Auditory model responses to harmonic and inharmonic complex tones: Effects of the cochlear amplifier

VÁCLAV VENCOVSKÝ*

Musical Acoustics Research Center, Academy of Performing Arts in Prague, Prague, Czech Republic

Hopkins and Moore [J. Acoust. Soc. Am. 122, 1055-1068 (2007)] measured the ability of hearing-impaired (HI) listeners to discriminate harmonic (H) from inharmonic (I) – all harmonics shifted upwards by the same amount in Hz – complexes. The complexes were composed of many bandpass filtered harmonics (shaped stimuli) or five equal-amplitude harmonics (non-shaped stimuli). HI listeners performed worse with the shaped stimuli than with the non-shaped stimuli. Since shaping of the complexes should minimize envelope and spectral cues, listeners should discriminate H from I stimuli mainly using temporal fine structure (TFS) cues even when the harmonics are not resolved. This ability seems to be worsened in HI listeners. This study employed an auditory model with a physical cochlear model to show how the cochlear amplifier affects responses to H and I stimuli. For the shaped stimuli, the TFS of the simulated neural signals for H and I stimuli differed, represented by low cross-correlation coefficients computed from the shuffled cross-polarity correlograms. However, for the passive auditory model (simulating HI), the inter-spike intervals smaller than half of the stimulus period were similar. This could explain the poor performance for HI listeners. For the non-shaped stimuli, differences in the inter-spike intervals were observed even for the passive model, which could contribute to the improved performance.

INTRODUCTION

Hopkins and Moore (2007) investigated the ability of hearing-impaired (HI) listeners to discriminate harmonic (H) complex tones from inharmonic (I) complex tones, i.e. the complexes created by shifting all the spectral components in the H complexes upwards by the same amount in Hz. They used two types of stimuli: "shaped" and "non-shaped". The shaped stimuli were composed of many harmonics filtered by a bandpass filter. The bandpass filter should eliminate changes in the excitation patterns (spectral cues), which could be used to discriminate H from I complexes. Thus, if the stimuli contain unresolved spectral components (high relative to the fundamental frequency F0), listeners should discriminate between H and I complexes by using temporal fine structure (TFS) cues, i.e., the intervals between peaks of the TFS close to envelope maximums. On the other hand, the non-shaped stimuli were composed

---

*Corresponding author: vaclav.vencovsky@gmail.com

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

of five equal-amplitude harmonics. The listeners could thus use spectral cues to discriminate between H and I complexes. Hopkins and Moore (2007) showed that the performance of HI listeners in the H-I discrimination tasks was much poorer for the shaped than for the non-shaped stimuli. They therefore interpreted the results such that the HI listeners could not use TFS cues to discriminate the stimuli. However, some researchers have questioned that TFS information may code fundamental frequency (e.g., Oxenham *et al.*, 2009).

The aim of this study was to analyze the H and I complexes by an auditory model. The auditory model was composed of known models of the outer/middle ear processing, a physical cochlear model and algorithms simulating the physiology of inner hair cells and auditory-nerve synapses. The physical cochlear model simulated the active function of outer hair cells by a feedback force. This cochlear amplifier was removed to simulate hearing impairment.

METHODS

Stimuli

The stimuli were the same as those used in Hopkins and Moore (2007). The shaped stimuli were composed of the fundamental and higher harmonics up to 20 kHz, each starting in sine phase. The complexes were filtered by a bandpass filter centered at nominal harmonic number $N = 11$. The bandwidth of the bandpass filter was set such that the components between $N-2$ and $N+2$ had an amplitude of 1 and the amplitude of the remaining components decreased at a rate of 30 dB/octave. The non-shaped stimuli contained only five harmonics centered at $N = 11$, each starting in sine phase. The fundamental frequency, F0, of the stimuli was 100 Hz. The I complexes (shaped and non-shaped) were created by shifting all the spectral components in the H complexes upwards by $\Delta f = 35$ Hz, which should allow normal-hearing (NH) listeners to discriminate between the H and I complexes (Hopkins and Moore, 2007). The overall level of the shaped and non-shaped stimuli was 65 dB SPL.

