Loudness of dynamic stimuli in acoustic and electric hearing[a)]

Chaoying Zhang and Fan-Gang Zeng[b)]
Auditory Perception Laboratory, Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 31 March 1997; revised 14 July 1997; accepted 14 July 1997)

Traditional loudness models have been based on the average energy and the critical band analysis of steady-state sounds. However, most environmental sounds, including speech, are dynamic stimuli, in which the average level [e.g., the root-mean-square (rms) level] does not account for the large temporal fluctuations. The question addressed here was whether two stimuli of the same rms level but different peak levels would produce an equal loudness sensation. A modern adaptive procedure was used to replicate two classic experiments demonstrating that the sensation of ''beats'' in a two- or three-tone complex resulted in a louder sensation [E. Zwicker and H. Fastl, *Psychoacoustics—Facts and Models* (Springer-Verlag, Berlin, 1990)]. Two additional experiments were conducted to study exclusively the effects of the temporal envelope on the loudness sensation of dynamic stimuli. Loudness balance was performed by normal-hearing listeners between a white noise and a sinusoidally amplitude-modulated noise in one experiment, and by cochlear implant listeners between two harmonic stimuli of the same magnitude spectra, but different phase spectra, in the other experiment. The results from both experiments showed that, for two stimuli of the same rms level, the stimulus with greater temporal fluctuations sometimes produced a significantly louder sensation, depending on the temporal frequency and overall stimulus level. In normal-hearing listeners, the louder sensation was produced for the amplitude-modulated stimuli with modulation frequencies lower than 400 Hz, and gradually disappeared above 400 Hz, resulting in a low-pass filtering characteristic which bore some similarity to the temporal modulation transfer function. The extent to which loudness was greater was a nonmonotonic function of level in acoustic hearing and a monotonically increasingly function in electric hearing. These results suggest that the loudness sensation of a dynamic stimulus is not limited to a 100-ms temporal integration process, and may be determined jointly by a compression process in the cochlea and an expansion process in the brain. A level-dependent compression scheme that may better restore normal loudness of dynamic stimuli in hearing aids and cochlear implants is proposed. [S0001-4966(97)05810-4]

PACS numbers: 43.66.Ba, 43.66.Cb, 43.66.Mk, 43.66.Nm [WJ]

INTRODUCTION

Models of loudness have been mostly based on steady-state sounds, and rarely validated using dynamic sounds with large temporal fluctuations. For pure tones, loudness (L) has been measured directly using the magnitude estimate technique and found to grow as a power function of sound intensity (I), $L=kI^{0.3}$ (Stevens, 1957). Fletcher (1940) established that the ear can be modeled as a bank of frequency filters with the filter bandwidth being termed ''critical bandwidth.'' Fletcher and Munson (1933, 1937) used the critical band concept implicitly in either an equal-loudness-contour-based empirical formula or a masking audiogram to calculate the loudness of complex sounds with broadband spectra. Based on the pioneering work by both Fletcher and Stevens, Zwicker formulated an excitation-pattern model to predict the loudness of complex sounds (e.g., Zwicker and Scharf, 1965). In Zwicker's model, the physical spectrum of a complex sound is first subject to a critical-band analysis, in which the ''critical-band intensity'' is calculated by adding spectral intensities within each critical band; then an ''excitation pattern'' of the complex sound is derived using the shape of masked thresholds; finally, the ''specific loudness'' is calculated using a power law transformation of the excitation level for each critical band and is added across critical bands to obtain the ''total loudness'' of the complex sound.

Zwicker and his colleagues used a two-tone complex to test the validity of the spectrally based loudness model. The loudness of a two-tone complex was found to be determined by adding the intensity of the two tones if they were within one critical band and by adding the loudness if they were in two widely separated critical bands (Zwicker *et al.*, 1957; Zwicker and Fastl, 1990). However, a small but consistent deviation from the spectrally based model was also noticed, in which the loudness of two tones exceeded the predicted value based on intensity addition when the frequency separation of the two tones was within 10 Hz and produced a sensation of ''beats.'' Under such conditions, the loudness of the two-tone complex appeared to be determined by the peak amplitude rather than the rms (root-mean-square) amplitude.

[a)]Parts of this paper were presented at the 19th Midwinter Research Meeting of the Association for Research in Otolaryngology, February 1996, St. Petersburg, Florida and the Third Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, December 1996, Honolulu, Hawaii.

[b)]Author to whom correspondence should be addressed. Electronic mail: zeng@hei.org

Zwicker *et al.* further explored this phenomenon using a three-tone complex in which the phase relationship of three tones was manipulated to produce different temporal waveform envelopes. Again, when the total bandwidth of the three-tone complex was less than 10 Hz, the waveform with greater peak amplitude (amplitude modulation) produced a louder sensation than the waveform with smaller peak amplitude (quasifrequency modulation). Perhaps because this deviation from the spectrally based model produced a relatively small effect and occurred at very low-frequency temporal fluctuations (less than 10 Hz or longer than 100 ms), the loudness of dynamic stimuli has been assumed to be due to the temporal integration process, with a time constant of about 75 ms (Zwicker and Fastl, 1990). Zwicker's model is now widely accepted and has formed the basis on which the international standard for calculating the loudness of various complex sounds is established (ISO532). Recently, Moore and Glasberg (1996) revised Zwicker's model to account more accurately for equal loudness contours and the loudness of partially masked sounds.

We are motivated to revisit the issue of loudness of dynamic stimuli based on questions arising from fitting prosthetic devices to hearing-impaired listeners. The first question relates to hearing aid gain prescription. To compensate for reduced audibility due to hearing loss, one desirable goal for a hearing aid is to amplify speech sounds appropriately in order to restore loudness to the most comfortable listening level. Most fitting procedures currently in use select the gain and frequency response from pure-tone thresholds and are based on empirical formulas (e.g., Byrne and Dillon, 1986). Although attempts have been made to use Zwicker's loudness model in hearing aid gain prescription, the results are not satisfactory because the model-predicted gain always overcompensates for the amount of hearing loss, and produces ''too loud'' sensations (Launer and Bachler, 1996). One reason for this failure is that Zwicker's loudness model calculates the rms level and does not account for temporal waveform fluctuations in dynamic stimuli such as speech sounds, which can have a peak amplitude of 10–20 dB above the rms level (Fletcher, 1953; Boothroyd *et al.*, 1994).

