Time interval representations of a consonant-vowel syllable in computer models of normal and hearing-impaired listeners

T. Jürgens*§, R. Meddis§, G.J. Brown‡, N.R. Clark§ and M. Robertson‡

*Medizinische Physik, Carl-von-Ossietzky Universität Oldenburg, Germany

§Department of Psychology, University of Essex, UK

‡ Department of Computer Science, University of Sheffield, UK

tim.juergens@uni-oldenburg.de, rmeddis@essex.ac.uk, g.brown@dcs.shef.ac.u

Motivation

- Physiological studies in (normal-hearing) cats show that formants of synthetic speech are represented in the time-domain as bands of phase-synchronous auditory nerve (AN) discharge rates (Miller and Sachs, 1983).
- This “synchrony capture“ phenomenon weakens in cats after acoustic trauma, especially for higher-frequency formants (Miller *et al.,* 1997).

- **Does the auditory model of Meddis (2006) match the physiological data on temporal coding when configured with normal hearing parameters, and how does temporal coding differ in an impaired auditory model?**

Goals of the study

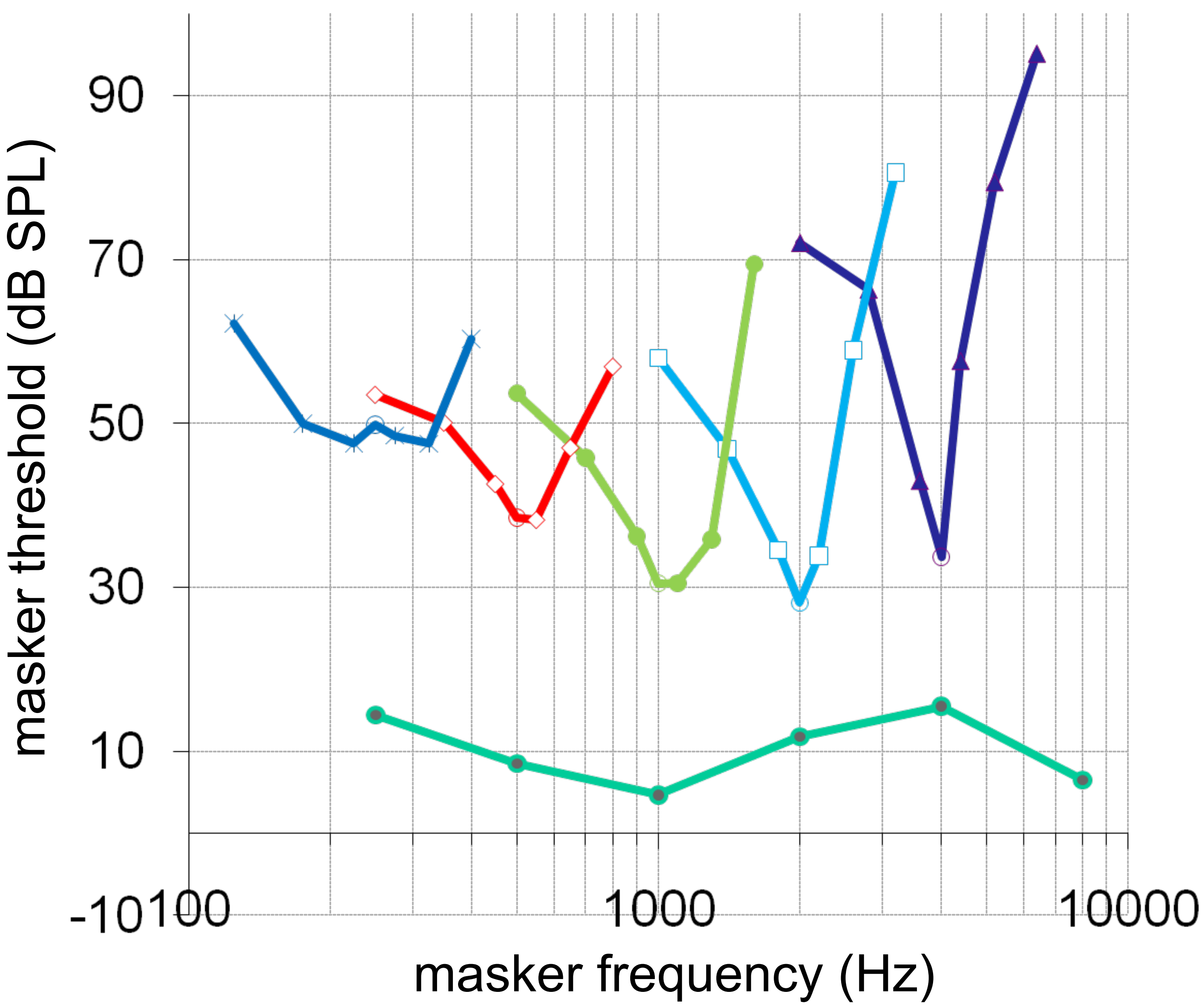
•To determine whether the auditory model of Meddis (2006) set to normal hearing encodes speech formants in the time intervals of AN firing patterns (synchrony capture), as shown in cat data

•To determine whether synchrony capture is different in the normal-hearing and an impaired version of the Meddis (2006) model