Auditory model

The auditory model was composed of known algorithms simulating different parts of the peripheral ear: an outer- and middle-ear model from the Matlab Auditory Periphery (MAP), a physical cochlear model designed by Nobili *et al.* (2003) and algorithms simulating the function of inner hair cells and auditory-nerve synapse (Meddis, 2006). The model input was an acoustic waveform at the entrance of the outer ear; the model outputs were neural discharges (spikes) fired into the auditory nerve.

The individual stages of the auditory model were used with the parameters described in the above given references. The physical cochlear model had 300 channels with characteristic frequencies (CFs) distributed between 20 Hz and 17 kHz. The model simulated the active function of outer hair cells by a feedback force. This

level (dB SPL)	characteristic frequency (kHz) 0.125	0.25	0.5	1	2	4
	active model (NH), ERB (Hz)					
20	43	62	89	141	225	390
40	43	62	90	148	245	521
60	43	70	122	201	337	818
80	54	98	168	307	528	1107
	passive model (HI), ERB (Hz)					
	90	170	311	529	982	1640
$\mathrm{ERB_{GM}}$(Hz)	38	52	79	133	241	456

Table 1: Equivalent rectangular bandwidth (ERB) of simulated cochlear filters (physical cochlear model). $\mathrm{ERB_{GM}}$ are psychoacoustical data given in Glasberg and Moore (1990).

cochlear amplifier was removed to simulate hearing impairment. Since the model with the parameters given in Nobili *et al.* (2003) had wider cochlear filters than was measured psychophysically (e.g., Glasberg and Moore, 1990), its damping parameter was adjusted here. Table 1 shows the measured equivalent rectangular bandwidth (ERB) of the simulated cochlear filters: active model (with the cochlear amplifier); and passive model (without the cochlear amplifier). $\mathrm{ERB_{GM}}$ are the psychoacoustically measured ERBs given in Glasberg and Moore (1990).

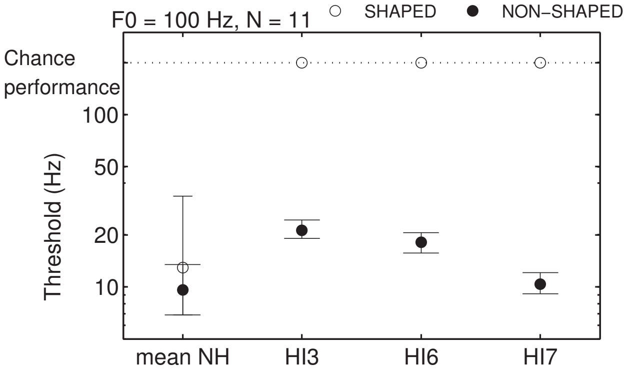

Fig. 1: Thresholds (Δf) in H-I discrimination tasks. The data were reproduced from Hopkins and Moore (2007). Open circles are the thresholds for the shaped stimuli, filled circles for the non-shaped stimuli, and the error bars are standard deviations of the mean. Abscissa denotes the listeners: “mean NH” are the mean values across the NH listeners; HI3, HI6, and HI7 are data from the HI listeners. The notation is the same as in Hopkins and Moore (2007).

