Perception & Psychophysics
2005, 67 (4), 699-712

Sweep-induced acceleration in loudness change and the "bias for rising intensities"

ROBERT TEGHTSOONIAN and MARTHA TEGHTSOONIAN
Smith College, Northampton, Massachusetts

and

GEORGES CANÉVET
Laboratoire de Mécanique et d'Acoustique, CNRS, Marseille, France

A pure tone changing continuously in intensity shows sweep-induced fading (SIF) of loudness as intensity sweeps down and may show a lesser degree of sweep-induced enhancement (SIE) as intensity sweeps up (Canévet & Scharf, 1990); the former effect has been called *decruitment*, the latter *upcruitment*. An opposite effect—upsweeps being judged to show more loudness change than downsweeps—has been reported by Neuhoff (1998). These disparate results might stem from several procedural differences. We found that differences in the sweep's duration and intensity level did not account for the disparity, nor did the presence of a steady tone preceding the sweep. In a second experiment, direct judgments of sweep size, such as those Neuhoff's (1998) listeners made, were affected not only by sweep size itself, but also by the intensity at the end of the sweep. The latter effect was especially marked for upsweeps. Neuhoff's (1998) proposed "bias for rising intensities" was found only with a method for judging sweep size that is more sensitive to end level than to sweep size.

There is now a substantial body of evidence illustrating the accelerated loss of loudness that occurs when a tone starting at a moderate level decreases in intensity in a continuous manner: The end level of such a sweep is judged to be softer than that level presented alone, a phenomenon that has been called *decruitment* (Canévet, 1986) but, in this report, will be called *sweep-induced fading* (SIF). We know that this occurs for pure tones, and, under certain conditions and perhaps to a lesser degree, for noise. We also know that for tones, the magnitude of the effect decreases sharply as the duration of the sweep decreases below the region of 10–20 sec, and that it disappears entirely by 1 sec (Teghtsoonian, Teghtsoonian, & Canévet, 2000). (These temporal spans might be termed *macro*-durations to distinguish them from those of 250 msec or less [*micro*-durations], used, for example, by Stecker & Hafter, 2000.) It is also known that the effect does not survive a test in the contralateral ear: When the sweep is presented to one ear and the end level alone is presented to the other ear, the latter is judged to be as great as if it had been presented without the contralateral sweep (Schlauch, 1992), indicating a phenomenon that occurs at the periphery. Yet Schlauch has also shown that the magnitude of the effect is diminished when the listener performs a second, distracting task, suggesting that we are dealing with a peripheral process governed, to some extent, by a central factor, attention. Theoretical speculation about SIF has focused primarily on various forms of adaptation (e.g., Canévet & Scharf, 1990; but see also Teghtsoonian et al., 2000), although there is still no definitive account of the process and how it occurs. Despite this uncertainty, the empirical phenomenon is robust and striking in its magnitude: A 40-dB 1-kHz tone heard at the end of a 20-sec downsweep starting at 70 dB will sound only about a quarter as loud as that same tone heard alone.

Another approach to the study of continuously changing sound intensities has been reported by Neuhoff (1998). His listeners were asked to judge the size of an intensity sweep with a duration of 1.8 sec, spanning 15-dB ranges in the interval between 60 and 90 dB for both increasing and decreasing conditions. For a 1-kHz tone and a synthetic vowel, although not for a white noise, he found that upsweeps were consistently judged to show greater change than corresponding downsweeps did. He offered the speculation that such a perceptual bias for rising intensities might have survival value by making the approach of a fixed-level sound source more salient than the receding of that source over an equal distance. (The merits of that speculation are not the main concern of this report, although we will offer some comments on it in our Discussion section.) Yet, in the work on SIF, rising sweeps never show a larger range of judged loudness than falling sweeps do. It seemed important to explore further an approach to the perception of dynamic signals that apparently had yielded the opposite result. In our study, we attempted to illuminate the roles that might have been played by four factors that distinguished Neuhoff's (1998) work from SIF studies.[1]

1. In some of the research on sweep-induced loudness change, the initial level of the signal has been maintained

---

Correspondence concerning this article should be addressed to R. Teghtsoonian, Clark Science Center, Smith College, Northampton, MA 01063 (e-mail: rtex@smith.edu).

for several seconds, to make it easier for the listener to judge its loudness (e.g., Teghtsoonian et al., 2000). This feature, absent from Neuhoff's (1998) procedure, might have contributed to an underestimation of sweep-induced enhancement (SIE), relative to the degree of SIF.

2. A factor deserving closer attention is the duration of the sweeps. Whereas Neuhoff (1998) used durations of 1.8 sec, durations in SIF studies have been much longer, with only one study extending the range of durations down to 1.0 sec (Teghtsoonian et al., 2000).

3. A third factor that could have been responsible for the differences in the two sets of findings is the region of the loudness scale being explored. In most studies of SIF, reductions in loudness were observed during downsweeps at levels below 60 dB SPL, whereas Neuhoff (1998) used intensities from 60 to 90 dB SPL. Is it possible that different principles apply to the perception of intensity sweeps in these different regions of the full range?

