I. INTRODUCTION

The operation of the cochlea has principally been studied using tones and/or clicks (e.g., von Békésy, 1960; Robles et al., 1976, 1986; Sellick et al., 1982). This has provided an enormous amount of information about the complicated nonlinear cochlear system. However, since the cochlea is nonlinear (e.g., Rhode, 1971; Sellick et al., 1982; Robles et al., 1986), one cannot predict the response to a novel stimulus in a straightforward manner. Hence, until accurate models are developed, it is necessary to present each stimulus of interest in order to determine the cochlear response.

It is of interest to determine which response properties of the auditory nerve express cochlear mechanical filtering and which are due to later stages of signal processing by the inner hair cells and rectification that occurs at the hair cell–auditory nerve synapse. Commonly used stimuli for the study of auditory-nerve fiber (ANF) responses include amplitude-modulated (AM) signals, harmonic complexes, and clicks (e.g., Javel, 1980; Joris and Yin, 1992; Pfeiffer and Kim, 1972). These signals are important as they approximate signals encountered in the everyday environment and also communication signals such as speech. Nerve responses to AM signals indicate an enhancement of the response as measured by the modulation gain over a range of frequencies and intensities. That is, phase-locked firings of the ANF at the modulation frequency are greater than would be expected based on the modulation of the signal. The present study indicates that a portion of this enhancement is present in basilar-membrane (BM) vibration.

A common characteristic of all the stimuli used in this study is that they produce a pitch percept at the difference frequency between the individual components. The temporal discharge patterns of auditory-nerve fibers have been shown to contain information about the frequency content of a stimulus through both spatial and temporal patterns (Evans, 1978; Kiang et al., 1965; Rose et al., 1967). Early studies showed that the interspike intervals in response to two harmonically related tones corresponded to the fundamental frequency (Rose et al., 1969). Other studies found interspike intervals corresponding to perceived pitches when amplitude-modulated stimuli were presented (Evans, 1978; Javel, 1980; Rhode, 1995), two-tone complexes (Greenberg and Rhode, 1987), and synthetic speech sounds (Delgutte, 1980; Miller and Sachs, 1984; Palmer et al., 1986). In an extensive series of auditory nerve (AN) studies, Cariani and Delgutte (1996a, b) provided strong support for the hypothesis that the dominant interspike intervals are capable of explaining pitch perception for a variety of complex stimuli similar to those used in psychophysical experiments.

Horst et al. (1986, 1990) studied the representation of multicomponent (N=4 to 64) octave band stimuli centered at the characteristic frequency of an auditory-nerve fiber in cat. They observed that the edges of the stimulus spectrum were dominant in the response with increasing N and that the center component was reduced under these conditions. Here, we report that similar behavior is observed at the level of basilar-membrane mechanics for N as low as 3 and it appears that the frequency spacing of the stimulus components is the most important factor.

In order to explore a portion of the stimulus space used in auditory nerve studies, 1-, 3-, 5-, 6-, and 7-component stimuli were used to study basilar-membrane vibratory re-
II. METHODS

Methods are essentially those detailed in Cooper and Rhode (1992). Eight chinchilla cochleas were studied at approximately 3.5 mm from the basal end of the basilar membrane [characteristic frequency (CF) = 6.3–9 kHz]. All procedures were approved by the Animal Care and Use Committee of the University of Wisconsin.

Each animal was anesthetized with pentobarbital using a dose rate of 75 mg/kg. Additional doses were administered to maintain the animal in a deeply areflexive state. All anesthetics were administered intraperitoneally. A tracheotomy was performed to ensure an open airway and to place the animal on a respirator if necessary, though one was never used. After the ear was surgically removed, four screws were implanted in the skull and cemented in with dental cement in order to form a rigid base. A bolt was then cemented to the base to provide a stable fixation of the skull to a head holder with six degrees of freedom for the purpose of positioning the cochlea under the microscope.

The bulla was opened widely and a silver ball electrode was positioned so as to touch the edge of the round window for the purpose of recording the compound action potential (CAP) of the auditory nerve in response to short-duration tones (16 ms) for each animal. Since we recorded only in the high-frequency region of the cochlea, the stimulus was stepped in 2-kHz increments from 2 to 20 kHz. At each frequency a visual detection threshold for CAP was determined by viewing an average of 20 repetitions as the stimulus level was varied in 1-dB steps. If the thresholds were above our best threshold curve by more than 30 dB, no data were collected, as high CAP thresholds equated to little or no compression in the hook region (Sellick et al., 1982). CAPs were not typically recorded after mechanical measurements were initiated except to verify they had increased whenever mechanical sensitivity decreased.

The overlying cochlear bone in a region with CFs between 6 and 10 kHz was shaved down using a microchisel until the remaining tissue and/or bone debris could be removed with a pick fabricated out of a microelectrode. Gold-coated polystyrene beads 25 μm in diameter served as retroreflectors. They were placed in the perilymph and allowed to sink to the basilar membrane. They have a specific gravity of 1.05 that is near that of water (1.0), and therefore any loading of the basilar membrane by the bead should be minimal. A glass cover slip was placed over the cochlear opening with no hydromechanical seal. The cover glass served to avoid the problem of an unstable air–fluid interface.

An opening in the bony ear canal, immediately over the tympanic membrane, was made so that an acoustically calibrated probe tube to which a 4-in. Bruel & Kjaer condenser microphone is adjoined could be visualized as it was positioned parallel to the tympanic membrane within 1 mm of the tip of the malleus. The opening was sealed with a glass cover after a 45-μm bead was placed on the tympanic membrane at the tip of the malleus (or umbo). The bead was used as a retroreflector for the interferometer and allowed the measurement of the transfer function of the malleus. The sound source was a RadioShack super tweeter dynamic phone or a condenser microphone.