The second question relates to the loudness mapping problem in cochlear implants. In designing the implant processor, the amplitude of acoustic stimuli has to be compressed to accommodate the extremely narrow 10–20 dB dynamic range which is characteristic of the electric stimulation of the auditory nerve. Contrary to traditional compression circuits in hearing aids that have a fast attack time and a slow release time and are controlled by short-term averaging, the cochlear implant processor usually employs a zero attack- and release-time compression and employs a nonlinear compressive map between the instantaneous acoustic envelope magnitude and the electrical pulse amplitude (Wilson *et al.*, 1991). The rationale for compressing the instantaneous peak amplitude in cochlear implants was partially verified by Zeng and Shannon (1995), who found that the loudness of amplitude-modulated sinusoids was determined by the peak amplitude at an uncomfortable loudness level, but by the rms amplitude at the threshold level. The significant effect of the 100-Hz sinusoidal modulator on the loudness of the amplitude-modulated stimuli in Zeng and Shannon's study could not be explained simply on the basis of the presumably much slower temporal integration process.

The failure of applying Zwicker's loudness model to the fitting of prosthetic hearing devices can be due to many differences in analyzing sounds between normal-hearing and hearing-impaired listeners. One significant difference is the broader auditory filter in hearing-impaired listeners, or in the extreme case of single electrode stimulation of the auditory nerve, no spectral analysis mechanisms exist at all. Generally, the extent to which the temporal fluctuation of a broadband stimulus can be preserved is proportional to the ear's analysis (or critical) bandwidth. Compared with normal-hearing listeners, the temporal envelope of a dynamic stimulus is more likely to be preserved with broader auditory filters in impaired listeners and is preserved entirely with electric stimulation in implant listeners. The other significant difference is the steeper loudness growth in impaired listeners. Loudness recruitment is presumably due to a reduced level of compressive nonlinearity in cochlear-impaired listeners and a total loss in implant listeners. Several studies have attempted to modify Zwicker's model to take into account the factors of broader filters and loudness recruitment in hearing-impaired ears, but none of them was designed to deal with dynamic sounds (Florentine and Zwicker, 1979; Launer, 1995; Moore and Glasberg, 1997). There has been evidence that loudness recruitment not only occurs for steady-state sounds, but also accentuates the difference between peaks and valleys in dynamic stimuli (Moore *et al.*, 1996). If the temporal fluctuations can make a sound louder, then a greater temporal effect might be observed in hearing-impaired listeners.

The present study used two different methods to separate the effects of the critical band analysis from temporal processing in loudness sensation. First, we asked normal-hearing listeners to perform a loudness balance task between a white noise and a sinusoidally amplitude-modulated (SAM) noise. Both the white noise and the SAM noise had identical long-term spectra, and the only difference between them was the greater temporal fluctuations in the SAM noise. Thus if the temporal envelope had no effect on loudness sensation, then the white noise and the SAM noise would be equally loud as long as their rms levels were kept the same. Second, we asked cochlear implant listeners to judge the loudness of two harmonic stimuli with the same magnitude spectra but different phase spectra (Schroeder, 1970). Because manipulating phase relationships did not change the stimulus rms level, and the cochlear implant listeners were not able to resolve the harmonics at all, hypotheses regarding whether the rms level or the peak level determines loudness could be tested directly.

I. GENERAL METHODS

A. Subjects

Six normal-hearing listeners, including three males and three females, ranging from 25 to 35 years old, participated in this study. Four participated in the two- and three-tone experiments (exp. 1), after which two of them dropped out

and two additional listeners were recruited to participate in the SAM noise experiment (exp. 2). All subjects had normal hearing (10 dB HL or better pure-tone thresholds for frequencies between 125 and 5000 Hz). The subjects were seated in a double-walled acoustic booth and tested individually.

Three cochlear implant listeners participated in the study of balancing loudness between two harmonics of different phases (exp. 3). All cochlear implant listeners used the Ineraid device, which has a percutaneous plug that connects directly to intracochlear electrodes and allows undistorted electric stimulation with essentially any waveform. At the time of the test, the three implant listeners were 35–65 years old, and had used the device for more than seven years. All implant listeners had extensive experience in various psychophysical tasks. The most apical electrode was used as the stimulation electrode and the lead in the temporalis muscle was used as the ground electrode.

B. Stimuli

All stimuli were generated digitally by a PC using TDT (Tucker Davis Technologies) system II hardware. The stimuli were delivered through a 16-bit D/A converter at a 40-kHz sampling rate and smoothed by a 14-kHz low-pass filter. All stimuli had a 500-ms duration and 10-ms cosine-squared rise and fall times. In acoustic stimulation, the sound level was controlled digitally and calibrated periodically by a B&K sound level meter linked to a Zwislocki coupler and an ER-2 insert earphone (Etymotic Research). In electric stimulation, the output of the D/A converter was routed to a voltage-controlled current source that could deliver a maximal peak current of 1 mA (Vurek *et al.*, 1981). The speech processor of the cochlear implant was not used in the experiment. A foot-operated safety switch was provided between the current source and the percutaneous plug and could disconnect the circuitry immediately by either the experimenter or the listener in case of unpleasant or too loud stimulation.

C. Procedure

A two-interval, two-alternative, forced-choice, adaptive procedure was employed in which two randomly interleaved sequences with different decision rules bracketed the point of subjective equality in the loudness balance task (Jesteadt, 1980; Schlauch and Wier, 1987; Zeng and Turner, 1991). In these two sequences, the upper sequence used a 2-down, 1-up decision rule to track the 71% louder response level, while the lower sequence used a 1-down, 2-up decision rule to track the 29% louder level. In each trial, a standard sound with a fixed level and a comparison sound with a variable level that was selected randomly from one of two sequences were presented randomly in either the first interval or the second interval. Each interval was signaled by a light on the computer monitor and the two intervals were separated by 500 ms. The subject's task was to judge which one of the two intervals contained the louder sound by pressing one of two buttons on a PC mouse. No feedback regarding the ''correct'' response was given to the subject after each trial. Each sequence was terminated by either having a total of 75 trials or having reached 12 reversals, which corresponded to a change in the direction of louder response from the standard to the comparison stimulus or vice versa. The step size was 5–10 dB for the first four reversals and was reduced to 1 dB afterwards. The 10-dB step size was used initially in combination with a wide separation of the starting levels for the upper and lower sequences in order to locate quickly the point of subjective equality. The level at which a reversal occurred was recorded and the reversals, excluding the first four, were averaged at the end of the run to estimate the 71% and 29% louder levels. The point of subjective equality was estimated as the average of the 71% and the 29% levels. Each run took about 5–10 min and each data point reported in this study was estimated from 3–6 runs. The averaged standard deviation across all runs in all subjects was about 1.5 dB.

