

Representations of the Pitch of Complex Tones in the Auditory Nerve

Leonardo Cedolin^{1,2}, Bertrand Delgutte^{1,2,3}

¹ Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary

² Harvard-MIT Division of Health Science and Technology, Speech and Hearing Bioscience and Technology Program, cedro@mit.edu

³ Research Laboratory of Electronics, MIT, bard@epl.meei.harvard.edu

1 Introduction

Previous studies of the coding of the pitch of complex tones in the auditory nerve and cochlear nucleus have documented a robust temporal representation based on interspike interval distributions (Cariani and Delgutte, 1996; Rhode, 1995; Palmer and Winter, 1993). However, these studies have largely neglected possible rate-place cues to pitch available when individual harmonics are resolved by the peripheral auditory system. Stimuli used in these studies had fundamental frequencies in the range of human voice (100-300 Hz), which may produce few, if any, resolved harmonics in typical experimental animals, which have a poorer cochlear frequency selectivity compared to humans (Shera, Guinan and Oxenham, 2002). Human psychophysical studies suggest that the low pitch produced by stimuli with resolved harmonics is stronger and less dependent on phase relationships among the partials than the pitch based on unresolved harmonics (Shackleton and Carlyon, 1994).

Here, we investigate the resolvability of harmonics of complex tones in the cat auditory nerve, and compare the effectiveness of rate-place and interval-based representations of pitch over a much wider range of fundamental frequencies (110-3520 Hz) than in previous studies.

2 Method

2.1 Stimuli and recording techniques

We recorded from auditory-nerve (AN) fibers in dial-anesthetized cats using glass micropipettes filled with 2-M KCl. Upon contact with a fiber, we measured the pure-tone tuning curve to determine the characteristic frequency (CF).

Stimuli were complex tones whose fundamental frequency (F0) stepped up and down over a two-octave range. Each of the 25 F0 steps lasted 200 ms. The har-

monics of each complex spanned a two-octave range around the fiber's CF, and were all of equal amplitude, in cosine phase. The fundamental frequency was always missing. The sound pressure level of each harmonic was usually 15-20 dB above a fiber's threshold, and ranged from 10 to 70 dB SPL.

2.2 Data Analysis

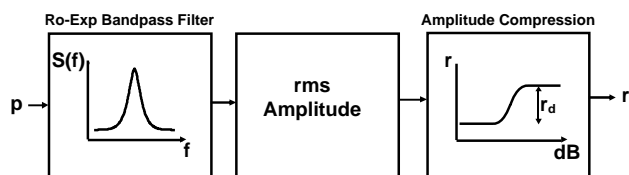


Fig.1. Block diagram of single-fiber rate model.

Simple phenomenological models were used to analyze average-rate responses to the complex-tone stimuli. Specifically, a single-fiber model was used to fit responses of a given fiber as a function of stimulus F0, while a population model was used to fit profiles of average rate against CF for a given F0. The model parameters provide quantitative measures of the ability of AN fibers to resolve individual harmonics.

The single-fiber model (Fig. 1) is a cascade of 3 stages. The band-pass filtering stage, representing cochlear frequency selectivity, is implemented by a symmetric rounded exponential function (Patterson, 1976). The Sachs and Abbas (1974) model is used to express the mean discharge rate as a function of the r.m.s. amplitude at the output of the band-pass filter. The model has 6 free parameters, considerably fewer than the 25 F0 values for which responses are obtained in each fiber.

The population model is an array of single-fiber models indexed on CF so as to predict the entire auditory-nerve rate response as a function of cochlear place. The bandwidths of the band-pass filters are constrained to be a power function of the CF (Shera et al., 2002). The population model has no free parameters; rather, parameters of the stimulus (F0 and SPL) are selected to fit the measured "rate-place profiles" expressing the normalized driven discharge rate as a function of CF (Sachs & Young, 1979). The resulting best-fitting F0 gives a rate-based estimate of pitch that does not require *a priori* knowledge of the actual F0.

