An effect of temporal asymmetry on loudness

G. Christopher Stecker and Ervin R. Hafter
Department of Psychology, University of California at Berkeley, 3210 Tolman Hall #1650, Berkeley, California 94720-1650

(Received 22 May 1999; revised 1 December 1999; accepted 6 March 2000)

A set of experiments was conducted to examine the loudness of sounds with temporally asymmetric amplitude envelopes. Envelopes were generated with fast-attack/slow-decay characteristics to produce F–S (or ''fast–slow'') stimuli, while temporally reversed versions of these same envelopes produced corresponding S–F (''slow–fast'') stimuli. For sinusoidal (330–6000 Hz) and broadband noise carriers, S–F stimuli were louder than F–S stimuli of equal energy. The magnitude of this effect was sensitive to stimulus order, with the largest differences between F–S and S–F loudness occurring after exposure to a preceding F–S stimulus. These results are not compatible with automatic gain control, power-spectrum models of loudness, or predictions obtained using the auditory image model [Patterson *et al.*, J. Acoust. Soc. Am. **98**, 1890–1894 (1995)]. Rather, they are comparable to phenomena of perceptual constancy, and may be related to the parsing of auditory input into direct and reverberant sound. © *2000 Acoustical Society of America.* [S0001-4966(00)02606-0]

PACS numbers: 43.66.Cb, 43.66.Mk [RVS]

INTRODUCTION

Sounds in the world can often be characterized as the result of two kinds of processes: production (the act of driving an acoustic medium to vibrate) and filtering (the effects of the medium upon the sound). Each of these processes imbues specific characteristics to the sound. Driving forces give a sound its attack; striking a cymbal produces a rapid attack, while bowing the same cymbal produces a very gradual attack. Filtering, on the other hand, primarily shapes the decay; a very stiff cymbal will decay more slowly than one composed of softer metal. The stiffness and mass of the cymbal affect its temporal envelope as well as its spectral characteristics. Of course, there are many cases where filtering influences a sound's attack, and where driving forces control the decay, but for a wide variety of dynamic sounds (e.g., percussion), attack time is controlled primarily by the driving force, while decay time is controlled by filtering.

In fact, when we consider sounds from the perspective of a listener, we realize that the ''filter'' affecting any particular sound is actually a combination of many interacting filters, including the source object (e.g., a violin string) and coupled structures (the violin body) and air masses (the violin's interior, the room containing both violinist and listener). Embedded in the sound's decay is information about the characteristics of each of these filter components. The size of a room, for instance, can be estimated from its reverberation time, the materials which cover its walls, by the spectral changes in the decay, etc. Similarly, the geometric shape of struck objects can be discriminated based on the vibrational modes apparent in their acoustic decay (Lakatos *et al.*, 1997). The ability to ''parse'' a sound in such a way as to recover information about these different characteristics would certainly be advantageous for any organism. Not only would this ability assist in identifying sound sources by their physical characteristics, but it would also help orient the listener in reverberant space.

The term ''temporal asymmetry'' (Patterson, 1994a) is used to describe sounds with differing attacks and decays. Such sounds can be useful in any study attempting to describe auditory processes involved in perceiving dynamic stimuli, and especially those involved in processing attacks and decays. For instance, if a tone with an instantaneous attack and a gradual decay is reversed in time, the result is a new tone which shares many characteristics with the first (e.g., Fourier spectrum, peak amplitude, etc.). However, the attack and decay times of the two tones are radically different. In terms of driving forces and filtering, the first tone resembles an impulse delivered to a linear filter; interpreting the second in a similar manner implies a very different filter (broadband, with a short time constant), as well as a very different driving force (gradually increasing). Differential processing of attacks and decays by the auditory system should result in major perceptual differences between the two stimuli.

It has been shown that temporal asymmetry can affect a stimulus' timbre (Cutting and Rosner, 1974; Rosen and Howell, 1981; Patterson, 1994a), its perceptual timing (Morton and Marcus, 1976; Vos and Rasch, 1981; Gordon, 1987), and even its pitch (Hartmann, 1978). Best known is the study of Cutting and Rosner (1974), who showed that changing the attack time of tonal stimuli resulted in a change in timbral categorization, with subjects describing fast-rising tones as ''plucked'' and slow-rising tones as ''bowed.''

Patterson (1994a, 1994b) introduced temporal asymmetry into the envelope of an amplitude-modulated tone by reversing the order of attack and decay of each cycle of the modulator. In his so-called ''damped'' tones, each cycle had an instantaneous attack followed by an exponential decay. Simply reversing the modulator in time produced ''ramped'' tones, with exponential rises and instantaneous decays. Despite the fact that the two stimuli had identical energy spectra, Patterson's subjects reported that ramped tones had a stronger sinusoidal quality than damped tones, which sounded ''hollow'' and ''percussive.'' Clearly, the exponen-

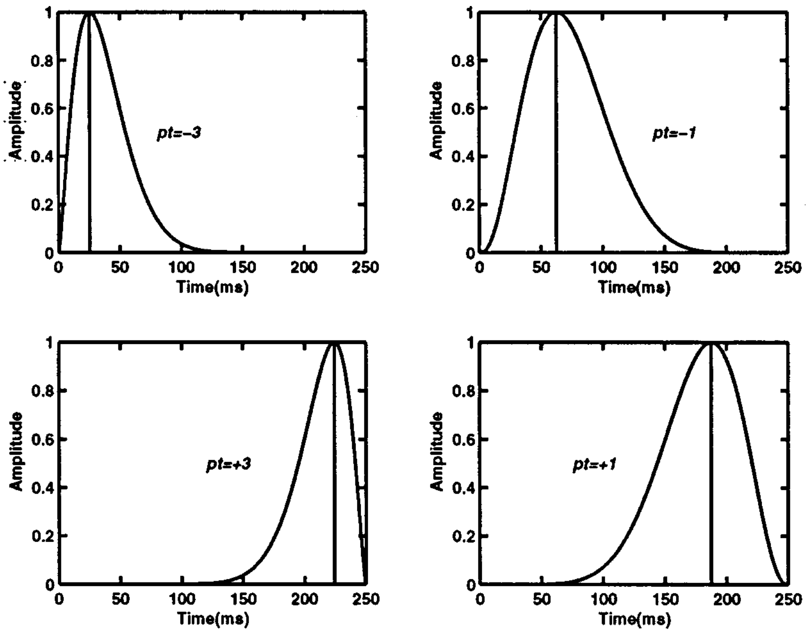

FIG. 1. Gating envelopes used in generating experimental stimuli. The top two plots (with negative values of *pt*) show envelopes for F–S stimuli. On the bottom, with positive values of *pt*, are S–F envelopes. Note that the left- and right-side envelope pairs differ in both total energy and rise/fall slope, but that F–S and S–F envelopes within each pair are equal-energy, temporally reversed versions of one another. The four envelopes shown here were used to gate pure tones (in experiments 1 and 3) and broadband noise (in experiment 2).

tial portion of the stimulus was perceived differently based upon its role as attack or decay.

Vos and Rasch (1981) studied the perceptual onset time (the moment at which a stimulus is judged to begin) of temporally asymmetric stimuli. Two such stimuli were generated by gating 20-component harmonic complexes with trapezoidal envelopes having identical decays, but different rise times. Pairs of these stimuli were repeated every 800 ms, and subjects manually adjusted the time between first and second tones to produce perceptual isochrony; that is, rhythmic regularity. From this study, Vos and Rasch (1981) concluded that the perceptual onsets of these stimuli occurred at the moment when their levels exceeded a relative threshold of about 6 to 15 dB below their peak levels. Thus, perceptual onset was strongly related to attack time.