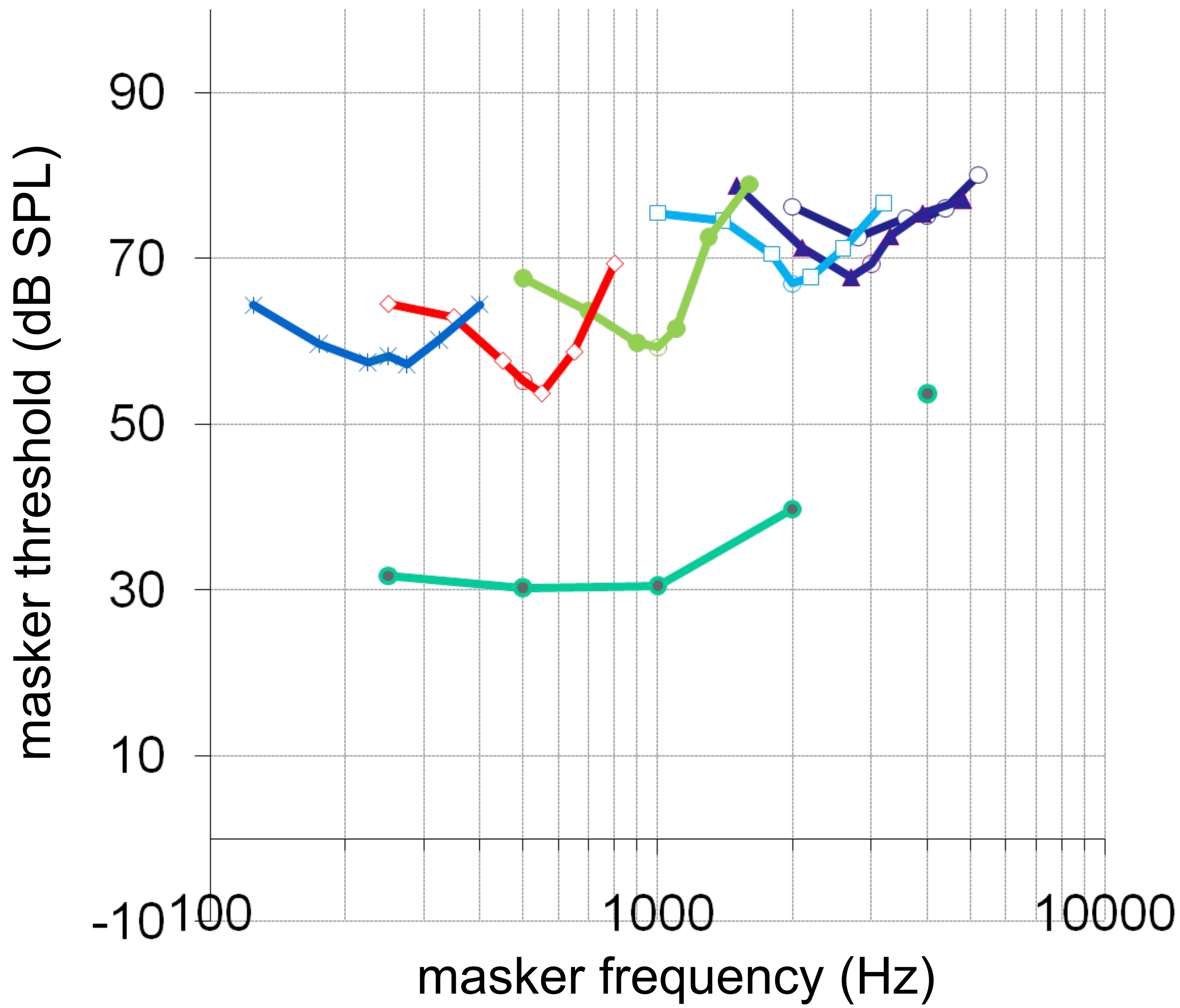
•To evaluate the level-dependency of synchrony capture in the model in both normal-hearing and hearing-impaired listeners

Auditory profiles of listeners

Normal-hearing listener



Hearing-impaired listener



Auditory profiles of a normal-hearing listener (left) and a hearing-impaired listener (right): Iso-Forward-Masking-Level-Contours displaying frequency tuning at different target frequencies, and absolute thresholds.