3 Results: Rate-Place Representation

3.1 Single-fiber rate responses

Figure 2 shows the average discharge rate plotted against complex-tone F0 for two auditory-nerve fibers. The horizontal axis represents the dimensionless "harmonic number" CF/F0, so that higher F0s are towards the left. For both fibers, the mean rate shows a peak when the CF is a small integer multiple of F0, and a valley when the CF falls halfway between two harmonics. However, the oscillations are more pronounced for the higher-CF fiber on the right.

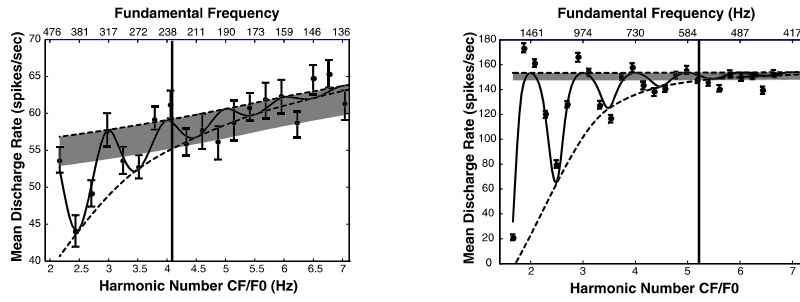


Fig. 2. Average discharge rate as a function of F_0 for two fibers with CFs of 952 Hz (left) and 2922 Hz (right). Error bars show ± 1 standard deviation of the discharge rate obtained by bootstrap resampling over stimulus trials. The solid lines are least-squares fits to the data using the model of Fig. 1.

For both fibers, the response of the best-fitting single-fiber model captures the main trend in the data. The harmonics of F_0 are considered to be resolved so long as the oscillations in the fitted curve exceed two typical standard deviations (gray shading). We call N_{\max} the maximum resolved harmonic number. Here, N_{\max} for the low-CF fiber is 4.1, smaller than N_{\max} for the high-CF fiber (5.3). The ratio N_{\max}/CF gives $F_{0\min}$, the lowest fundamental frequency whose harmonics can be resolved in the rate response of a given fiber.

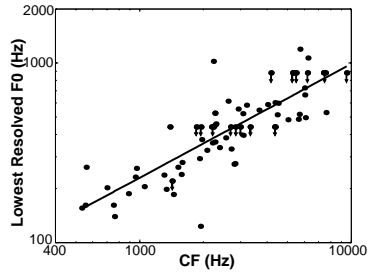


Fig. 3. Lowest F_0 whose harmonics can be resolved as a function of CF. For some fibers (arrows), $F_{0\min}$ was bounded by the lowest F_0 presented and was therefore somewhat overestimated.

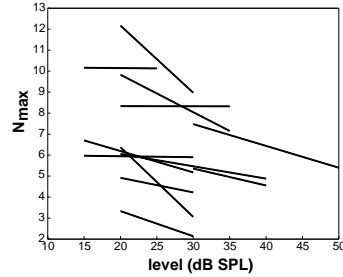


Fig. 4. Maximum resolved harmonic number for 11 fibers as a function of level expressed in dB SPL per component.

Figure 3 shows that $F_{0\min}$ increases systematically with characteristic frequency. It also suggests that harmonics of F_0 s in the range of human voice (100-200 Hz) are rarely, if ever, resolved by AN fibers in the cat. The increase in $F_{0\min}$ is well fit by a power function of the CF with an exponent of 0.63 (solid line). This result is consistent with the progressive sharpening of peripheral tuning with increasing CF when expressed as a Q factor. The exponent for Q would be 0.37, which closely

matches the 0.37 exponent found by Shera et al. (2002), based on pure-tone tuning curves from AN fibers in the cat.