Based on a suggestion by Wessel (1995), Stecker (1995) used a similar method to examine the perceptual onset of smoothly varying 330-Hz tones of 250-ms duration. Attacks and decays were sigmoidal, as opposed to the linear form used in Vos and Rasch (1981). Alternate tones were time-reversed, such that a stimulus with a fast attack and slow decay (hereafter referred to as fast–slow, or F–S) was followed by one with a slow attack and fast decay (slow–fast, or S–F). Each F–S, S–F pair was repeated every 1000 ms. When subjects adjusted the division of this period to produce isochrony, the time from F–S to S–F was shorter than that from S–F to F–S, again suggesting that perceptual onsets were earlier for stimuli with rapid attacks than for those with slow attacks. In terms of the relative threshold model suggested by Vos and Rasch (1981), the results indicate a threshold approximately 15 to 20 dB below peak amplitude.

The S–F and F–S stimuli in Stecker (1995) were equated in loudness using a perceptual adjustment task prior to the temporal judgments. Unexpectedly, S–F stimuli were judged to be louder than F–S stimuli of equal amplitude. The current study was designed to examine this result more explicitly. In experiment 1, the relative loudnesses of F–S and S–F tones were judged in a two-interval comparison (2IC) procedure. Subsequently, in experiment 2, a magnitude estimation (ME) procedure was used to obtain absolute judgments of loudness for tones and broadband noises. Finally, the role of carrier frequency was examined, again using the ME procedure, in experiment 3.

I. EXPERIMENT 1

In this experiment, subjects made paired comparisons between the loudness of fixed-level standard tones and variable-level test tones. Both standard and test tones were gated with temporally asymmetric envelopes.

A. Methods

1. Subjects

Three college-aged, paid subjects with normal auditory thresholds from 250 to 4000 Hz participated in the experiment. Although experienced at listening in psychoacoustics (signal detection and localization) experiments, all subjects were unaware of the purpose of the present study at the time of their participation.

2. Stimuli

The stimulus set consisted of four pure-tone (330-Hz) stimuli, each gated by a different temporally asymmetric envelope, 250 ms in duration. Four envelopes were generated using the equations in the Appendix, which define a function with the following properties: (1) It is smooth at all points in time, with a continuous change in slope. (2) Attack and decay portions are divided by a momentary amplitude peak, with no intervening steady-state period. (3) The degree of temporal asymmetry is described by the peak time of the envelope, which is set using the parameter *pt*. As shown in Fig. 1, four values of *pt* were used to generate two pairs of envelopes: $-3/+3$ and $-1/+1$. Note that the wider envelopes generated with $pt=\pm 1$ present greater total energy, shallower rise/fall times, and (by definition) more centralized peak times than those with $pt=\pm 3$ (see Fig. 1). However, the comparisons of interest will be between equal-energy F–S and S–F envelopes with the same absolute value of *pt*. Carriers and envelopes were multiplied with the phase of the carrier set to 0 deg at the beginning of each envelope. This multiplication was performed in software prior to running the experiment.

3. Procedure

A two-interval comparison (2IC) procedure was used to obtain relative loudness judgments among the four stimuli. Trials presented a standard stimulus in the first interval, followed after 300 ms by a test stimulus for comparison. For each trial, a new test stimulus was selected, at random, from among the four experimental stimuli, while the standard remained the same for an entire run of 108 trials. Standards were always presented at a level of 80 dB SPL (levels throughout this paper are of equivalent 1-kHz sinusoid at signal's peak amplitude), while the levels of test stimuli were selected at random from a set of nine possibilities: 70, 74, 76, 78, 80, 82, 84, 86, and 90 dB SPL. Subjects were told to indicate, by button press, which interval contained the louder stimulus. They were specifically instructed to base each judgment on the ''total 'sound energy' contained over the duration of the sound,'' ignoring other aspects, such as sharpness and timbre. Stimuli were delivered via Stax model SR-5 electrostatic headphones in a double-walled sound-proof booth located in the basement of Tolman Hall on the Berkeley campus.

Blocks of 108 trials were organized such that each of the four envelopes was presented as the test tone three times at each of the nine presentation levels, in random order. The standard remained the same for all of the 108 trials within a particular block; a new standard was selected from among the four potential envelopes for each new block. Each subject completed 12 blocks with each of the four standards. The resulting data were analyzed using logistic regression. (SPSS, 1999).

B. Results

As can be seen in Fig. 2, the range of levels used for test tones was adequate to evoke nearly a full range of potential responses. That is, subjects nearly always judged the test tone as being the louder when presented at 90 dB SPL (10 dB above the standard), and nearly always judged it to be the quieter when presented at 70 dB SPL. The shapes of the loudness functions (relating response probability to presentation level) do not differ among the four stimuli used; differences in loudness between the stimuli can be seen as lateral shifts of the loudness functions. Overall, S–F stimuli were judged louder than the standard significantly more often than F–S stimuli ($p<0.05$); this is discussed in more detail below. Additionally, the effect of overall stimulus energy ($pt=\pm 3$ vs $pt=\pm 1$) is also significant; this is to be expected due to the relationship between energy and loudness. There is no significant interaction between envelope type (F–S vs S–F) and stimulus energy. Therefore, subsequent analyses combine stimuli across magnitude of *pt*, instead focusing on differences in its sign.

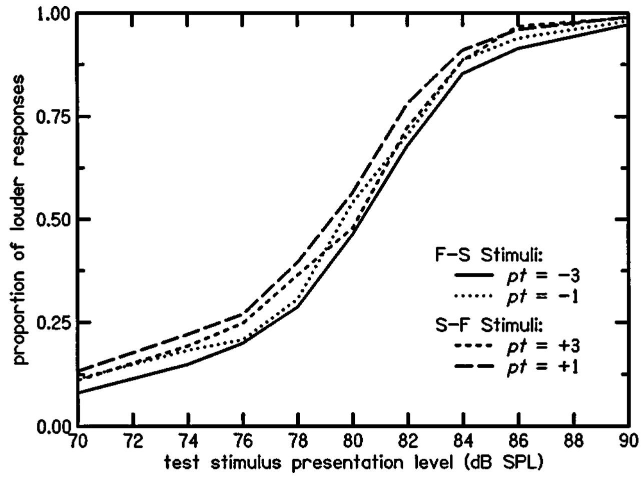

FIG. 2. Results of experiment 1 (2IC procedure), indicating proportion of responses (averaged across three listeners) indicating the test tone to be louder than the standard, as a function of test-tone level. The standard was always presented at 80 dB SPL. Overall, responses indicate greater loudness for S–F tones than for F–S tones. Test tones with $pt=\pm 1$ were louder than their $pt=\pm 3$ counterparts, consistent with the energy difference between them. However, the pattern of (F–S vs S–F) results is the same for the two pairs, and the shapes of the loudness functions do not differ among the four stimuli.