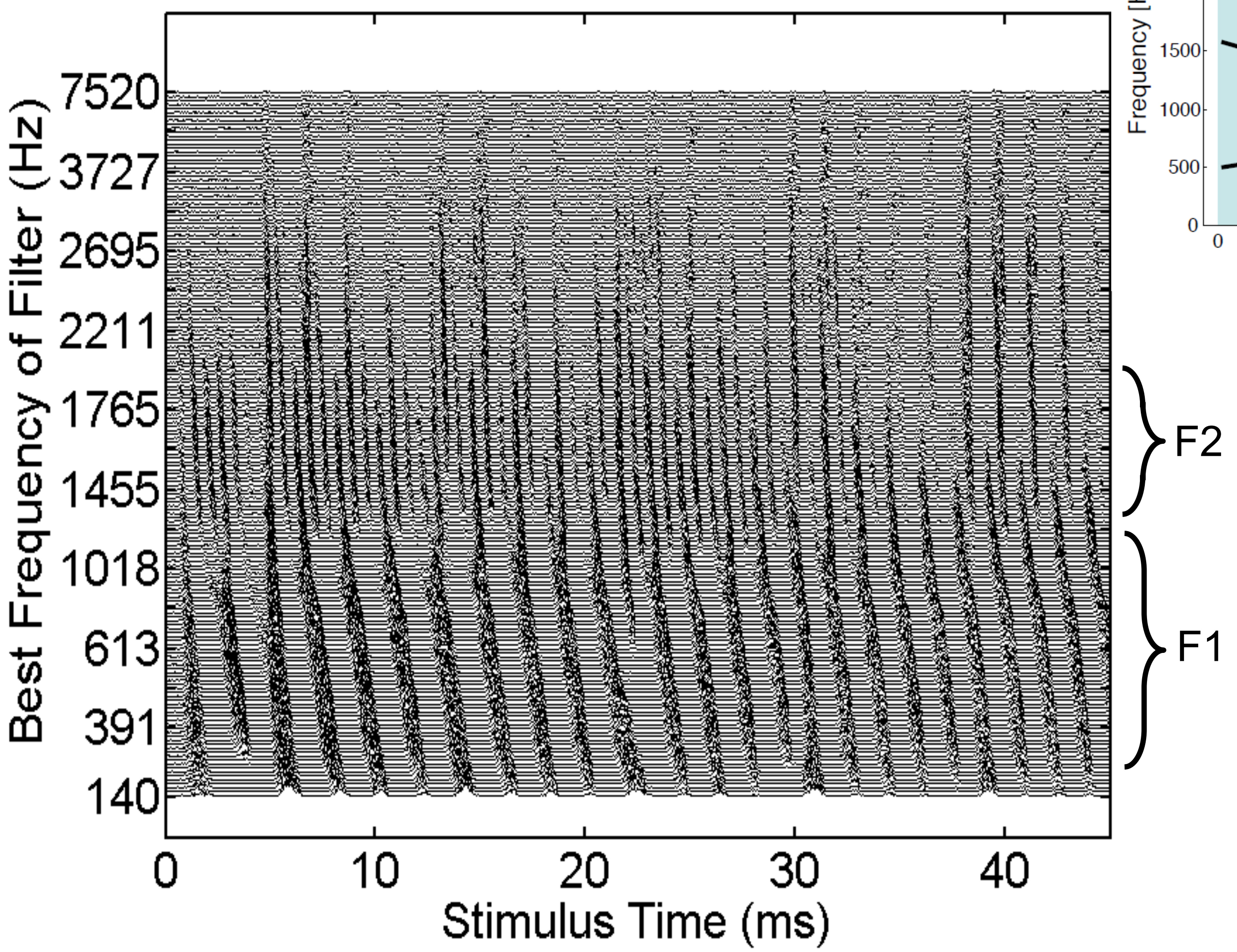
The hearing-impaired listener shows raised absolute thresholds and a broadening of tuning.

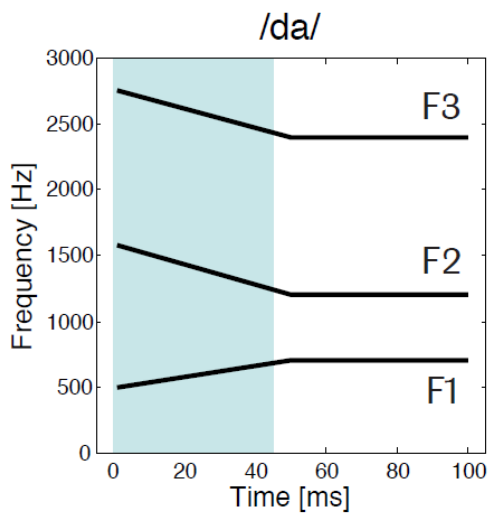
The auditory model was set to reproduce as closely as possible these profiles (modeled auditory profiles can be found on poster #78 from N.R. Clark *et al.*, University of Essex).