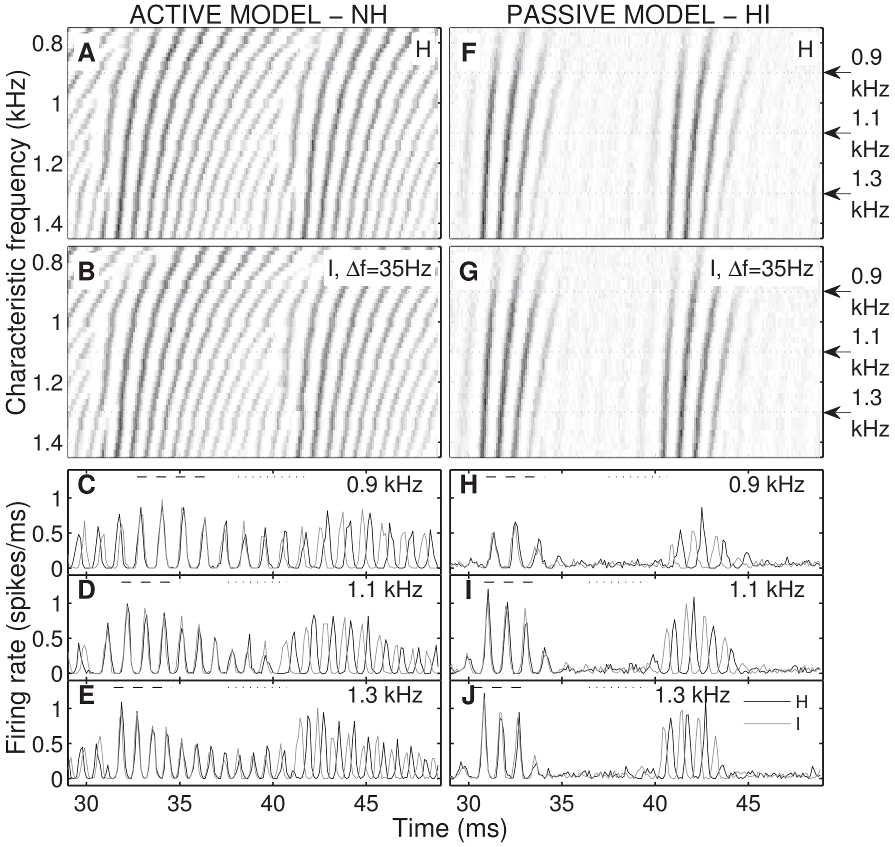

Fig. 2: Auditory model responses to the shaped stimuli. Panels in the left column (A to E) show the data for the active auditory model (NH); panels in the right column (H to J) for the passive auditory model (HI). (A, F) The responses to the H complexes. (B, G) The responses to the I complexes ($\Delta f = 35$ Hz). (C, D, E) The responses of the active model (in the channels with CF of 0.9, 1.1 and 1.3 kHz, respectively). (H, I, J) The responses of the passive model (in the same channels as in C, D, E).

RESULTS AND DISCUSSION

Fig. 1 shows psychoacoustical thresholds (Δf) from the H-I discrimination tasks, measured by Hopkins and Moore (2007). Open circles show the thresholds for the shaped stimuli, filled circles for the non-shaped stimuli, error bars show standard deviation of the mean. The abscissa of the graphs shows the listeners: the mean values calculated across the NH listeners are denoted as “mean NH”; HI3, HI6, and HI7 denote the data from the HI listeners. The notation is the same as used in Hopkins and Moore (2007). For the shaped stimuli, the performance of the HI listeners was at a chance level, which may indicate their inability to use TFS cues in these tasks.

Figures 2 and 3 show the auditory model responses – post stimulus time histograms (PSTHs), binwidth = 10/(sample frequency), 600 repetitions – to the shaped and non-