4. Finally, there were differences in the methods of judgment used in the two bodies of work. Neuhoff's (1998) listeners were asked to mark, on a line, the point representing the loudness change of the sweep, where the left end was defined as *very small* and the right end as *very large*. In contrast, in much of the SIF literature, the method used was magnitude estimation, in which the listener assigns a number to represent the loudness of any given target. Thus, whereas in Neuhoff's (1998) paradigm the loudness span reported was a single judgment made by the listener, the corresponding value in the SIF literature has been a calculation based on separate judgments of the loudness of the start and end levels. We wished to know whether the methods themselves might account for the different results.

In an earlier report (Canévet, Teghtsoonian, & Teghtsoonian, 2003) we showed that classic decruitment effects could be obtained in the absence of an initial plateau if, in its place, an 800-msec burst at the same level as the initial level of the sweep was presented 3 sec before the sweep itself. A judgment of the loudness at the beginning of the sweep could be obtained by asking the listener to judge the 800-msec burst and report it before the onset of the sweep. This ensured that the acts of judging and reporting could not interfere with attending to the sweep itself, an important consideration when very short sweeps are studied. But we were still uncertain how our previously studied 10-sec initial plateau might affect the judgment of start level, so, in the first study to be reported here, we specifically compared plateau and no-plateau conditions.

In that same earlier report, we used sweeps that started or ended no higher than 75 dB SPL and found little evidence for loudness enhancement during upsweeps. But the results reported by Neuhoff (1998) included higher levels, so, in the first experiment to be reported here, we explored start and end levels up to 90 dB SPL.

EXPERIMENT 1

The Effects of an Initial Plateau, Short Sweep Times, and Higher Intensities on the Range of Judged Loudness

In this study, we addressed two main questions. First, could our original finding of SIF for a 20-sec sweep be regarded as an artifact of the initial plateau? Second, using our method of loudness judgment, could we find the bias for rising tones reported by Neuhoff (1998)?

Method

Observers. There was a total of 11 observers. Eight were Smith College undergraduates with no prior experience in the judgment of loudness or hearing continuous tone sweeps; they were paid for their participation. Three (including the first two authors) were experienced in both respects and were volunteers.

Procedure. We followed the same basic procedure as that reported earlier (Teghtsoonian et al., 2000). The listeners heard a 1-kHz tone monaurally via earphones and were given standard magnitude estimation instructions to judge loudness at the beginning and end of each sweep. We studied both a 40- to 70-dB span (the condition we used in earlier research) and a 75- to 90-dB span (an intensity range for which Neuhoff, 1998, reported the strongest effect). We employed both 20-sec sweep times (the condition in which we had previously reported strong SIF effects) and 2-sec sweep times (the condition in which we had obtained some SIF, but much less than for longer durations) close to the 1.8-sec value used by Neuhoff. We ran all the conditions as in our original studies, with a 10-sec plateau prior to the beginning of the sweep, and a no-plateau replication. In all the conditions, the listeners were tested with both upsweeps and downsweeps. Thus, there were 16 conditions, each presented twice in each of two sessions, in two successive random orderings. The sessions were separated by 3–5 days. Each sweep yielded loudness judgments of both start and end levels.

Certain features are unique to this study and should be noted. Plateau and no-plateau trials were randomly intermixed, but the listeners were given a verbal warning signal ("plateau" or "no plateau") to identify the coming condition. They were instructed to make their judgments within the 10-sec period for the plateau conditions. In the no-plateau conditions, they were to make their judgment of the starting level of the sweep "as quickly as possible" in the 20-sec duration condition; for 2-sec durations, they were advised to make the judgment of the initial level as soon as possible but to wait until the sweep was completed before reporting it, along with the judgment of the end level.

Results

Figure 1 shows the geometric mean magnitude estimates for the start and end levels under each condition for sweeps between 40 and 70 dB. Comparable plateau conditions are shown in vertical alignment; comparable sweep-duration conditions, in horizontal alignment. Judgments during upsweeps are shown by upward-pointing triangles; during downsweeps, by downward-pointing triangles. (This and the subsequent figures present the judgments for start and end levels themselves. However, for purposes of statistical analysis, the *ratio* of these start and end judgments, which eliminates variability due to individual differences in modulus, was taken as the measure of change in each sweep condition.[2] Details of those