Comparison of simulated auditory nerve firing patterns

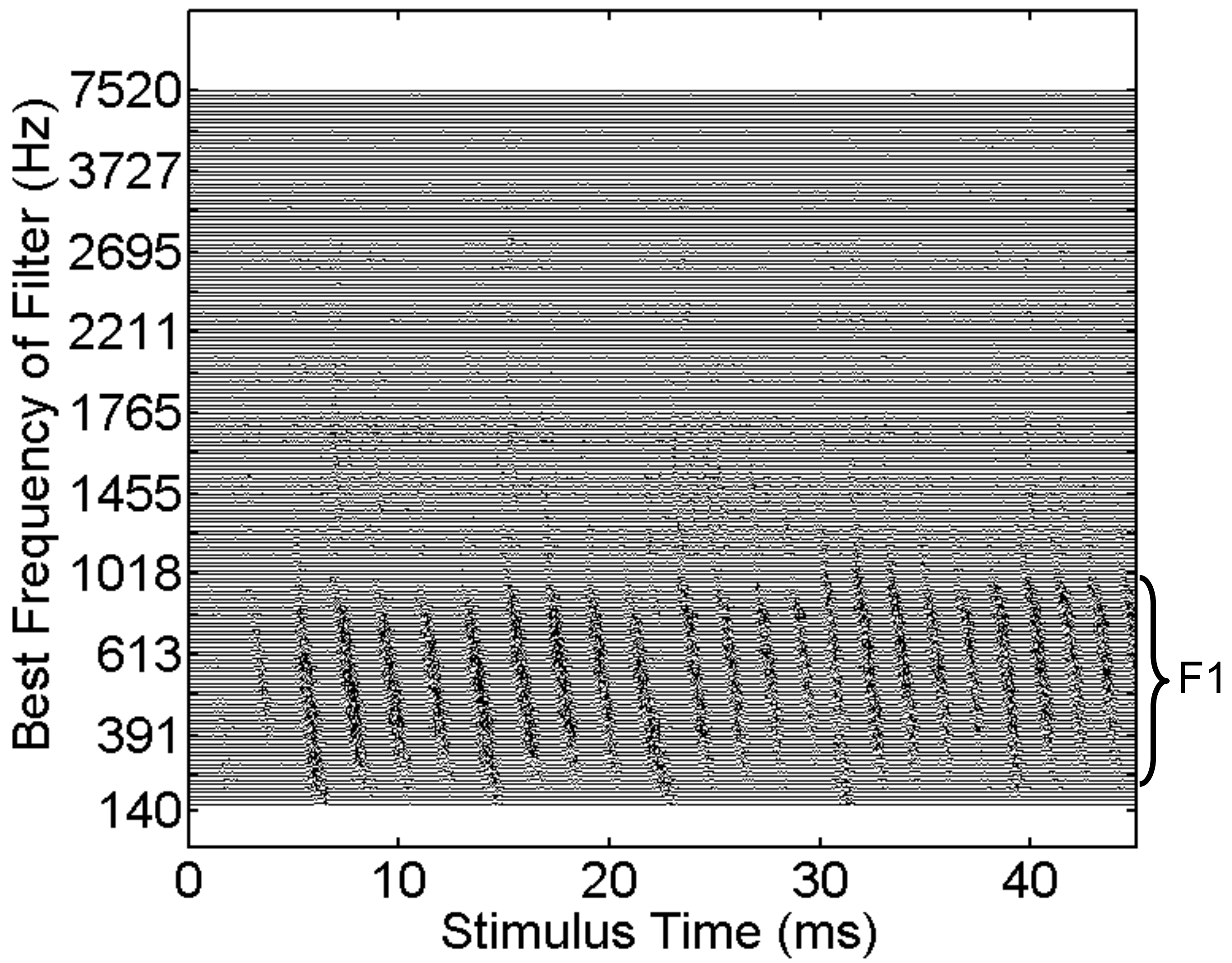
- A replica of the three-formant syllable /da/ used by Miller and Sachs (1983) was synthesized in Praat (100 ms in length, with formant transitions over the first 50 ms).
- Auditory nerve firing patterns were simulated at various levels, here shown for 50 dB SPL.

Normal-hearing listener





Hearing-impaired listener



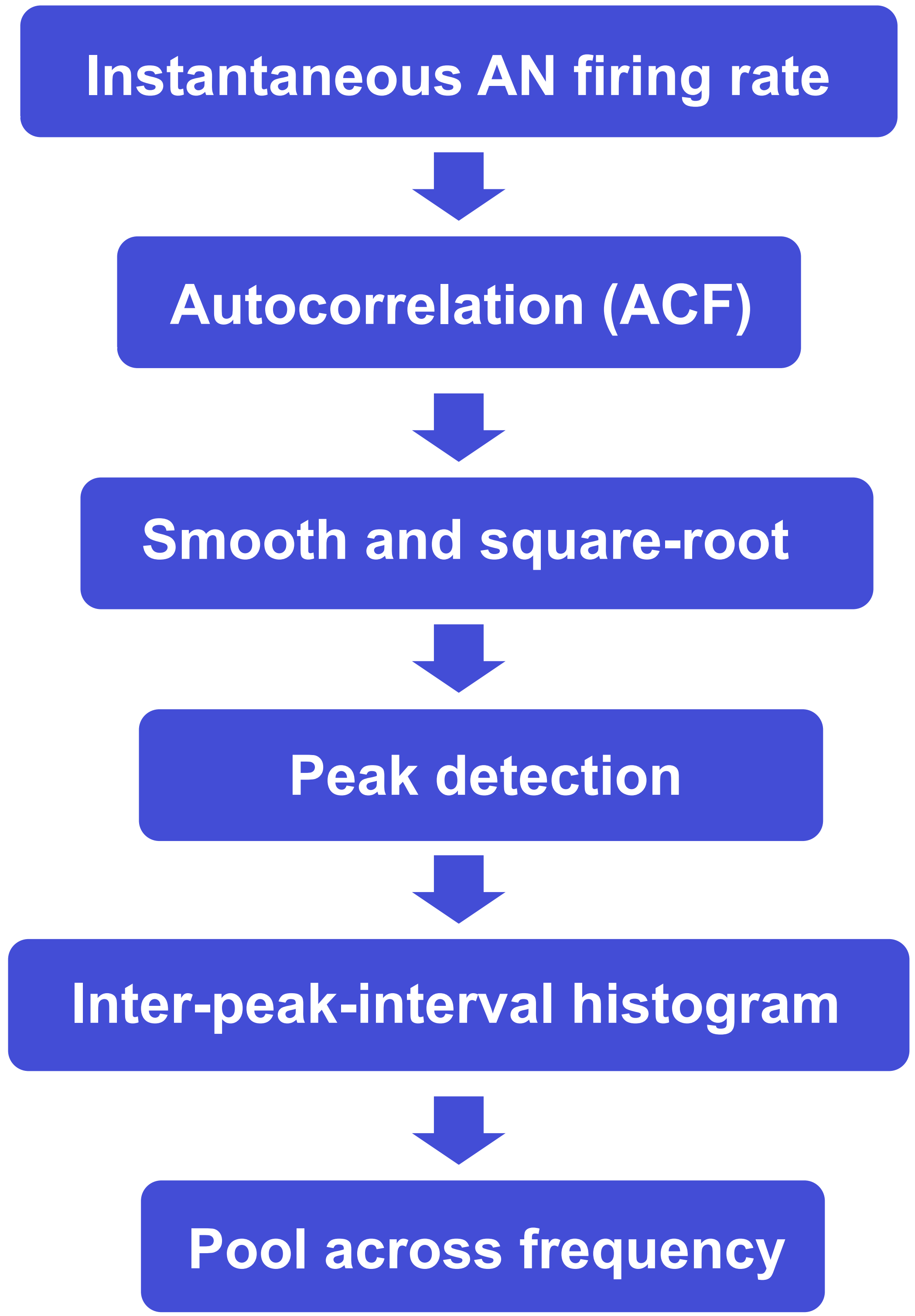
The normal model response to the syllable exhibits the phenomenon of 'synchrony capture'. The formants F1 and F2 are each represented by a band of AN fibres with synchronised firing activity.