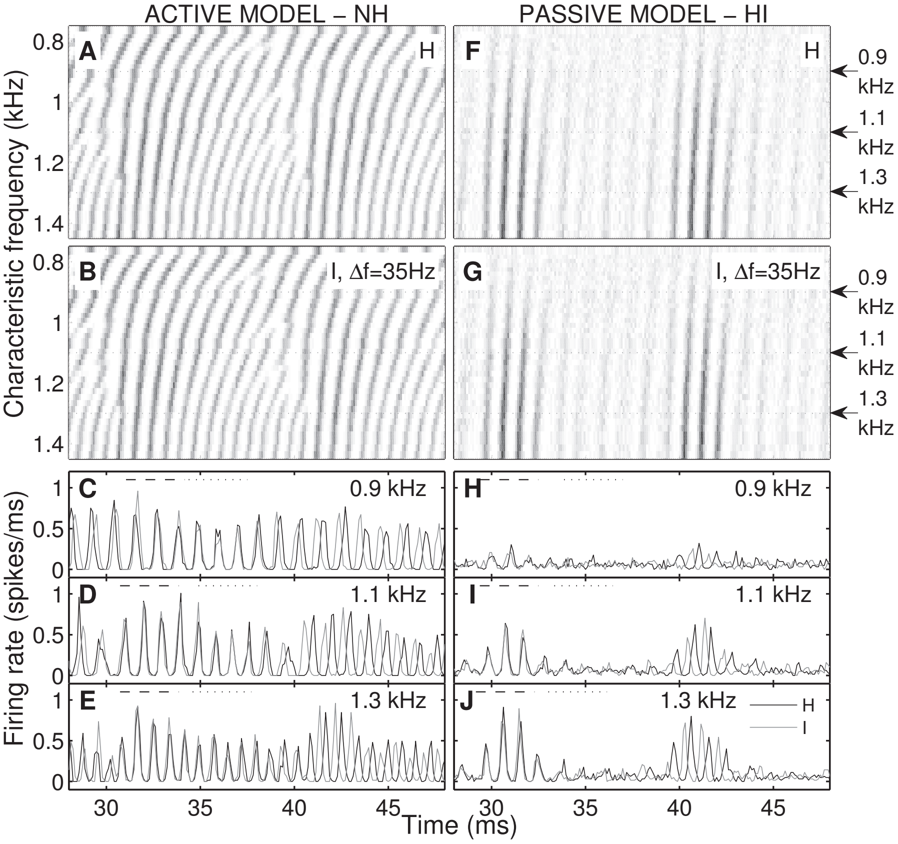

Fig. 3: Auditory model responses to the non-shaped stimuli. Description of the panels is similar to Fig. 2.

shaped stimuli, respectively. In each figure, the panels in the left column show the responses of the active model (simulating NH); the panels in the right column show the responses of the passive model (simulating HI). Panels A and F show the PSTHs (in the several adjacent channels) for the H complexes; panels B and G show the PSTHs for the I complexes ($\Delta f = 35\,\text{Hz}$). Panels C, D, E (for the active model) and H, I, J (for the passive model) show the PSTHs in three discrete model channels with CF given in each panel; the black solid lines show the PSTHs for the active model, and the gray solid lines for the passive model.

The responses to the H and I complexes differ – the corresponding PSTHs do not exactly overlap (see Figs. 2 and 3). This difference leads to low (close to zero) across-stimulus neural cross-correlation coefficient calculated according to the method described in Kale *et al.* (2014), which indicates perceptible changes in the stimuli. However, the TFS cues suggested by Hopkins and Moore (2007) are also visible in the responses of the passive model. These results would indicate that HI listeners also have TFS information in the activity patterns.