II. EXPERIMENT 1: LOUDNESS OF COMPLEX TONES IN ACOUSTIC HEARING

The purpose of this experiment was to use the modern adaptive procedure to replicate Zwicker's original two-tone and three-tone experiments. This adaptive procedure obtained relatively unbiased measures in the subjective task of balancing loudness between two sounds of different qualities (Jesteadt, 1980; Schlauch and Wier, 1987; Zeng and Turner, 1991). The adaptive procedure not only allowed us to estimate the mean magnitude of both the spectral analysis and the temporal envelope effects on loudness, but also allowed us to estimate the variability of these effects, because the 21% and the 79% levels approximated the plus and minus one-standard deviation from the mean in a 2-AFC task (Green, 1964).

A. Stimuli

The stimuli as described in the original two- and three-tone complex experiments (Zwicker and Fastl, 1990) were replicated in this experiment. For the two-tone complex standard, the two tone frequencies, f_1 and f_2, were geometrically centered on 1 kHz and had a frequency separation ($f_2 - f_1$) of 2, 5, 10, 20, 50, 100, 200, 500, 1000, and 2000 Hz. Each of the two tones was presented at a fixed level of 60 dB SPL, independent of the frequency separation. An example of the two-tone complex waveform is shown on the bottom panel of Fig. 1. Also shown is the waveform of the 1000-Hz comparison stimulus.

For the three-tone experiment, two types of standard stimuli were constructed to produce the same spectrum but different temporal envelopes. First, an amplitude-modulated (AM) stimulus was constructed with a modulation depth of 0.5 and an overall level of 45 dB SPL. The frequency spectrum of the AM stimulus had three components, including the carrier (f_c) and two side bands ($f_c - f_m$ and $f_c + f_m$, with f_m being the modulation frequency). Second, a quasi-frequency-modulation (QFM) stimulus was constructed identical to the AM stimulus, except that the carrier was 90 deg out of phase relative to the two side bands. Examples of the AM and QFM stimuli are shown on the bottom panel of Fig. 2. Notice the greater (1.9 dB) temporal envelope fluctuation

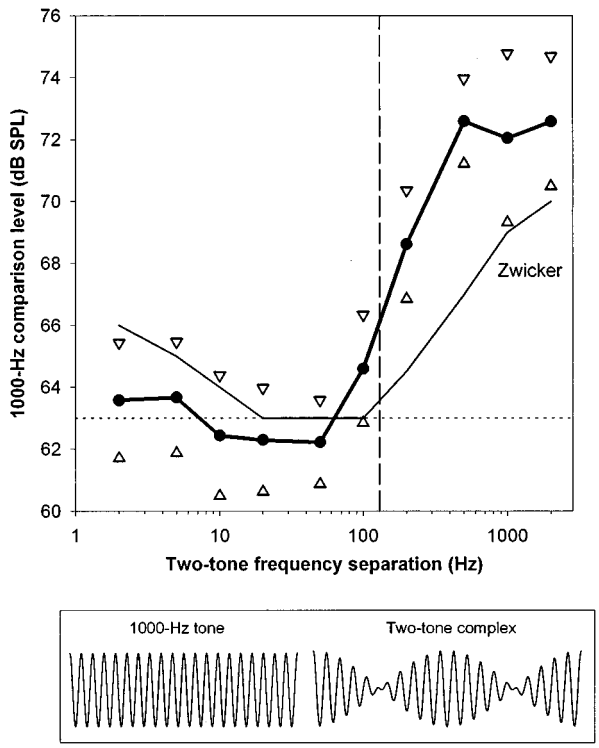

FIG. 1. Averaged loudness balance data in normal-hearing listeners between a 1000-Hz comparison tone (y axis) and a two-tone complex standard as a function of the frequency separation (x axis). The data were obtained from an adaptive procedure which tracked the 29% (regular triangles) and 71% (inverted triangles) louder response to the 1000-Hz tone on the psychometric function. The points of equal loudness (filled circles) were the average of the 29% and 71% response levels. Zwicker and Fastl's (1990) data are represented by the solid thin line. Examples of the waveforms are shown on the bottom panel.

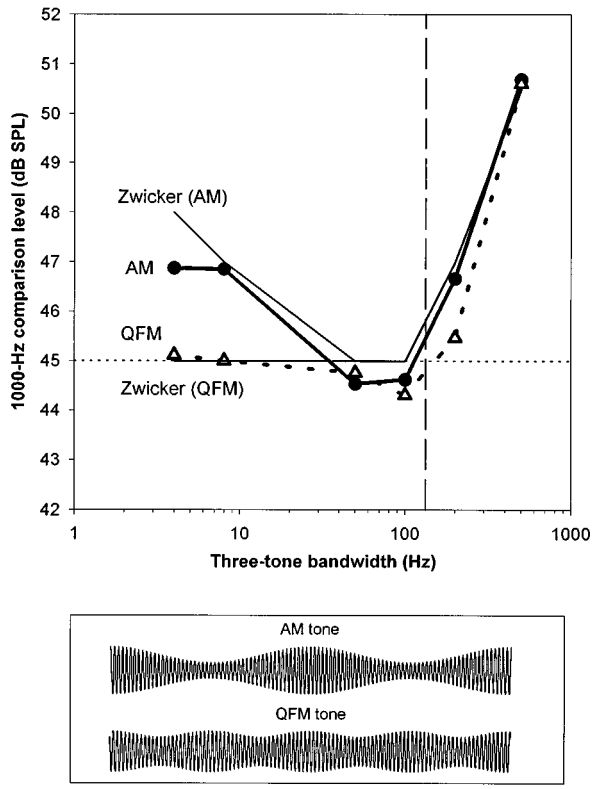

FIG. 2. Averaged loudness balance data in normal-hearing listeners between a 1000-Hz comparison tone (y axis) and a three-tone complex standard as a function of the total bandwidth (x axis). Filled circles represent data for the amplitude modulated (AM) stimuli and open triangles represent data for the quasi-frequency-modulated (QFM) stimuli. Zwicker and Fastl's (1990) data are represented by two solid thin lines. Examples of the waveforms are shown on the bottom panel.

in the AM stimulus than the QFM stimulus. The total bandwidth of the three-tone complex ($2\,f_m$) was selected at 4, 8, 50, 100, 200, and 500 Hz. The comparison stimulus was again the 1000-Hz tone.

B. Results and discussion

Since all four subjects had a similar pattern of results, only the averaged data for the two-tone loudness balance experiment are shown in Fig. 1. The filled circles represent the sound pressure level that was required for the 1000-Hz comparison (the y axis) to balance the loudness to the two-tone complex as a function of the frequency separation (the x axis). The upward and downward triangles are the averaged results from the lower and the upper sequences, which correspond to the 29% and 71% louder response to the 1000-Hz tone, and approximate the minus and plus one standard deviation from the mean. For comparison, the data from Zwicker and Fastl (1990, p. 189) are reproduced as the solid thin line. The horizontal dotted line represents the rms level (63 dB SPL) of the two-tone complex. The vertical broken line represents the critical band value of 133 at 1000 Hz (Glasberg and Moore, 1990).

