positions in the cochlea). The amplification causes the filters to be almost symmetrical near CF. On the low-frequency side of the CF, in the so-called tail of the response, the slope becomes shallow as a result of the asymmetric travelling wave response.

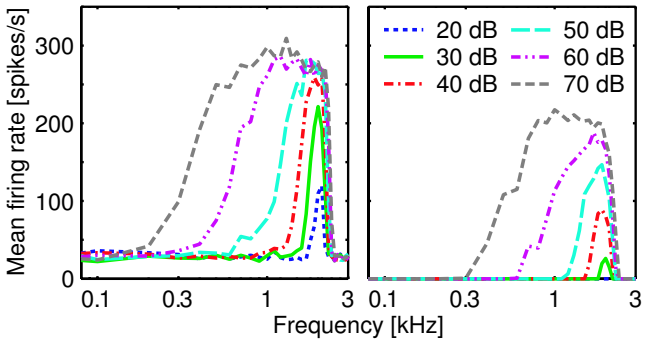

Figure 2: Auditory nerve response areas of a modelled HSR- (left panel) and LSR-fiber (right panel), both with CF 2 kHz. The HSR-fiber has low threshold and limited dynamic range. The LSR-fiber has a large dynamic range, and the widening of cochlear filters and shift of CF with stimulus level is apparent.

In Fig. 2, AN fiber responses to pure tones with varying intensity are plotted (CF 2 kHz). The left panel shows a HSR-fiber with low threshold and a spontaneous rate of 30 spikes/s. The fiber saturates within 30 dB, because BM vibration grows almost linearly with intensity at low levels. The right panel shows the response of an LSR-fiber, innervating the same IHC. Threshold is higher, causing the fiber to react only in the compressive part of BM response. The response area widens because of broadening cochlear filters. Also apparent in Fig. 2 is the fiber's CF shift towards lower frequencies in accordance with psychoacoustic and physiological measurements.

Figure 3 shows the whole LSR AN population (2000 fibers in total, 20 per frequency channel) response to the vowel "a" (male speaker, 63 dB_{SPL}). The increasing delays towards low CFs is a result of the travelling wave. Four formants (F1: 0.7, F2: 1.7, F3: 2.4 and F4: 3.8 kHz) are distinguishable, most clearly at onset. As is obvious already from Fig. 2, a rate-place code cannot resolve individual harmonics of spoken language; at normal intensity levels the cochlear filters are too wide. In Fig. 3 phase-locking to the signal envelope is apparent. The fundamental frequency (signal periodicity) is preserved for all formants. Each stroke of the glottis elicits synchronous firing of a wide range of frequency channels (indicated by arrows in Fig. 3).

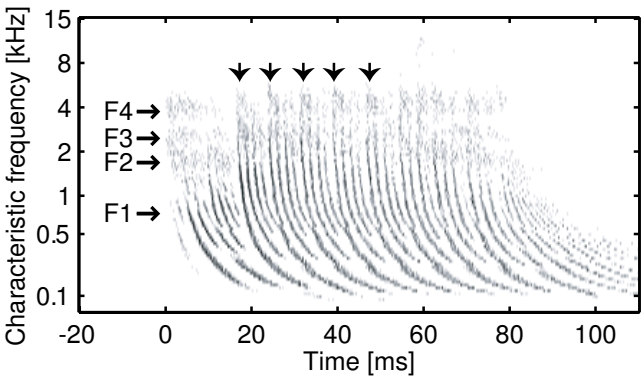

Figure 3: Auditory nerve population response to vowel "a" (male speaker, 63 dB_{SPL} (rms)). From each section 20 LSR fibers were derived. Periodicity, caused by the strokes of the glottis (arrows), is preserved in the representation of the formants (F1-F4).

Conclusions

We have constructed a model of the auditory periphery that reproduces human auditory filter shapes and the high dynamic compression found in physiological and psychoacoustic experiments. Because the dynamic range of the sensory cells is low, large compression is essential to code sound signals into action potentials of the auditory nerve without information loss. In addition, the presented model also achieves good temporal resolution which conserves temporal traits such as onset of plosives (data not shown) and the periodicity of vowels. We postulate this fine-grained temporal information to be crucial for human speech intelligibility in noisy environments, and a way to achieve more robust automatic speech recognition.

References

[1] D.D. Greenwood. J Acoust Soc Am **87** (1990), 2592–2605.

[2] E.A. Lopez-Poveda, C.J. Plack, and R. Meddis. J Acoust Soc Am **113** (2003), 951–960.

[3] A.J. Oxenham, and C.A. Shera. J Assoc Res Otolaryngol **4** (2003), 541–554.

[4] A. Robert, and J.L. Eriksson. J Acoust Soc Am **106** (1999), 1852–1864.

[5] H.W. Strube. Acustica **58** (1985), 207–214.

[6] C.J. Sumner, E.A. Lopez-Poveda, L.P. O'Mard and R. Meddis. J Acoust Soc Am **111** (2002), 2178–2188.