A. AM stimuli

Signals were synthesized and presented using a TDT system (Tucker-Davis Technologies ®) system. The formula for an AM signal is provided in Eq. (1)

\[ S(t) = (1 + m \cdot \sin(2\pi f_{\text{mod}})) \sin(2\pi f_{\text{car}}). \]  

This can be expanded as a sum of three sinusoids

\[ S(t) = m/2 \cdot \cos(2\pi(f_{\text{car}} - f_{\text{mod}})) + \sin(2\pi f_{\text{car}}) \]
\[ - m/2 \cdot \cos(2\pi(f_{\text{car}} + f_{\text{mod}})) \]
\[ = f_{\text{lsb}} + f_{\text{car}} + f_{\text{usb}}. \]

where \( f_{\text{lsb}} \) = lower sideband, \( f_{\text{car}} \) = carrier, and \( f_{\text{usb}} \) = upper sideband.

The carrier frequency, \( f_{\text{car}} \), was set equal to the characteristic frequency of the basilar membrane. The modulation frequency, \( f_{\text{mod}} \), was varied in 100-Hz steps from 100 to 1000 Hz, and was set to 1250, 1500, and 2000 Hz beyond 1 kHz. The modulation coefficient or depth, \( m \), was set to 0.25, 0.5, 1, or 2. When \( m = 2 \) the signal consists of three equal-amplitude tones and is also described as a 200% modulated signal. The stimulus level was varied from 0 to 90 dB SPL in 5-dB steps. Stimuli were 30 ms in duration, repeated 8 times at a rate of 10 per second.

The response was passed through a zero-phase high-pass filter to remove low-frequency noise (function filtfilt in MATLAB™, filter corner frequency \( f_{\text{mod}}/2 \)) before the envelope of the AM signal was recovered through the use of the Hilbert transform (Bennett, 1970). The envelope was then filtered using a fifth-order zero-phase Butterworth filter (corner frequency \( = 5f_{\text{mod}} \)). The minimum (min) and maximum (max) of the resulting envelope were determined by the use of a phase-locked loop technique that computes dc and the Fourier component (sine wave fit, abbreviated sinfit) at \( f_{\text{mod}} \). When \( m < 1 \), modulation was then computed by the relation

\[ \text{BM}_{\text{mod}} = (\max - \min)/\max \cdot \min = a/dc, \]  

where \( a = \) amplitude of the first Fourier component.

Gain for AM coding was defined as

\[ \text{gain} = 20\log_{10}(\text{BM}_{\text{mod}}/m). \]

A difficulty in determining modulation depth arises whenever \( m > 1 \) as this implies there will be a phase reversal in the envelope that can easily be missed upon visual inspection of the response. One solution used when \( m = 2 \) was to determine the ratio of the smaller peak/largest peak of the envelope whenever two distinct peaks per modulation period were present (e.g., Fig. 3). A curve was generated for this relation for \( 1 < m < 5 \) that was used to determine the modulation based on the measured ratio (e.g., Fig. 4). This method broke down for modulation frequencies > 400 Hz, at which point it became difficult to ascertain that there were two distinct peaks in the envelope. Modulation gain was also measured by determining the ratio between the two largest spec-
tral components, ‘‘two-component analysis,’’ in the response (usually the carrier and the lower sideband). For example, regardless of which component was largest, a ratio of 0.5 was considered an indication of \( m = 1 \) based on Eq. (1). This procedure was done as a check on the waveform analysis technique. The two-component situation arises because the cochlear filter often eliminates the upper sideband component, especially as \( f_{\text{mod}} \) increases. The two-component analysis generally resulted in modulation gain usually greater than the waveform technique and often resulted in a bandpass temporal modulation transfer function \( (t\text{MTF}=\text{gain as a function of modulation frequency}) \). A resolution of the difference between these approaches to estimate modulation gain was not obtained. It was decided that the waveform technique was most similar to methods previously employed to estimate modulation depth in auditory-nerve studies which rarely found bandpass tMTFs (e.g., Joris and Yin, 1992).

B. Multicomponent stimuli

A subset of these stimuli consisted of five equal-amplitude sine waves with the first, third, or the fifth component frequency set equal to the CF of the basilar membrane location under study. The separation of the harmonics was varied in steps of 100 Hz up to 1 kHz and set to 1250, 1500, and 2000 Hz above 1 kHz. \( F_{\text{mod}} \approx 2000 \text{ Hz} \) was not explored since ANFs do not show any AM coding for this condition. The stimulus level was varied from 0 to 90 dB SPL in 5-dB steps. Stimuli were 30 ms in duration, repeated 8 times at a rate of 10 per s. Analysis consisted of determining the amplitude of nine response components around CF using the sinfit procedure described above. The amplitude at the frequency difference was also determined but was insignificant or buried in the noise except for high levels and large frequency separations.

Also used were seven-component stimuli that consisted of seven equal-amplitude sine waves and six-component stimuli that consisted of the same complex with the center component deleted. The center component (i.e., the fourth) was always centered at CF. These stimuli had the same parameters as the five-component stimuli and were analyzed in the same manner except that 11 components were analyzed using the sinfit technique.

The amplitude of each component of all the stimuli used (AM, 5-, 6-, and 7-component stimuli) was compensated by the acoustic calibration. The starting phase of all the terms in Eq. (1b) as well as all the terms in the multicomponent stimuli was zero.

C. Single-tone basilar-membrane and middle-ear vibration measurements

Basilar membrane input–output (I/O) functions were determined using 30-ms tone bursts with 1-ms raised cosine rise and fall times and presented every 100 ms. The stimulus level covered a 100-dB SPL range in 5-dB SPL steps. A minimum of eight basilar-membrane and four middle-ear responses was averaged for each stimulus condition. Analysis consisted of Fourier decomposition of the steady-state portion of the averaged response at the stimulus frequency.

Measurements of the basilar-membrane I/O function at CF were made throughout the experiment to monitor the preparation’s stability. Vibration of the ossicles was measured at the tip of the manubrium (umbo) or at the incudo-stapedial joint or both locations either before or after (sometimes both before and after) the basilar-membrane measurements.