Qualitatively, Zwicker and Fastl's results of the two-tone experiment are replicated by the present study using an adaptive procedure. When the frequency separation between the two tones is within the critical band, the rms intensity is added $[10\log(I+I)=3\text{ dB}+10\log I]$ to produce a 3-dB effect on the loudness sensation of the two-tone complex (the dotted line). When the frequency separation is markedly wider than the critical band, the loudness is added $[I^{0.3}+I^{0.3}=(10*I)^{0.3}]$ to produce a 10-dB effect ($10\log 10=10$ dB) on the loudness sensation of the two-tone complex. However, three deviations from Zwicker and Fastl's results are also apparent. First, for frequency separations below 10 Hz, Zwicker and Fastl observed a 6-dB effect which they attributed to the fact that loudness was determined by the peak level of the two-tone complex. The present data show only a 3.5-dB effect. Second, for frequency separations between 10 and 100 Hz, Zwicker and Fastl observed an intensity summation effect of precisely 3 dB. The present data show a 2-dB effect, suggesting that the two-tone complex actually sounds softer than the 1000-Hz tone at equal rms levels. Third, for frequency separations above the 133-Hz critical bandwidth, Zwicker and Fastl observed a shallow growth in loudness and a maximal 10-dB effect at the greatest frequency separation. In contrast, the present data show a much steeper loudness growth and a maximal 12-dB effect at the 1000-Hz separation.

The three-tone loudness balance data from the same four subjects are averaged and presented on the top panel of Fig. 2. The filled circles represent the loudness balance data for

the AM stimulus and the open triangles represent the data for the QFM stimulus. To avoid crowding the figure, the data from the upper and lower sequences are not plotted. The averaged difference between the 29% and 71% levels (approximating two standard deviations) across all bandwidth conditions is 3.0 dB for the AM stimulus and 2.4 dB for the QFM stimulus. Zwicker and Fastl's data (represented by two solid thin lines) and the present data are more consistent in this experiment than in the above two-tone experiment. Note the 2-dB difference between the AM and QFM stimuli for the 10-Hz or less total bandwidth conditions, which can be accounted for entirely by the 1.9-dB difference in the peak level between the AM and QFM stimuli. Note also the -0.5-dB difference between the present data and Zwicker and Fastl's data for three-tone bandwidths of 50 and 100 Hz. To summarize, although exp. 1 has replicated qualitatively Zwicker and Fastl's two- and three-tone experiments, severe deviations were also apparent, particularly in the two-tone experiment, where a much smaller peak-level effect, a less than perfect intensity summation effect, and a much steeper transition across the critical band were observed.

III. EXPERIMENT 2: LOUDNESS OF THE SAM NOISE IN ACOUSTIC HEARING

Experiment 1 demonstrated a temporal effect on loudness for slow temporal fluctuations. Because of the critical band analysis in normal hearing, a broadband stimulus is likely to be resolved into multiple bands, each of which will have less temporal fluctuations than the original overall fluctuations. For example, in the three-tone experiment, when the component frequency separation is much wider than the critical band, the ''brain'' perceives only the three individual components at the output of the periphery, with no difference in the temporal envelope between the AM and QFM stimuli. For this reason, and also because of the more dominant cross-critical-band loudness addition, the effect of the temporal envelope on loudness is difficult to observe in normal-hearing listeners except for the within-critical-band temporal fluctuations. In the present experiment, we used a sinusoidally amplitude-modulated (SAM) noise to circumvent the critical band analysis in the sense that the SAM noise effectively prevents spectral sidebands from being detected. Because of the continuous spectrum of the noise, amplitude modulation only increases the overall rms level, but does not change the spectral distribution of the noise, as in the case of the above three-tone experiment. We note, however, that the SAM noise does not prevent the reduction in modulation depth produced by the critical band analysis in normal-hearing listeners. A similar approach has been used to remove spectral cues in normal-hearing listeners in order to study exclusively the role of temporal cues in pitch perception (Burns and Viemeister, 1976) and in speech recognition (Schroeder, 1968; Van Tasell *et al.*, 1987; Rosen, 1992; Shannon *et al.*, 1995).

A. Stimuli

The comparison stimulus was white noise with a bandwidth of 20 Hz to 14 kHz. The standard stimulus was the SAM noise, which was constructed according to the formula: $S(t)=A[1+m\ \cos(2\pi f_m t)]n(t)$, where A was the carrier amplitude, m was the modulation index or depth, f_m was the modulation frequency, and $n(t)$ was the waveform of the white noise. Given fixed values of A and m, the rms level of the modulated noise was increased by a factor of $(1+m^2/2)^{0.5}$ relative to the unmodulated noise, whereas the peak amplitude was increased by a factor of $(1+m)$ (Viemeister, 1979). For example, in the case of a 100% amplitude modulation ($m=1.0$), the peak amplitude of the SAM noise was increased by 6 dB, whereas the rms level increased by only 1.8 dB. In other words, if the peak level determined loudness, a maximal difference of 4.2 dB in the rms level would be observed between the white noise and the SAM noise. The absolute thresholds for both the white noise and the SAM noise were at about 20 dB SPL. Three overall rms levels for the standard SAM noise were selected of 30, 45, and 70 dB SPL, corresponding to soft, comfortably loud, and loud sensations, respectively. Three modulation depths at 0.25, 0.5, and 1.0 were used for the 45-dB SPL condition, two depths of 0.5 and 1.0 for the 30 dB SPL, and only one depth of 1.0 for the 70 dB SPL. The modulation frequencies used were 4, 10, 20, 40, 60, 80, 100, 200, 400, and 1000 Hz.

B. Results and discussion

Figure 3 shows the averaged loudness balance data as a function of modulation frequency (x axis) and as a function of the overall level of the SAM noise (three panels). To have the same scale for the y axis in all three level conditions, the results are plotted as the difference in the rms level between the white noise and the SAM noise at the point of subjective equality in loudness. For example, if a 48-dB white noise is loudness balanced to a 45-dB SAM noise, then the y coordinate in Fig. 3 is 3 dB. The dotted line on each panel indicates no difference in the rms level between the SAM noise and the white noise at the point of subjective equality. The filled circles on all three panels represent the loudness balance data in the 100% modulation condition. The open squares represent the data obtained for 50% modulation (the upper and middle panels) and the open triangles represent the data for 25% modulation (the middle panel only). The measurement variability (the difference between the mean and the upper or the lower sequence) was consistent in all conditions and had an average value of 1.2 dB.