Figure 3 shows the same data averaged over presentation level and magnitude of *pt*, but plotted as a function of the F–S/S–F envelope characteristic of the standard in each trial. Overall analyses of the response proportions given for each combination of standard and test tone indicate two things: first, as noted above, S–F stimuli were judged as being louder than any particular standard more often than were the corresponding (equal-energy) F–S stimuli. Second, the difference between S–F and F–S versions of the envelopes was greatest when the standard itself was one of the F–S stimuli. Both the main effect of test-tone envelope and the interaction with the envelope characteristic of the standard were statistically significant for both pairs ($pt=\pm 1$, $pt=\pm 3$) of envelopes ($p<0.05$). Finally, linear regression estimates based on the psychometric functions for one listener (subject CD) indicate the magnitude of the loudness

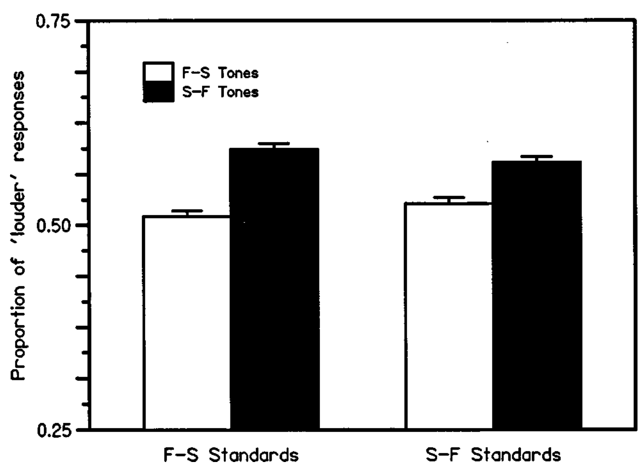

FIG. 3. Results of experiment 1 (2IC procedure), continued. Data are plotted separately for F–S ($pt<0$, left bars) and S–F ($pt>0$, right bars) standards. An overall loudness advantage for S–F test tones appears for both types of standard, but the effect is significantly attenuated for S–F standards.

difference to be approximately 1–3 dB following F–S standards, but less than 1 dB following S–F standards.

C. Discussion

The main finding of experiment 1 is that S–F stimuli were judged to be louder than F–S stimuli of equal energy. A second finding is that the judgments were influenced by the stimulus ''context'' immediately preceding the judged tone: the difference in loudness was greatest when judged tones were compared to stimuli with F–S envelopes.

II. EXPERIMENT 2: MAGNITUDE ESTIMATION

While the main findings of experiment 1 clearly demonstrate that S–F tones are judged to be louder than equal-energy F–S tones, the finding that this difference depends upon the choice of standard is troubling. No theory which assigns loudnesses to single stimuli in a context-free manner can account for this type of finding. Perhaps the context effect indicates a procedural influence, with listeners altering their response criteria based on the identity of the standard. In experiment 2, we sought to reduce this context effect by eliminating the standard from each trial. Instead, we used a magnitude estimation (ME) procedure, where each response indicates an *absolute* judgment for a single stimulus. While the context for relative judgments in the 2IC task consists of the previous stimulus (the standard), context in ME tasks consists of the entire set of stimuli [see, e.g., Marks (1988)].

We adopted a procedure quite similar to that of Stevens (1956), with a few critical differences. In one case, Stevens presented listeners with a standard or ''modulus,'' set to a fixed level and assigned a magnitude, for instance, ''10.'' Listeners were instructed to estimate the loudness of subsequent test stimuli by assigning numbers to each, such that the ratio of the response to the modulus equaled the perceived ratio of the two loudnesses. That is, a stimulus 4 times as loud as the modulus should be called ''40,'' one half as loud should be ''5,'' etc. In an alternative procedure, Stevens (1956) omitted the modulus but maintained the ratio instructions by requiring listeners to respond with numbers whose ratios matched the relative loudnesses of the test stimuli.

Since we hoped to eliminate the effects of relative judgment in this experiment, we omitted Stevens' (1956) ratio instructions, instead asking subjects to simply assign numbers which matched their perceived loudnesses, using any scale with which they felt comfortable. In an effort to reduce intersubject variability, however, we included a single ''reminder'' stimulus at the start of each run. Subjects were told that this reminder, presented at a moderate fixed level, should be called ''50,'' and that its purpose was to help them anchor their scales of loudness from run to run and day to day. However, subjects were also told that they were free to choose any particular values for their ratings, so long as they tried to be consistent across trials (the same loudnesses should produce the same ratings later in the experiment). As a result, most subjects used scales centered around 50 (e.g., 0–100, 20–80), although a few used very different scales (see Fig. 4).

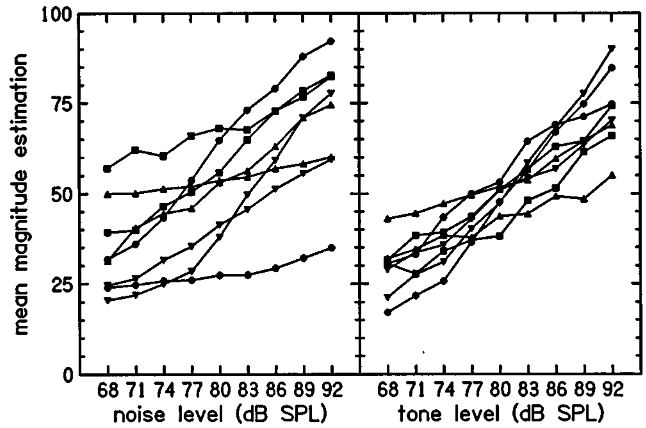

FIG. 4. Intersubject variability in rating data (ME procedure). The mean value of magnitude estimations made by each subject is plotted as a function of stimulus level. Ratings for noise stimuli appear on the left; ratings for tonal stimuli are plotted on the right. Separate lines represent individual subjects, who varied in both the range (indicated by the slopes of their loudness functions) and bias (indicated by their intercepts) of their responses.

We were also interested in whether these results would generalize to stimuli with different characteristics, in terms of tonality, frequency, bandwidth, and regularity. Irino and Patterson (1996), for instance, found that temporal asymmetry for ramped/damped modulation was reduced when using noise, compared to tonal carriers. As such, we ran two conditions in experiment 2: a pure-tone condition with the same stimuli as in experiment 1, and a noise condition using a white-noise carrier which changed from trial to trial.

A. Methods

1. Subjects

One group of eight undergraduate volunteers, participating to obtain course credit, was included in the pure-tone condition. A second group of eight undergraduates, recruited separately, participated in the noise condition. All subjects reported having normal hearing; however, audiometric testing was not performed in conjunction with the experiment. None was informed about the purpose of the study prior to participation.

2. Stimuli

In the pure-tone condition, stimuli were exactly as described for experiment 1, although their levels were changed slightly (see below). However, carrier and envelope signals were synthesized separately and combined in real time using an analog multiplier (voltage-controlled amplifier), just prior to amplification and stimulus delivery via earphones. Carrier phase (at the envelope's initial sample) was set to 0 deg.

In the noise condition, the carrier signal was a white (1 Hz–20 kHz$\pm$0.1 dB) noise, generated by an analog noise generator. As in the pure-tone condition, this carrier was gated in real time using an analog multiplier, thus causing each trial to contain a different noise sample. The four envelopes (see Fig. 1) used for gating were identical to those used in experiment 1.

3. Procedure

At the start of each experimental run, subjects were presented with a ''reminder'' stimulus. A single envelope (F–S, $pt=-3$) and level (80 dB SPL) was used for this in all conditions. Subjects were instructed to consider the perceived loudness of this stimulus to be ''50,'' and then rate the magnitude of each of 72 subsequent test stimuli on a scale of their choosing. The purpose of the reminder was to designate to listeners the median presentation level at the start of each run; subjects were not instructed to use it as a reference, but to rate each trial independently. Each trial contained a single stimulus followed by a prompt on the computer screen and the subject's response. Test stimuli were drawn randomly from the set of four envelopes and nine presentation levels, spaced every 3 dB from 68 to 92 dB SPL. Note that this is a slightly larger range of levels than that used in the 2IC procedure, and that the purpose of randomizing levels is the same as in experiment 1: namely, to provide a source of variation in stimulus loudness. Randomization was constrained to produce an equal number of presentations, within each run, for each envelope and level. Here, stimuli were delivered to subjects via Etymotic ER-4S earphones within a single-walled soundproof booth.

At the prompt, subjects indicated the perceived loudness of the test tone by entering a numerical value on a computer keyboard. As in experiment 1, subjects were instructed to base each judgment on the total sound energy contained over the duration of the sound, ignoring other aspects, such as sharpness and timbre. Each subject finished six runs, including an initial (discarded) ''training'' run, and runs 2–6, which were used in the subsequent analysis. The entire procedure was identical for the two conditions (pure-tone and noise).