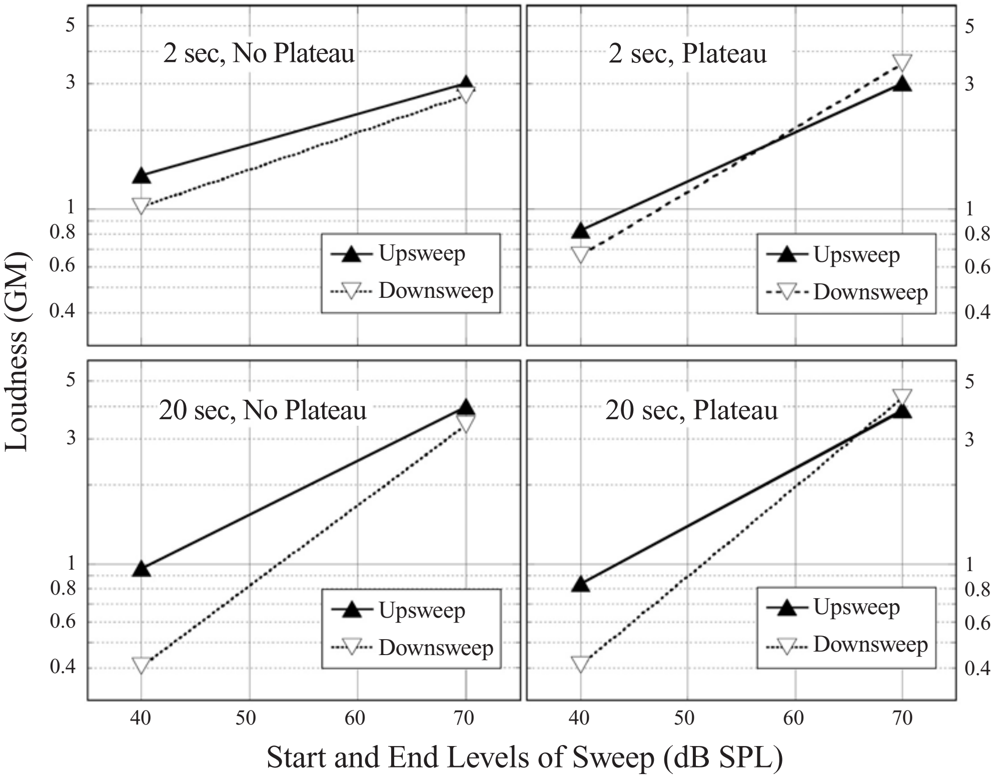

Figure 1. Loudness (geometric mean magnitude estimates, $N = 9$) for start and end levels of sweeps between 40 and 70 dB, lasting 2 or 20 sec, with or without an initial 10-sec plateau, in Experiment 1.

analyses are presented later; in the description of the results below, all differences noted were reliable [$\alpha = .05$] unless otherwise stated.)

The first question is whether the absence of a plateau eliminated SIF when the tone swept down between 70 and 40 dB in 20 sec. As can be seen in the two bottom panels in Figure 1, the answer is *no*: Both with no plateau and with a 10-sec plateau, loudness judgments span a greater range in the downsweep than in the upsweep. The slopes are similar between the two plateau conditions: In each, the downsweep range is about 10 to 1; the upsweep range is less than 5 to 1.

A second, related, question is whether the presence of a plateau produced SIF for a 2-sec downsweep between 70 and 40 dB. The top panels in Figure 1 show that, in each plateau condition, the slope is the same for both directions of the sweep.

Overall, then, we conclude that SIF does not depend on the presence of a plateau, since its absence before a 20-sec sweep does not erase SIF, nor does its presence before a 2-sec sweep create SIF.

The data for sweeps between 75 and 90 dB are shown in Figure 2. Comparable plateau conditions are vertically aligned; comparable sweep-duration conditions, horizontally aligned. The major question here is whether, at this higher range, with a 2-sec sweep and no plateau, the relation between up- and downsweeps will be reversed. Inspection of the top left panel in Figure 2 shows that the slopes for up- and downsweeps are parallel: The higher value is judged to be two to three times greater than the lower, regardless of sweep direction. There is no indication that rising intensities span a greater subjective range even when the stimulus conditions most closely duplicate those in Neuhoff (1998). In fact, in all four panels in Figure 2, it is evident that the judgment ratios are the same for upsweeps and downsweeps. Overall, judgment ratios are higher at longer durations and with a 10-sec plateau, but the same in each sweep direction.

The statistical analyses were two analyses of variance (ANOVAs) of ratios of judgment of end level to judgment of start level, where the main factors were sweep duration (2 and 20 sec), plateau duration (0 and 10 sec), and sweep direction (up and down). Judgments of the 15-dB span, from 75 to 90 dB, were analyzed separately from those of the 30-dB span, from 40 to 70 dB. (Two outlying observers were discarded from the statistical analyses—one who showed SIF effects an order of magnitude greater than those of any other observer, and a second who showed no SIF effects. It should be noted that (1) the means shown in Figures 1 and 2 are little