The impaired model response at the same level is different. F1 is well represented in the AN discharge rates, F2 is very poorly represented.

Time-domain analysis of auditory nerve firing

•The interpeak-interval analysis proposed by Secker-Walker and Searle (1990) is used to extract timing information from the AN firing rate pattern.

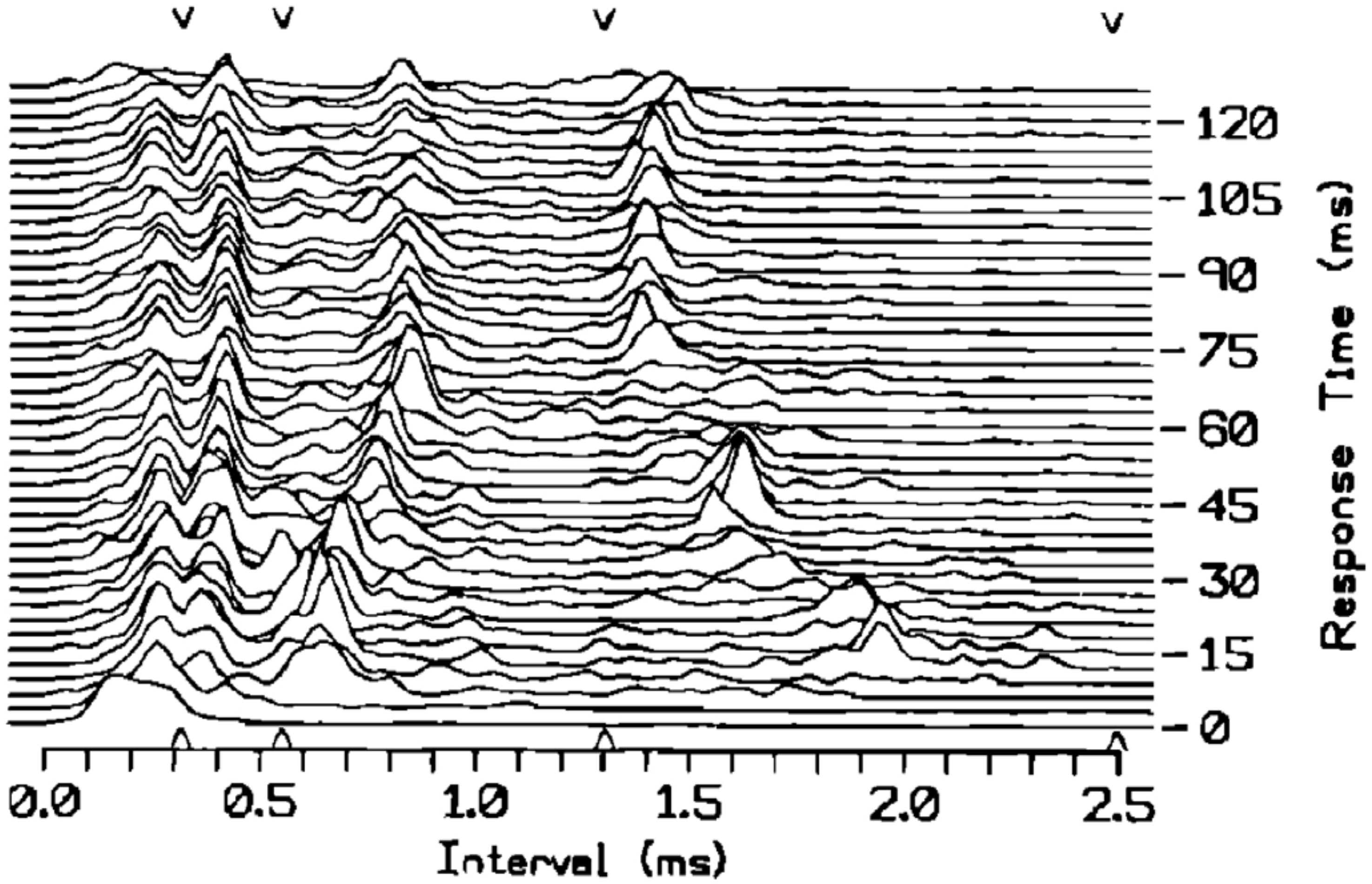
- Short sections (10 ms) of the instantaneous firing rate of each channel are autocorrelated.
- Square root of each autocorrelation is smoothed with an 11-point Hamming window.
- Peaks are found in the resulting smoothed root autocorrelation (SRA).
- The first three intervals in the SRA of each frequency channel are pooled to give an inter-peak-interval histogram (IPIH).