B. Results

Figure 4 shows the mean magnitude estimations for each subject, as a function of stimulus SPL. Notice that the ordinate is plotted in linear, not logarithmic, coordinates, and that the loudness functions are all approximately linear. Normally, magnitude estimation experiments produce ratio responses, and the obtained loudness functions are power functions. The linear functions shown in Fig. 4 indicate that our subjects did not use ratio scales in estimating loudness; indeed, they resemble functions obtained when the scale is preassigned (Stevens, 1956). Stevens argued that such linear functions resulted from overly constraining the listeners' responses and that his magnitude estimation method, free of such constraints, revealed the natural judgments to be based on ratios, rather than linear scales. Without discounting the converging evidence in support of Stevens' power law, we might say that, based on our own subject's tendency to use linear scales in the absence of specific ratio instructions, the idea that magnitude estimations are inherently ratios may follow from procedural effects rather than any fundamental perceptual law.

Figure 4 also reveals large intersubject variation in both the range and bias of their responses. To facilitate the pooling of data across subjects, we normalized the slope and intercept of each subject's loudness function using the z transform: each rating (y) was transformed into a z score

$$z=\frac{y-\overline{y}_{\mathrm{subj}}}{\sigma_{\mathrm{subj}}},$$

based on the individual subject's overall mean ($\overline{y}_{\mathrm{subj}}$) and standard deviation (σ_{subj}) of responses (across all presentation levels). Z-score ratings were treated as raw data points for subsequent analysis by analysis of variance (ANOVA). As stated above, this transformation (which follows from the decision to treat the obtained loudness functions as linear) normalizes both the slope and intercept of the loudness function. While it is not useful for visualizing the relationship between intensity and loudness, by reducing intersubject variability, it allows the pooling of data across subjects and facilitates the comparison of loudness functions for different stimuli. A difference in loudness between F–S and S–F stimuli, for instance, should result in a lateral displacement of the normalized loudness functions, and a resulting difference between the mean z scores, averaged across SPL, as shown in Fig. 5. Mean z-score loudness ratings are shown for 330-Hz tones and broadband noises, plotted as a function of the envelope shapes of both the test stimulus and that of the test stimulus on the previous trial (i.e., the local context). This partitioning of the data allows us to examine the influence of prior stimulation, despite the fact that judgments were not intended to incorporate this local context, as they were under the 2IC procedure.

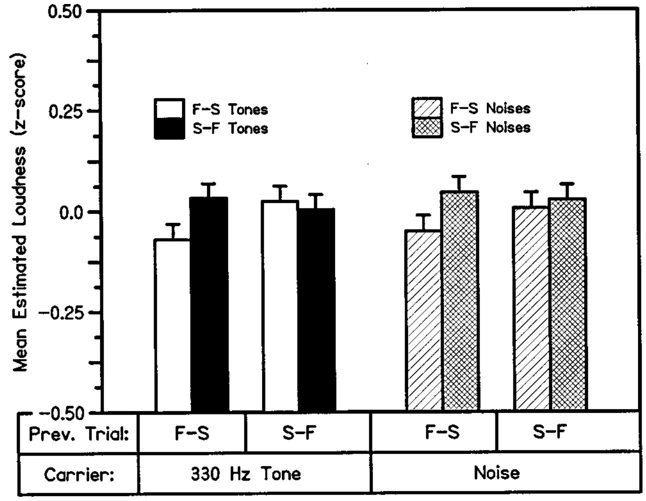

FIG. 5. Results of experiment 2 (ME procedure). The mean loudness ratings for 330-Hz tones (black and white bars) and broadband noises (hatched bars), are plotted by test-tone envelope shape, separately for trials immediately preceded by F–S trials and those preceded by S–F trials. Each listener's ratings are normalized, using a z transform, prior to combining data across subjects (see the text). As in experiment 1, an overall loudness advantage is observed for S–F stimuli, and is significantly reduced following previous S–F stimuli. There are no statistically significant differences between the results for tones and noises.

For pure tones, the main effect observed in experiment 1 was reproduced using the ME procedure. Overall, stimuli with S–F envelopes were rated louder than F–S stimuli ($p<0.05$). In addition, there was a significant interaction ($p<0.05$) between the loudness ratings for a given stimulus and the envelope shape of the context stimulus. Examination of the simple effects of test-stimulus envelope for each context reveals a significant loudness difference following F–S ($p<0.05$), but no significant difference following S–F context stimuli.

There was no significant difference between the results for noise and those for pure tones. As in the pure-tone condition, S–F noise stimuli received significantly higher ratings ($p<0.05$) than F–S stimuli, and the difference was significantly greater ($p<0.05$) when preceded by F–S context stimuli.

C. Discussion

First, the main finding of experiment 1, that S–F tones were louder than F–S tones of equal energy, was reproduced in experiment 2, using the ME procedure. This was true for both tones and noises, indicating that the effect did not depend upon the tonal quality or the spectral makeup of the stimuli used.

Second, although the ME procedure was designed to eliminate the effects of local context, it did not. It is clear that subjects did, to some extent, judge all test stimuli in the context of previous stimuli. Specifically, S–F stimuli were rated louder than F–S stimuli when the previous stimulus had F–S characteristics.

Previously, two kinds of context effect have been reported in magnitude estimation and ratio-scaling experiments. Marks (1988), for instance, demonstrated that manipulating the overall range of presentation levels for each of two stimulus types can systematically alter loudness matches between the stimuli. This does not explain the short-term effect of stimulus context observed in this study, since the effect Marks described reflects differences across the entire stimulus set. Jesteadt *et al.* (1977) demonstrated an influence of response history on magnitude estimations, whereby large responses tend to be followed by large responses and small responses tend to follow small responses (an *assimilative* sequential effect). The current study, however, shows an assimilative effect only when both the just-prior stimulus and the test stimulus have F–S envelopes. Responses to S–F test stimuli in this case are not assimilated. Thus, the effect is not simply assimilative or contrastive, but depends upon the envelope of the stimulus being judged. Additionally, the context effect is asymmetric, in that F–S contexts affect subsequent judgments, but S–F contexts do not. These facts, along with the agreement of results from both 2IC and ME procedures, suggest that the observed context effect is not procedural in origin, but may instead relate to the mechanism responsible for the difference between S–F and F–S loudness. We will return to this issue in the general discussion below.

III. EXPERIMENT 3: EFFECTS OF CARRIER FREQUENCY

Patterson (1994a, 1994b) showed that ramped tones produce a stronger sinusoidal percept than do corresponding damped tones, a finding not unlike that observed in the current study. In addition, he noted that this perceptual asymmetry was diminished for high-frequency (>3 kHz) carriers, and related that finding to the loss of phase-locking in auditory-nerve fibers at high frequencies (Patterson, 1994b). He applied the auditory image model (AIM) (Patterson *et al.*, 1995), a modular framework for modeling the early auditory system, to explain his results. While carrier frequency was not a focus of Patterson's modeling work, there are at least three temporally asymmetric processing stages in AIM, each related to frequency in a different manner. These stages are auditory filtering, neural encoding, and temporal integration (Irino and Patterson, 1996).

AIM's representation of basilar-membrane activity includes narrow-band filtering, rectification, and compression (Patterson *et al.*, 1995). Filters slightly mistuned from the carrier frequency of a stimulus respond differently to ramped and damped tones. The output waveform in response to a ramped tone is extended in time, while that to a damped tone is more sharply peaked, with an amplified peak and attenuated decay.[1] Compression, which follows auditory filtering, enhances this difference by ''squashing'' the peaked response to damped tones. This results in greater overall activity in response to ramped tones. This mechanism is frequency dependent. In general, increasing carrier frequency (without changing the stimulus envelope) will reduce the difference between responses to ramped and damped stimuli. This difference is naturally related to the similarity between the decay of the stimulus and that of the filter's impulse response.

