Phase effects on the perceived elevation of complex tones

William M. Hartmann[a)]
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

Virginia Best,[b)] Johahn Leung, and Simon Carlile
School of Medical Sciences, University of Sydney, Sydney, New South Wales 2006, Australia

(Received 17 August 2009; revised 4 March 2010; accepted 4 March 2010)

Free-field source localization experiments with 30 source locations, symmetrically distributed in azimuth, elevation, and front-back location, were performed with periodic tones having different phase relationships among their components. Although the amplitude spectra were the same for these different kinds of stimuli, the tones with certain phase relationships were successfully localized while the tones with other phases led to large elevation errors and front-back reversals, normally growing with stimulus level. The results show that it is not enough to have a smooth, broadband, long-term signal spectrum for successful sagittal-plane localization. Instead, temporal factors are important. A model calculation investigates the idea that the tonotopic details that mediate localization need to be simultaneously, or almost simultaneously, accessible in the auditory system in order to achieve normal elevation perception. A qualitative model based on lateral inhibition seems capable in principle of accounting for both the phase effects and level effects.
 [DOI: 10.1121/1.3372753]

PACS number(s): 43.66.Nm, 43.66.Qp, 43.66.Ba [JCM] Pages: 3060–3072

I. INTRODUCTION

It is now well established that sound localization in the sagittal planes, specifically the perception of elevation and the distinction between front and back, depends on asymmetrical filtering by the human anatomy, especially by the head and pinnae (Roffler and Butler, 1968; Blauert, 1969–70; Middlebrooks, 1992). The elevation cues reside in the pattern of peaks and valleys in the spectrum as it appears at the eardrum and at higher stages of the tonotopically organized auditory system. A broadband stimulus such as an extended noise burst causes the peaks and valleys established by the anatomy to be available to an auditory central processor, which can then use the information to identify elevation.

Recently, it has also become evident that a broad bandwidth is not the only stimulus requirement for elevation perception. There are also temporal factors. Impulsive stimuli lead to anomalously large errors in the estimate of the elevation of a source of sound. Single clicks or brief bursts of broadband noise cannot be localized in elevation nearly as well as less impulsive stimuli with the same long-term spectra. The effect appeared in front-back-overhead confusions (Hartmann and Rakerd, 1993) and in the compression of perceived elevation toward the horizon (Hofman and Van Opstal, 1998; Macpherson and Middlebrooks, 2000; Vliegen and Van Opstal, 2004). These studies observed that the errors in elevation estimates tend to grow with increasing signal level—an effect called the "negative level effect" by Hartmann and Rakerd.

The experiments by Hofman and Van Opstal (1998) included 500-ms complex tones with fundamental frequencies between 24 and 781 Hz and with equal-amplitude harmonics up to 16 kHz. The phases of the harmonics were selected according to Schroeder's minimum crest factor algorithm called "Schroeder-plus" (Schroeder, 1970). This tone consists of chirps—a series of linear downward frequency sweeps from the highest frequency in the spectrum to the lowest—one sweep per cycle. Therefore, the sweep rate is the same as the fundamental frequency. In this article such sweeps will be called "chirp↓" to indicate the direction of frequency change with time. Hofman and Van Opstal found that when the fundamental frequency became less than about 200 Hz, there was a dramatic reduction in perceived elevation.

In a similar observation, Zhang and Hartmann (2010) found that listeners made a large number of front-back confusions in free-field listening to broadband complex tones of the chirp↑ form. The tones had 250 equal-amplitude harmonics of a 65-Hz fundamental, spanning the range from 195 to 16 250 Hz.

Recently, Brungart and Simpson (2008) discovered that a negative level effect applies to pulse trains. They compared 100-Hz complex tones where all the spectral components were in cosine-phase (pulse trains), 100-Hz complex tones where the phases were randomized, and bursts of noise. They found that elevation errors were largest for the cosine-phase stimuli even at a level of 40 dB sound pressure level (SPL), and grew dramatically as the level increased to 80 dB. By contrast, there was little level effect for the other two stimuli.

The experiments reported in this article explored elevation localization, including front-back reversal, for complex tones in greater detail. We particularly wanted to know how

---

[a)]Author to whom correspondence should be addressed. Electronic mail: hartmann@pa.msu.edu
[b)]Present address: Hearing Research Center, Boston University, Boston, MA 02215.

elevation perception for chirp tones compares with that for cosine- and sine-phase tones, and whether the perceived elevations show a level dependence.

We also wanted to compare elevation judgments for chirp↑ tones and chirp↓ tones. These two Schroeder-phase stimuli produce very different temporal patterns at particular places on the basilar membrane (Recio and Rhode, 2000) or the model basilar membrane (Smith *et al.*, 1986; Kohlrausch and Sander, 1995). Chirp↓ tones have a positive phase curvature in that the second derivative of the frequency-dependent phase is a positive constant. This curvature tends to cancel the large negative phase curvature on the basilar membrane near the place of resonance, causing the tone to be especially impulsive as it appears at that place. By contrast, there is no special relationship to basilar membrane mechanics for the chirp↑ tones. Because the duty factor of the effective stimulus for chirp↓ tones is particularly small, the ability of a chirp↓ tone to mask a sine tone with the corresponding frequency can be greatly reduced. Lentz and Leek (2001) and Oxenham and Dau (2001) found differential masking effects as large as 30 dB. We wondered whether a similarly dramatic effect might be present for elevation perception.

II. EXPERIMENT 1

All experiments were done in an anechoic chamber with a single loudspeaker whose azimuth and elevation were controlled by robotic hoops. The system was described by Carlile *et al.* (1997).

A. Protocol

Thirty source locations were used in each experimental run. There were 15 source locations in front of the listener and 15 in back. The front locations were at lateral angles of −30°, 0°, and +30° crossed with polar angles of −40°, −20°, 0°, +20°, and +40°, where the definition of lateral and polar angles follows Macpherson and Middlebrooks (2000). The back locations were the same as the front except that they were reflected in the median frontal plane. Therefore, the pattern of locations was completely symmetrical left-right, up-down, and front-back.

During an experiment run, the listener was standing in the dark with his or her head at the center of the hoops. Head orientation was monitored with an Intersense IC3 head tracker. Before the presentation of a stimulus, the listener's head needed to be correctly positioned at the center. An array of light-emitting diodes indicated the required adjustments in position to the listener. The stimulus was delivered, and the listener responded, indicating the perceived location of the sound by pointing his or her nose at the source and then pressing a handheld button. A single run consisted of one presentation from each of the 30 locations in random order. A run typically required about 10 min, and three runs were normally done between breaks.

B. Listeners

There were five listeners, two females, V and R, and three males, G, J, and S, with ages between 19 and 51. All had normal hearing through 8 kHz according to pure-tone audiometry tests. Listeners V, J, and S were coauthors of this article and were highly experienced in sound localization tasks; listeners G and R were inexperienced. The experimental protocol and listener consent forms had been approved by the University of Sydney Human Research Ethics Committee.

C. Stimuli

All stimuli had an overall duration of 220 ms and a temporal window with 20-ms raised-cosine edges. Such a window preserves the spectrum of a periodic waveform having a fundamental frequency f_o=5 Hz, or an integer multiple of 5 Hz (Hartmann and Wolf, 2009). The stimuli were generated by a Tucker-Davis DD1 16-bit digital to analog converter using a sample rate of 80 ksps. All the stimuli had the same spectral ranges with the lowest component always at 325 Hz and the highest component at 16 250 Hz. Two different levels were explored, 50 and 70 dB SPL, as read on an *A*-weighted sound level meter at the position of the listener's head. During the course of an experimental run, the levels were varied randomly by 0 or ±2 dB to discourage listeners from trying to use overall level as a cue to location.

1. Noise

Noise stimuli were made with 3186 components, harmonics of 5 Hz with equal amplitudes and random-phases. There were six different noises, made with different randomizations of the phases, and these were presented during noise experiments according to a random schedule.

2. Random-phase tones

The random-phase tones had a 65-Hz fundamental and 246 harmonics with equal amplitudes. The phases were random, uniformly distributed over the range 0 to 2π. Six different random-phase tones were made with different randomizations, and these were presented according to a random schedule. The six tones were selected to have a similar crest factor, 3.22 (±0.01).

3. Chirp tones

Like the random-phase tones, the chirp tones consisted of 246 harmonics of f_o=65 Hz with equal amplitudes

$$x(t) = \sum_{n=N_b}^{N_t} \cos(2\pi n f_o t + \phi_n). \quad (1)$$

Phases were chosen according to Schroeder's algorithm. For the condition known as chirp↑ the phase of the *n*th harmonic was given by the Schroeder-minus relation

$$\phi_n = -\pi \frac{n(n-1) - N_b(N_b-1)}{N_t - N_b + 1}, \quad (2)$$

where N_b=5 is the bottom harmonic and N_t=250 is the top. For the condition known as chirp↓, the leading minus sign was changed to a plus sign. The crest factor of $x(t)$ was 1.57, for both signs.