D. Recording system

Mechanical responses were measured using a custom-built, displacement-sensitive heterodyne laser interferometer (Cooper and Rhode, 1992). The laser was coupled to the preparation using a long working distance lens (Nikon SLWD 5X, NA 0.1). The laser was focused to an area of \( \sim 5\mu\text{m} \) diameter on the reflective beads. The interferometer was not sensitive enough to measure basilar-membrane vibration without the gold-coated beads. Instantaneous phase was measured using two single-cycle phasemeters that worked in quadrature. The phasemeter outputs were sampled at 250 kHz and the phase was unwrapped using software. Response amplitudes were corrected for the frequency response of the recording system. The noise floor was \(< 5 \text{ pm/\sqrt{Hz}} \).

III. RESULTS

A. AM response variation with modulation frequency

A portion of the basilar membrane response to an AM signal \( (m=1) \) is shown in Fig. 1. At \( f_{\text{mod}} = 200 \text{ Hz} \), the modulation of the vibration is nearly 100%, while with increasing \( f_{\text{mod}} \) the modulation depth of basilar-membrane motion decreased systematically to 0.58 for \( f_{\text{mod}} = 800 \text{ Hz} \). The relation between modulation depth and modulation frequency is portrayed in the tMTF [Fig. 2(A)]. The modulation depth of basilar-membrane vibration is \( >80\% \) up to 400 Hz with a reduction in the modulation depth that is a complicated function of frequency and stimulus level for \( f_{\text{mod}} \).
400 Hz. Similar behavior was observed in data from seven other cochleae. The gain-level curve for $f_{\text{mod}} = 500$ Hz increases at low levels, peaks near 25 dB SPL, and then decreases slightly.

For $f_{\text{mod}} = 100$ Hz or lower, the motion of the basilar membrane is overmodulated as evidenced by a secondary peak in the BM response that is not present in the stimulus waveform (indicated by the arrow in Fig. 3, top row, center panel, $m = 1$). Overmodulation also occurs when 200% modulated AM signals are presented ($m = 2$, row 2 of Fig. 3). The ratio of the secondary-to-primary peak in the stimulus waveform is 0.34 for a 200% modulated signal, while it is 0.64 at 60 dB SPL in the BM response (Fig. 3, second row, middle panel). This represents a gain of ~7 dB in the BM response. The amount of BM overmodulation (and thus gain) is highest at midrange levels (~60 dB SPL), as is apparent in the gain-level curves for four low-modulation frequencies in Fig. 4, all of which exhibit similar level-dependent behavior.

The tMTFs for 25% and 50% modulation are similar in form (Fig. 5) to each other but differ from those for higher modulation depths in their frequency and level dependence. For lower modulation stimuli, tMTFs vary in a systematic way from low pass to high pass with increasing stimulus level. Modulation gain is near 0 dB for low-modulation frequencies but increases to 10 dB at high $f_{\text{mod}}$ [Fig. 5(A)]. The gain-level functions increase monotonically when $f_{\text{mod}} > 700$ Hz (panels A and B). This pattern deviates from the nonmonotonic gain-level function in the AN where the gain decreases as level increases beyond 10–15 dB SPL above neural threshold (Joris and Yin, 1992). Of course, the AN will discharge at all modulation phases with equal probability as the stimulus level is raised sufficiently, resulting in no modulation of the firing. In contrast, the mechanical modulation gain increases for all $f_{\text{mod}} > 100$ Hz for levels >50 dB SPL.

The tMTFs for 200% modulation in Fig. 6, as well as results shown in Fig. 2, raise an interesting issue: how should the modulation gain be measured? Below 500 Hz the gain was recovered by measuring the height of the two peaks in a modulation period and determining what modulation is nec-

![FIG. 2. (A) Basilar-membrane temporal modulation transfer functions (tMTFs) when $f_{\text{carr}} = 8000$ Hz, $f_{\text{mod}}$ is varied from 100 to 2000 Hz, and $m = 1$. The symbols attached to the four curves correspond to stimulus level divided by 10 (e.g., symbol 3 refers to 30 dB SPL stimulus level. This labeling is used throughout). (B) Gain I/O functions at the three modulation frequencies indicated. Increasing line width corresponds to increasing modulation frequency. C07.]

![FIG. 3. Basilar-membrane response to AM when the carrier is set equal to CF = 8000 Hz with the modulation frequency = 100 Hz. The modulation depth is 100% for row 1 and 200 % for row 2. The level of the stimulus was varied from 40 to 90 dB as indicated above each column. The data were part of those collected over a 90-dB SPL range in 5-dB steps.]

![FIG. 4. Modulation gain-level functions for 100 to 400 Hz, $m = 2$. The curves were obtained by determining what the modulation depth corresponds to the ratio of primary-to-secondary peak ratio observed in the basilar-membrane response. Increasing line thickness corresponds to increasing modulation frequency. C07.]

that at high levels (usually >85 dB SPL) the basilar-membrane filter is broadened and becomes increasingly linear. A possible explanation may lie in an increase in two-tone suppression resulting from the use of equal-amplitude components. In contrast, at lower modulation depths of 25% and 50%, the gain did increase with level.

Similar tMTFs were obtained in seven other experiments. In all instances, the amplitude of the carrier in the basilar-membrane response did not vary significantly as a function of modulation frequency.

B. AM I/O functions

The I/O functions for a subset of the AM conditions studied show that when the carrier frequency is dominant (m < 2), the response to it remains the largest component (symbol -○-) at all levels except when \( f_{\text{mod}} > 600 \) Hz and the stimulus level is >70 dB SPL (e.g., Fig. 7, column 1, \( f_{\text{mod}} = 1000 \) Hz). This latter result is probably a consequence of broadening of BM filter and lowering of “CF” of the filter, whereby the lower sideband (1) becomes the largest component in the basilar-membrane response. The upper sideband (1) does not play a role in the response when \( f_{\text{mod}} > 600 \) Hz since it is usually smaller than the largest response component by 30 dB or more, having been reduced by the cochlear mechanical filter and suppressed by the lower frequency components.