Later stages explicitly incorporate mechanisms of temporal adaptation which are asymmetrical in time, responding rapidly to increases in input level, but lagging behind rapid decreases. The end result of this adaptation is that the exponential attack of a ramped sound has more influence on activity in the model than does the corresponding exponential decay of a damped stimulus.[2] Irino and Patterson (1996) suggest that these mechanisms, rather than the behavior of auditory filters, play the dominant role in producing temporal asymmetry in AIM, for two reasons: first, they exhibit a much larger effect of temporal asymmetry (Irino and Patterson, 1996), and second, they provide a better explanation for temporal asymmetry in nonperiodic sounds, such as damped and ramped noises (Akeroyd and Patterson, 1995). While a comprehensive treatment of these mechanisms is beyond the scope of this paper, it is important to point out that these adaptive mechanisms do not appear to be frequency dependent.

We became interested in AIM as a potential mechanism for producing temporal asymmetry in the loudness of F–S and S–F tones, and posed the following questions: Can AIM predict differences in the *loudness* of F–S and S–F tones? If so, will the predicted difference be affected by carrier frequency? Finally, will the loudness of high-frequency F–S and S–F tones be judged by human listeners in agreement with predictions from AIM?

As it stands, AIM does not generate direct predictions of loudness; however, the software package itself[3] includes a loudness function which calculates a loudness measure based on instantaneous activity in the temporally integrated auditory image. This measure is essentially the mean of activation across frequency channels, and represents the instantaneous loudness of a stimulus as it changes over time. Figure

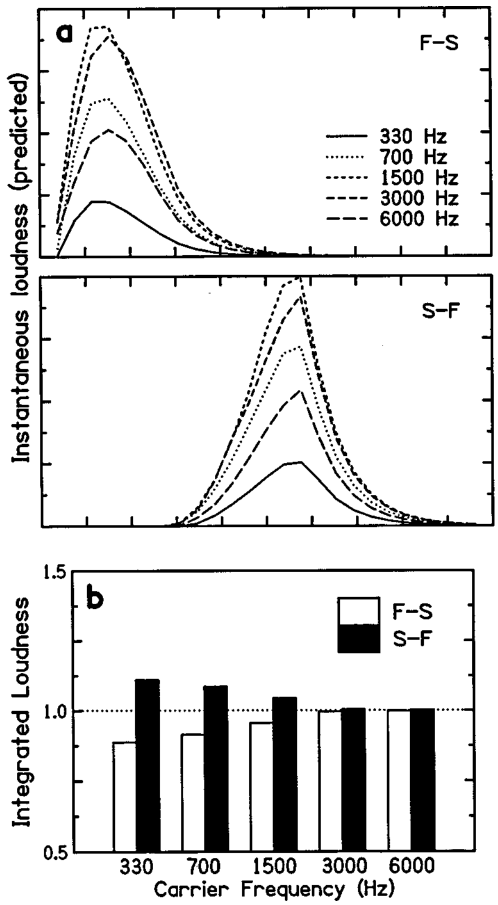

FIG. 6. Instantaneous loudness as a function of time (a) for F–S (upper panel) and S–F (lower panel) sounds. Values are predictions from AIM (Patterson *et al.*, 1995) and the parameter is carrier frequency. There are two points to note: first, middle frequencies (1500 and 3000 Hz) are louder overall than either high or low frequencies. Second, while F–S stimuli have functions which resemble their envelopes (see Fig. 1), S–F stimuli produced continued activity following their offsets. In (b), predicted loudness has been integrated over time, producing a single loudness estimate for each stimulus. Since the overall magnitudes of these estimates vary across frequency, each value was normalized (divided) by the mean of the two estimates for stimuli at its carrier frequency. The normalized mean is indicated by the dotted line. As can be seen in the plot, AIM predicts a larger difference between F–S and S–F stimuli at low frequencies (330–1500 Hz) than at higher frequencies.

6 displays these predictions as functions of time for each stimulus. By integrating these instantaneous estimates across time, we generated overall loudness values for pure-tone signals with envelopes as in experiment 1 and carrier frequencies from 330 to 6000 Hz. These showed that, as we found, AIM predicts greater loudness for S–F than F–S stimuli at low frequencies (330–1500 Hz), but a marked decline in this temporal asymmetry above 3 kHz (Stecker and Hafter, 1998). The purpose of experiment 3 was to test this result further by examining the loudnesses of F–S and S–F tones over a range of carrier frequencies (330–6000 Hz), again using the ME procedure.

A. Methods

1. Subjects

Subjects were a new group of ten undergraduate volunteers, participating to obtain course credit. All reported having normal hearing, although audiometric testing was not done. None was informed of the study's purpose prior to participating, and none had participated previously in any of the earlier experiments in this study.

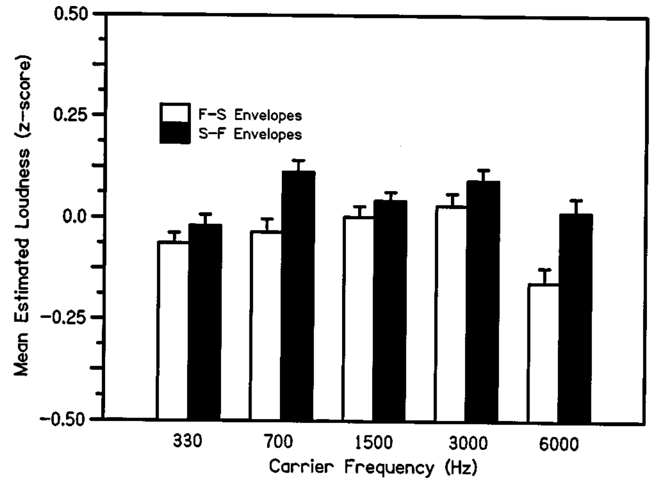

FIG. 7. Results of experiment 3: loudness ratings for F–S and S–F tones across carrier frequency. While AIM (Patterson *et al.*, 1995) predicts an elimination of temporal asymmetry above approximately 3000 Hz, the observed loudness advantage is consistently in favor of S–F tones across carrier frequencies from 330–6000 Hz.

2. Stimuli

Sinusoidal carriers, 250 ms in duration, were synthesized at 330, 600, 1500, 3000, and 6000-Hz carrier frequencies, using a starting phase of 0 deg. Envelopes were generated using the same procedure described in the Appendix, and were of the same form used in the prior experiments. As in experiment 2, carriers were gated in real time using an analog multiplier (voltage-controlled amplifier) prior to amplification for earphone presentation.

3. Procedure

Within each frequency condition, the stimulus presentation and subject response procedures were identical to those used in the ME procedures in experiments 1 and 2. Each subject completed between six and eight experimental runs at each of two randomly assigned carrier frequencies (resulting in four subjects per frequency condition). Since not all subjects completed all eight runs, only runs 2–6 were used for subsequent analysis. A single carrier frequency was used for all stimuli in a single run, including the initial standard-level presentation and all of the test-stimulus presentations. Frequency conditions were not intermixed; all runs at one carrier frequency were completed before moving on to another. Loudness data were analyzed according to envelope type (S–F/F–S) and carrier frequency.

B. Results

The results of experiment 3 can be seen in Fig. 7. As in the previous experiments, S–F tones were rated as louder than F–S tones ($p<0.05$). This difference was consistent across all tested carrier frequencies, with no significant interaction between carrier frequency and envelope type.