Three major points can be noted from Fig. 3. First, the temporal effect on the loudness is nonmonotonic as a function of the overall level. The greatest temporal effect on each panel changes from 2 dB at 30 dB SPL to 3.5 dB at 45 dB SPL to less than 1 dB at 70 dB SPL. The 3.5-dB magnitude of the temporal effect approaches the 4.2-dB maximum possible effect under the 100% modulation condition, if the peak level is presumed to determine the loudness. Second, the louder sensation produced by the amplitude-modulated noise, when present (30 and 45 dB SPL), occurs well beyond the 10-Hz range observed in Zwicker and Fastl's original two- and three-tone experiments. This louder sensation effect is relatively constant for modulation frequencies below 100, gradually rolls off from 100 to 1000, and eventually disappears at the 1000-Hz modulation frequency. This low-pass

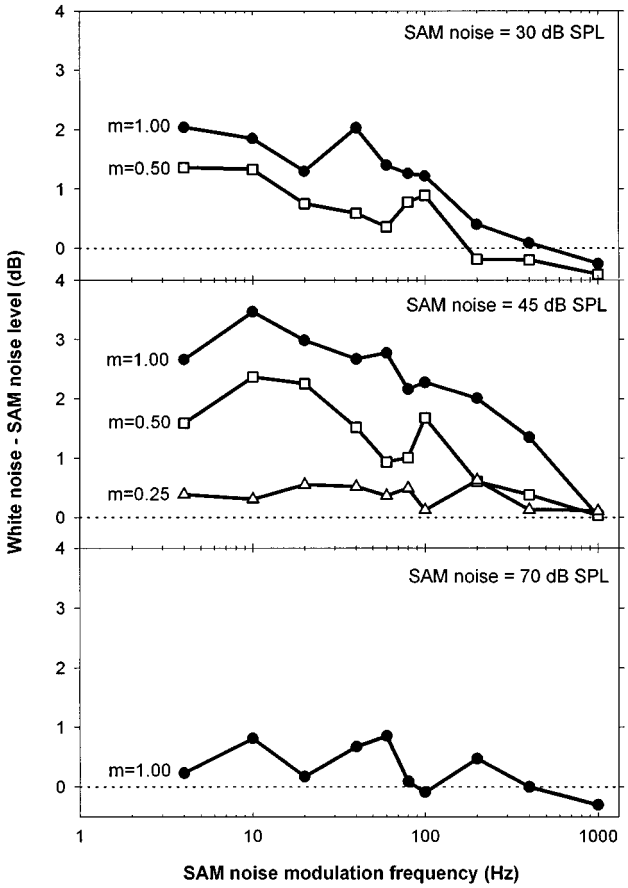

FIG. 3. Averaged loudness balance data in normal-hearing listeners represented by the rms level difference (y axis) between white noise and SAM noise as a function of modulation frequency (x axis), overall stimulus level (three panels), and modulation depth (100% = circles, 50% = squares, and 25% = triangles).

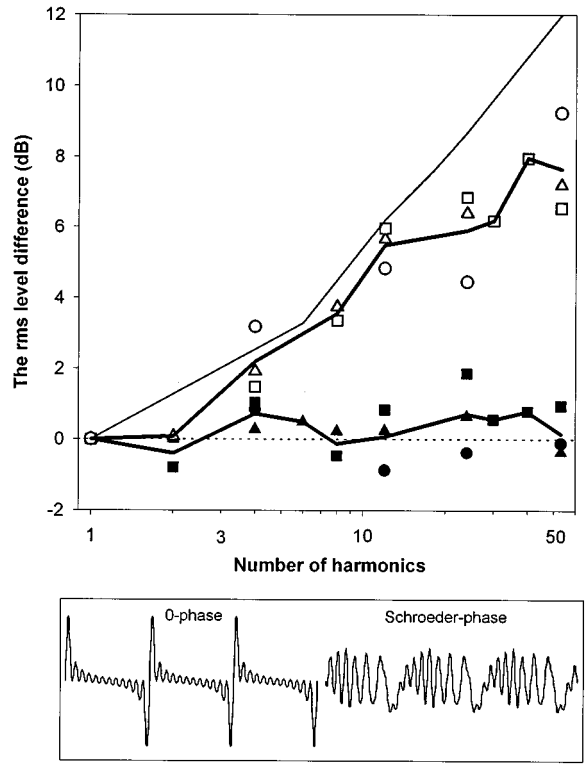

FIG. 4. Individual and averaged threshold and uncomfortable loudness data for three cochlear implant listeners. The data are represented by the rms level difference (y axis) between the zero- and Schroeder-phase stimuli as a function of number of harmonics (x axis). The individual data are represented by filled symbols for the threshold difference and open symbols for the uncomfortable level difference. The two thick lines represent the averaged data. The solid thin line represents the rms level difference between the zero- and Schroeder-phase stimuli when their peak level was equated. The dotted thin line indicates equal rms level for the two stimuli. Examples of waveforms (number of harmonics = 12) are shown in the bottom panel.

characteristic pattern of the temporal effect on loudness is qualitatively reminiscent of the temporal modulation transfer function measured by the modulation detection method (Viemeister, 1979). Third, the temporal effect on loudness was reduced significantly as the modulation depth decreased from 100% to 50% and to 25%. At 30 dB SPL for modulation frequencies between 4 and 200 Hz (the top panel), the 50% modulation produced an averaged temporal effect that was 0.7 dB lower than the 100% modulation ($t[df=7]=5.6$, $p<0.01$). At 45 dB SPL (the middle panel), the difference was 1.0 dB ($t[df=7]=3.7$, $p<0.01$) between the 100% and 50% modulations and was 0.43 dB ($t[df=7]=3.9$, $p<0.01$) between the 50% and 25% modulations.

IV. EXPERIMENT 3: LOUDNESS OF HARMONIC STIMULI IN ELECTRIC HEARING

To separate entirely the critical band analysis from the temporal envelope effects on loudness sensation of dynamic stimuli, we measured the loudness of two harmonic stimuli with the same magnitude spectrum but different phase spectra in three cochlear implant listeners. The different phase spectra used dramatically changed the temporal envelope or the peak level, but produced no change in the rms level. Direct stimulation of the auditory nerve bypassed the normal cochlear filtering processing, so that the overall temporal envelope of the harmonic stimuli was preserved in cochlear implant listeners.