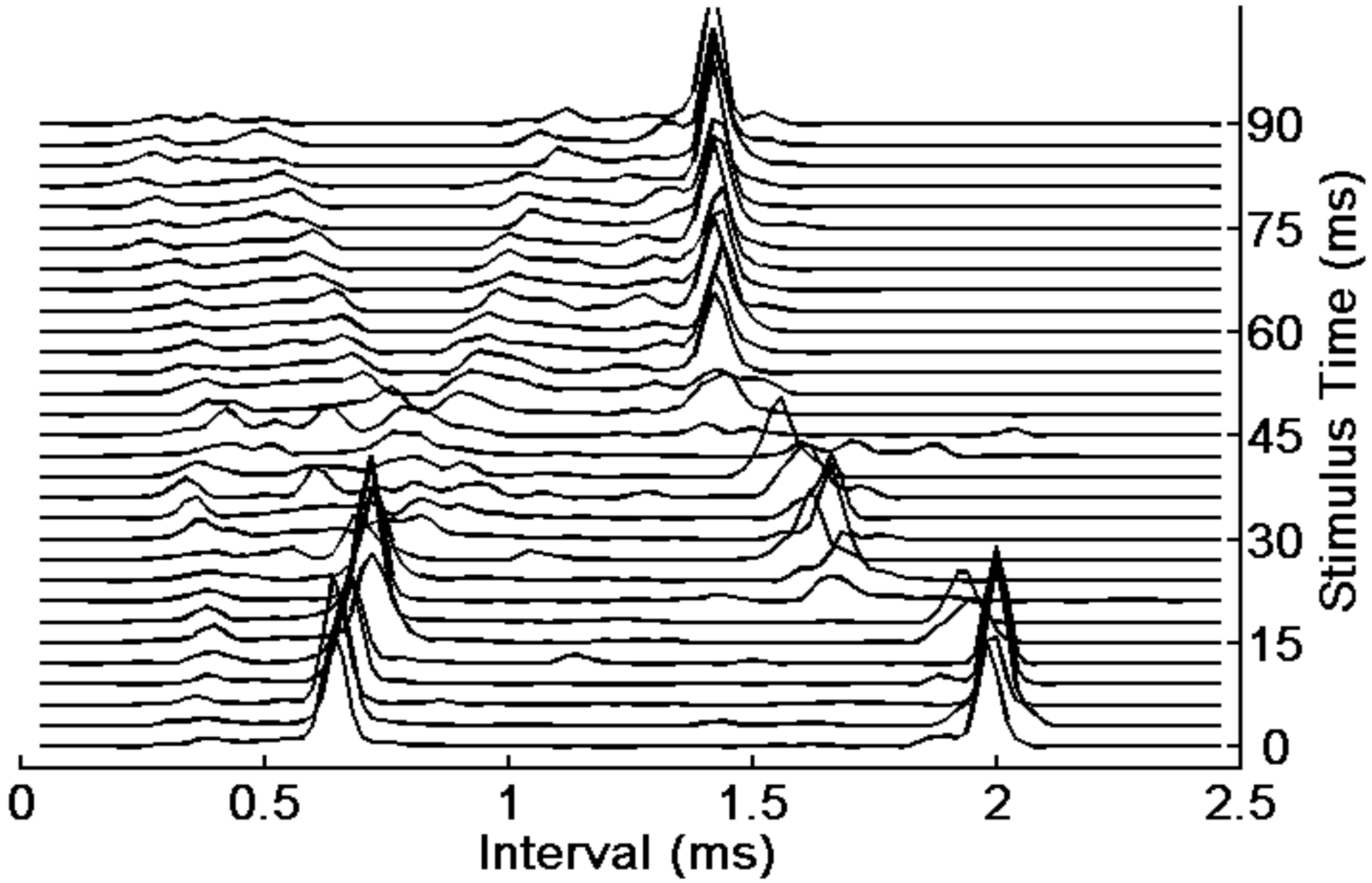
Comparison of physiological and simulated IPIH data

- Pooled Inter-Peak-Interval Histograms (IPIH) of the /da/ stimulus at 49 dB SPL were derived from the simulated AN firing patterns (slide 3).
- These IPIHs were compared to IPIHs derived from firing-rate data of the auditory nerve of a cat (Miller and Sachs,1983, and Secker-Walker and Searle,1990).