Digital recordings made with a microphone (Sennheiser 4 mm) at the listener's position verified that the chirp stimuli retained the character of the frequency sweeps. The maximum and minimum of the envelope during a chirp, caused by the imperfect frequency response of the signal chain, were in a ratio of about 8/5. The component phase shifts of chirp tones vary so rapidly with frequency that the overall character of the up and down chirp tones is preserved when the tones are subjected to the phase shifts of head-related transfer functions, normally thought to be minimum phase plus a constant delay (Mehrgardt and Mellert, 1977; Kulkarni *et al.*, 1999).

D. Results

Results were based on two runs for each listener and condition, i.e., two repetitions of each location, a total of 60 trials. Localization errors in polar angle and lateral angle were measured.

1. Polar-angle errors

Polar-angle errors were evaluated by collapsing the data across the three lateral angles of the sources. Data from the experiment were represented in two ways: First, source locations and responses were reflected in the median frontal plane so that all locations in back were mapped to locations in front having the same elevation as per Wightman and Kistler (1989). The mean response and root-mean-square error (rms error) were computed for the remapped locations. Second, the effect of the remapping on response errors was taken into account by separately counting the number of reversals across the median frontal plane, i.e., front-to-back or back-to-front reversals.

Polar-angle responses are shown by Fig. 1 for listener V as an example. Different stimulus phase configurations appear in different rows of the figure, and the two levels correspond to the two columns. Symbols F and B refer to the number of reversals when the source was in front (F) or in back (B), and the total (T) indicates the sum of the two. Numbers F and B are limited to a maximum of 30. Random guessing between front and back would lead to an expected total number $T=30$, half the number of trials for a given phase and level combination.

The mean response $R(k)$ for each of the five source polar angles ($k=-40°, -20°, 0°, 20°, 40°$) is shown by the plotted points connected by straight-line segments. The error bars are standard errors; i.e., they are standard deviations divided by the square root of the number of trials per source location, namely, by $\sqrt{12}$. If the responses all agreed perfectly with the polar angles of the sources, the $R(k)$ plot would fall on the 45° line, shown dashed. The slope parameter is the slope of the best-fitting straight line, similar to the "elevation gain" reported by Hofman and Van Opstal (1998). Ideally, the slope is 1.0. Compression of the perceived polar angle is indicated by a slope less than 1.0. It is evident from Fig. 1 that the perceived polar angles were severely compressed for the two chirp tones. Mean responses tended to be on the horizon, with a polar angle (and elevation) near zero.

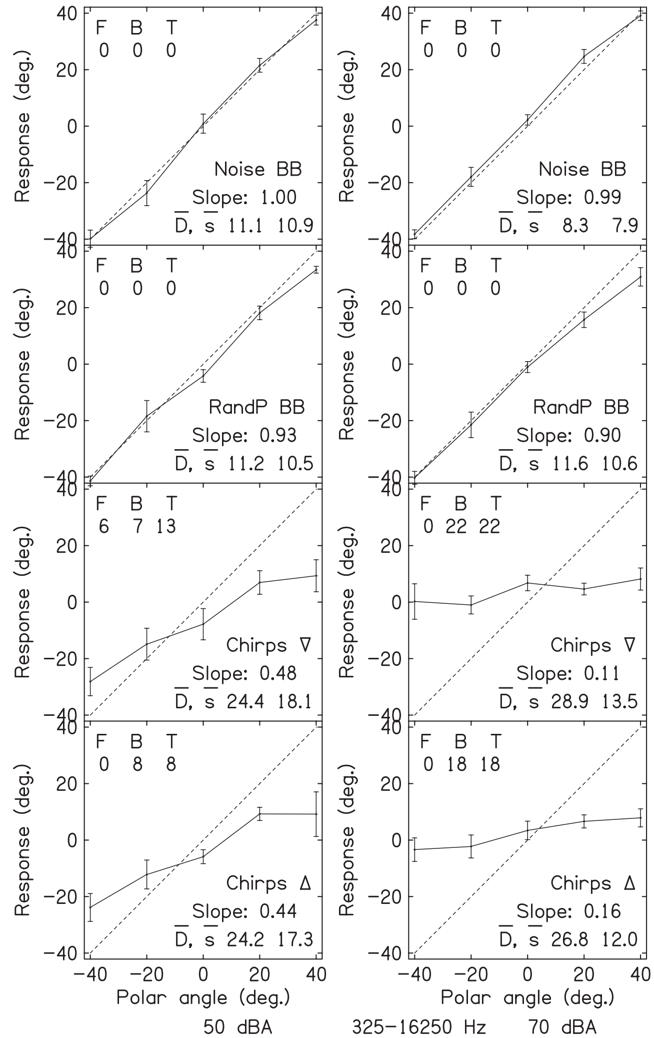

FIG. 1. Results for the broadband noise, random-phase tones, and chirp tones, both chirp↓ and chirp↑. The plot shows the mean response to each of the five polar angles of the source locations for one listener, listener V. Perfect responses would fall along the dashed 45° line. Error bars show standard errors. The number of front-back reversals for sources in front (F) and in back (B) are given together with the total (T). The slope is otherwise known as the "polar-angle gain." Parameters D and s are the rms error and standard deviation (see text). The left and right column are for 50 and 70 dBA SPLs, respectively.

Summary statistics $\bar{D}$ and $\bar{s}$ also appear in the panels. Quantity $\bar{D}$ refers to the rms error in polar angle, averaged over all the trials contributing to the data in a panel. Given the severe compression of elevation observed in $R(k)$ it is useful to know that if all the responses were on the azimuthal plane (polar angle of zero) then $\bar{D}$ would be 28.3°, a value called the "horizon limit." The horizon limit corresponds to no elevation information at all. This limit is a large rms error, but it is actually a minimum in the sense that any other constant polar-angle response would lead to a larger value of $\bar{D}$. Values of $\bar{D}$ greater than this limit occurred frequently for some listeners.

Quantity $\bar{s}$ is the standard deviation of the responses about the mean. It is the average, across the five source polar angles, of the standard deviations that led to the error bars

TABLE I. Total number of front-back reversals expressed in the form [total for 50 dB]|[total for 70 dB]. 100% errors for both levels would be 60|60. Values in parentheses for listeners G and R are for 40 dB.

Broadband	V	G	J	R	S
Noise	0|0	0|1	0|1	2|1	1|0
Random-phase	0|0	0|0	1|1	3|1	0|1
Chirps ↓	13|22	26|13	13|12	10|13	12|15
Chirps ↑	8|18	22|21	7|10	6|11	6|9
High-passed	**V**	**G**	**J**	**R**	**S**
Noise	0|0	0|1	1|8	0|0	1|3
Random-phase	0|1	3|1	3|3	0|1	1|5
Chirps ↓	17|29	(10)26|25	12|14	(5)23|24	10|18
Chirps ↑	14|25	(11)14|29	14|26	(8)22|19	11|30
Cosine-phase	5|12	(19)18|8	13|10	(5)14|15	2|4
Sine-phase	8|12	(15)22|9	10|8	(7)11|17	6|3

TABLE III. rms error $\bar{D}$, rounded to integer degrees and expressed in the form [$\bar{D}$ for 50 dB]|[$\bar{D}$ for70 dB]. Values in parentheses for listeners G and R are for 40 dB.

Broadband	V	G	J	R	S
Noise	11|8	13|12	9|12	16|14	17|14
Random-phase	11|12	13|14	14|18	15|17	14|12
Chirps ↓	24|29	28|27	23|47	24|28	23|31
Chirps ↑	24|27	28|28	27|49	26|27	29|29
High-passed	**V**	**G**	**J**	**R**	**S**
Noise	9|12	13|14	9|16	10|17	18|14
Random-phase	13|11	13|12	13|12	14|15	19|18
Chirps ↓	27|33	(25)31|29	39|51	(29)31|30	31|31
Chirps ↑	27|31	(22)26|31	33|57	(20)27|37	25|52
Cosine-phase	20|31	(22)25|26	20|31	(33)30|32	26|38
Sine-phase	22|24	(21)25|27	23|27	(27)27|33	18|27

shown on $R(k)$. Quantities $R(k)$, $\bar{D}$, and $\bar{s}$ for describing locations and location errors were defined as per Rakerd and Hartmann (1986).

The data for all listeners were summarized by three statistics: the total number of front-back reversals, the polar-angle gain (slope), and the rms polar-angle error ($\bar{D}$). The summary statistics are given by the top portions of Tables I–III.

Table I shows the number of front-back reversals (F/B) in the form (number for 50 dB)|(number for 70 dB). If a listener reversed every trial, the table would show 60|60; random guessing would have an expected value of 30|30. If a listener reported every stimulus to be the same, either all in front or all in back, the table would again show 30|30. Summed over all listeners and levels, for noise and random-phase tones (1200 trials) the total number of reversals was 13. The same sum for the two chirp tones (also 1200 trials) was 267.

Different listeners had distinctly different patterns of front-back reversals. Out of her total of 61 reversals, listener V reversed only 6 chirp tones with the source in front compared to 55 chirp tones with the source in back, a pattern to be notated here 6:55. Similarly, listener R's pattern was 6:34. Listener G was effectively the opposite, 69:13, and so was listener S, 34:8. Listener J was more even, 25:17.