A. Stimuli

Two types of phase manipulations were used in this experiment. One was called the zero-phase stimulus, in which all harmonics had the same 0° starting phase and the resulting waveform had the largest peak factor. The other was called the Schroeder-phase stimulus, in which the starting phase was manipulated to result in a waveform with the smallest peak factor (Schroeder, 1970). The Schroeder phase was defined as: $\theta_n=\pi n(n-1)/N$, where n is the harmonic number and N is the total number of harmonics. In this experiment, the fundamental frequency was 100 Hz and the number of harmonics ranged from 1 to 53. The bottom panel of Fig. 4 shows typical waveforms of the zero-phase and Schroeder-phase stimuli with 12 harmonics. Although the rms level was the same, the zero-phase stimulus had greater peak level than the Schroeder-phase stimulus.

TABLE I. Average threshold and uncomfortable loudness levels (ULL) in three cochlear implant listeners. The data were collected for both the zero-phase and Schroeder-phase (S-phase) stimuli as a function of the harmonic number, and are expressed as the rms level in μA.

Number of harmonics	Threshold		ULL	
	O-phase	S-phase	O-phase	S-phase
1	0.8	0.8	52.8	52.8
2	1.4	1.4	59.7	60.2
4	1.5	1.7	56.7	75.0
8	3.1	2.9	59.9	89.4
12	2.5	2.6	55.1	104.8
24	3.4	4.0	56.4	111.8
30	3.1	3.3	59.6	121.5
40	3.6	3.9	55.4	138.5
53	6.1	6.0	51.1	124.0

B. Procedure

In addition to the adaptive procedure in the loudness balance task, the method of limits was used to measure the threshold and uncomfortable loudness levels of the zero-phase and Schroeder-phase stimuli. In each trial of the threshold estimate, the level of the harmonic stimulus was adjusted by the listener to reach a criterion of just audible from a subthreshold level and was then adjusted to a criterion of just inaudible from a suprathreshold level. The just ''audible'' and ''inaudible'' levels were averaged to result in one estimate of the threshold. The uncomfortable level was estimated from only the ascending sequence in which the stimulus level was increased from soft to loud to uncomfortably loud. In a typical run, 10 to 12 such estimates were obtained. After determining the threshold and uncomfortable level for all harmonic conditions, the loudness balance task was performed from the threshold to the uncomfortable level between the zero- and Schroeder-phase stimuli for the 12-harmonic condition only.

C. Results and discussion

Table I shows both the zero-phase and the Schroeder-phase data for the threshold and uncomfortable loudness measures as a function of the number of harmonics. Notice the similar rms level between the zero- and Schroeder-phase stimuli for the threshold measure and the increasing difference between the two stimuli for the uncomfortable loudness measure. Another interesting point is that all zero-phase uncomfortable loudness levels have similar rms values. At present, we do not know the physiological mechanisms underlying these phenomena.

Figure 4 shows two theoretical predictions (Zeng and Shannon, 1995) and the difference in the threshold and uncomfortable levels between the zero- and Schroeder-phase stimuli. The solid thin line represents the ''equal-peak, equal-loudness'' prediction corresponding to the calculated rms level difference between the zero-phase and Schroeder-phase stimuli when their peak amplitudes are made equal. This rms level difference under the ''equal-peak'' condition increases monotonically from 0 to 6 and 12 dB as the number of harmonics is increased from 1 to 12 and 53, respectively. On the other hand, the dotted thin line represents the

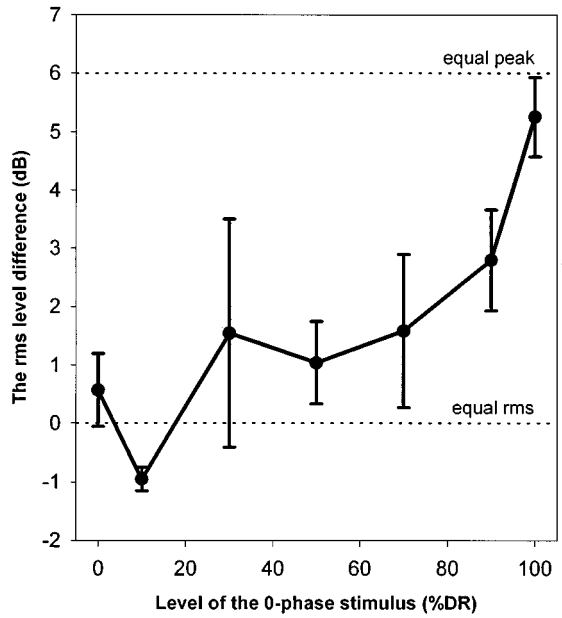

FIG. 5. Averaged loudness balance data between the zero-phase and Schroeder-phase stimuli for the 12-harmonic condition in three cochlear implant listeners. The balance data are represented by the rms level difference (y axis) between the zero- and Schroeder-phase stimuli as a function of the standard zero-phase stimulus level plotted as a percentage of dynamic range (x axis). The ''equal rms'' dotted thin line shows predictions of the ''equal-rms, equal-loudness'' hypothesis, while the ''equal peak'' dotted thin line shows predictions of the ''equal-peak, equal-loudness'' hypothesis.

''equal-rms, equal-loudness'' prediction, indicating no temporal envelope effect on loudness. The filled symbols represent the individual listener's threshold difference in the rms level between the zero-phase and Schroeder-phase stimuli and the solid thick line through these filled symbols represents the averaged results. Correspondingly, the open symbols represent the individual data for the uncomfortable level and the solid thick line through these open symbols represents the averaged results. It is obvious from Fig. 4 that the threshold data are consistent with the ''equal-rms, equal-loudness'' hypothesis, whereas the uncomfortable loudness data are significantly different from this ''equal-rms'' hypothesis. The trend of the uncomfortable loudness data is more closely predicted by the ''equal-peak, equal-loudness'' hypothesis, despite a greater deviation from the prediction as the number of harmonics was increased. In general, the present data are consistent with and extend the findings of Zeng and Shannon (1995). Both sets of data suggest that the temporal effect of loudness in electric hearing is level dependent, wherein the threshold is determined by the rms level and the uncomfortable loudness is determined more by the peak level of the temporal envelope.

To study the transition of this level dependence from the threshold to the uncomfortable loudness level, loudness balance data were obtained between the zero-phase and Schroeder-phase stimuli for the 12-harmonic complex only. Figure 5 shows the rms difference between the zero- and Schroeder-phase stimuli (y axis) as a function of the standard zero-phase level, which has been normalized according to each individual listener's dynamic range (x axis). The lower

dotted line represents the ''equal-rms, equal-loudness'' prediction, while the upper dotted line represents the ''equal-peak, equal-loudness'' prediction, corresponding to a maximal 6-dB difference in the rms level between the zero-phase and Schroeder-phase stimuli (see Fig. 4). Figure 5 shows that as the standard level increases from threshold (0% dynamic range) to uncomfortable loudness (100% dynamic range), the rms-level difference between the two stimuli increases monotonically from 0 to 5.3 dB. We have not attempted to model this loudness balance function, but this monotonic level dependence of the temporal effect on loudness in electric hearing is in great contrast to the nonmonotonic effect in acoustic hearing.