When the modulation depth is 200%, the case of three equal components, the carrier is suppressed by the sidebands (column 2, Figs. 7 and 8, \( f_{\text{mod}} = 100 \) Hz). That is, the carrier amplitude is lower than the sideband amplitudes by about 1 dB. As \( f_{\text{mod}} \) increases, the lower sideband becomes dominant at ~50 dB SPL and at the same time the upper sideband is at least 20 dB smaller than the largest response component. When \( f_{\text{mod}} \) is increased to 2000 Hz, the carrier is dominant up to 55 dB SPL, at which level compression at \( f_{\text{carr}} \) and suppression by the lower sideband results in the lower sideband response component being larger than the carrier component.

Three distortion products above and three below the sideband frequencies were analyzed. For \( f_{\text{mod}} > 500 \) Hz, the only distortion component of significant level (≥30 dB re the largest component) was \( f_{\text{carr}} - 2 f_{\text{mod}} \). This component (symbol -□-) has been referred to as the cubic difference tone (CDT, Goldstein, 1967) and is \( = 2 f_{\text{1sb}} - f_{\text{carr}} = f_{\text{carr}} - 2 f_{\text{mod}} \) and is the most prominent distortion component perceived psychophysically. The CDT is largest for small \( f_{\text{mod}} \) and large modulation depths, a result likely due to the stimulus frequencies being located within the nonlinear region of the cochlear filter. When the three AM-response components are nearly the same amplitude, the CDT can be within 6–10 dB of the carrier amplitude [Fig. 7(4) □]. Distortion on the high-frequency side of the carrier can also be of similar amplitude under these conditions (panels A and D, curve for \( f_{\text{carr}} + 2 f_{\text{mod}} \), □).

The amplitudes of nine response components for each of the conditions in Fig. 7 are superimposed on the isolevel-single-tone curves in Fig. 8 (note the differing frequency scales). In each instance of complex signals, it is the level of the center component that is specified. The frequency loca-
tions of the stimulus components are indicated by the symbols along the top axis of the graph, with a triangle marking the location of the carrier frequency. For small $f_{\text{mod}}$, AM-response amplitudes are lower than for the single-tone response (indicated by the dashed lines with the tone levels marked by numbered symbols). This result is likely due to mutual suppression effects. Sideband amplitudes for 200% modulation are larger than for the carrier, as noted above (Fig. 8A). The large distortion components indicated by arrows at $f_{\text{carr}} + 4f_{\text{mod}}$, larger than the adjacent distortion components at $f_{\text{carr}} + 3f_{\text{mod}}$, a result that is opposite from the 100% modulation case (Fig. 8A). The large difference between the components at $f_{\text{carr}} + 3f_{\text{mod}}$ between the 100% and 200% cases suggests that there may be cancellation of distortion products arising from several locations. Magnitudes of the distortion products decrease as $f_{\text{mod}}$ increases for both 100% and 200% modulation. As the I/O functions for the lower sideband become linear, the lower sideband amplitude increases faster than the carrier’s amplitude, resulting in suppression of the carrier at levels $>50$ dB SPL. The lower component tracks the single-tone-isolevel curves, while the carrier and upper sideband components increasingly deviate from the isolevel curves with increasing level. At the lower levels, the carrier amplitude tracks the single-tone curves, but by 60 dB the lower sideband suppresses the carrier by 10–20 dB (panels E and F, 200%). The amplitude of the lower sideband tracks the single-tone data better as $f_{\text{mod}}$ is

FIG. 7. Column 1: I/O curves for an AM signal with $m=1$ and the indicated modulation frequencies. Column 2: Same as row 1 except that $m=2$. The individual components are indicated by numbers: $1 = f_{\text{carr}} + f_{\text{mod}}$, $2 = f_{\text{carr}} - 2f_{\text{mod}}$, etc. $\triangle = f_{\text{carr}}$, $\Diamond = f_{\text{carr}} + 2f_{\text{mod}}$, etc. Ten components were analyzed: the three frequencies of the AM signal, three distortion frequencies on either side of the AM signal, and the modulation frequency. The response at the modulation frequency was always in the system noise and hence is not shown. Components that were deemed in the noise were removed from the display. Any data for levels $<20$ dB SPL or amplitudes $<\pm 20$ dB re 1 nm are not shown. The modulation frequency is indicated in each panel. The modulation depth was 100% for the left column and 200% for the right column. Cf07.

FIG. 8. The I/O data of Fig. 5 presented along the frequency axis (solid lines) with the isolevel BM data superimposed (dash-dot lines). Column 1: $m=1$. Column 2: $m=2$. Data below $-25$ dB re 1 nm were omitted. The use of the symbols is the same as in Fig. 7. The locations of the stimulus frequencies are indicated by solid circle symbols at the top of each panel and a diamond symbol indicates the CF frequency. This same method of indicating the stimulus components is used in all subsequent similar illustrations.
increased (panel F) since it is positioned in the linear portion of the basilar-membrane response.

C. Envelope distortion in AM signals

Visual examination of the envelope of the waveforms, especially at high levels, shows the presence of distortion by virtue of the deviation of the actual response envelope for the ideal envelope indicated by the thick solid line in Fig. 1. Fourier analysis (results not shown) performed on the envelopes of both waveforms reveals the presence of a second-order harmonic whose level is about 11 dB below that of the fundamental frequency. The amount of distortion increases with level until it reaches a maximum value, then decreases for higher levels of stimulation, a nonmonotonic relation. Similar distortions were found in the waveforms used in the results shown in Fig. 5.

There is an asymmetry between the rise and fall of the AM envelope that may be related to the envelope distortion. This asymmetry was best observed by using the Hilbert transform to calculate the time derivative of phase of the analytic signal representation of the waveform (Bennett, 1970), which is defined as the instantaneous frequency (IF) of the AM signal. For low $f_{\text{mod}}$, the IF is near the carrier frequency for most of the modulation period, except in the low-amplitude portion where for low levels the IF moves in the direction of the upper sideband frequency [Figs. 9(A), (D)]. For higher levels the IF drops to a frequency lower than the lower sideband frequency [Figs. 9(B), (E)]. This occurs for both 100% and 200% modulation. As $f_{\text{mod}}$ is increased, there is an increasing asymmetry in IF as a function of envelope phase; it appears to take longer times for IF to attain its final frequency [see the arrows in Fig. 9(F) fast decrease and slow increase in IF]. Further, for $f_{\text{mod}}<1000$ Hz the maximum frequency attained is never close to the carrier frequency with IF lower for 200% than 100% modulation. Zero-crossing analysis results in the same outcome, only with fewer sample points per period. A direct analysis of waveforms obtained by passing AM signals through a gammatone filter led to similar results without the asymmetry in the IF at the envelope minimum. Passing the AM signal through a filter that consisted of the isolevel amplitude and phase curves for the basilar membrane showed that the change in IF is much smaller than observed on the basilar membrane and was symmetric about the envelope minimum [dashed line, Fig. 9(E)]. IF also decreased with increasing levels for all $f_{\text{mod}}$.