Tables II and III show that the chirp tones also gave rise to shallower slopes and larger rms errors when compared to noise and random-phase tones. The large differences in performance between the different stimuli did not come as a surprise to the listeners. Informally, they remarked that their judgments were much more uncertain for chirp tones than for noise or random-phase tones. Some listeners even felt that the chirp tones were badly externalized, and all listeners were aware that the source elevations seemed to have collapsed toward the horizon.[1]

Figure 1 and all the tables indicate a strong negative level effect for listeners V, S, and J for chirp tones. Listeners V and S showed an increase in the number of front-back reversals and a decrease in the slope as the level increased by 20 dB. For listener J, the negative level effect emerged as a large upward displacement of the images for high levels and, consequently, a large rms error $\bar{D}$, much larger than the ho-

TABLE II. Polar-angle gain (slope) expressed in the form [slope for 50 dB]|[slope for 70 dB]. Values in parentheses for listeners G and R are for 40 dB.

Broadband	V	G	J	R	S
Noise	1.00|0.99	0.80|0.74	1.06|1.02	0.60|0.76	1.03|1.27
Random-phase	0.93|0.90	0.87|0.68	0.95|0.83	0.56|0.66	1.14|1.27
Chirps ↓	0.48|0.11	0.11|0.19	0.83|0.26	0.22|0.24	0.57|0.41
Chirps ↑	0.44|0.16	0.06|0.07	0.98|0.82	0.29|0.39	0.61|0.33
High-passed	**V**	**G**	**J**	**R**	**S**
Noise	0.93|0.92	0.79|0.69	1.07|1.19	0.89|0.92	1.24|1.29
Random-phase	0.86|0.97	0.69|0.75	1.15|1.09	0.92|0.91	1.11|1.12
Chirps ↓	0.21|−0.03	(0.20)0.02|0.03	0.70|0.14	(0.25)0.05|0.20	0.86|0.26
Chirps ↑	0.23|0.05	(0.31)0.13|−0.01	1.04|0.10	(0.71)0.17|0.12	0.85|−0.04
Cosine-phase	0.48|0.16	(0.31)0.21|0.19	1.06|0.22	(0.41)0.26|0.12	0.71|0.45
Sine-phase	0.44|0.33	(0.33)0.17|0.12	0.86|0.30	(0.50)0.54|0.08	0.96|0.51

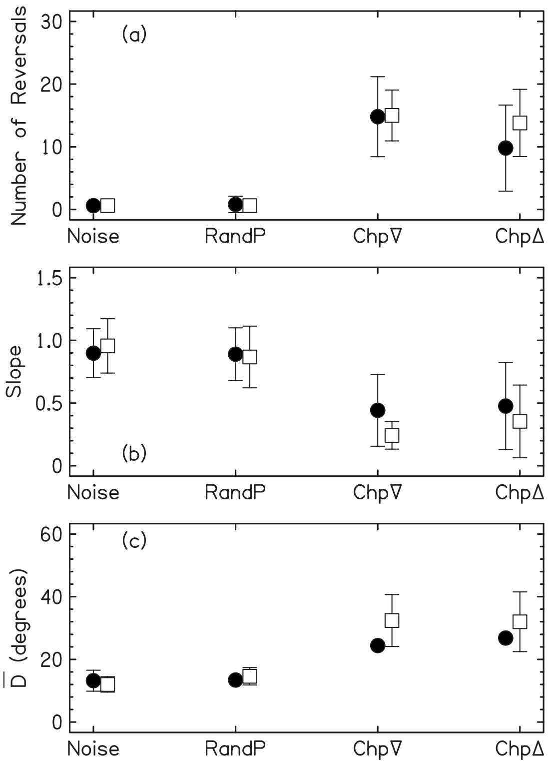

FIG. 2. Three statistics, averaged over five listeners for four waveforms in experiment 1: broadband noise, random-phase tones, chirp↓, and chirp↑. High (□) and low (●) levels are 70 and 50 dB SPL. Error bars are two standard deviations in overall length.

rizon limit. Compared to noise, $\bar{D}$ doubled for 50-dB chirp tones, and it doubled again when the level was raised to 70 dB—a strong negative level effect.

Relatively inexperienced listeners G and R also showed a strong effect of phase, with reduced polar-angle gains and values of $\bar{D}$ close to the horizon limit of 28.3° for the chirp tones. However, these listeners exhibited no negative level effect in experiment 1.

To assess the effect of waveform and level on the group data, a two-way repeated-measures analysis of variance (ANOVA) was conducted on the number of reversals, on the slopes, and on the values of $\bar{D}$.

- For front-back reversals, there was a significant effect of waveform $[F(3,12)=28.1,\ p<0.001]$, no significant effect of level $[F(1,4)=0.6,\ p=0.5]$, and no significant interaction $[F(3,12)=1.1,\ p=0.4]$.
- For slopes, there was a significant effect of waveform $[F(3,12)=22.7,\ p<0.001]$, no significant effect of level $[F(1,4)=1.6,\ p=0.3]$, and no significant interaction $[F(3,12)=2.5,\ p=0.1]$.
- For $\bar{D}$, there was a significant effect of waveform $[F(3,12)=45.6,\ p<0.001]$, no significant effect of level $[F(1,4)=1.7,\ p=0.3]$, and a significant interaction $[F(3,12)=3.8,\ p<0.05]$.

Plots of the three statistics averaged over listeners in Fig. 2 support the conclusions of the ANOVAs. While the effect of waveform is clear, the large individual differences obscured the negative level effect for the group as a whole. This issue was addressed further in experiment 2, where listeners G and R were tested at a still lower level of 40 dB (see below).

Is the polar-angle localization different for chirp↑ and chirp↓? Table I and Fig. 2 indicate more front-back reversals for chirp↓ tones. The difference (pooled across the two levels) is significant at the 0.05 level in a paired t-test. However, the individual listeners responded to these difficult conditions in different ways, and the differences in slope and in $\bar{D}$ for the two signs are not statistically significant. Most important, the comparison reveals nothing like the enormous difference seen in masking experiments. Instead, what is evident is that elevation perception for both chirp signs was greatly degraded compared to that for noise or random-phase tones.

2. Lateral-angle errors

Lateral response errors for the three lateral angles were evaluated by collapsing the data across trials with different stimulus elevations. For all the stimulus conditions, the listeners correctly identified the lateral angle of the stimulus in the sense that the mean responses were closer to the actual azimuth of the source than to the azimuth of a different source. A measure of variability is $\bar{D}_L$, the rms error in lateral angle averaged over the three source angles. There was a tendency for $\bar{D}_L$ to be larger for stimuli that also lead to poorer elevation localization. That tendency was quantified by averaging the lateral-angle errors first over noise and random-phase tones (good elevation localization) and then over chirp↑ tones and chirp↓ tones (poor elevation localization). Each average comprised eight runs when summed also over the two levels. For each of the five listeners, the average was greater for chirps. Averaged over listeners, $\bar{D}_L$ was 7.7° for noise and random-phase tones and $\bar{D}_L$ was 11.1° for chirps. Brungart and Simpson (2008) also found a small effect like this—azimuth errors tended to grow slightly as elevation errors grew rapidly with increasing stimulus level for click trains.

E. Discussion

Hofman and Van Opstal (1998) employed chirp↓ tones with different periods, ranging from 1.3 to 41 ms, corresponding to fundamental frequencies from 769 to 24 Hz. They used a sound pressure level of 70 dBA. They found a significant reduction in elevation gain when the period exceeded 5 ms, i.e., for fundamental frequencies less than 200 Hz. The period used in experiment 1, approximately 15 ms, is in a range where Hofman and Van Opstal (1998) found a depressed elevation gain, averaging about 0.16 [standard deviation (sd)=0.06] , and in this range their measured elevation gain was insensitive to the period. Averaged over our five listeners and both signs of chirp, experiment 1 found slopes of 0.45 and 0.30 at 50 and 70 dB, respectively. However, the average was skewed by listener J, for whom the effect of chirps was mainly observed in the rms error and not in the polar-angle gain. Excluding J, slopes averaged

0.35 and 0.23 at 50 and 70 dB, respectively. The value of 0.23 is closer to the 70-dB elevation gains reported by Hofman and Van Opstal (1998).

Hofman and Van Opstal (1998) interpreted their results in terms of a linear model of the auditory system beginning with a short-term analysis of the spectrum with a duration of 5 ms. The 5-ms analysis window was set by the observed period dependence of their chirp↓ experiment. They reasoned that if the analysis duration is shorter than the chirp duration then successive analyses will return quite different spectra, leading to confusion about the true spectrum. Therefore, in order to obtain a consistent internal representation of the spectrum, the analysis duration needs to be at least as long as the period. The results of our experiment 1 are consistent with that model, although that model does not account for a level effect, as seen for some of our listeners.

No clear difference emerged in the data for chirp tones with the two different signs. Therefore, the phase effect observed in experiment 1 is not at all the same kind of thing observed in masking experiments by Lentz and Leek (2001), by Oxenham and Dau (2001), and by their predecessors. The masking experiments require the phase curvature of the masker to cancel the mechanical phase curvature at the signal place in the cochlea. By contrast, the unimportance of chirp sign in experiment 1 suggests that chirp tones disrupt the broad structure of the pattern of excitation all along a critical tonotopic coordinate, or else disrupt the readout of that structure.