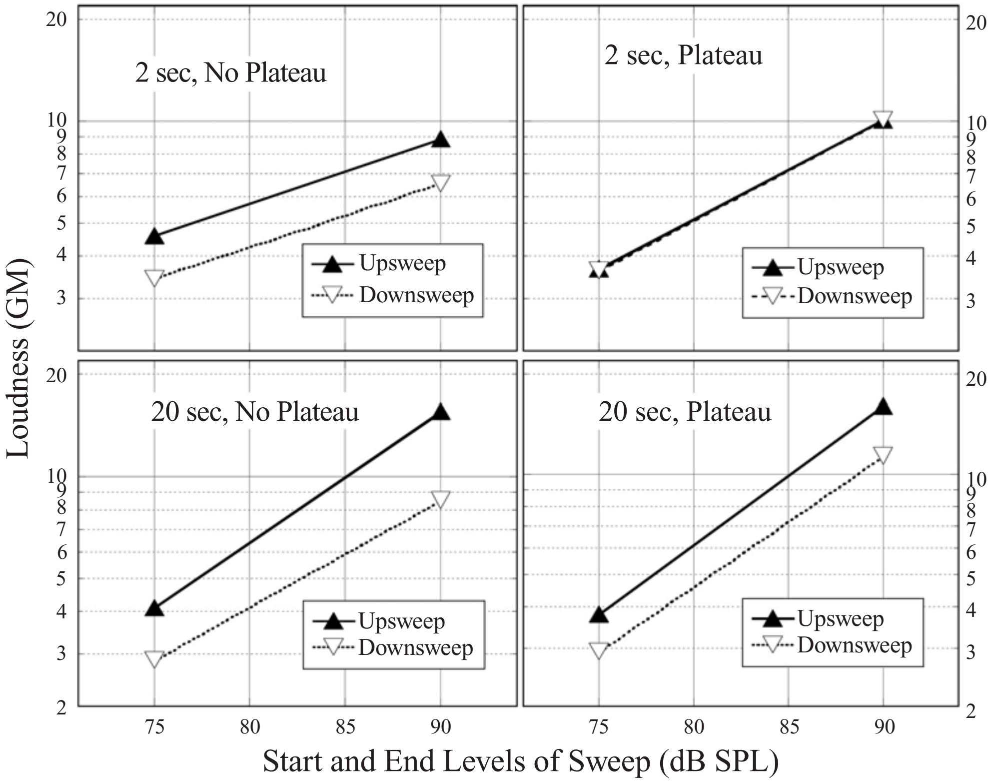

Figure 2. Loudness (geometric mean magnitude estimates, $N = 9$) for start and end levels of sweeps between 75 and 90 dB, lasting 2 or 20 sec, with or without an initial 10-sec plateau, in Experiment 1.

changed if these 2 subjects are included and (2) eliminating these subjects on the basis of SIF effects is neutral with regard to the effect of plateau.)

For sweeps from 40 to 70 dB, the results are easily summarized. All three main effects are reliable. The effect of plateau duration—a judgment ratio of 6.6 with a 10-sec plateau, but 4.6 without a plateau—is uncomplicated, since it interacts with neither sweep duration nor direction [$F(1,8) = 13.96$, $MS_e = 5.25$]. The interaction between sweep duration and direction is of the following nature [$F(1,8) = 10.10$, $MS_e = 12.05$]: At a sweep duration of 20 sec, the downsweep ratio of 5.3 is much larger than the upsweep ratio of 2.1, whereas the difference is small at a 2-sec sweep duration (2.0 vs. 1.6 for downsweep and upsweep, respectively). A test of the simple main effect of direction shows it to be reliable at 20 sec [$t(4) = 6.32$, $t_{crit}(4) = 2.13$], but not at 2 sec ($t = 0.85$). The three-way interaction is not reliable.

For sweeps from 75 to 90 dB, only the main effects of plateau duration and sweep duration are reliable; no interactions are reliable. For a 10-sec plateau, the judgment ratio is 4.3, whereas for no plateau, it is 3.2 [$F(1,8) = 16.08$, $MS_e = 0.915$]. For a sweep duration of 20 sec, the judgment ratio is 4.8, whereas for a duration of 2 sec, it is 2.7 [$F(1,8) = 12.31$, $MS_e = 4.33$]. For sweep direction, the judgment ratio for downsweeps is 3.9, whereas for upsweeps it is 3.6 [$F(1,8) < 1$].

One effect that is evident in both Figures 1 and 2 is that the initial judgment on upsweeps is higher with no plateau than with a 10-sec plateau and that, on downsweeps, it is lower. This effect was not anticipated, but post hoc analyses confirm its reliability. Separate ANOVAs of log start judgments (2 sweep durations × 2 plateau conditions × 2 start points) for 40–70 dB and 75–90 dB revealed only a single notable effect—a significant interaction between plateau duration and start point [for 40–70 dB, $F(1,8) = 23.60$, $MS_e = 0.012$; for 75–90 dB, $F(1,8) = 34.67$, $MS_e = 0.0064$]. Table 1 shows mean values and significance levels for tests of the simple main effects. Without an initial plateau to provide a basis for judging the loudness of the true level at the start of a sweep, up or down, the listener may instead report on his or her experience at a later point in the sweep, either the loudness at that point or a sort of averaging of the loudness expe-

Table 1
Judgments (Geometric Means) of Start Level on Upsweeps and Downsweeps

	40–70 dB		75–90 dB	
	No Plateau	Plateau	No Plateau	Plateau
Start low	1.14	0.84*	4.35	3.74
Start high	3.01	3.95*	7.49	10.73*

*Plateau mean differs from no-plateau mean [$t < t_{crit}(4) = 2.13$].

rienced over the initial portion of the sweep. Elevating the start point of an upsweep and depressing the start point of a downsweep will produce the displacement of the two functions seen in Figures 1 and 2.