IF asymmetry versus envelope phase is reminiscent of the “glide” that has been described for both AN and BM responses to clicks (e.g., de Boer and Nuttall, 1997; Carney et al., 1999). In the present instance, it appears that cochlear nonlinearity is necessary for the asymmetry and may be related to asymmetry in the envelope of the AM response.

D. Multicomponent stimuli

In a linear system, the presentation of five equal-amplitude sinusoids would result in a sample of the mechanical transfer function at the five stimulus frequencies. The component frequency separation or $f_{\text{diff}}$ was varied in 100-Hz steps between 100 and 1000 Hz, while above 1000 Hz, frequency separations of 1250, 1500, and 2000 Hz were typically used. These stimuli are similar to the N-component octave band stimuli used by Horst et al. (1990) in studies of the auditory nerve. The frequency of the first, third, or the fifth component was set equal to the CF of the location of the basilar membrane under study since natural stimuli are not normally centered on a particular auditory unit’s CF. The analysis was the same as employed for AM stimuli.

The response to varying frequency separation between components is illustrated in Figs. 10(A)–(C) along with waveforms (D)–(E). There is only a small component amplitude variation for $f_{\text{diff}}=100$ Hz. However, as in the AM case, the central component is reduced relative to the edge frequency components. In fact, the three central components are all reduced relative to the edge frequency components for levels between 20 and 80 dB SPL. A similar effect occurs at $f_{\text{diff}}=200$ Hz (results not shown), while at $f_{\text{diff}}=300$ Hz there is a reduction of the central components only at 70 dB SPL (panel B). As the frequency separation is increased, the high-frequency components are attenuated by the cochlear mechanical filter, while the low-frequency components located in the “tail” region of the filter, grow at a linear rate, hence faster than the center component. Initially, the CF component grows at the same rate as the single-tone response did up to 40 dB SPL, but its growth slows beyond that as a result of suppression by the lower-frequency components. At 70 dB SPL, the CF component is 20 dB below the single-tone
amplitude, while all the lower-frequency components are at the single-tone level. The CF component is also ~20 dB lower than the 6-kHz component. When the frequency separation is >1000 Hz (not shown), high-frequency components extend beyond the compressive region of the basilar-membrane filter and are not suppressed by lower-frequency components. Figures 10(D)–(F) shows a portion of the BM response waveforms to the five-component stimuli (at 80 dB SPL) with $f_{\text{diff}}$ equal to that in the panels immediately to the left.

An effect of varying which component frequency is set to CF when $f_{\text{diff}} = 1000$ Hz is illustrated in Fig. 11. When the first component = CF, several of the components are beyond the cochlear filter and hence don’t appear in the response. The component at CF is dominant and its amplitude closely tracks the single-tone amplitude [Fig. 11(A)]. Higher-frequency components that are within the passband of the cochlear filter are suppressed in this instance. When the third component = CF, the lower-frequency components respond to level increases with linear growth and suppress the CF and higher components [Fig. 11(B)]. At 70 dB SPL the CF component is suppressed over 20 dB relative to the single-tone response. With the fifth component at CF, the CF component is suppressed ~10 dB at 70 dB SPL. This is about 10 dB less suppression than in the previous case, suggesting that components on both sides of CF are required to achieve maximum suppression of the CF component. At low levels, all five components track the single-tone response (panel C).

The I/O functions corresponding to the data in Fig. 10 show the growth of the five individual components along with two upper- and two lower-frequency distortion components (Fig. 12). The distortion components nearest in frequency to the stimulus frequencies [3, 8] where CF = 8000

![FIG. 10. Frequency analysis of the response of five-component stimuli for three frequency differences ($f_{\text{diff}} = 100, 300$, and 1000 Hz) is illustrated. The third component frequency was set to CF. Dotted lines correspond to the single-tone-isolevel amplitude transfer function for this basilar-membrane location. Ct07.](image)

![FIG. 11. The first, third, and fifth components of the five-component stimulus were set to CF (A, B, and C, respectively). Data shown for the case where $f_{\text{diff}} = 100$ Hz. Single-tone-isolevel functions indicated with dash-dot lines. Level in SPL indicated by a numbered symbol (level = 10^* symbol number). Ct07.](image)
Hz, Figs. 12(A), (B)] are the largest distortion products and can equal or exceed the amplitude of the stimulus components [best seen in Fig. 13(B), 3]. The distortion tone, likely the cubic difference tone generated by the two lowest-frequency components, is normally the largest distortion component and attains maximum amplitude for small frequency differences. For \( f_{\text{diff}} > 1000 \) Hz, only the distortion component labeled 3 is present in the response at a significant amplitude and only at higher levels (panel C). All other distortion components are relatively small.

The response to seven-tone stimuli was similar to that of five-component stimuli (Fig. 13). For small \( f_{\text{diff}} \) there is a reduction in the amplitude of the center components relative to the stimulus edge frequencies of 5900 and 6600 Hz for levels \( > 20 \) dB SPL [Fig. 13(B)]. There is substantial mutual suppression of each component that decreases as \( f_{\text{diff}} \) increases (not shown). When the center component is omitted the response at that center frequency is negligible at low stimulus levels and reaches parity with the other components at midlevels [ ○ in Fig. 13(C) and the diamond symbol along the top of the panel in Fig. 13(D)].