It is often supposed that the spectral cues to elevation are at high frequencies, above 5 kHz (Butler and Belendiuk, 1977), because the anatomical features that lead to these cues are so small that they do not scatter sounds of longer wavelength and lower frequency. We suspected that the phase effect in experiment 1 came from the high-frequency region normally associated with sagittal-plane localization where components of a 65-Hz tone are not resolved. Experiment 2 was done to check that suspicion.

III. EXPERIMENT 2, HIGH-PASSED STIMULI

Experiment 2 used tones and noise with no components below 4030 Hz, corresponding to harmonic 62 of a 65-Hz tone. As in experiment 1, the upper frequency was 16 250 Hz. High-passing the stimulus in this way reduced the level by a small amount, 1.1 dB. Computations with gammatone auditory filters having equivalent rectangular bandwidths given by Glasberg and Moore (1990) showed that no harmonics of 65 Hz in this frequency range were resolved. Therefore, experiment 2 (range of two octaves) offered simplifications compared to experiment 1 (five octaves).

Apart from the change in the low-frequency boundary, experiment 2 was identical to experiment 1. The chirp tones and random-phase tones had 189 harmonics with equal amplitudes. The chirp tones had a crest factor of 1.90. Again listeners heard six different random-phase tones, randomly presented. Those tones were selected to have similar crest factors, 3.16 (±0.01). As in experiment 1, two levels were used for all the stimuli, 50 and 70 dBA, randomized by ±2 dB.

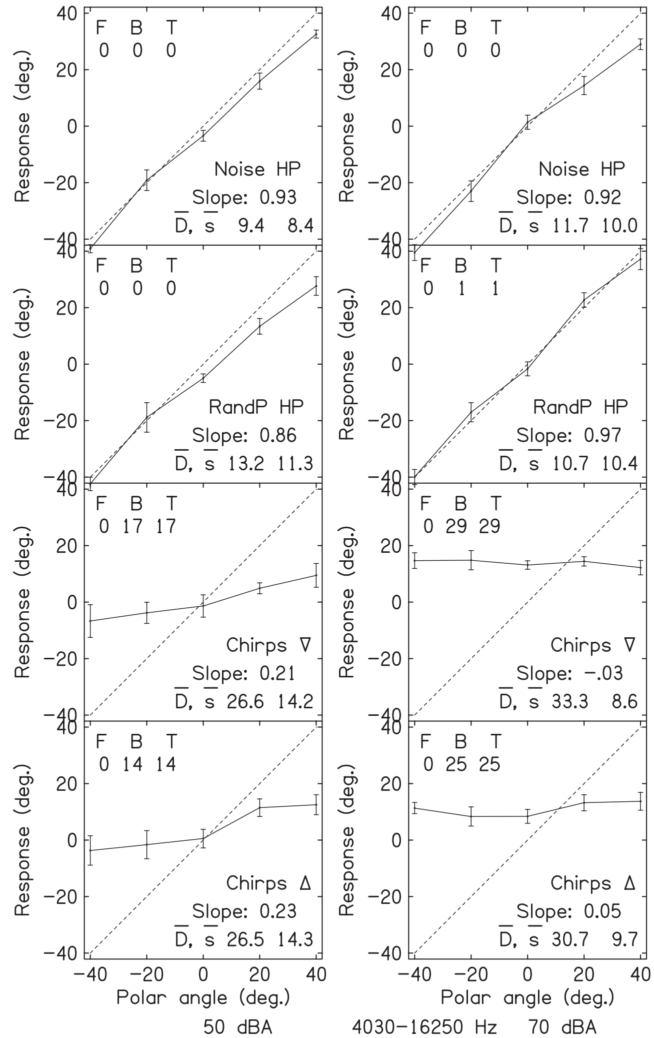

FIG. 3. Same as Fig. 1 except for high-pass noise and tones in experiment 2.

Additional runs at 40 dBA were done for listeners G and R to see whether a negative level effect could be found with the inclusion of this lower level. Therefore, listeners G and R completed runs at all three levels: 40, 50, and 70 dB.

A. Results

Figure 3 shows the results of experiment 2 for listener V. That figure is entirely parallel to Fig. 1, with noise and periodic tones in the same order. A comparison between experiment 2 (high-passed) and experiment 1 (broadband) for all the listeners can be made by comparing the bottom and top halves of Tables I–III. The comparison shows that performance was somewhat worse for the high-passed condition compared to broadband. For the high-passed stimuli, the number of front-back reversals was greater and $\bar{D}$ values were greater. A two-way repeated-measures ANOVA was conducted on each of the statistics with factors of bandwidth and waveform/level condition.

- For reversals, there was a significant effect of bandwidth $[F(1,4)=44.9,\ p<0.005]$ and waveform/level condition $[F(7,28)=36.7,\ p<0.001]$, and a significant interaction $[F(7,28)=3.3,\ p<0.02]$.

- For slopes, the main effect of bandwidth was not significant [$F(1,4)=0.9$, $p=0.4$], but the effect of waveform/level condition was [$F(7,28)=32.3$, $p<0.001$] as was the interaction [$F(7,28)=2.9$, $p<0.02$].
- For $\bar{D}$, there was a significant effect of bandwidth [$F(1,4)=12.8$, $p<0.05$] and waveform/level condition [$F(7,28)=25.2$, $p<0.001$], and a significant interaction [$F(7,28)=4.3$, $p<0.005$].

Overall, these analyses and inspection of the tables indicate that high-pass filtering impaired polar-angle performance under some waveform/level conditions. At least some subjects gained useful elevation information from spectral cues below 4 kHz. Asano *et al.* (1990) and Zhang and Hartmann (2010) also found that some listeners can use cues below 4 kHz to distinguish front from back, and Blauert (1969–70) found sagittal-plane localization bands even at 1 kHz and below.

An observation made from the polar-angle plots that is not captured in our summary statistics is that there was some tendency for the mean polar-angle response to increase when the chirp tones were high-passed (e.g., for listener V, compare Figs. 1 and 3), consistent with the general rule that a high spectral centroid is associated with higher elevation (Pratt effect: Pratt, 1930; Cabrera and Morimoto, 2007). The mean response angle, collapsed across chirp directions, levels, and listeners increased from 3.4° to 11.7° (approaching significance in a paired t-test, $p=0.06$). An increase in response polar angle was not seen for the noise and random-phase tone conditions, where mean response angles were −1.0° and −1.6° for broadband and high-passed conditions, respectively ($p=0.36$). These results agree and extend the results of Cabrera and Morimoto in the sense that the Pratt effect occurred when listeners were uncertain about elevation but not when elevation was clear.

Despite some quantitative effects of high-pass filtering, a principal conclusion from experiment 2 is that the broad features of experiment 1 were retained. Noise and random-phase tones continued to be well localized, and chirp tones of both signs were poorly localized. Further, Tables I–III show that those listeners who showed a negative level effect for the chirp tones in experiment 1 usually continued to show that effect in experiment 2. Similarly, listeners G and R, who showed no negative level effect in experiment 1, also showed no negative level effect at 50 dB in experiment 2. Knowing that the results are largely unchanged when the frequency region of resolved components is removed simplifies the analysis of the observed phase effects. The remainder of this article will be entirely concerned with the spectral region between 4 and 16 kHz.

The special experiment for listeners G and R was successful. Although those listeners showed no consistent negative level effect when the level was reduced from 70 to 50 dB, a negative level effect did appear in experiment 2 at 40 dB. The tables show that the performance statistics improved for both these listeners when the level was reduced from 50 to 40 dB. The average number of front-back reversals decreased from 21 to 8, and the slope increased from 0.09 to 0.37. Therefore, both G and R (the youngest listeners in the experiment by 10 years) exhibited a negative level effect like the other three listeners, but at a lower level.

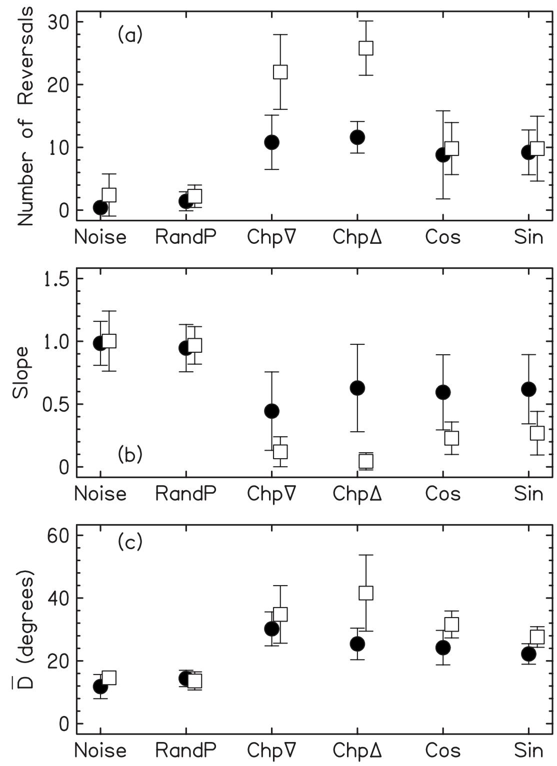

FIG. 4. Same as Fig. 2 except for high-pass noise and tones in experiments 2 and 3. High (□) and low (●) levels were 70 and 50 dB SPL.