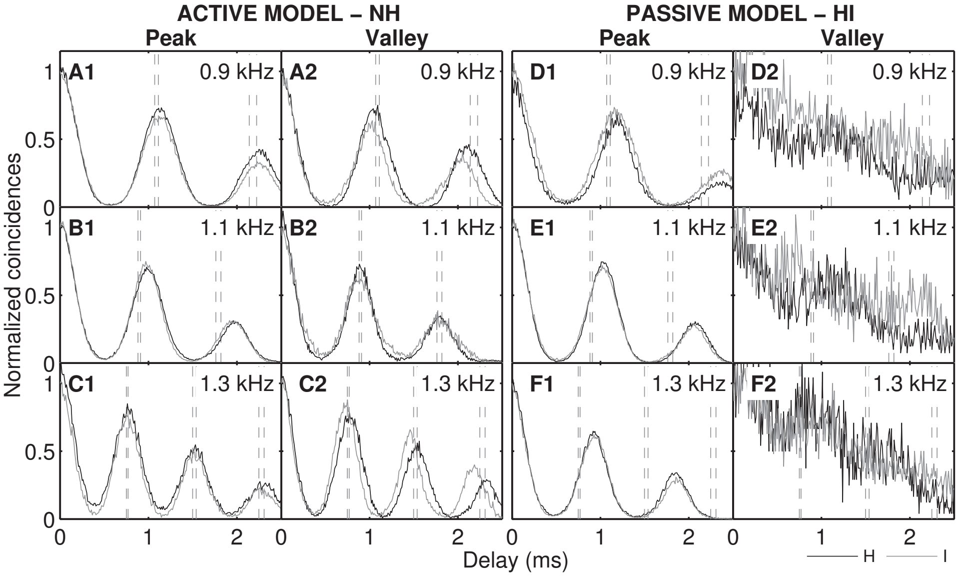

Fig. 4: SACs of the responses to the shaped stimuli.

Removing the cochlear amplifier affected the responses to the shaped and non-shaped stimuli in quite similar way: the valleys of the PSTHs became very pronounced with much smaller amplitude in comparison to the envelope maximum. The responses seem to lack TFS information in the valleys. To find out whether the intervals between successive peaks in the PSTHs are involved, shuffled auto-correlations (SACs) across the spike trains were calculated by tallying inter-spike intervals in the responses (Joris, 2003).

Figures 4 and 5 show the SACs for the shaped and non-shaped stimuli, respectively. The columns denoted “Peak” show the SACs calculated from the portions of the auditory model responses (shown in Figs 2 and 3) marked by the horizontal dashed lines; those denoted “Valley” show the SACs calculated from the responses marked by the horizontal dotted lines. The vertical dashed lines in each panel of Fig. 4 and Fig. 5 show the delays corresponding to k/CF and $k/(\mathrm{CF}+\Delta f)$, where k is 1, 2, or 3.

The positions of the first peak in the SACs (at delay about 1/CF) calculated from the active model responses to the shaped H and I complexes seem to differ only in the valleys. Since only the information from the peaks is available in the responses of the passive auditory model, this could explain why the HI listeners performed poorly in the H-I discrimination tasks. In contrast to this, the SACs at “Peak” portions of the responses which were calculated for the non-shaped stimuli are not overlapping even for the passive auditory model. Therefore the HI listeners may use these TFS cues in addition to spectral cues to discriminate between the H and I non-shaped stimuli. This could contribute to their better performance (see Fig. 1). However, all these results only show that TFS of the responses may differ, but do not relate the TFS information with the perceived pitch of the complexes.

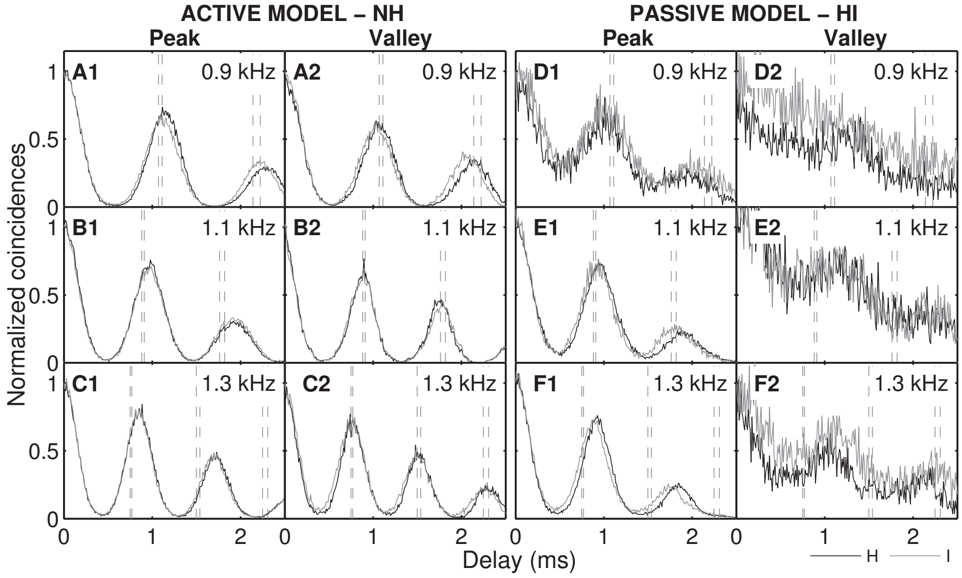

Fig. 5: SACs of the responses to the non-shaped stimuli.