C. Discussion

While AIM correctly predicts that S–F tones should be louder than F–S tones, it incorrectly predicts that this difference should be reduced at high frequencies. Based on the AIM literature (Patterson *et al.*, 1995; Patterson, 1994a; Irino and Patterson, 1996; Carlyon, 1996), we assume that the predicted frequency difference arises primarily due to processes involved in auditory filtering. However, Irino and Patterson (1996) suggest that the greatest amount of temporal asymmetry arises later, in the temporal integration stage, which is not frequency dependent; this raises an interesting question: why, if most of AIM's temporal asymmetry is attributable to mechanisms which are *not* frequency dependent, does AIM predict such a large reduction in temporal asymmetry at high frequency? Regardless of the answer to that question, it seems reasonable to assume that AIM could be modified to eliminate the effects of carrier frequency, but we have not undertaken such a modification in our work.

Regardless of the mechanisms producing frequency dependence in AIM, our results show that the difference in loudness between F–S and S–F tones exists across frequencies from 330–6000 Hz. Among other things, this implies that phase-locking in auditory-nerve fibers is not necessary to produce the effect.

IV. GENERAL DISCUSSION

A. Main effect of greater loudness for S–F stimuli

The primary finding reported here is the greater loudness of S–F over F–S stimuli, with slow-rise/fast-decay stimuli perceived as being louder than fast-rise/slow-decay stimuli of equal energy. In effect, loudness mechanisms treat the two ends of the stimulus differently based on their roles as either attack or decay. This effect was observed in every experiment, under both two-interval (2IC) and magnitude-estimation (ME) procedures.

We believe that one may rule out several possible mechanisms which might have caused this finding. One is automatic gain control (AGC), which predicts exactly the opposite, specifically that F–S stimuli should have been louder than S–F stimuli. Since AGC mechanisms must utilize estimates of a stimulus' envelope to adjust its level, they naturally respond with some delay. When presented with rapid increases in stimulus level, extra energy is allowed into the system before the AGC can respond. Thus, the output of an AGC will be greater in magnitude for stimuli with fast attacks than for those with slower attacks.

Spectral models which calculate loudness by integrating activity across frequency bands prior to (or in the absence of) compression (Zwicker and Scharf, 1965; Moore and Glasberg, 1996) predict little or no difference between the loudnesses of F–S and S–F stimuli, since no difference exists between their energy spectra. Similarly, temporally symmetric mechanisms such as the autocorrelogram (Slaney and Lyon, 1990) cannot predict a difference between F–S and S–F stimuli [see Patterson and Irino (1998) for a discussion of temporal asymmetry in autocorrelogram models]. Thus, these models are also incompatible with our findings.

Finally, modeling results obtained with AIM (Patterson *et al.*, 1995), show that, while it correctly predicts the main finding of experiments 1 and 2 (though not the effect of context), the model is insufficient for explaining the results of experiment 3. Surprisingly, AIM predicts a reduction or elimination of temporal asymmetry above 3 kHz. With human listeners, the loudness difference between F–S and S–F tones remains for carrier frequencies up to 6 kHz.

We should point out that, although this study has focused on judgments of *loudness*, results from other researchers suggest a more general phenomenon of decay suppression (for lack of a better term) in temporally asymmetric stimuli. Patterson's (1994a, 1994b) initial ramped/damped studies, for instance, demonstrate a *timbral* difference between ramped and damped sinusoids: damped tones produce a weaker ''sinusoidal'' component, as if the carrier contained in the damped decay were perceptually attenuated. More recently, Schlauch *et al.* (1998) observed that the perceived *duration* of damped tones is shorter than that of corresponding ramped tones, again as if the damped tails were somehow reduced in perceptual strength. Both of these results are congruent with those of the current study; in all three cases, exponential decays seem to be perceptually suppressed, making smaller contributions to the judgments.

Two other phenomena which may be related to the current findings are deserving of mention here. First is the ''decruitment'' phenomenon reported by Canevet and colleagues (Canevet and Scharf, 1990), whereby tones and noises which continuously decrease in level are perceived to decrease in loudness more rapidly than would be expected from the change in level. This would seem to be in agreement with our finding that F–S signals (which are decreasing in level for most of their duration) are less loud, overall, than they should be. However, decruitment appears to require stimuli with durations greater than 1 s (Canevet *et al.*, 1999), much longer than any stimuli used in the current study.

Using a similar stimulus, Neuhoff (1998b) recently showed that loudness *change* is judged greater for simple and complex tones increasing in level, relative to tones decreasing by the same amount. While this phenomenon could also be related to our current findings, several key points make comparing the studies directly difficult: most importantly, Neuhoff argues that the perception of loudness change is fundamentally different from the perception of loudness itself (Neuhoff, 1998a) and specifically instructs his listeners to ignore the overall loudness of the stimuli. Second, as in the decruitment phenomenon (Canevet and Scharf, 1990), stimulus durations in Neuhoff's experiment were much longer (1.8 s) than in ours. The amount of level change was also limited to 15 dB, resulting in a much slower rate of change. Third, while Neuhoff found the results quite similar for sinusoids and complex tones, he found no such asymmetry between white noises increasing and decreasing in level, whereas the differences in loudness found here were similar for the two carrier types.

B. Effect of stimulus context

In addition to finding greater loudness for S–F tones and noises, the current data also show significant effects of prior

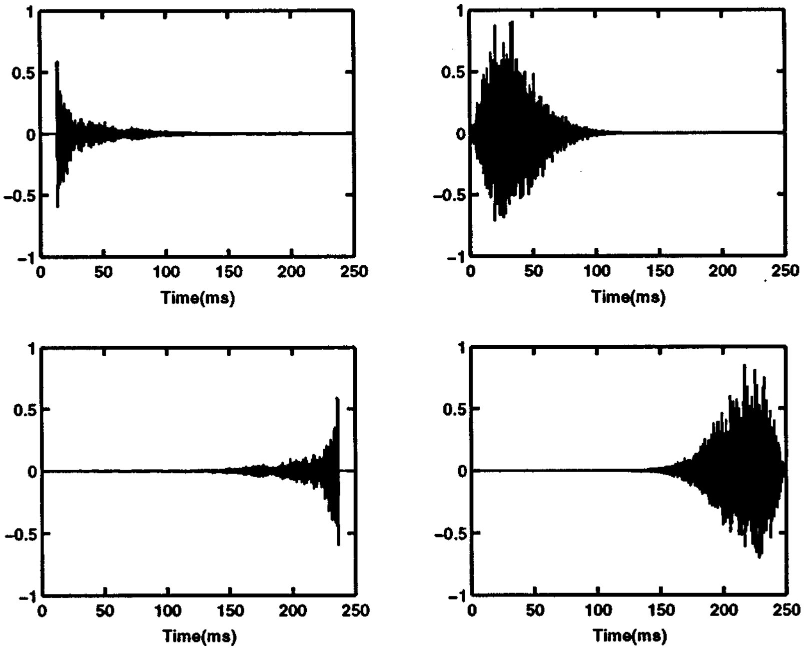

FIG. 8. Decay of reverberant sound. On the left, waveforms of a brick being struck by a hammer in a conference room (Houtsma *et al.*, 1987). On the right, examples of noise stimuli used in experiment 2. The F–S stimulus (top right) has a decay similar to the hammer/brick sound (top left). Playing the hammer/brick recording backwards (bottom left) produces a stimulus similar to the S–F stimuli (bottom right) used in this experiment.

stimulus context. Specifically, preceding the test stimulus with an S–F context reduces or eliminates the loudness difference (i.e., the difference is greatest when stimuli are preceded by F–S contexts). It is important to note that, to our knowledge, no existing models of loudness can account for this finding, as they normally assign loudnesses based on the characteristics of single sounds, independent of previous stimuli. In other words, the context effect we have demonstrated suggests a mechanism with a time constant spanning hundreds of milliseconds or even seconds. Aside from Neuhoff's explanation of perceptual bias for rising tones (Neuhoff, 1998b, 1998a), none of the models discussed above possesses time constants this long, and so they cannot make predictions regarding the effect. Perhaps models which can accommodate our main finding (AIM, for low frequency carriers) could be extended through the addition of higher-order mechanisms that operate on this longer time scale.