It is also possible that, with no plateau, the end judgment on upsweeps might be elevated by SIE and the end judgment on downsweeps might be depressed by SIF. Comparison of side-by-side panels in Figure 2 shows that this is not the case. For both 2- and 20-sec upsweeps, the end judgment with no plateau is lower than the end judgment with a 10-sec plateau, just the opposite of SIE. For a 2-sec downsweep, the end judgment with no plateau is about the same as that with a 10-sec plateau, providing no evidence of SIF; the same is true of a 20-sec downsweep. (No statistical tests of these comparisons seemed warranted.)

If we compare results for 40–70 dB and 75–90 dB, other similarities and differences are suggested, although their reliability was not evaluated.

1. *Plateau duration*. The presence of a 10-sec plateau, as compared with no plateau, increases the ratio of end to start loudness—the judgment ratio—by about 45% for 40–70 dB and about 33% for 75–90 dB. This is because, as was noted above, in the absence of a plateau, start judgments are elevated on an upsweep and depressed on a downsweep, an effect particularly marked for 2-sec sweeps.
2. *Sweep duration*. The judgment ratio is higher for 20-sec sweeps than for 2-sec sweeps. For 40–70 dB, the amount of increase depends on the direction of sweep; it is about 30% for upsweeps, but about 100% for downsweeps. For 75–90 dB, the judgment ratio is about 60% higher for 20-sec sweeps at both sweep directions.
3. *Sweep direction*. The judgment ratio is higher for downsweeps than for upsweeps for 40–70 dB, but only for 20-sec sweeps, indicating the occurrence of SIF at 20 sec, but not at 2 sec. The judgment ratio is the same for both directions for 75–90 dB, indicating the absence of SIF at both durations.

What emerges clearly from these data is that the subjective size of the upsweep, as measured by the ratio of end to start judgment, is never greater than the subjective size of the corresponding downsweep, regardless of the presence or absence of a plateau, a short or long sweep duration, the range of intensities, or the absolute intensity level. Indeed, the only conditions in which there is any reliable difference are those with a 20-sec sweep duration, and the difference there is greater for the downsweep. This confirms our previous finding (Canévet et al., 2003) with magnitude estimations of the start and end of a sweep. It is the classic result called *decruitment* or, as we have termed it here, sweep-induced fading or SIF. We can find no indication of the "bias for rising intensities" reported by Neuhoff (1998), if that is understood as a greater ratio of loudness judgments made at the start and end of an upsweep.

EXPERIMENT 2
Direct Judgment of Loudness Change

In our second experiment, we considered the case in which degree of change in loudness is judged directly. We wished to test the assumption that direct judgments of sweep size provide a valid measure of the loudness change experienced by the listener. Since Neuhoff (1998) used only a single sweep size, 15 dB, his data cannot provide any information about the sensitivity of his measure to the nominal independent variable—actual sweep size. We note, too, that we have suggested that, in such a task, the observer may be responding not only to the size of the sweep itself, but also to the level at the end of the sweep (Canévet, Scharf, Schlauch, Teghtsoonian, & Teghtsoonian, 1999; Canévet et al., 2003). To evaluate the relative contributions of sweep size and end level to judgments of change, we explored several start and end levels (from 30 to 90 dB SPL) and two sweep sizes (15 and 30 dB SPL). Sweep duration was 1.8 sec, as in Neuhoff's (1998) report, and, as in that study, no plateau was used at either end of the sweep.

Neuhoff (1998) used a visual analogue scale, with the observer instructed to judge the amount of change by positioning a cursor along a scale marked "no change" and "large change" at opposite ends. In studies of SIF, observers are usually asked to judge the loudness of sweep end levels by the method of magnitude estimation. In this study, we asked the observers to judge the amount of change in each sweep (Neuhoff's task) by making magnitude estimations (our method).

Method

Observers. Students and members at the CNRS laboratory (N = 13, including 4 women and 9 men, with ages ranging from 19 to 57 years, with a mean of 31) participated in the study. All had normal hearing thresholds, as measured by a Békésy tracking audiometer (B&K 1800). The members of the laboratory and some of the students were experienced in magnitude estimation of loudness.

Stimuli. The stimuli were a 1-kHz tone and a broadband noise (0.1–15 kHz). For loudness sweep measurements, stimuli were ramped, with linear level variation in decibels; rise and fall times were set to 30 msec, and duration to 1.8 sec. There was no plateau at either the beginning or the end of the sweep. There were seven sweep ranges of either of two sizes—15 or 30 dB: 30–45, 45–60, 60–75, and 75–90 dB, and 30–60, 45–75, and 60–90 dB. For loudness function measurements, the stimuli had a constant amplitude over a duration of 800 msec and the same 30-msec rise and fall times. Seven levels were tested, from 30 to 90 dB, in steps of 10 dB.