cat data



simulated (normal-hearing) data

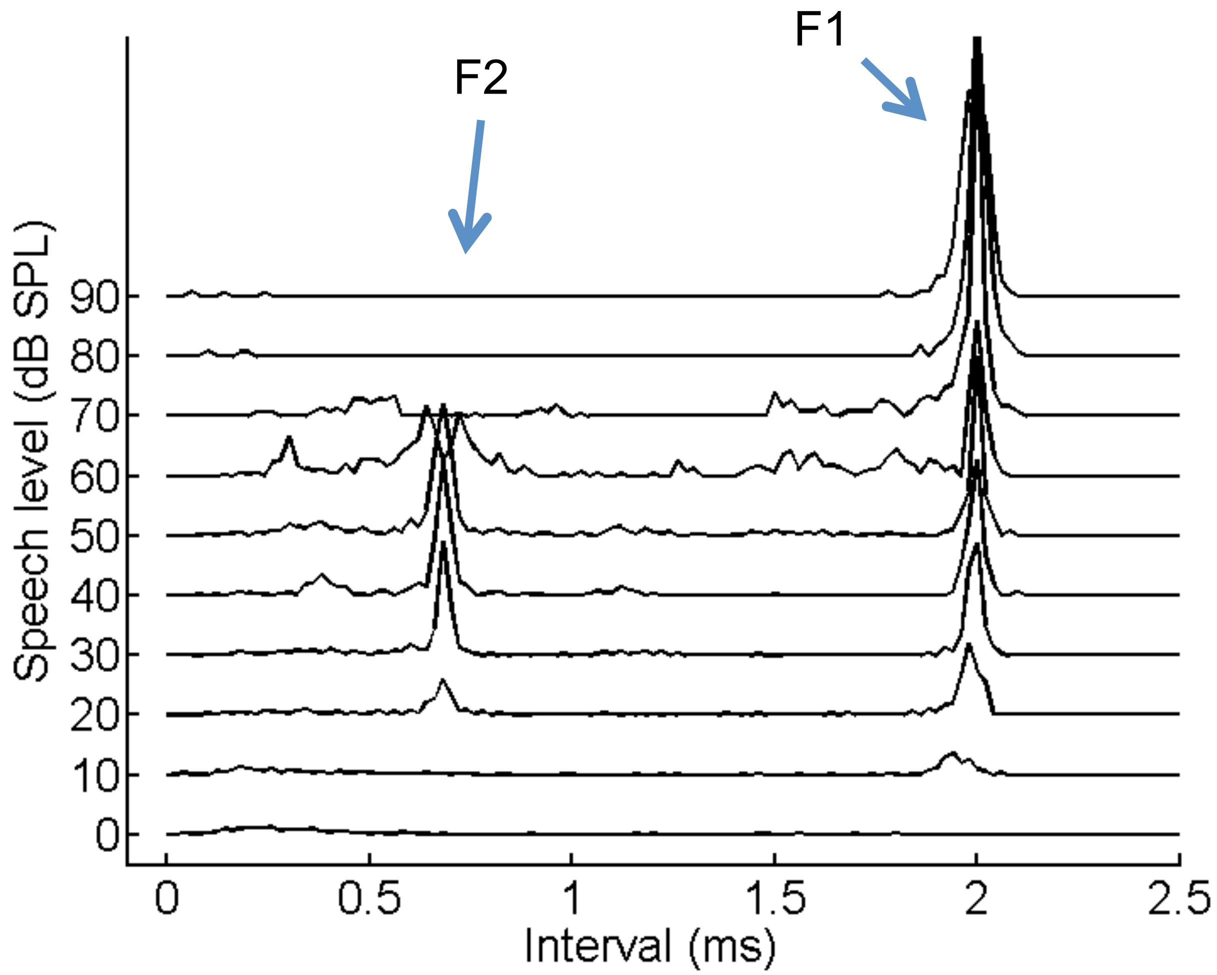


- Despite the difference in species, there is a good agreement between the cat data and the simulated (normal-hearing) data, except for the highest formant F3 (in the cat data represented by peaks at about 0.4 ms Interval).