Rate responses of AN fibers to complex stimuli are known to depend strongly on stimulus level (Sachs and Young, 1979). To address this issue, responses to complex tones were recorded at two different stimulus levels in a few fibers. In these cases, the maximum resolved harmonic number N_{\max} tended to decrease with increasing stimulus level (Fig. 4). This decrease could reflect either broadened cochlear tuning with increasing level, or rate saturation. Preliminary analysis suggests that the latter may be the dominant factor, since the bandwidths of the model auditory filters stay essentially constant with level.

3.2 Pitch estimation from rate-place profiles

Figure 5 shows the normalized driven discharge rate in response to a complex tone with an F0 of 541.5 Hz as a function of CF. Despite some scatter in the data, the normalized rate tends to show a local maximum when the CF is an integer multiple of F0 and a minimum when the CF falls half-way between two harmonics.

The oscillatory pattern in the rate-place profile of Fig. 5 can be used to estimate the fundamental frequency of the stimulus. To quantitatively derive such estimates, we determined the F0 of a complex tone with equal-amplitude harmonics for which the response of the population model (see Method) best fits the observed rate-place profile. Here the estimated pitch was 547.4 Hz, only 1.1% above the actual F0.

Figure 6 shows measures of the accuracy and precision of rate-based pitch estimates as a function of F0. With few exceptions, median pitch estimates only deviate by a few percent from the true F0 (top panel). For F0s above 400-500 Hz, the interquartile range of the pitch estimates over 100 bootstrap resamplings of the data are all below 5% (middle panel). However, the model produces few reliable estimates for F0s below 400 Hz. The bottom panel of Fig. 6 shows a measure of pitch salience derived from the amplitude of the oscillations in the model rate-place profiles. The salience is very low below 800 Hz, then

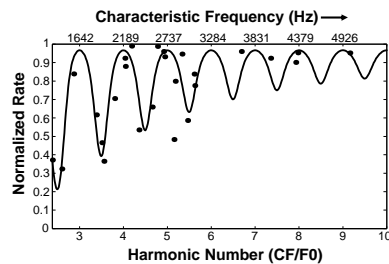


Fig. 5. ••: normalized discharge rate as a function of CF; —: model rate-place profile. F0 = 541.5 Hz.

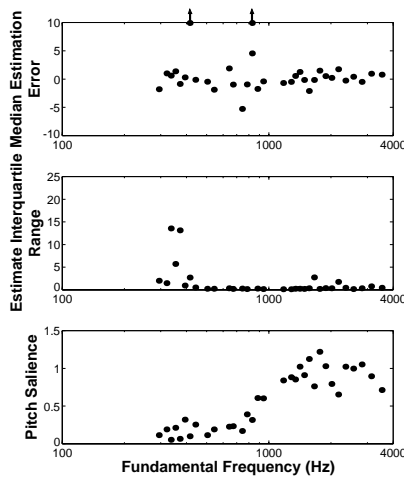


Fig. 6. Median pitch estimation error, interquartile range of the estimates, and pitch salience as a function of fundamental frequency.

increases rapidly to saturate above 2 kHz. Overall, pitch estimation from rate-place profiles works best for F0s above 400 Hz.

4 Pitch estimation from pooled interspike interval distributions

As in previous studies of the neural coding of pitch (Cariani and Delgutte, 1996; Rhode, 1995), we derived pitch estimates from pooled interspike interval distributions. The pooled interval distribution is the sum of the all-order interspike intervals for all sampled auditory-nerve fibers, and is closely related to the summary autocorrelation in the Meddis and Hewitt (1991) model.