To assess the effect of waveform and level on the group data for experiment 2 a two-way repeated-measures ANOVA was conducted on the number of reversals, on the slopes, and on the values of $\bar{D}$. For the factor of level, the highest (70 dB) and lowest (40 or 50 dB) levels tested on a given listener were used. The results were significant in all respects.

- For reversals, there was a significant effect of waveform [$F(3,12)=57.7$, $p<0.001$], a significant effect of level [$F(1,4)=138.5$, $p<0.001$], and a significant interaction [$F(3,12)=10.32$, $p<0.005$].
- For slopes, there was a significant effect of waveform [$F(3,12)=115.7$, $p<0.001$], a significant effect of level [$F(1,4)=13.5$, $p<0.05$], and a significant interaction [$F(3,12)=11.4$, $p<0.005$].
- For $\bar{D}$, there was a significant effect of waveform [$F(3,12)=29.2$, $p<0.001$], a significant effect of level [$F(1,4)=16.9$, $p<0.05$], and a significant interaction [$F(3,12)=8.0$, $p<0.005$].

Therefore, including the 40-dB data for listeners G and R along with the 50-dB data from the other listeners revealed the expected negative level effect. These results are illustrated in Fig. 4 (the four leftmost clusters in each panel).

Paired t-tests were done on the measures in the high-passed portions of Tables I–III to discover whether there was a significant effect attributable to the signs of the chirps in experiment 2. None of the three tests was significant. Consistent with most of the evidence from experiment 1, the sign of chirps had little influence on elevation localization.

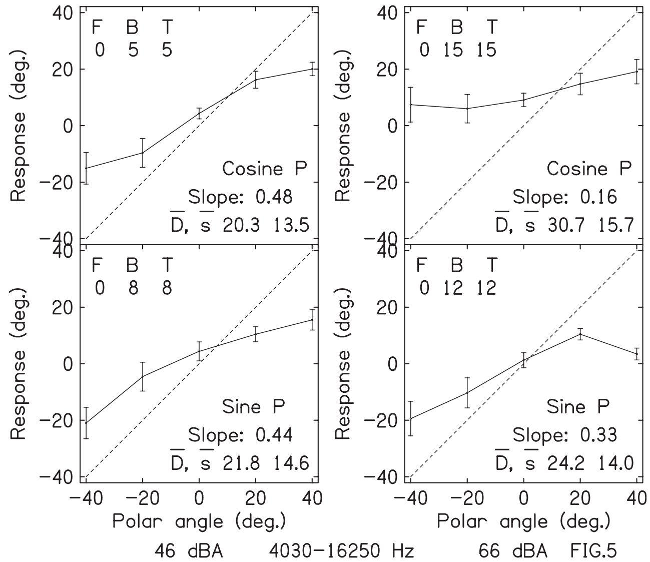

FIG. 5. Results for high-pass stimuli as in Fig. 3, but for different phase relationships, cosine-phases, and sine-phases in experiment 3.

IV. EXPERIMENT 3. ALTERNATIVE PHASES

Experiment 3 was like experiment 2 in that all the stimuli were high-passed, having spectra extending from 4030 to 16 250 Hz. Experiment 3 used different phase conditions.

(1) *Cosine-phase tones.* The cosine-phase tones added the 189 harmonics such that all harmonics contributed maximally once during each cycle. Such a tone has the largest possible crest factor ($\sqrt{2 \cdot 189} = 19.4$), and it resembles the periodic pulses presented by Brungart and Simpson (2008). Because of experimental limitations, this tone was presented at levels of 46 and 66 dBA, randomized by ±2 dB.

(2) *Sine-phase tones.* The sine-phase tones added the 189 harmonics in sine-phase, which led to a peaky waveform with a crest factor of 16.9—slightly less peaky than cosine-phase. Sine-phase tones were also presented at nominal levels of 46 and 66 dBA and randomized by ±2 dB.

Runs for experiment 3 were done contemporaneously with runs for experiment 2 to permit the fairest comparison of the results. Lower sound level runs were again included for listeners G and R.

A. Results

Figure 5 shows the results from experiment 3 for listener V. For this listener, the cosine- and sine-phase stimuli led to worse performance than the random-phase stimuli. The number of front-back reversals was larger, the elevation gain was depressed, and the rms error was larger with the sine- and cosine-phases. This was true of all listeners (Tables I–III). For cosine- and sine-phase tones, a negative level effect was evident for across all listeners for slopes and for $\bar{D}$, but not for the number of front-back reversals. Data for listeners G and R indicate a negative level effect for cosine- and sine-phases when the level was reduced from 50 to 40 dB, as expected from the effect seen for the chirp tones, though the tables show that the effect did not hold in every possible comparison.

The critical comparison between experiment 3 and experiment 2 can be made by examining the last four lines of Tables I–III. By these measures, the chirp tones led to worse elevation localization than the cosine- or sine-phase tones. Most notably, listeners on the whole made half as many reversals for cosine- and sine-phases as compared to both chirp signs (Table I).

A difference in slopes was observed at the high level, where, averaged over listeners, the slope for chirps was 0.08 and the slope for cosine- and sine-phases was 0.25 (Table II). Table III indicates larger $\bar{D}$ for chirps compared to cosine- and sine-phases, 38° and 30°, respectively.

To look for a difference between sine, cosine, and chirp waveforms and to assess the effect of level on the group data, a two-way repeated-measures ANOVA was conducted on the number of reversals, on the slopes, and on the values of $\bar{D}$, again using the highest and lowest levels tested per listener.

- For reversals, there was a significant effect of waveform [$F(3,12)=10.7$, $p<0.005$], a significant effect of level [$F(1,4)=14.2$, $p<0.05$], and a significant interaction [$F(3,12)=7.0$, $p<0.01$].
- For slopes, there was no significant effect of waveform [$F(3,12)=3.4$, $p=0.05$], a significant effect of level [$F(1,4)=16.1$, $p<0.05$], and no significant interaction [$F(3,12)=2.8$, $p=0.09$].
- For $\bar{D}$, there was no significant effect of waveform [$F(3,12)=3.0$, $p=0.07$], a significant effect of level [$F(1,4)=26.7$, $p<0.01$], and a significant interaction [$F(3,12)=4.1$, $p<0.05$].

Thus, a negative level effect was consistently found in experiments 2 and 3. Improved localization of cosine- and sine-phase tones as compared to chirps appeared in the reversal statistic and otherwise only at 70 dB. The results of the ANOVAs agree with the visual impression of the three statistics averaged over the five listeners shown in Fig. 4 (the four rightmost clusters in each panel).

V. DISCUSSION AND CONCLUSIONS

The experiments reported in this article explored sagittal-plane localization for periodic complex tones having different component phases: random-phases, Schroeder-phases (chirp tones), and cosine- or sine-phases. The chirps are frequency sweeps, but these were not perceivable as such; the listener was not aware of a rising or falling pitch. The chirp tones have 65 sweeps/s and sound like buzzy tones. Perceptually, they differed from random-phase, 65-Hz tones only in timbre. The same can be said for the cosine- and sine-phase tones.[2]

The experiments found that the Schroeder-phase tones of both sweep directions and also cosine- and sine-phase tones led to dramatically poorer sagittal-plane localization compared to noise or random-phase tones. A second conclu-

sion of the experiments is that chirp tones led to somewhat worse elevation localization than the impulsive tones using cosine- and sine-phases. The relative advantage of sine- and cosine-phases seems paradoxical. The cosine- and sine-phase tones are the most impulsive waveforms of all; the Schroeder-phases are intended to make the chirp tones smooth, with minimal crest factor. Evidently the distribution of energy over time renders the chirp tones harder to localize than even the peakiest of physical waveforms.

Finally, the localizability of the chirps and impulsive tones improved considerably at low stimulus levels, an effect known as the negative level effect. There are several different theoretical approaches that address these results.

A. Excitation pattern broadening

Hartmann and Rakerd (1993) attributed the difficulty that listeners have with the elevation of intense impulses to tonotopic broadening, probably beginning on the basilar membrane. Such broadening would reduce the availability of detailed spectral information at the central processor and consequently reduce the localizability of the impulses. According to this interpretation, the tonotopic broadening grows with increasing intensity, which would account qualitatively for the negative level effect. The failure of elevation localization thus resembles the paradox introduced by Sachs and Young (1979) in connection with the perception of intense vowels. The paradox is that given the known cochlear tuning, as evidenced by the rate-place coding in the auditory nerve of cat, listeners should not be able to discriminate between different vowel sounds because distinct formant bands would be largely obliterated even at moderately high sound levels. However, in practice, listeners have no difficulty discriminating vowels at all reasonable levels. Although the original paradox is now less stark, because of synchronous firing rate (Young and Sachs, 1979), or high-threshold auditory nerve fibers (Liberman, 1978), or cochlear differences between cat and human (Recio *et al.*, 2002), tonotopic broadening is known to occur (Glasberg and Moore, 2000), and it may well account for failures to encode elevation.

Using 3-ms noise bursts, Macpherson and Middlebrooks (2000) also found a negative level effect for stimulus levels greater than 40–45 dB, in agreement with Hartmann and Rakerd (1993) and with the observations of the present experiments. Macpherson and Middlebrooks (2000) also found that adding continuous, spatially diffuse noise to intense, impulsive stimuli led to release from the negative level effect. On the basis of that experiment, they suggested that a continuous signal would lead to an adaptation of the rate pattern on the auditory nerve, perhaps via the olivocochlear efferents. However, periodic impulsive tones as used by Brungart and Simpson (2008) and in the present experiments would presumably lead to that kind of adaptation as well, and yet the negative level effect was strong for these tones.