We originally assumed that this effect of stimulus context was related to the direct use of standard-test comparisons in the 2IC procedure; however, a nearly identical effect was obtained using ME in experiment 2. This suggests that the influence of local context on loudness judgments is not methodological in nature. Instead, we argue that this influence reflects the perceptual processing of these stimuli, independent of the judgment task. Specifically, we suggest that our results [along with those of Patterson (1994a, 1994b, 1995) and Schlauch (1998)] point to a general process of decay suppression, and that the effects of local context reflect the natural dynamics of this process.

What would a general mechanism for perceptually suppressing decays be good for? One possibility is that decays are normally more informative about the characteristics of the listening environment (primarily due to reflections, reverberation, and room resonances) than about the sound sources themselves. As shown in Fig. 8, even very short stimuli can produce rather lengthy decays due to room effects. Perhaps the auditory system ''parses'' the stimulus into direct and reverberant sound, providing the listener with a cleaner version of the direct stimulus, i.e., the source.

Although the stimuli used here were not generated purposely to resemble reverberated sounds, F–S stimuli are similar to those observed in real rooms, with an initial transient followed by a slower decay. Conversely, an S–F stimulus is more like a room recording played in reverse (Houtsma *et al.*, 1987), which does not provide a sense of reverberation (see Fig. 8). The reduced loudness of F–S relative to S–F stimuli may then be the result of the auditory system compensating for the apparent effects of reverberation in F–S stimuli; eliminating the tail from the judgment of loudness because it is interpreted as an acoustic by-product, rather than a meaningful part of the stimulus being judged. In other words, listeners perceptually ''interpret'' an F–S stimulus as the output of a system of reverberant filters introduced by the environment. They then make judgments based upon their estimate of the input to that system. The manner in which listeners acquire knowledge of the reverberant system is an important issue which we will discuss below.

Such compensation for environmental effects has been termed ''perceptual constancy'' and results in the perception of distal ''object'' properties, rather than proximal stimulus properties (Boring, 1952; Jameson and Hurvich, 1989). For instance, a white piece of paper appears to be the same color in sunlight and fluorescent lighting, despite large changes in the proximal stimulus, including both the magnitude and spectrum of reflected light (Gilchrist, 1977). While perceptual constancies can be very strong and apparently automatic, they are not absolute or instantaneous; Rock (1997) writes of a two-stage process in which a stimulus is initially perceived in a ''literal mode'' corresponding to its proximal features, and then in a subsequent ''constancy mode,'' where the visual system has ''reinterpreted'' the stimulus to match its distal form.

According to Rock (1997), a good example of this process occurs in amodal completion, depicted in Fig. 9(a). The

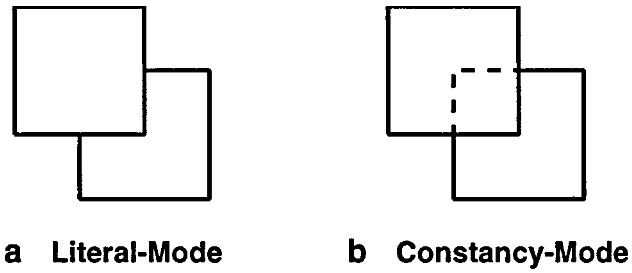

FIG. 9. Amodal completion. The ambiguous figure in (a) can be seen as either a square lying next to an L-shaped object (a literal mode interpretation), or (b) two overlapping squares (a constancy-mode interpretation).

literal-mode solution to this figure sees two objects (one square and one L-shaped) abutting, but not overlapping, while the constancy-mode solution ''fills in'' the occluded region, resulting in the perception of two overlapping squares. While the literal-mode interpretation is correct, it is superseded by the constancy-mode solution, which represents a more likely state of affairs in the world. Rock does point out that although the literal-mode solution is ''superseded'' by the constancy mode, both perceptions may in fact be available to other cognitive mechanisms simultaneously. In addition, the formation of a constancy-mode interpretation takes time. Sekuler and Palmer (1992) showed that, for visual displays similar to Fig. 9(a), it requires approximately 200 ms. As argued by Rock (1997), perception remains in the literal mode during this time.

We may distinguish the effects of these two perceptual modes on stimuli in the current study. For instance, in the literal mode the F–S stimulus is perceived as a whole, and so appears equally loud to the equal-energy S–F stimulus. But in constancy mode, the F–S stimulus is heard as composed of two parts, source and reverberation, with loudness based only on the source component. Accordingly, the loudness of the F–S will be reduced in constancy mode. In contrast, the short decay of an S–F stimulus indicates little or no effect of reverberation, and parsing the stimulus in constancy mode produces a ''source'' which is nearly identical to the stimulus itself. Thus, the loudness of S–F stimuli will be the same in both perceptual modes.

The formation of the constancy-mode interpretation is influenced by previous stimulus exposure. In essence, one perceptual interpretation of the stimulus (matching a particular reverberant environment) is built up, or ''primed,'' by experience. This facilitates the constancy-mode perception of subsequent stimuli which agree with that interpretation. Initial exposure to an F–S stimulus, for instance, primes the formation of a reverberant solution; this ultimately reduces the loudness of subsequent stimuli which match the primed interpretation. A mismatch between this interpretation and subsequent stimuli (as would occur, for instance, when moving from one room to another, differently arranged, room) causes perception to switch back to literal mode, and a new constancy-mode interpretation must be formed.

We will make two points regarding this explanation before closing. First, it must be noted that, rather than specifying a model, we have provided only a description of a potential mechanism. A ''general mechanism for decay suppression'' could take many forms, and it is possible that existing or future psychoacoustic models of temporal asymmetry, such as AIM (Patterson *et al.*, 1995), may provide a framework which accomplishes what we have outlined. The effect of prior stimulation, which we have described as adaptation to changing reverberant environments, could be modeled by incorporating dynamic parameters controlled by adaptive, memory-based, or explicitly cognitive functions.

V. CONCLUSIONS

In this paper, we have described an effect of temporal asymmetry on the loudness of short-duration (<250 ms) stimuli. In short, stimuli with fast attacks and slow decays (F–S) are heard as less loud than comparable stimuli with slow attacks and fast decays (S–F). This effect is observed across a range of experimental procedures (two-interval, magnitude-estimation) and spectral composition (high- and low-frequency pure tones, wideband noise). Additionally, the degree of effect is modulated by the envelope characteristics of prior stimuli (the local context). Neither of these effects is predicted by spectral mechanisms or AGC. The auditory image model (AIM) (Patterson *et al.*, 1995) correctly predicts the overall loudness advantage for S–F stimuli, but incorrectly predicts that the effect should be eliminated at high frequencies, and makes no prediction regarding the influence of previous stimuli.

We have argued, based upon the results of this and other studies (Patterson, 1994a, 1994b; Akeroyd and Patterson, 1995; Schlauch *et al.*, 1998), for a general mechanism of ''decay suppression,'' which may act to reduce the perceptual effects of room responses, and that this mechanism acquires a representation (a constancy-mode ''solution'') of the reverberant context. It uses this representation for tuning its responses to subsequent stimuli. However, regardless of the actual mechanism responsible, the loudness difference between F–S and S–F stimuli demonstrates that the perception of loudness is sensitive to the shape of a stimulus' temporal envelope, as well as to the local context of successive stimulus presentations.

ACKNOWLEDGMENT

This research was supported by the National Institutes of Health (NIDCD Grant No. 00087), which the authors gratefully acknowledge.