Procedure. The experiment was run in an isolated soundproof room. The stimuli were presented through an appropriately calibrated Sennheiser HD 545 headphone to the right ear. Judgments were entered and recorded via a VT 320 terminal (see Canévet et al., 2003, for further details).

Before loudness sweeps were presented for judgment in the first test session, absolute thresholds and loudness functions (by magnitude estimation) were collected. For loudness function measurements, the seven test stimuli were presented seven times, with a different order each time and for each listener. After each presentation, a judgment was requested: The listener was instructed to assign a number to the loudness; there was no designated standard and no assigned modulus. Once the judgment was entered, a fixed pause of 2 sec was inserted before the next signal presentation.

Loudness sweeps were measured in three sessions, on 3 different days. In each test session, half the listeners were presented with all increasing sweeps first, followed by all decreasing sweeps; the other half received the reverse order. Pure tone and noise alternated from one trial to the next. The sequence of sweep ranges was selected at random for each listener. Each test signal was preceded by a warning light appearing on the listener's screen for 350 msec; then the listener was instructed to assign a number to *the amount of loudness change* perceived in each sweep. Again, there was neither designated standard nor assigned modulus. No plateau of any sort was used.

Results

The average absolute thresholds for the group were 11 and 16 dB SPL (at the eardrum) for the tone and the noise, respectively. Loudness judgments (for 12 listeners only) for the 1-kHz tone follow a power function of sound pressure with an exponent of 0.43, as is shown in Figure 3 (left panel). Those for the broadband noise have the usual curvilinear shape (Figure 3, right panel).

Tone. We will consider first the sweep judgments for the tone. These are presented in the following way. The geometric mean magnitude estimates of the amount of loudness change in a given sweep are plotted on a logarithmically scaled ordinate; the end levels of the sweeps in decibels are plotted on the abscissa. In such a plot, the effect of the end level is reflected in the slope of a line fitted to the points. At one extreme, if end level has no effect on the judgment, the slope will be zero; at the other extreme, if end level controls the judgment, the slope will approximate the magnitude estimation exponent for loudness. The intercept, although not independent of the slope, reflects the effect of sweep size; other things being equal, a difference in intercepts reflects a difference in sweep sizes.

Figure 4 shows the judgments of loudness change in downsweeps (left panel) and upsweeps (right panel). Open symbols represent judgments of a 15-dB sweep; filled symbols, those of a 30-dB sweep. The lines show least-square fits. The dotted line is the function fitted to magnitude estimates of loudness shown in Figure 3. As in reporting Experiment 1, differences noted in the description of results are reliable ($\alpha = .05$) unless otherwise stated. A detailed description of the statistical analyses follows the presentation of the results.

To summarize, Figure 4 shows that, whereas judgments of amount of change in these 1.8-sec sweeps were influenced by both the sweep size and the end level of the sweep, the relative balance between these two factors differed markedly between down- and upsweeps. Surprisingly, sweep size itself, the ostensible target of judgment, exerted little control in the upsweep: A 30-dB change was judged, on average, to be only about 18% greater than the 15-dB sweep (see the vertical separation of the two functions in the right panel of Figure 4), whereas, given the loudness function for the tone shown in Figure 3, we would expect the 30-dB sweep to seem about twice as great as the 15-dB sweep. In contrast, sweep size exerted considerably more control on the downsweeps: A 30-dB change was indeed judged to be about twice as great as a 15-dB change—about what would be predicted from the loudness function.

Perhaps more surprisingly, end level, which was not anticipated to affect perceived change, exerted strong control in the upsweep, as well as some control in the downsweep. On the upsweep, the relatively steep slopes of the fitted functions indicate that for every 15-dB increase in end level, there was about a doubling in judged change—close to what would be predicted from the loudness function if these judgments were based entirely on the loudness of the end level. As can be seen in the left