B. Excitation patterns: Temporal response

An alternative viewpoint focuses on the integration window requirement suggested by Hofman and Van Opstal (1998). To study the temporal properties of the stimuli as transformed by a model auditory system, we passed several of our stimuli through a cochlear model with three features: cochlear delay, gammatone filters, and half-wave rectification as described in the Appendix.

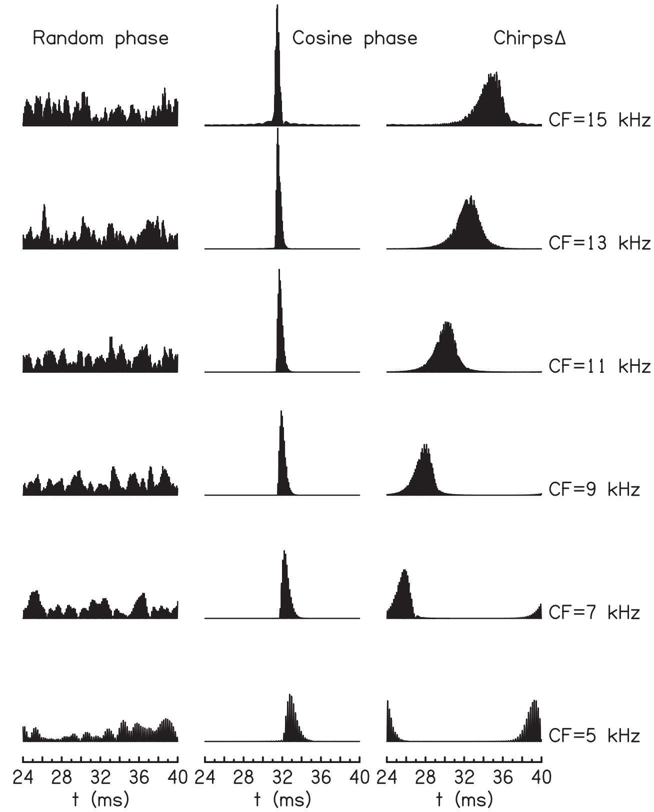

FIG. 6. Linear response of a gammatone filter bank to periodic tones used in experiments 2 and 3 represented at six different places on the model basilar membrane. The small effect of cochlear delays can be seen for cosine-phase where excitation at the 15-kHz place occurs about 1 ms before excitation at the 5-kHz place. For the chirp↑ tones there is negligible temporal overlap between excitation at 15- or 13-kHz places and excitation at the 9-kHz and lower places. The filter bank has equivalent rectangular band widths given by Glasberg and Moore (1990).

The results of the calculation are shown in Fig. 6 for six representative auditory filters, centered on f_c=5, 7, 9, 11, 13, and 15 kHz. The time axis begins 24 ms after the onset of the signal. Because the impulse responses are shorter than 24 ms, the figure shows the steady-state response. The time axis continues for an additional 16 ms, slightly longer than one period of the tones (1/65 s). The amplitude scale factors are the same for all plots.

Figure 6 can be studied column by column. The detailed response to a *random-phase tone* depends on the specific phases of the tone, but any set of randomly chosen phases leads to a continuous and stochastically homogeneous output, $y(t)$ as shown in the left column of Fig. 6.

The response to each cycle of a *cosine-phase tone* resembles the response to a single impulse. A small effect of cochlear delay is seen because the excitation arrives at the 5-kHz place about 1 ms after arriving at the 15-kHz place. The excitation at the 5-kHz place has the longest duration because the bandwidth is the smallest so that the ring time of the auditory filter is the longest.

The response to a *Schroeder-minus-phase tone* (chirp↑)

shows the expected pattern for a rising frequency sweep. Excitation appears first at the 5-kHz place and is progressively delayed toward places with higher frequency.

Although Fig. 6 shows that the cosine-phase tone leads to a more impulsive response than the chirp tone, the chirp tone poses the special difficulty that the excitation at different places occurs at different times. Throughout the stimulus duration, there never is a time when excitation near 15 kHz and excitation at 9 kHz or below occur simultaneously. This is the point originally made by Hofman and Van Opstal (1998) and by Vliegen and Van Opstal (2004). In this model, the profile analysis that occurs along the tonotopic axis to mediate elevation localization has a short integration time. An integration time of 5 ms was suggested by Hofman and Van Opstal, meaning that excitations separated by more than 5 ms cannot be compared.

It is interesting to speculate that the negative level effect might be predictable from a variation on our elementary filter bank model. It is expected that lower levels would lead to cochlear filters with narrower bandwidths (Glasberg and Moore, 2000). We wondered what effect that would have on the duration and temporal overlap of excitations at different places.

For chirp stimuli, reducing the bandwidth might have two opposite effects. It might increase the duration of the responses at all the places because decreasing bandwidth leads to longer *ring time*. According to the view of Hofman and Van Opstal (1998), this would predict better performance with reduced level. On the other hand, reducing the bandwidth might decrease the duration of the responses at all places because the frequency of the sweep spends less time within the *passband* of each filter.

The gammatone filter bank calculation that produced Fig. 6 was rerun with the filter bandwidths successively reduced, paying particular attention to temporal overlaps for excitations with chirps. The calculations are described in the Appendix. When the model bandwidth was reduced by a factor of 2 or by a factor of 4 compared to Fig. 6, there was little effect on the duration of the excitations and their temporal overlaps. The ring-time effect and the passband effect seemed to offset one another. But when the bandwidth was reduced by a factor of 8, the ring-time effect dominated dramatically, as shown in Fig. 7. The responses to chirp tones were so extended that the excitation patterns overlapped despite the frequency sweep. Also the responses to cosine-phase tones were extended so that the overall duration approached the 5-ms integration window suggested by Hofman and Van Opstal (1998).

The filter bank model, which appears to account for elevation localization in a satisfactory way, raises a number of questions. The most obvious problem with it is the reduction in bandwidth by a factor of 6 or 8. There is evidence for considerable reduction in bandwidth with decreasing level in the low-level forward masking experiments by Oxenham and Shera (2003) and the otoacoustic emissions experiments by Shera *et al.* (2002). Bandwidth reduction compared to Glasberg and Moore (1990) became as large as a factor of 2 at the highest frequency studied in the masking experiments, 8 kHz. However, a factor of 2 is not yet a factor of 6 or 8. It could be noted that the standard measures of auditory tuning do not extend above 10 kHz. Because sagittal-plane localization involves frequencies up to about 16 kHz, it includes about an octave of unknown territory. The power law extrapolation by Shera *et al.* says that bandwidths grow as the 0.7 power of the place center frequency. That would suggest a bandwidth reduction by a factor of 2.5 at 16 kHz. However, a factor of 2.5 is also not yet a factor of 6 or 8.

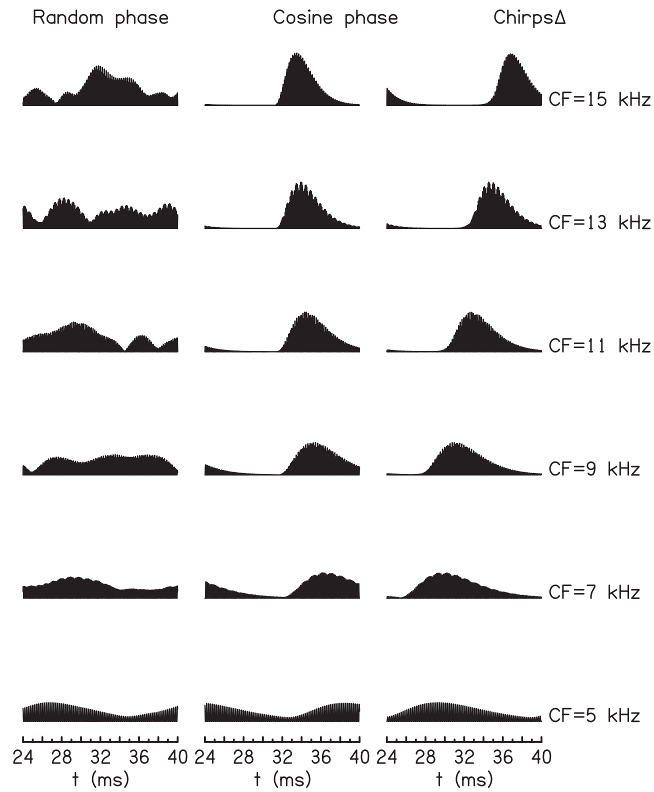

FIG. 7. Same as Fig. 4 except that the auditory filter bandwidths have been divided by a factor of 8. Also, the amplitudes of excitation patterns have been multiplied by a factor of 3 for better comparison with Fig. 4. For the chirp↑ tones there is considerable temporal overlap of excitation at all places, and the response to the pulse train is extended in time.