To verify that level-dependent behavior observed in Fig. 13(D) wasn’t present in the stimulus, i.e., a stimulus artifact, responses to multicomponent stimuli were measured in the vibration of the umbo. The umbo I/O curves were linear [Fig. 14(A)] and the amplitude distribution shows that the missing center component remains so at all stimulus levels [Fig. 14(B)]. Other tests of either AM stimuli or the five-component stimuli verified that component interaction and suppression is only present in the cochlea.

**FIG. 12.** I/O functions corresponding to the five-component data shown in Fig. 10. (A) \( f_{\text{diff}} = 100 \) Hz. The five I/O curves are largely superimposed. (B) \( f_{\text{diff}} = 300 \) Hz. (C) \( f_{\text{diff}} = 1000 \) Hz. Third component frequency (○)=CF = 8000 Hz. Symbols are as indicated in Fig. 7. C07.

**FIG. 13.** Example of the basilar-membrane response to the seven-component stimulus where CF = 6300 Hz. (A) I/O functions for \( f_{\text{diff}} = 100 \) Hz for 11 response components: 5700 to 6800 Hz in 100-Hz steps. Stimulus frequencies: 5900 to 6600 Hz. (B) Isolevel responses for the complex versus single-tone curves. Example of the basilar-membrane response to the seven-component stimulus with the center component stimulus deleted (C and D). (C) I/O functions for \( f_{\text{diff}} = 100 \) Hz for nine response components with the center component stimulus deleted. (D) Isolevel responses for the complex versus single-tone curves with the center component stimulus deleted. Symbols are as indicated in Fig. 7. C26.
E. Phase versus level for multicomponent stimuli

There are well-known phase versus level effects in basilar-membrane mechanics (e.g., Geisler and Rhode, 1982) that are similar to those initially demonstrated in the auditory nerve; phase lags with increasing level for frequencies below CF and phase leads for frequencies above CF (Anderson et al., 1971). However, for the mechanical results this is somewhat of a simplification since above 70–80 dB SPL there is a phase lag with increasing level for all frequencies (Ruggero et al., 1997; Rhode and Recio, 2000).

The relative phase versus level behavior for five-component stimuli at four difference frequencies largely recapitulates the behavior of the single-tone stimuli (Fig. 15). The smallest level (<70 dB SPL)-dependent phase changes occur for small $f_{\text{diff}}$ (Fig. 15(A), which is similar to the neural results where there is relatively little phase change at CF. By $f_{\text{diff}}=600$ Hz, there are phase changes as large as 180° between 60 and 80 dB SPL (panel C). For larger $f_{\text{diff}}$ there is almost no phase change with level for components lower than CF (Fig. 15(D), 1 and 2). The lower-frequency components have smaller phase changes that likely are due to their being in the linear part of the basilar-membrane response. The phase-level pattern is fairly consistent regardless of the difference frequency. There is a phase lead up to ~60 dB SPL and a phase lag from there to 90 dB SPL that approaches $\pi/2$ radians (90°).

By superimposing the single-tone phase relations on the multitone phase data, it is apparent that there is an effect of
multiple tones simultaneously being presented on the phase (Fig. 16, multitone data—solid lines; single-tone data—dashed lines). The principal effect of multiple tones is a faster rate of phase lag accumulation with increasing level except for the lowest-frequency component (7000 Hz, 2). For the two components above CF, there is nearly a π radians (180°) phase lag that occurs for a 20-dB level change (60–80 dB SPL). In addition, the level at which the phase lag begins appears to be about 10 dB lower than for the single-tone data.

The principal effect of inducing separate amounts of phase change in the individual components is to alter the waveform; the individual components are no longer in sine wave phase, and hence the peak of the waveform would be reduced. This effect likely contributes to the reduction in modulation gain for AM stimuli with increasing modulation frequency.

IV. DISCUSSION

A. AM

There has been relatively little in the literature as to the response of the basilar membrane to complex stimuli (cf. Schroeder masksers, Recio, and Rhode, 2000). However, the response of the auditory nerve to AM stimuli has been well documented (Javel, 1980; Smith and Brachman, 1980; Joris and Yin, 1992). Interpreting the relation between neural and mechanical behavior is complicated, since the inner hair cell’s membrane properties and its synapse with the auditory nerve are interposed and limit the bandwidth of the transduction of mechanical motion to the neural domain (Kidd and Weiss, 1990). Rectification at the synapse primarily introduces even-order nonlinearities (Kiang et al., 1965; Pfeiffer and Kim, 1972; Horst et al., 1986) into the process that adds to the distortion resulting from cochlear mechanics that is primarily third order (cubic), although recent evidence indicates the presence of second-order harmonics in BM responses to single tones (Cooper, 1998).

The high-frequency component makes the smallest contribution to the basilar-membrane response because its response is reduced by the cochlear filter as the modulation frequency and stimulus level increase. The spectrum of the BM response is then reminiscent of a single sideband-modulated signal (Schwartz, 1959). It is noteworthy that there is no AM coding theory that applies to cochlear mechanical response since the sidebands are not of equal amplitude. The remaining two components vary in a predictable manner based on the properties of the cochlear filter and nonlinear compression (a simulation was also run using a gammatone filter with the stimulus set that yielded similar results). As the level is increased, the lowest-frequency component is the least compressed by the cochlear nonlinearity; therefore, it increases faster with increasing level than the other components. Therefore, for modulation depths less than 200%, modulation gain increases as the lower component amplitude approaches the amplitude of the carrier component. The maximum gain occurs when the two are equal in amplitude (assumes that the upper sideband component is significantly smaller than the carrier). The exception to the latter situation occurs for small $f_{mod}$, when all three components are relatively unfiltered and are similarly affected by the cochlear nonlinearity. In this latter instance there is edge enhancement of the spectrum; that is, both sideband components are larger than the carrier. This is reminiscent of Mach bands, where the edge of the band is enhanced relative to the center component (Carterette et al., 1969) and the AN responses to the multicomponent stimuli of Horst et al. (1990). Edge enhancement also implies that the modulation gain is $>0$ dB. Edge component enhancement was also found in response to five- and seven-component stimuli.