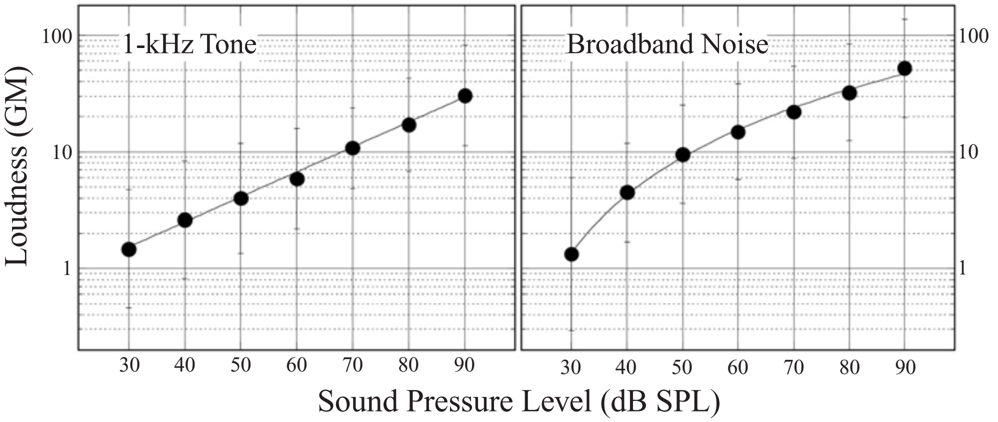

Figure 3. Loudness (geometric mean magnitude estimates, $N = 12$) of a 1-kHz tone (left panel) and a broadband noise (right panel), in Experiment 2. The fitted line (for 12 listeners only) for the 1-kHz tone shows a power function of sound pressure with an exponent of 0.43; for the broadband noise, the fitted line shows a power function of sound pressure level with a correction for threshold.

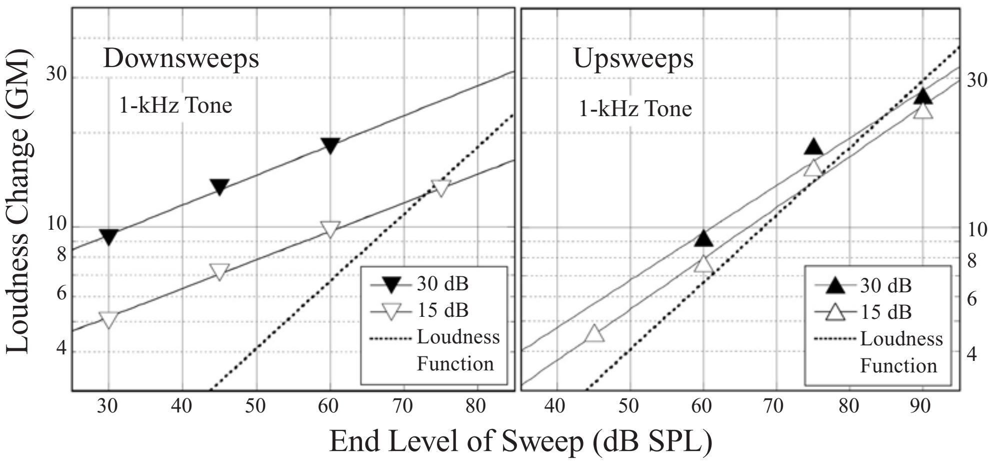

Figure 4. Loudness change (geometric mean magnitude estimates, $N = 13$) of a 1-kHz tone that swept down 15 or 30 dB in 1.8 sec (left panel) or swept up 15 or 30 dB in 1.8 sec (right panel), in Experiment 2. The dotted line represents the magnitude estimation function for loudness of a 1-kHz tone, as shown in the left panel of Figure 3.

panel in Figure 4, the fitted slopes are somewhat flatter: For downsweeps, each successive 15-dB increase in end level produced an increase in judged change of only about 40%. End level is clearly important in determining judged change in downsweeps but is not nearly as potent as in the case of upsweeps.

Even experienced subjects were unaware of the effect of end levels and expressed surprise that their judgments were influenced by any attribute of the signal other than the change in decibel level from the start of the sweep to the end. Nonlinearities in the relation between a physical attribute and its perceptual correlate are commonplace, but it is unusual for judgments of one attribute to be so highly dependent on the values of another, quite different, feature of the stimulus; indeed, such cases are often classified as illusions. Nonetheless, our data are clear in showing that, at least for 15- and 30-dB sweeps in the range between 30 and 90 dB, judged sweep size is governed to some degree by the level of the end point, and the magnitude of the effect depends on the direction of the sweep: It is moderate for downsweeps and quite large for upsweeps.

We are now in a position to understand how a *bias for rising tones* could emerge from this phenomenon. In Figure 5, the data are plotted to show a direct comparison of upsweeps and downsweeps between the same end levels. As in Figure 4, judged sweep size is plotted against the level at the end of the sweep, but here the upsweep points are displaced along the abscissa toward the origin by 15 dB (left panel in Figure 5) or 30 dB (right panel), so that for any given sweep (e.g., from 30 to 45 dB or from 60 to 90 dB), the judgments for upsweep and downsweep are plotted at the same locus on the abscissa. The fitted lines are the same as those shown in Figure 4.

Given the fact that sweep direction alters the relative weight of sweep size and end level in determining judgments of change (as can be clearly seen in Figure 4), it follows that the tendency to judge upsweeps as greater than downsweeps will depend on the location of the sweeps (as is apparent in Figure 5). Judgments of change follow one function for downsweeps and another, steeper, function for upsweeps. At high levels, the advantage for upsweeps grows large, whereas at lower levels, it disappears, and the advantage may even be reversed. *Any attempt to determine the effect of sweep direction on judged change that relies on a single combination of sweep size and sweep location cannot reveal the relative contributions of sweep size and end level in determining those judgments*.