The filter bank model also assumes linearity in the sense that excitation y doubles if the input to the auditory filter doubles. The model fails to agree with our localization data if the filter output is highly compressed. To examine the effect of compression, we considered a compressive exponent of 0.3 (i.e., 0.3 dB/dB) because exponents observed in pure-tone physiology and psychoacoustics are often this small or smaller (Bacon *et al.*, 2004). Then the excitation overlap (which can be compared to Fig. 8 in the Appendix) ranges from 0.5 to 1.8 for all phase relationships; the choice of bandwidth has very little effect on overlap. Such a small range of overlaps is inconsistent with our modeling of the Hofman and Van Opstal (1998) overlap theory. Because the bandwidth matters little when compression is added, there would be no negative level effect. Because overlaps are substantial even for normal auditory filter bandwidths, there would be no effect of phase on perceived elevation. Both these results are contrary to the results of our experiments.

Ultimately one cannot be satisfied with the computational version of the Hofman and Van Opstal (1998) tempo-

ral model presented in this section. The model cannot account for the negative level effect without unreasonable assumptions about the dependence of auditory filter widths on intensity. With standard auditory filter bandwidths, the model does not tolerate the compression that is expected physiologically.

C. Excitation patterns: Lateral inhibition

An alternative to the linear filter model is a model with lateral inhibition. In this model it is assumed that cochlear tuning is adequate to resolve the high-frequency spectral structure encoded by a head-related transfer function (HRTF) for low-level stimuli, but cochlear tuning is inadequately sharp at high levels. Nevertheless, the auditory system is capable of encoding the high-frequency structure that mediates sagittal-plane localization at high levels because of subsequent lateral inhibition which sharpens the tonotopic pattern of neural firing rates. However, lateral inhibition can operate only if temporal features of the stimulus allow it to do so.

A lateral inhibitory model of this kind was presented by Shamma (1985), and mainly applied to low-frequency excitation where phase locking occurs on the auditory nerve. One can imagine a high-frequency application where only a time-dependent rate code is available. In that model, a neural firing probability is determined by a membrane potential that obeys a differential equation that is first-order in time, like a low-pass filter. The model membrane is driven by a tonotopically local input excitation and is inhibited by the firing rate of neighboring neurons (recurrent model). The inhibitory connections were given their own dynamics by Hancock and Voigt (1999). If the excitation is broadband and stable in time, or if the excitation varies rapidly compared to the time constants of the dynamical system, lateral inhibition will occur and will sharpen the output neural firing pattern. Particularly, the firing pattern in spectral valleys of excitation will be inhibited by the greater firing rate of neighboring neurons, thus accentuating the peak-and-valley structure from the HRTF. However, if the excitation sweeps relatively slowly in frequency, the solution to the membrane differential equation will be entirely local in space. In particular, the solution in a spectral valley region will depend only on the excitation in the valley because all memory of the firing rate in neighboring neurons will be lost. For such a slowly moving stimulus, normal lateral inhibition will not occur, firing rate in the valleys will not be inhibited, the final representation of the spectrum will not adequately reflect the HRTF at high level, and elevation localization will be impaired.

As presented here, the lateral inhibition model is entirely qualitative. It has not been tested against our data. Lateral inhibition is known to occur in the dorsal cochlear nucleus (DCN), a likely site in the auditory pathway that mediates the perception of spectral properties (Young *et al.*, 1992). The lateral inhibition model has a chance of plausibility if the critical temporal parameters of the psychophysical observations agree with the dynamical time constants observed in the DCN physiology. The best estimate for a psychophysical temporal parameter is the 5-ms value obtained in the chirp experiments by Hofman and Van Opstal (1998). Membrane time constants in the model by Hancock and Voigt (1999) that mimics DCN electrophysiology are also of this order, ranging from 1 to 10 ms.

It seems possible that a filter bank model that incorporates lateral inhibition could account for both the phase effects and the level effects observed in our experiments. In this model all of the impulsive stimuli are localized poorly due to excitation pattern broadening. However, this effect is stronger with chirp stimuli than with cosine- and sine-phase stimuli because the inhibitory system is too fast for the stimulus; its operations are completed while the only relevant stimulus excitation is tonotopically local.

ACKNOWLEDGMENTS

We are grateful to Craig Jin, André van Schaik, Douglas Brungart, and Brian Simpson for useful discussions. Yun Jin Cho helped with calculations. This work was supported by the NIDCD Grant No. DC-00181 and by the Australian Research Council. V.B. was supported by a University of Sydney postdoctoral research fellowship.

APPENDIX: OVERLAP CALCULATIONS

To study the temporal properties of the stimuli as transformed by a model auditory system, we passed the random-phase, cosine-phase, and Schroeder-phase (minus) tones from experiment 2, as described by Eq. (1), through a bank of fourth-order gammatone filters.

The filters had impulse responses for positive time t given by

$$h_c(t) = b_c^4 t^3 e^{-2\pi b_c t} \cos(2\pi f_c t), \qquad \text{(A1)}$$

where subscript c refers to the filter channel and f_c is the center frequency of the channel. Filters of this form provide reasonable fits to human auditory filters as determined from notched-noise masking data (Patterson *et al.*, 1992) and to the impulse responses measured on auditory nerve fibers (Carney and Yin, 1988). Parameters b_c were chosen to match the equivalent rectangular bandwidths of human auditory filters (Cambridge ERBs) determined by Glasberg and Moore (1990), $b_c = 1.018\ (\mathrm{ERB}_c)$.

The responses to the tones were given by the convolution between the impulse response and the signal

$$y_c(t) = \mathcal{R}_{\mathcal{H}}\left[\int_0^t dt'\, h_c(t - t_c - t')x(t')\right], \qquad \text{(A2)}$$

where $\mathcal{R}_{\mathcal{H}}$ indicates a half-wave rectification, and parameter t_c is the cochlear delay for channel c, here taken to be $6/f_c$ as a fit to the data from Goodman *et al.* (2009) (e.g., if $f_c = 6$ kHz then $t_c = 1$ ms).

The measures of overlap, prominent in Figs. 6 and 7, were made quantitative by computing normalized product moments of excitations for 5-ms time slices. The overlap in excitation between patterns at places c and c' in the 5-ms interval beginning at initial time t_i is

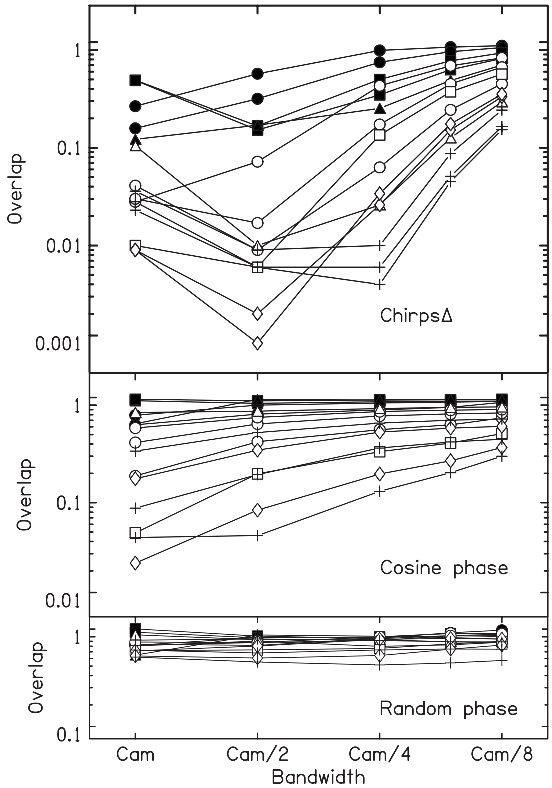

FIG. 8. Overlap $V_{\mathrm{av},c,c'}$ for 15 pairs of bands $[c,c']$ as a function of auditory filter bandwidth. A bandwidth of 1 Cam corresponds to Cambridge equivalent rectangular bands, called "normal" in the text. Other bandwidths are smaller by factors of 2, 4, 6, and 8. Different symbols indicate different band pairs, described by center frequencies in kilohertz: Filled circles 7–5 and 9–7. Filled squares 11–9 and 13–11. Filled triangles 15–13. Open circles 9–5, 11–7, and 13,9. Open squares 11–5. Open triangles 15–11. Open diamonds 13–7 and 13–5. Plus signs, 15–9, 15–7, and 15–5. Overlaps computed for bandwidths of Cam and Cam/8 correspond to Figs. 6 and 7, respectively.

$$V_{c,c'}(t_i) = \frac{10}{\mathrm{norm}} \int_{t_i}^{t_i+5\ \mathrm{ms}} dt\, y_c(t) y_{c'}(t), \tag{A3}$$

where the normalization is computed over a period, approximately

$$\mathrm{norm} = \max\left\{ \int_{t=24}^{t=39\ \mathrm{ms}} dt\, y_c^2(t), \quad \int_{t=24}^{t=39\ \mathrm{ms}} dt\, y_{c'}^2(t) \right\}. \tag{A4}$$

The scale factor of 10 was for convenience, to put the largest values of $V_{c,c'}$ in the vicinity of 1.0. The calculation was done for 16 time intervals with t_i given by 24, 25, 26, …, 39 ms. The average overlap was computed by averaging $V_{c,c'}(t_i)$ over those 16 intervals

$$V_{\mathrm{av},c,c'} = \frac{1}{16} \sum_{t_i=24}^{t_i=39\ \mathrm{ms}} V_{c,c'}(t_i). \tag{A5}$$

Because there are six bands (c) in the calculation there are 15 band pairs (c,c') for overlap computation. Plots were made of $V_{\mathrm{av},c,c'}$ for the 15 pairs as a measure of the overlap. The plots are shown in Fig. 8. As expected, the maximum overlap normally occurred between neighboring places, such as 15- and 13-kHz places. The minimum overlap occurred between widely separated places, such as 15 and 5 kHz. Although the plots normally behaved as expected, the competition between ring-time and bandwidth effects led to some unexpected non-monotonic behaviors.