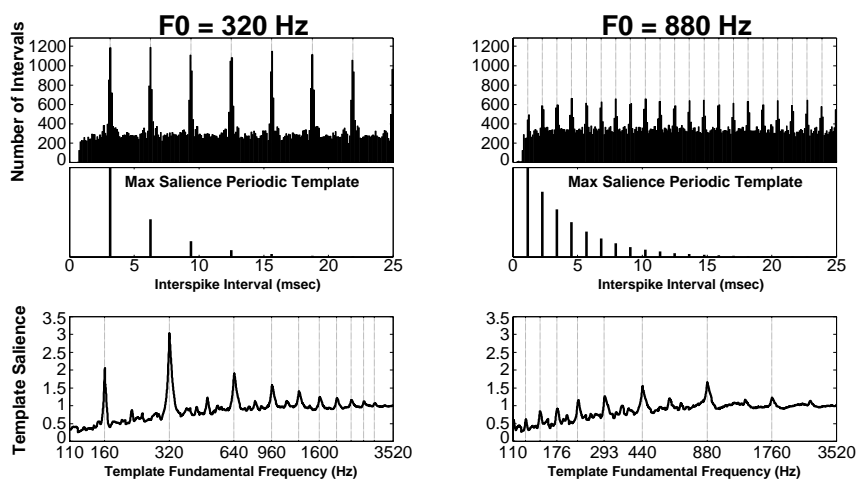


Fig. 7. Top row: pooled interval distributions for complex tones at two F0s. Middle row: periodic template with maximum salience. Bottom row: salience of periodic template as a function of its fundamental frequency. Results are shown for two F0s: 320 Hz (left column) and 880 Hz (right column).

Figure 7 (top) shows pooled interval distributions for two complex-tone stimuli with F0s of 320 and 880 Hz. For both stimuli, the pooled distributions show modes at the period of F0 and its integer multiples (dotted lines). However, these modes are less prominent at the higher F0.

To derive pitch estimates from pooled interval distributions, we used “periodic templates” that select intervals at a given period and its multiples (middle panels of Fig. 7). Specifically, the salience of a periodic template is defined as the ratio of the weighted mean number of intervals within the template to the mean number of intervals per bin. A pitch estimate is obtained by finding the template whose fundamental period maximizes the salience. Templates with exponentially-decaying weights were found to give fewer octave and suboctave errors than flat templates.

The bottom panels of Fig. 7 show the template salience as a function of template F0 for the same two stimuli as on top. For both stimuli, the salience reaches an absolute maximum when the template F0 is very close to the actual stimulus F0. However, the maximum salience is larger for the lower F0.

Figure 8 shows measures of the accuracy, precision and strength of the interval-based pitch estimates as a function of F0. The estimates are very accurate below 1300 Hz, where their medians are within 1-2% of the true F0. The bootstrap interquartile ranges of the estimates are also below 1-2% for F0s up to 1600 Hz. However, the interval-based estimates of pitch abruptly break down above 1300 Hz due to the degradation of phase locking at harmonic frequencies near the CF. Since the stimuli have missing fundamentals, the lowest harmonic actually present is always above 2600 Hz, in a range where the degradation of phase locking is already substantial (Johnson, 1980).

The salience of the estimated pitch is highest below 400 Hz, then decreases gradually with increasing F0, to reach essentially zero at 1300 Hz. Thus, the salience of interval-based estimates of pitch is highest in the F0 range where rate-based pitch estimates are the least reliable due to the lack of strongly resolved harmonics. Conversely, the salience of rate-based estimates of pitch is highest above 2000 Hz, where the interval-based estimates break down.

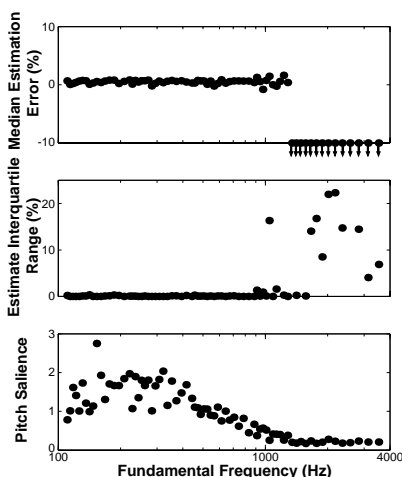


Fig. 8. Median pitch estimation error, interquartile range of the estimates and pitch salience as a function of F0.