Modulation gain in the AN-neural representation of AM signals varies with level but in a manner that is different than the mechanical gain relation to level. The neural gain reaches a maximum within 10–20 dB of the neural threshold and then decreases with increasing levels (Khanna and Teich, 1989; Joris and Yin, 1992). Mechanical modulation gain-level curves increase at low levels but do not show the same decrease with increasing levels as the neural curves [Fig. 2(B), $f_{mod}=500$ Hz]. In fact, for modulation depths $<1$, gain continues to increase with level for $f_{mod}=800$ Hz (Fig. 5). It appears that the mechanical tMTFs can be largely explained by the individual level curves for the three components. That is, the upper sideband’s influence on the basilar-membrane vibration decreases with increasing $f_{mod}$ so that the lower sideband and carrier are left to interact to produce the modulation. At low levels the carrier is usually larger (Fig. 7). Maximum gain occurs when the two lower-frequency components are equal in amplitude, which is analogous to 200% modulation. Therefore, if $m<2$, the gain is $>0$ dB. The lower sideband grows faster with level than the carrier ($=CF$), resulting in it becoming the largest component at high levels and also suppressing the carrier (and the higher-frequency component) at high levels. While there are instances of mechanical modulation gain-level curves with a nonmonotonic shape similar to the neural curves, there are also differences (Figs. 2 and 4). Their peak values occur between 60–70 dB SPL, substantially higher than those for the neural curves and for higher modulation depths than were typically used for the AN studies. A related study of neural coding of 200% modulated signals in primarylike units in the cochlear nucleus showed the peak in the modulation gain occurred between 25–40 dB SPL (Rhode, 1994). While there is a species difference between the neural and mechanical studies, it is likely that the primary determinant of any decrease in neural modulation gain with level results from auditory-nerve fibers being driven to saturation throughout the entire modulation period.

Comparison of a neural tMTFs for a cat-AN fiber (CF = 6.9 kHz, at the best modulation level for the unit, Joris and Yin, 1992) with several mechanical tMTFs (CF=8 kHz) shows they are similar (Fig. 17). Mechanical tMTFs are low pass at low levels (e.g., Fig. 2), as are neural tMTFs. The neural I/O curves are nonmonotonic, peaking between 20 and 40 dB SPL and decreasing rapidly with increasing levels. The decrease is due to the nerve discharging throughout the modulation period and the resulting decrease in firing synchronized to $f_{mod}$. Mechanical modulation gain (modulation depth=1) varies relatively little for low $f_{mod}$.
The gains of mechanical tMTFs bracket the neural tMTFs. The neural curve and mechanical tMTFs are for \( m = 0.5 \) (dashed line) and \( m = 1 \) (solid line) from the same experiment as shown in Figs. 2(A) and 5(B) but at levels of 30 and 65 dB SPL (indicated by a thin and thick line, respectively, Ct07).

while for higher \( f_{\text{mod}} \)'s they have a nonmonotonic shape similar to the neural curves but with the maximum value at higher levels.

Corner frequencies of the neural tMTFs decrease with decreasing CF and can vary nearly an octave with a mean value \( \sim 900 \) Hz at 10 kHz. For a 6.9-kHz fiber, the cutoff frequency was 400–500 Hz and around 1 kHz for higher CFs (Joris and Yin, 1992). In comparing BM tMTFs with neural tMTFs, the former have higher cutoff frequencies, suggesting that the neural frequency limit is a result of the filtering properties of hair cells (Fig. 17).

Joris and Yin found that the modulation of neural discharges increased as modulation depth increased while the average rate and phase remained nearly constant. In contrast, the BM response phase varies with level, both increasing and decreasing in various experiments. Little phase change in AN firing occurs with level changes for stimulus frequencies=CF; this suggests that the carrier frequency may not have been set to the exact CF of the basilar membrane (Fig. 16). Neural phase at CF is known to vary little but does so off CF (Anderson et al., 1971). For modulation depths <1, modulation gain was generally >1. For modulation depths <1 and \( f_{\text{mod}} < 800 \) Hz, mechanical gain ranged between \(-2 \) and \( +2 \) dB, which is near that found for the AN response. However, for 100% modulation the gain was never >1, implying that there are other factors involved in determining the neural tMTF. An exception occurs for low \( f_{\text{mod}} \) (=100 Hz or less) where the occurrence of a second peak in the envelope implies modulation >100% (Fig. 3, top row).

Mechanical modulation gain decreases 2–8 dB for \( f_{\text{mod}} > 1000 \) Hz depending on stimulus level and modulation depth. Regardless of modulation depth and stimulus level, the mechanical tMTFs are relatively flat in comparison to the neural tMTFs. The gains of mechanical tMTFs bracket the neural gain if one compares the \( m = 0.5 \) functions at 20 and 50 dB SPL. However, with the same modulation used for the neural curve and mechanical tMTFs (solid lines) in Fig. 17, the mechanical gain is less than the neural. Therefore, some of the neural gain is likely due to hair cell and hair cell synapse properties. However, neural modulation gains also vary considerably and overlap the mechanical gains.

Some neural tMTFs can have a high-pass shape, as can mechanical tMTFs for low modulation depths. However, the corner frequency of neural tMTFs does not increase for CFs >10 kHz, which is likely due to low-pass filtering in the hair cells plus other mechanisms (Kidd and Weiss, 1990). The effect of adaptation may contribute to producing modulation gains >1 since faster stimuli will result in a larger transient response than that due to stimuli with slower rise times. These features suggest that neural effects modify the mechanical tMTFs and are necessary to fully account for the AN response to AM stimuli.