The plots in Fig. 8 demonstrate the effects of reduced bandwidth in the filter bank model. For random-phase tones, the overlap values spanned a small range from 0.5 to 1.2 and it did not matter whether the auditory filter was wide or narrow. By contrast, the overlaps for cosine-phase tones were quite small for the most widely separated places for normal bandwidths (Cam/1). These small overlaps grew by a factor of more than 10 when the bandwidth was divided by 8.

The computations for Schroeder-phase tones (chirp tones) were even more dramatic. Both the minimum and the median overlaps were small for normal bandwidths, and they grew by more than a factor of 15 when the bandwidth was divided by 8. The large differences between the overlaps for cosine-phase and chirp tones compared to random-phase tones seen for normal bandwidths became much smaller when the bandwidth became one-eighth as large. Dividing the bandwidth by a factor of 6 was almost as effective as a factor of 8, increasing the minimum by a factor of 5 and the median by a factor of 10 for chirp tones.

The overlap calculations shown in Fig. 8 are in line with the elevation localization, and its level dependence, seen experimentally for the three different phase relationships. For random-phase tones, localization did not depend on level. For moderate levels (presumably leading to normal filter bandwidths), cosine-phase and Schroeder-phase tones were poorly localized, especially Schroeder-phase. For low levels (narrow filter bandwidths) tones with all the different phase relationships were often almost equally localizable.

[1]Mathematically, the responses near the horizon (zero polar angle) seen in Table II can occur in two ways: The individual responses can mostly be near zero or the individual responses can be widely spread over positive and negative values such that they average to zero. These two cases can be compared for our depressed slopes by remembering that the variance, averaged over sources, is $\overline{s^2}$, and considering that the squared constant error, or bias, is $\overline{D^2 - s^2}$. The case in which individual responses are mostly near zero can be recognized if the bias is greater than the variance, or $\bar{D} > \bar{s}\sqrt{2}$. This condition held for all chirps for all listeners where the slope was less than 0.3.

[2]Informal discrimination experiments on high-passed signals were done by listeners V, G, and R in the anechoic room with the loudspeaker overhead. Chirps could easily be discriminated from cosine- or sine-phase tones and from random-phase tones because chirps produced almost no sense of low pitch. Chirp↑ tones could usually be discriminated from chirp↓ tones because the latter have a slightly brighter timbre.

Asano, F., Suzuki, Y., and Sone, T. (**1990**). "Role of spectral cues in median plane localization," J. Acoust. Soc. Am. **88**, 159–168.

Bacon, S. P., Fay, R. R., and Popper, A. N. (**2004**). *Compression From Cochlea to Cochlear Implants* (Springer, New York).

Blauert, J. (**1969**). "Sound localization in the median plane," Acustica **22**, 205–213.

Brungart, D. S., and Simpson, B. D. (**2008**). "Effects of temporal fine structure on the localization of broadband sounds: Potential implications for the design of spatial audio displays," Proceedings of the 14th International Conference on Auditory Display, Paper No. ICAD08-1-08–5.

Butler, R. A., and Belendiuk, K. (**1977**). "Spectral cues utilized in the location of sound in the median sagittal plane," J. Acoust. Soc. Am. **61**, 1264–1269.

Cabrera, D., and Morimoto, M. (**2007**). "Influence of fundamental frequency and source elevation on the vertical localization of complex tones and

complex tone pairs," J. Acoust. Soc. Am. **122**, 478–488.
Carlile, S., Long, P., and Hyams, S. (**1997**). "The nature and distribution of errors in sound localization by human listeners," Hear. Res. **114**, 179–196.
Carney, L. H., and Yin, T. C. T. (**1988**). "Temporal coding of resonances by low-frequency auditory nerve fibers: Single-fiber responses and a population model," J. Neurophysiol. **60**, 1653–1677.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched noise data," Hear. Res. **47**, 103–138.
Glasberg, B. R., and Moore, B. C. J. (**2000**). "Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise," J. Acoust. Soc. Am. **108**, 2318–2328.
Goodman, S. S., Fitzpatrick, D. F., Ellison, J. C., Jesteadt, W., and Keefe, D. H. (**2009**). "High-frequency click-evoked otoacoustic emissions and behavioral thresholds in humans," J. Acoust. Soc. Am. **125**, 1014–1032.
Hancock, K. E., and Voigt, H. F. (**1999**). "Wideband inhibition of dorsal cochlear nucleus type IV units in cat: A computational model," Ann. Biomed. Eng. **27**, 73–87.
Hartmann, W. M., and Rakerd, B. (**1993**). "Auditory spectral discrimination and the localization of clicks in the sagittal plane," J. Acoust. Soc. Am. **94**, 2083–2092.
Hartmann, W. M., and Wolf, E. (**2009**). "Matching the waveform and the temporal window in the creation of experimental signals," J. Acoust. Soc. Am. **126**, 2580–2588.
Hofman, P. M., and Van Opstal, A. J. (**1998**). "Spectro-temporal factors in two-dimensional human sound localization," J. Acoust. Soc. Am. **103**, 2634–2648.
Kohlrausch, A., and Sander, A. (**1995**). "Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets," J. Acoust. Soc. Am. **97**, 1817–1829.
Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (**1999**). "Sensitivity of human subjects to head-related transfer function phase spectra," J. Acoust. Soc. Am. **105**, 2821–2840.
Lentz, J. J., and Leek, M. R. (**2001**). "Psychophysical estimates of cochlear phase response: Masking by harmonic complexes," J. Assoc. Res. Otolaryngol. **2**, 408–422.
Liberman, M. C. (**1978**). "Auditory-nerve response from cats raised in a low-noise chamber," J. Acoust. Soc. Am. **63**, 442–455.
Macpherson, E. A., and Middlebrooks, J. C. (**2000**). "Localization of brief sounds: Effects of level and background noise," J. Acoust. Soc. Am. **108**, 1834–1849.
Mehrgardt, S., and Mellert, V. (**1977**). "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. **61**, 1567–1576.
Middlebrooks, J. C. (**1992**). "Narrow-band sound localization related to external ear acoustics," J. Acoust. Soc. Am. **92**, 2607–2624.
Oxenham, A. J., and Dau, T. (**2001**). "Towards a measure of auditory-filter phase response," J. Acoust. Soc. Am. **110**, 3169–3178.
Oxenham, A. J., and Shera, C. A. (**2003**). "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," J. Assoc. Res. Otolaryngol. **4**, 541–554.
Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M., "Complex sounds and auditory images," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford, **1992**), pp. 429–446.
Pratt, C. C. (**1930**). "The spatial character of high and low tones," J. Exp. Psychol. **13**, 278–285.
Rakerd, B., and Hartmann, W. M. (**1986**). "Localization of sound in rooms III: Onset and duration effects," J. Acoust. Soc. Am. **80**, 1695–1706.
Recio, A., and Rhode, W. S. (**2000**). "Basilar membrane response to broadband stimuli," J. Acoust. Soc. Am. **108**, 2281–2298.
Recio, A., Rhode, W. S., Kiefte, M., and Kluender, K. R. (**2002**). "Responses to cochlear normalized speech stimuli in the auditory nerve of cat," J. Acoust. Soc. Am. **111**, 2213–2218.
Roffler, S. K., and Butler, R. A. (**1968**). "Factors that influence the localization of sound in the vertical plane," J. Acoust. Soc. Am. **43**, 1255–1260.
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.
Schroeder, M. R. (**1970**). "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation," IEEE Trans. Inf. Theory **16**, 85–89.
Shamma, S. A. (**1985**). "Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve," J. Acoust. Soc. Am. **78**, 1622–1632.
Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Smith, B. K., Sieben, U. K., Kohlrausch, A., and Schroeder, M. R. (**1986**). "Phase effects in masking related to dispersion in the inner ear," J. Acoust. Soc. Am. **80**, 1631–1637.
Vliegen, J., and Van Opstal, A. J. (**2004**). "The influence of duration and level on human sound localization," J. Acoust. Soc. Am. **115**, 1705–1714.
Wightman, F. L., and Kistler, D. J. (**1989**). "Headphone simulation of free-field listening. II: Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.
Young, E. D., and Sachs, M. B. (**1979**). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers," J. Acoust. Soc. Am. **66**, 1381–1403.
Young, E. D., Spirou, G. A., Rice, J. J., and Voigt, H. F. (**1992**). "Neural organization and responses to complex stimuli in the dorsal cochlear nucleus," Philos. Trans. R. Soc. London, Ser. B **336**, 407–413.
Zhang, P. X., and Hartmann, W. M. (**2010**). "On the ability of human listeners to distinguish between front and back," Hear. Res. **260**, 30–46.