5 Conclusions

We examined two possible representations of the pitch of complex tones with a missing fundamental in the cat auditory nerve for low and moderate sound levels.

1. A rate-place representation based on resolved harmonics was found to be viable for fundamental frequencies above 400 Hz.
2. Fundamental frequencies up to 1300 Hz were precisely represented in pooled interspike-interval distributions of the entire auditory nerve.
3. In the 400-1300 Hz range where both representations overlap, estimated pitch salience varies in opposite directions with F0 for the two representations.

The range of F0 over which the rate-place representation is viable in the cat does not include the 100-300 Hz region which is the most important for human voice. This failure may reflect the poorer frequency selectivity of the cat cochlea compared to the human (Shera et al., 2002). On the other hand, the frequency range

of rate-based pitch estimates does include the F0 range of most cat vocalizations, which is centered around 600 Hz. An interesting question is whether this relationship between frequency ranges of vocalizations and resolved harmonics would hold in other species.

The range of F0s over which interval-based estimates of pitch are reliable in the cat roughly covers the entire perceptual range of the missing fundamental in humans (Moore, 1997). However, the salience of these estimates is strongest below 400 Hz, where individual harmonics are not strongly resolved in the cat. Thus, interval-based models of pitch may have trouble predicting the greater salience of pitch based on resolved harmonics compared to that based on unresolved harmonics (Shackleton & Carlyon, 1994).

In conclusion, neither representation of pitch is entirely satisfactory: the rate-place representation degrades at high sound levels and low frequencies, while the interval representation may have trouble accounting for the salience of pitch from resolved harmonics. This conclusion suggests a search for alternative neural codes for pitch such as those based on spatio-temporal patterns of discharge that would not rely on long interspike intervals (e.g. Shamma and Klein, 2000).

7 References

- Cariani, P. A. and Delgutte, B. (1996) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* 76, 1698-1716.
- Johnson, D.H. (1980) The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J. Acoust. Soc. Am.* 68, 1115-1122.
- Meddis, R. and Hewitt, M.J. (1991) Virtual pitch and phase sensitivity of a computer model of the auditory periphery. *J. Acoust. Soc. Am.* 89, 2866-2882.
- Moore, B. C. J. (1997) *An introduction to the psychology of hearing*. Academic Press.
- Palmer, A.R. and Winter, I.M. (1993) Coding of the fundamental frequency of voiced speech sounds and harmonic complexes in the cochlear nerve and ventral cochlear nucleus. In Merchán, M.A., Juiz, J.M., Godfrey, D.A. and Mugnaini, E. (Eds.) *The mammalian cochlear nuclei: Organization and Function*. Plenum Press, New York, pp. 373-384.
- Patterson, R. D. (1976) Auditory filter shapes derived with noise stimuli. *J. Acoust. Soc. Am.* 59, 640-654.
- Rhode, W.S. (1995) Interspike intervals as correlates of periodicity pitch in cat cochlear nucleus. *J. Acoust. Soc. Am.* 97, 2414-2429.
- Sachs, M. B. and Abbas, P. J. (1974) Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli. *J. Acoust. Soc. Am.* 56, 1835-1847.
- Sachs, M. B. and Young, E. D. (1979) Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. *J. Acoust. Soc. Am.* 66, 470-479.
- Shackleton, T.M. and Carlyon, R.P. (1994) The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.* 95, 3529-3540.
- Shamma, S. and Klein, D. (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *J. Acoust. Soc. Am.* 107, 2631-2644.
- Shera, C. A., Guinan, J. J., Jr. and Oxenham, A.J. (2002) "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements". *Proc. Natl. Acad. Sci. USA* 99, 3318-23.