APPENDIX: STIMULUS ENVELOPE GENERATION

Envelopes for these experiments were generated using the following function of time:

$$y(t)=t^{z-1}(1-t)^{w-1},$$

where

$$z=\frac{e^{s}}{-\left(\dfrac{1/2+\arctan(pt)}{\pi}\right)^{-1}},$$

and

$$w = \frac{e^s}{z}.$$

The parameter *pt* is used to control the envelope's peak time. When $pt = 0$, the resulting function is temporally symmetric, with equal rise and fall times. Negative values of *pt* result in temporally asymmetric functions, with shorter rise times and longer fall times. Exactly the opposite case holds for positive values of *pt*. In the limit, where $pt = \pm\infty$, the attack or decay becomes instantaneous. Two functions with equal absolute values but opposite sign of *pt* are temporally reversed versions of one another.

The parameter *s* controls the envelope ''sharpness,'' or overall width. Varying *s*, by itself, produces changes in the steepness of both attacks and decays. For all envelopes used in these experiments, *s* was set to a value of 3.

[1]This is caused by cancellation in the output of the filter: the abrupt onset of the tone causes the filter to ''ring'' at its best frequency, while simultaneously passing the decay portion of the stimulus. Since the frequencies of the carrier and impulse response differ, they drift out of phase, and their addition results in a shortened (and attenuated) decay at the filter's output. For ramped tones, the initial ringing response is small relative to the stimulus coming in, and remains so as both grow in magnitude over the attack duration. Following the abrupt offset, the filter continues to ring, causing the output to become extended in time (Patterson, 1994a).

[2]Adaptation in AIM's temporal integration stage affects the frequency of ''strobe'' generation, which is triggered by local maxima in the input waveform. The strobe threshold adapts over time, responding rapidly to increases, but slowly to decreases in level. With each strobe occurrence, activity in the neural encoding stage is copied into a leaky buffer, providing AIM with a mechanism for temporal integration. Since strobing takes place frequently during the slow-rising portion of a ramped tone, and rarely during the corresponding portion of a damped tone, the ramped tone will produce more activity within the integration buffer.

[3]AIM Release 8.2, 1997, obtained at ftp://ftp.mrc-apu.cam.ac.uk/pub/aim

Akeroyd, M. A., and Patterson, R. D. (**1995**). ''Discrimination of wideband noises modulated by temporally asymmetric function,'' J. Acoust. Soc. Am. **98**, 2466–2474.

Boring, E. G. (**1952**). ''Visual perception as invariance,'' Psychol. Rev. **59**, 141–148.

Canevet, G., and Scharf, B. (**1990**). ''The loudness of sounds that increase and decrease continuously in level,'' J. Acoust. Soc. Am. **88**, 2136–2142.

Canevet, G., Scharf, B., Schlauch, R., Teghtsoonian, M., and Teghtsoonian, R. (**1999**). ''Perception of changes in loudness,'' Nature (London) **398**, 673.

Carlyon, R. P. (**1996**). ''Spread of excitation produced by maskers with damped and ramped envelopes,'' J. Acoust. Soc. Am. **99**, 3647–3655.

Cutting, J. E., and Rosner, B. S. (**1974**). ''Categories and boundaries in speech and music,'' Percept. Psychophys. **16**, 564–570.

Gilchrist, A. L. (**1977**). ''Perceived lightness depends on perceived spatial arrangement,'' Science **195**(4274), 185–187.

Gordon, J. W. (**1987**). ''The perceptual attack time of musical tones,'' J. Acoust. Soc. Am. **82**, 88–105.

Hartmann, W. M. (**1978**). ''The effect of amplitude envelope on the pitch of sine wave tones,'' J. Acoust. Soc. Am. **63**, 1105–1113.

Houtsma, A. J. M., Rossing, T. D., and Wagenaars, W. M. (**1987**). ''Effects of echoes,'' in *Auditory Demonstrations* [Institute for Perception Research (IPO), Eindhoven, The Netherlands].

Irino, T., and Patterson, R. D. (**1996**). ''Temporal asymmetry in the auditory system,'' J. Acoust. Soc. Am. **99**, 2316–2331.

Jameson, D., and Hurvich, L. M. (**1989**). ''Essay concerning color constancy,'' Annu. Rev. Psychol. **40**, 1–22.

Jesteadt, W., Luce, R. D., and Green, D. M. (**1977**). ''Sequential effects in judgments of loudness,'' J. Exp. Psychol. **3**, 92–104.

Lakatos, S., McAdams, S., and Causse, R. (**1997**). ''The representation of auditory source characteristics: Simple geometric form,'' Percept. Psychophys. **59**(8), 1180–1190.

Marks, L. E. (**1988**). ''Magnitude estimation and sensory matching,'' Percept. Psychophys. **43**, 511–525.

Moore, B. C. J., and Glasberg, B. R. (**1996**). ''A revision of Zwicker's loudness model,'' Acust. Acta Acust. **82**(2), 335–345.

Morton, J., and Marcus, S. (**1976**). ''Perceptual centers (*p*-centers),'' Psychol. Rev. **83**(5), 405–408.

Neuhoff, J. G. (**1998a**). ''Perception of changes in loudness-reply,'' Nature (London) **398**, 673–674.

Neuhoff, J. G. (**1998b**). ''Perceptual bias for rising tones,'' Nature (London) **395**, 123–124.

Patterson, R. D. (**1994a**). ''The sound of a sinusoid: Spectral models,'' J. Acoust. Soc. Am. **96**, 1409–1418.

Patterson, R. D. (**1994b**). ''The sound of a sinusoid: Time-interval models,'' J. Acoust. Soc. Am. **96**, 1419–1428.

Patterson, R. D., and Irino, T. (**1998**). ''Modeling temporal asymmetry in the auditory system,'' J. Acoust. Soc. Am. **104**, 2967–2979.

Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). ''Time-domain modelling of peripheral auditory processing: A modular architecture and software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.

Rock, I. (**1997**). *Indirect Perception* (MIT Press, Cambridge, MA).

Rosen, S. M., and Howell, P. (**1981**). ''Plucks and bows are not categorically perceived,'' Percept. Psychophys. **30**, 156–168.

Schlauch, R., Ries, D., DiGiovanni, J., Elliot, S., and Campbell, S. (**1998**). ''Intensity discrimination of ramped and damped tones,'' in *16th International Congress on Acoustics* (Acoustical Society of America, Seattle).

Sekuler, A., and Palmer, S. (**1992**). ''Perception of partly occluded objects: A microgenetic analysis,'' J. Exp. Psychol. **121**(1), 95–111.

Slaney, M., and Lyon, R. (**1990**). ''A perceptual pitch detector,'' in ICASSP90, 1990 International Conference on Acoustics, Speech, and Signal Processing (IEEE, New York), pp. 357–360.

SPSS (**1999**). SPSS: Statistical package for the social sciences, Version 9.0, Chicago, IL.

Stecker, G. C. (**1995**). ''The effect of context on perceptual center,'' First-year research paper, Dept. of Psychology, U.C. Berkeley.

Stecker, G. C., and Hafter, E. R. (**1998**). ''Temporal asymmetry and loudness: Frequency and duration effects,'' in *Proceedings of the 21st ARO Mid-winter Meeting* (Association for Research in Otolaryngology, St. Petersburg, FL).

Stevens, S. S. (**1956**). ''The direct estimation of sensory magnitudes—loudness,'' Am. J. Psychol. **69**(1), 1–25.

Vos, J., and Rasch, R. (**1981**). ''The perceptual onset of musical tones,'' Percept. Psychophys. **29**(4), 323–335.

Wessel, D. (**1995**). Personal communication.

Zwicker, E., and Scharf, B. (**1965**). ''A model of loudness summation,'' Psychol. Rev. **72**(1), 3–26.