The amount of distortion in the envelope of responses to AM stimuli (\( m = 1 \)) was unexpected (Figs. 1 and 3). It is not a consequence of distortion in the stimuli, as this was verified. Perhaps more surprising is that Fourier analysis of the envelope showed second-order harmonic distortion (\( \sim 10 \) dB below fundamental) is larger than the distortion measured, in the same animal, in responses to single tones (\( \sim 20 \) dB below fundamental). Envelope distortion occurs mostly when two of the spectral components of the AM stimulus (usually the carrier and the lower sideband) dominate the BM response. This is similar to the spectrum of a single-sideband AM stimulus. Computer simulations showed that the envelope of a single-sideband AM stimulus (even in the case of a linear system) is distorted. Coherent demodulation of the waveform (i.e., multiplying by a sinusoidal whose frequency equals the frequency of the modulation frequency) yields a waveform with less distortion. It is, however, unlikely that the auditory system performs this type of analysis.

### B. Multicomponent stimuli

Horst et al. (1990) identified the short-term spectral content of the acoustic signal and instantaneous amplitude as being most important to acoustic coding. At low stimulus levels the response of the auditory nerve directly reflects a linear encoding of the stimulus. As the level is increased, nonlinear effects occur that can alter a component, such as decreasing or even eliminating the center frequency component of a harmonic complex. When a center component of the harmonic complex was left out, it was found restored in the output of the AN by the generation of distortion components due to the cochlear nonlinearity.

Present results indicate that basilar-membrane mechanics underlie much of the observed response of the AN to complex stimuli. The suppression of the center component for small \( f_{\text{diff}} \) is especially prominent. This effect is independent of the signal bandwidth and/or the number of stimulus components since it was observed for AM, five- and seven-component stimuli. The principal factor appears to be small \( f_{\text{diff}} < 300 \) Hz, at least in the frequency region that was studied here. Generation of distortion components is also greatest at low \( f_{\text{diff}} \) and decreases rapidly with increasing \( f_{\text{diff}} \).

Distortion component generation likely explains the restoration of a missing center component that could be a result.

FIG. 17. Superposition of mechanical and neural tMTFs. The neural curve (dashed line) is adapted from Joris and Yin (1992) that they describe as taken at the best modulation level (usually 20–40 dB SPL). Mechanical tMTFs are for \( m = 0.5 \) (dash-dot line) and \( m = 1 \) (solid line) from the same experiment as shown in Figs. 2(A) and 5(B) but at levels of 30 and 65 dB SPL (indicated by a thin and thick line, respectively, Ct07).
of the superposition of distortion products generated by both lower- and higher-frequency components. The missing component can be generated by combination two consecutive stimulus components interacting; that is, if the missing component is the center component, then $CF = 2f_1 - 2f_2$, $2f_1 - 2f_2$, $3f_1 - 3f_2$, and $2f_1 - 3f_2$ for the seven-component stimulus (the symbols correspond to the stimulus components as used in Figure 13, e.g., the first distortion component $= 2f_1 - f_2 = 2f_1 - 2f_2$, when $f_1$ and $f_2$ are the first two stimulus components immediately above CF). A distortion component at CF could also be generated by the components below CF. In addition, one could consider whether the distortion products that are generated can themselves interact with all other components present in the cochlea. Apparently restoring the missing component is a result of distortion product generation due to the nonlinear nature of the cochlea.

The interactions of components in the cochlea are dependent on: (1) the compressive nonlinearity of the cochlea and (2) filtering in the cochlea. The first was demonstrated by showing that the center component is not restored when the experiment was performed on the umbo. Filtering of the stimulus waveform by a filter constructed from the iso-intensity functions shown in Fig. 13 also produced a waveform without any spectral component at the frequency of the missing component. This indicates that the occurrence of a response at the frequency of the missing component is due to a distortion in the cochlea, as it could not be generated using a linear filter. The filtering of the cochlea removes the higher-frequency components when the component spacing increases, thereby eliminating two-tone interactions with the high-frequency component. Before the signal is transmitted via the AN there is additional filtering due to the hair cell membrane properties and the effect of neural refractoriness. As the number of components (that are harmonically related) increases, the crest factor of the signal increases. The high-amplitude portion (the signal becomes an impulse train as the number of components approaches infinity) has a broadband composition, while the low-amplitude portion is dominated by components near the low- and high-frequency cutoffs of the signal. It follows that at low levels the broadband response will be encoded and that with increasing level the peak will be compressed more relative to the low-amplitude portion by the cochlear nonlinearity. It follows that edge frequencies contained in the low-amplitude portion of the signal will be better represented at high levels.

There are, of course, other factors that affect the output of the AN such as half-wave rectification, refractoriness, filtering due to hair cell membrane properties, and synaptic properties. A major contributing factor is two-tone suppression that is largely explained at the level of basilar-membrane mechanics (e.g., Ruggero et al., 1992). However, suppression seen in AN fibers using low-frequency suppressors has never been completely explained and may involve nonmechanical processes (e.g., Hill et al., 1989; Cooper, 1996; Geisler and Nuttall, 1997). Horst et al. found significant differences between the encoding by low- and high-spontaneously active AN fibers and suggested either a center clipping nonlinearity or an expansive nonlinearity as possible explanations for this behavior. Neither of these latter phenomena has been observed in basilar-membrane mechanics.

Use of multiple component signals was motivated by their similarity to many common communication signals including speech and the fact that they result in the perception of a pitch. Signals generated by a pulsatile source and subsequently filtered have a spectrum consisting of a series of harmonically related signals. Formants in speech consist of a maximum peak in the spectrum surrounded by several components separated by the pulse frequency, often 80 to 500 Hz. The signals used here are an approximation of these signals. The responses to them indicate that there are complex interactions that largely depend on the nonlinearity of the cochlea as to how the final signal spectrum is shaped. The bandwidth separation of the components and amplitude all are major factors in determining the final representation of the signal spectrum in the cochlea. Suppression is one of the most important factors in determining the contribution that each component makes in exciting the auditory nerve. There is clearly mutual suppression of closely spaced components combined with distortion product generation that determines the response to a complex signal. Mutual suppression could play a role in limiting cochlear damage at CF due to complex stimuli.

ACKNOWLEDGMENTS

This work was supported by the National Institute of Deafness and Communication Disorders, Grant No. R01 DC 01910. Special thanks to C. Dan Geisler for reading an earlier version of this manuscript and the journal reviewers of this article, Wiebe Horst and an anonymous reviewer.