V. FINAL REMARKS

Traditional views on the temporal effect on loudness have been limited to slow changes in the dynamic stimuli; for example, the beats in a two-tone complex (Zwicker and Fastl, 1990). Recently, attention has been paid to the dynamics of loudness sensation as a result of much faster changes in two-tone noise complexes (Hellman, 1985), in the short-term temporal energy distribution (Stecker and Hafter, 1996), and in the temporal envelope of harmonic complexes (Carlyon and Datta, 1997) and amplitude-modulated tones (Moore *et al.*, 1997). The present work studied the loudness of dynamic stimuli by measuring the loudness of a sinusoidally amplitude-modulated (SAM) noise in normal-hearing listeners and the loudness of zero-phase versus Schroeder-phase stimuli in cochlear implant listeners. The results clearly demonstrate that fast temporal fluctuations up to 400 Hz can increase the loudness of dynamic stimuli. The increase in loudness has a low-pass characteristic qualitatively similar to that of the temporal modulation transfer function (Viemeister, 1979). This low-pass pattern suggests that the temporal effect on loudness is not determined by the temporal integration, as suggested by Zwicker and Fastl (1990), which has a time constant of about 100 ms (e.g., Zwislocki, 1960), but rather is determined by the absolute temporal resolution, which has a time constant 1–2 orders of magnitude lower (Viemeister, 1996).

The results of the present two- and three-tone experiments are generally, although not entirely, consistent with previous results (Zwicker and Fastl, 1990; Moore *et al.*, 1997). At low-modulation frequencies, Zwicker and Fastl observed a maximum 3-dB ''peak-level'' effect for both the two- and three-tone experiments. The present study observed only a 0.5-dB effect for the two-tone experiment and a 2-dB effect for the three-tone experiment. Moore *et al.* observed an essentially 0-dB effect for the 4-Hz modulation frequency (i.e., the rms level determines loudness). For modulation frequencies between 10 Hz and the critical bandwidth, Zwicker and Fastl found that the loudness was equal when the rms level was equal between the tone and the modulated stimuli; on the contrary, the present and Moore *et al.* studies found that the modulated stimuli were actually softer than the pure tone (about 1-dB effect) at equal rms levels. Since the present study used an adaptive procedure, Moore *et al.* used a ''bracket'' loudness balance procedure, and Zwicker and Fastl used an unreported procedure, the discrepancy among the three studies might be due to the procedural difference and/or other unknown factors.

Another important difference is noted between the two- or three-tone experiments and the SAM noise experiment: the experiments with sinusoidal carriers generally produced a relatively small and somewhat inconsistent temporal effect on loudness sensation, whereas the present experiment with a noise carrier demonstrated a consistently louder sensation for the amplitude-modulated noise than the unmodulated noise at equal rms levels, at least for low overall stimulus levels and over a large modulation frequency range. This difference in the temporal effect on loudness appears to be related to the nature of narrow-band versus wideband stimuli. Carlyon and Datta (1997) found that for five harmonics centered at 1100 Hz there was little difference in loudness between the positive and negative Schroeder-phase stimuli at equal rms levels (see also Preece and Wilson, 1988), but a significant phase effect occurred when the same five harmonics were presented simultaneously with a large number of ''off-frequency'' components. Carlyon and Datta attributed their findings to a fast-acting compression and interactions among these components at the basilar membrane level, which attenuated the ''peakier'' auditory filter output of the positive-phase stimulus more than the ''smoother'' output of the negative-phase stimulus (Kohlrausch and Sander, 1995). Both the present and Carlyon and Datta's studies suggest that wideband stimuli may be needed to illustrate reliably the temporal effect on loudness sensation.

The present study also demonstrates a level dependency of the temporal effect on the loudness of dynamic stimuli in both acoustic and electric hearing, but in opposite directions. In normal-hearing listeners, the modulated noise was louder than the unmodulated noise at low rms levels (30 and 45 dB SPL) but produced essentially equal loudness at a higher rms level of 70 dB SPL. This result suggests that loudness is determined more by the peak level at low sensation levels and more by the rms level at high sensation levels. Conversely, the result obtained from cochlear implant listeners stimulated with zero- and Schroeder-phase stimuli suggests that loudness is determined more by the rms level at low sensation levels and more by the peak level at high sensation levels. This apparent discrepancy between normal-hearing and implant listeners may be explained partially by the fast-acting basilar membrane mechanics which behaves more linearly at low levels and becomes compressive at medium and high levels (Ruggero, 1992; Oxenham and Plack, 1997), and partially by the expanding exponential loudness growth with direct stimulation of the auditory nerve (Zeng and Shannon, 1992, 1994). In normal-hearing listeners, the peak–trough difference in the modulated stimuli would be preserved in the low-level linear region and greatly reduced in the high-level compressive region, which can at least qualitatively produce the observed greater loudness effect at low levels. On the other hand, if we assume that the exponential loudness growth in electric stimulation is due to a ''fast-acting'' expansion, then the observed reversed level-dependent loudness effect in cochlear implant listeners could be explained. Thus the compression of the cochlea could explain the nor-

mal pattern of results, while the loss of that cochlear compression, coupled with the central expansion, could reverse the pattern in implants. Quantitatively, Moore *et al.* (1997) calculated the ''post-compression'' rms and peak-level difference between the modulated and unmodulated stimuli and found that a measure between the rms and peak level (after compression) could explain their results. We wonder whether a phenomenological model incorporating a peripheral compression and a central expansion (Zeng and Shannon, 1994; Zeng *et al.*, 1997) would produce such a measure that can better predict the present data and the Moore *et al.* (1997) data.

Although the physiological processes underlying the loudness sensation of dynamic stimuli remain unclear, the present results have direct clinical relevance to the design of the compression circuit in hearing aids and cochlear implants. If the purpose of the compression circuit is to restore normal loudness in hearing-impaired listeners, then the design should take into account the temporal fluctuation or the peak amplitude of dynamic stimuli such as speech sounds. For example, to restore normal loudness of speech sounds in cochlear implant listeners, the compression circuit should prescribe the gain for the soft consonants according to their short-term rms level while compressing the loud vowels according to their peak level. In other words, this compression circuit will need two different time constants that are dependent on the input sound level. At or near the threshold level, a long-term time constant (75–100 ms) is needed to take the temporal integration into account, whereas for loud sensations, a short or zero time constant is needed to compress the peak level of the speech sound. Investigations are under way to evaluate such a design strategy to see whether either improved speech recognition, better speech quality, or both can be achieved for hearing-impaired listeners.