The above shown results cannot rule out the possibility that listeners may use combination tones to discriminate between H and I complexes (Oxenham *et al.*, 2009) – hearing loss may eliminate the combination tones. Another possibility which cannot be ruled out is that the spectral components of the complexes were resolved in the NH listeners and unresolved – because of the wider cochlear filters – in the HI listeners (Santurette *et al.*, 2012). On the other hand, the above shown results would also explain the better performance of the HI listeners with the non-shaped stimuli. Since only the cochlear amplifier was removed from the model, all the simulation only holds for HI listeners with hearing deficits solely based on a loss of outer hair cells.

SUMMARY

Both, the shaped and the non-shaped H and I complexes were analyzed using the auditory model – the active auditory model and the passive auditory model (without a cochlear amplifier). The results can be summarized as follows:

1. The responses to the H and I complexes showed that the stimuli differed in the intervals between peaks of TFS (the intervals long as about the period of the complexes). As already suggested in Hopkins and Moore (2007), these TFS cues could be used to discriminate the H and I complexes with unresolved spectral components. This study showed that these TFS cues are available also in the responses to the non-shaped stimuli and that the cues may also be available for the HI listeners. However, this would not explain the poor performance of the HI listeners – with deficits based solely on a loss of outer hair cells – for the shaped stimuli.

2. The SACs calculated from the short portions of the responses showed the largest differences (shifts in the position of the first peak in the SACs) in the valleys of the responses to the shaped stimuli. Since there is no TFS information in the valleys of the passive auditory model responses, this could explain the poor performance of the HI listeners if their deficits were solely based on a loss of outer hair cells. However, for the non-shaped stimuli TFS information seem to be available in the envelope maximums of the passive auditory model responses. This may help the HI listeners to discriminate the H from I complexes.

ACKNOWLEDGMENTS

I thank Aleš Vetešník for sending parameters of the cochlear model and the ISAAR committee for the scholarship allowing me to attend the ISAAR symposium. This research was supported by the Ministry of Education, Youth and Sports of the Czech Republic in the Long Term Conceptual Development of Research Institutes grant of the Academy of Performing Arts in Prague: The “Sound quality”.

REFERENCES

Glasberg, B.R. and Moore, B.C.J. (**1990**). “Derivation of auditory filter shapes from notched-noise data,” Hear. Res., **47**, 103-138.

Hopkins, K. and Moore, B.C.J. (**2007**). “Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information,” J. Acoust. Soc. Am., **122**, 1055-1068.

Joris, P.X. (**2003**). “Interaural time sensitivity dominated by cochlea-induced envelope patterns,” J. Neurosci., **23**, 6345-6350.

Kale, S., Micheyl, C., and Heinz, M.G. (**2014**). “Implications of Within-Fiber Temporal Coding for Perceptual Studies of F0 Discrimination and Discrimination of Harmonic and Inharmonic Tone Complexes,” J. Assoc. Res. Otolaryngol., **15**, 465-482.

Matlab Auditory Periphery (MAP). University of Essex, Hearing Research Laboratory: `http://www.essex.ac.uk/psychology/department/hearinglab/modelling.html`.

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

Nobili, R., Vetešník, A., Turicchia, L., and Mammano, F. (**2003**). “Otoacoustic emissions from residual oscillations of the cochlear basilar membrane in a human ear model,” J. Assoc. Res. Otolaryngol., **4**, 478-494.

Oxenham, A., Micheyl, C., and Keebler, M.V. (**2009**). “Can temporal fine structure represent the fundamental frequency of unresolved harmonics,” J. Acoust. Soc. Am., **125**, 2189-2199.

Santurette, S., Dau, T., and Oxenham, A.J. (**2012**). “On the possibility of a place code for the low pitch of high-frequency complex tones,” J. Acoust. Soc. Am., **132**, 3883-3895.