Level-dependency of the Inter-Peak-Interval-Histograms

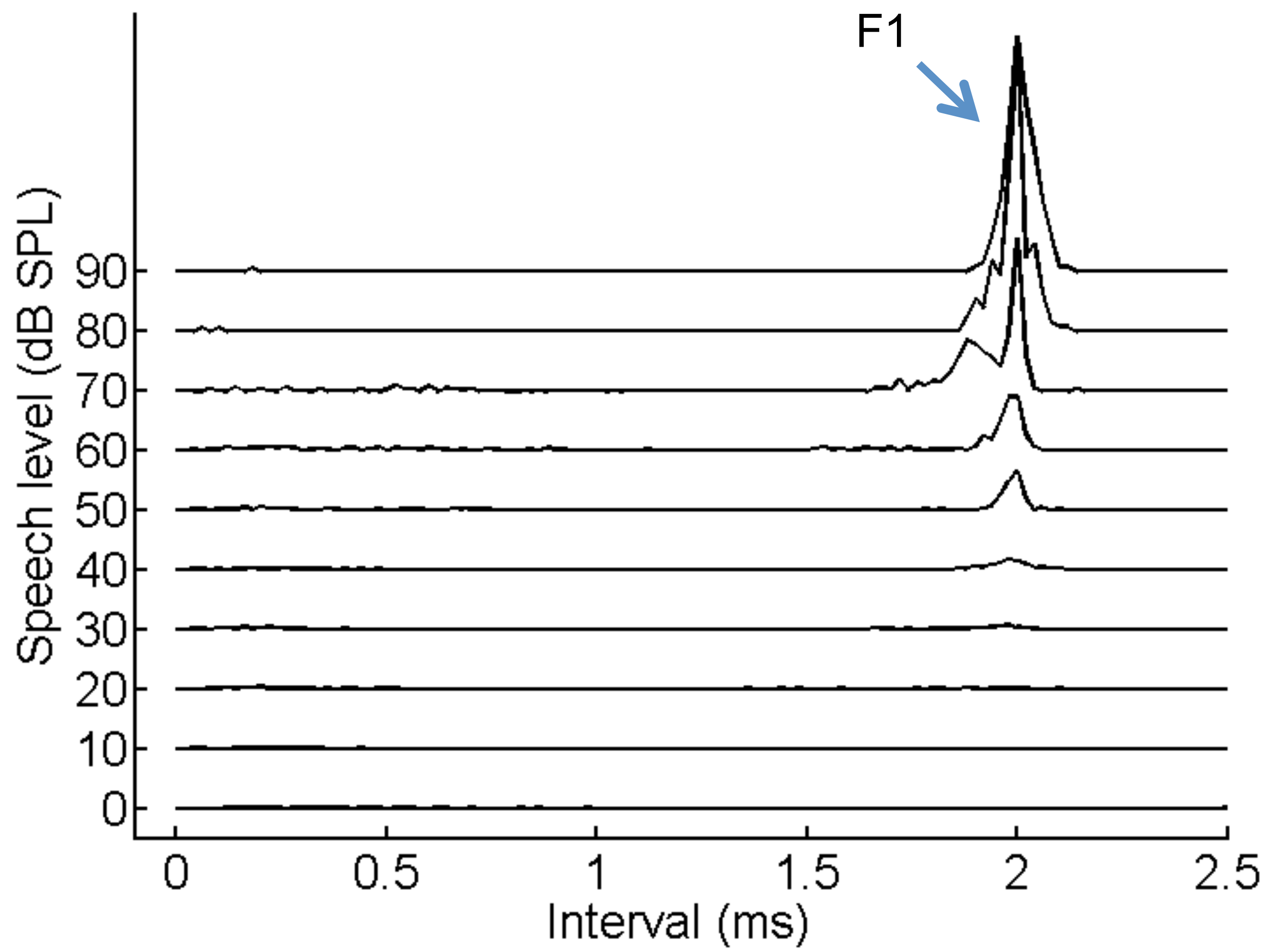
- To illustrate the dependency of the IPIH on speech level, one time segment of the IPIH (here: 12ms stimulus time) was chosen and plotted as line plot for different speech levels.
- The responses of the normal-hearing and the hearing-impaired listener are compared.

Normal-hearing listener



F1 and F2 are well represented across a broad range of speech levels.

Hearing-impaired listener



F1 is overrepresented at higher levels and F2 hardly represented.

Discussion

- Synchrony capture arises in the normal model, because the broad, asymmetric filter shapes allow a strong formant to capture the temporal response of filters near to weaker frequency components. This was also shown by Holmes *et al.* (2004) in guinea pig auditory nerve models.

- In a spectral valley, on-frequency harmonics get compressed in the normal model, whereas an off-frequency formant is processed linearly and thus can dominate the IPIH.

- The upward spread of the temporal response to F1 is greater in the hearing-impaired model, because the filters are wider.

- To discriminate speech sounds, both formants F1 and F2 are usually necessary. The missing of F2 in the time interval representation of the hearing-impaired listener could be one reason why some hearing-impaired listeners report that they hear the presence of speech, but cannot understand what was being said.

- At higher speech levels (≥ 70 dB SPL), the IPIH of the normal-hearing listener is also completely dominated by F1 (cf. slide 6). F2 is poorly represented in the temporal intervals of the AN discharge rates.

Software download

The model can be downloaded from http://tinyurl.com/5rmexxo.

Conclusions

- The formants F1 and F2 are well encoded in the IPIH of the model of normal-hearing listeners up to about 60 dB SPL, almost in agreement with the physiological data in cats.
- F1 was well represented in the IPIH of the model of a hearing-impaired listener; F2 is very poorly encoded.
- The data suggests that wider bandwidths and impaired compression lead to an exaggerated „synchrony capture“ effect for this hearing-impaired listener with F1 dominating the IPIH over almost all frequency channels.

Acknowledgements

The authors gratefully acknowledge the support of SFB TRR 31 “The active auditory system” and DFG (Jürgens), EPSRC (Brown, Meddis), and RNID (Robertson).

References

Holmes, S.D., Sumner, C.J., O`Mard, L.P., Meddis, R. “The temporal representation of speech in a nonlinear model of the guinea pig cochlea”, J. Acoust. Soc. Am., 116:3534-3545, 2004.

Meddis, R. “Auditory-nerve first-spike latency and auditory absolute threshold: A computer model”, J. Acoust. Soc. Am. 119:406–417, 2006.

Miller, M.I. and Sachs, M.B., “Representation of stop consonants in the discharge patterns of auditory-nerve fibres”, J. Acoust. Soc. Am., 74:502–517, 1983.

Miller, R.L, Schilling, J.R. , Franck, K.R, and Young, E.D., “Effects of acoustic trauma on the representation of the vowel /ε/ in cat auditory nerve fibers”, J. Acoust. Soc. Am., 101:3602-3616, 1997.

Secker-Walker, H.E. and Searle, C.L., “Time-domain analysis of auditory-nerve-fiber firing rates”, J. Acoust. Soc. Am., 88:1427–1436, 1990.