ACKNOWLEDGMENTS

The authors would like to acknowledge John J. Galvin III, Lendra Friesen, and Qian-Jie Fu for their help in data collection and in programming experiments. The authors also thank Bert Schlauch, Bob Shannon, Alena Wilson, Chris Plack, and Brian Moore for their comments on this paper. This work was supported by the National Institutes of Health (NIDCD-DC02267).

Boothroyd, A., Erickson, F. N., and Medwetsky, L. (**1994**). ''The hearing aid input: a phonemic approach to assessing the spectral distribution of speech,'' Ear Hear. **15**, 432–442.

Burns, M. E., and Viemeister, N. F. (**1976**). ''Nonspectral pitch,'' J. Acoust. Soc. Am. **60**, 863–869.

Byrne, D., and Dillon, H. (**1986**). ''The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid,'' Ear Hear. **7**, 257–265.

Carlyon, R. P., and Datta, A. J. (**1997**). ''Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,'' J. Acoust. Soc. Am. **101**, 3636–3647.

Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.

Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, New York), p. 87.

Fletcher, H., and Munson, W. A. (**1933**). ''Loudness, its definition, measurement, and calculation,'' J. Acoust. Soc. Am. **5**, 82–108.

Fletcher, H., and Munson, W. A. (**1937**). ''Relation between loudness and masking,'' J. Acoust. Soc. Am. **9**, 1–10.

Florentine, M., and Zwicker, E. (**1979**). ''A model of loudness summation applied to noise-induced hearing loss,'' Hearing Res. **1**, 121–132.

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

Green, D. M. (**1964**). ''Psychoacoustics and detection theory,'' in *Signal Detection and Recognition by Human Observers*, edited by J. A. Swets (Wiley, New York), pp. 58–94.

Hellman, R. P. (**1985**). ''Perceived magnitude of two-tone-noise complexes: Loudness, annoyance, and noisiness,'' J. Acoust. Soc. Am. **77**, 1497–1504.

ISO-Rec. 532 (**1975**). ''Acoustics-method for calculating loudness level'' (International Organization for Standardization, Geneva).

Jesteadt, W. (**1980**). ''An adaptive procedure for subjective judgments,'' Percept. Psychophys. **28**, 85–88.

Kohlrausch, A., and Sanders, 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.

Launer, S. (**1995**). '' Loudness perception in listeners with sensorineural hearing impairment,'' Ph.D. thesis, Olderburg, Germany.

Launer, S., and Bachler, H. (**1996**). ''Implications of modeling sensorineural loss to hearing instrument design, selection, and fitting,'' abstract from ''Issues in advanced hearing aid research'' p. 18, Lake Arrowhead, California.

Moore, B. C. J., and Glasberg, B. R. (**1996**). ''A revision of Zwicker's loudness model,'' Acustica **82**, 335–345.

Moore, B. C. J., and Glasberg, B. R. (**1997**). ''A model of loudness perception applied to cochlear hearing loss,'' Aud. Neurosci. **3**, 289–311.

Moore, B. C. J., Wojtczak, M., and Vickers, D. A. (**1996**). ''Effect of loudness recruitment on the perception of amplitude modulation,'' J. Acoust. Soc. Am. **100**, 481–489.

Moore, B. C. J., Launer, S., Vickers, D. A., and Baer, T. (**1997**). ''Loudness of modulated sounds as a function of modulation rate, modulation depth, modulation waveform, and overall level,'' Proceedings of the 11th International Symposium on Hearing, Grantham, England (in press).

Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.

Preece, J. P., and Wilson, R. H. (**1988**). ''Detection, loudness, and discrimination of five-component tonal complexes differing on crest factor,'' J. Acoust. Soc. Am. **84**, 166–171.

Rosen, S. (**1992**). ''Temporal information in speech: acoustic, auditory and linguistic aspects,'' Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.

Ruggero, M. A. (**1992**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Current Opinion. Neurobiol. **2**, 449–456.

Schlauch, R. S., and Wier, C. C. (**1987**). ''A method for relating loudness matching and intensity discrimination data,'' J. Speech Hear. Res. **30**, 13–20.

Schroeder, M. R. (**1968**). ''Reference signal for signal quality studies,'' J. Acoust. Soc. Am. **44**, 1735–1736.

Schroeder, M. R. (**1970**). ''Synthesis of low-peak-factor signals and binary sequences with low autocorrelation,'' IEEE Trans. Inf. Theory **16**, 85–89.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.

Stecker, G. C., and Hafter, E. R. (**1996**). ''An effect of temporal energy distribution on loudness,'' J. Acoust. Soc. Am. **100**, 2627.

Stevens, S. S. (**1957**). ''On the psychophysical law,'' Psychol. Rev. **64**, 153–181.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., and White, R. L. (**1987**). ''Speech waveform envelope cues for consonant recognition,'' J. Acoust. Soc. Am. **82**, 1152–1161.

Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.

Viemeister, N. F. (**1996**). ''Auditory temporal integration: What is being accumulated?,'' Current Direct. Psychol. Sci. **5**, 28–32.

Vurek, L. S. *et al.* (**1981**). ''Opto-isolated stimulators used for electrically evoked BSER,'' Ann. Otol. Rhinol. Laryngol. **90**, Suppl. 82, 21–24.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''Better speech reception with cochlear implants,'' Nature (London) **352**, 236–238.

Zeng, F.-G., and Shannon, R. V. (**1992**). ''Loudness balance between electric and acoustic stimulation,'' Hearing Res. **60**, 231–235.

Zeng, F.-G., and Shannon, R. V. (**1994**). ''Loudness-coding mechanisms

inferred from electric stimulation of the human auditory system,'' Science **264**, 564–566.

Zeng, F.-G., and Shannon, R. V. (**1995**). ''Loudness of simple and complex stimuli in electric hearing,'' Ann. Otol. Rhinol. Laryngol. **104**, 235–238.

Zeng, F.-G., and Turner, C. W. (**1991**). ''Binaural loudness matches in unilaterally impaired listeners,'' Q. J. Exp. Psychol. **43**A, 565–583.

Zeng, F.-G., Shannon, R. V., and Hellman, W. S. (**1997**). ''Physiological processes underlying psychophysical laws,'' Proceedings of the 11th International Symposium on Hearing, Grantham, England (in press).

Zwislocki, J. (**1960**). ''Theory of temporal auditory summation,'' J. Acoust. Soc. Am. **32**, 1046–1060.

Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics—Facts and Models* (Springer-Verlag, Berlin).

Zwicker, E., and Scharf, B. (**1965**). ''A model of loudness summation,'' Psychol. Rev. **72**, 3–26.

Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). ''Critical band width in loudness summation,'' J. Acoust. Soc. Am. **29**, 548–557.