It should be noted that, if it were the case that the relative balance between end level and sweep size were the same for both directions of sweep, the upsweep would always be judged greater, as long as end level has an effect at all. Obviously, for any pair of end levels, the upsweep ends at the high level, and the downsweep ends at the low level. What is perhaps not obvious is that this difference alone will produce a larger judgment for the upsweep, because of the end-level effect.

The differences reported above were evaluated by the following statistical analyses. In all cases, $\alpha = .05$.

1. The effects of sweep size and end level of sweep were evaluated in ANOVAs, separately for upsweeps and downsweeps. Note that one 15-dB sweep was omitted in each analysis (downsweep from 90 to 75 dB, upsweep from 30 to 45 dB), since there was no matching 30-dB sweep with the same end level. In each case, the effects of sweep size and end level were reliable, whereas their interaction was not. [For upsweeps, $F(1,12) = 13.37$, $MS_e = 0.0062$, for sweep size, and $F(2,12) = 50.83$,

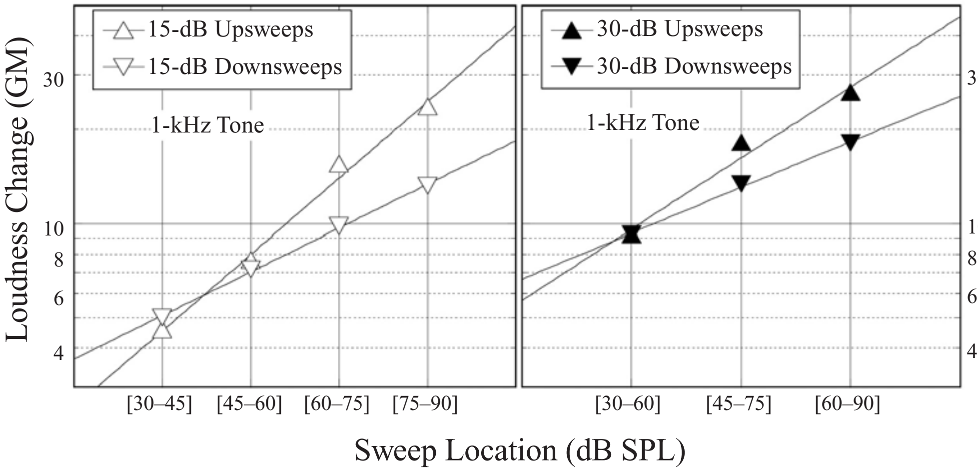

Figure 5. Comparison of loudness changes at the same sweep range for upsweeps and downsweeps of a 1-kHz tone, in Experiment 2. The data are the same as those shown in Figure 4, with data for 15-dB sweeps shown in the left panel and those for 30-dB sweeps in the right panel. As before, judged sweep size is plotted against the level at the end of the sweep, but here the upsweep points are displaced along the abscissa toward the origin by 15 dB (in the left panel) or 30 dB (in the right panel), so that, for any given sweep, the judgments for upsweep and downsweep are plotted at the same locus on the abscissa. The fitted lines are the same as those shown in Figure 4.

$MS_e = 0.029$, for end level. For downsweeps, $F(1,12) = 43.48$, $MS_e = 0.031$, for sweep size, and $F(2,12) = 9.63$, $MS_e = 0.057$, for end level. In each case, $F(2,24)$ for the interaction was less than 1.]

2. To determine whether the control exerted on judgments of change by sweep size differed between down- and upsweeps, we carried out the following analysis. For each subject, we found the geometric mean, for 15- and 30-dB sweeps, of the three judgments made at the three sweep ranges that entered into the ANOVAs. The ratio of the two values was taken, and the means of the resulting distributions for down- and upsweeps were compared in a *t* test. For downsweeps, a 30-dB change was judged to be about 1.8 times as great as a 15-dB change, reliably different from the 1.2 ratio for upsweeps [$t(12) = 4.45$].

3. To determine whether the control exerted on judgments of change by end level of the sweep differed between down- and upsweeps, we carried out the following analysis. Individual power functions were obtained by regressing each subject's log judgments on end level of sweep for the three values that entered into the ANOVAs, pooling over sweep size, separately for down- and upsweeps; *t* tests were used to evaluate differences between mean slopes of the resulting distributions. The exponent for upsweeps (0.31) was reliably greater than that for downsweeps (0.21) [$t(12) = 2.66$].

4. We compared the mean individual upsweep exponent (0.31) with the mean individual loudness exponent (0.47) over the range from 60 to 90 dB; a *t* test showed the difference to be reliable [$t(11) = 2.85$].

Noise. The picture for broadband noise differs somewhat, as is shown in Figure 6. The results are described first, with the details of statistical analysis following. The differences noted are reliable ($\alpha = .05$) unless otherwise stated. First, the judgments of a 30-dB change are consistently greater—by about 50%—than those of a 15-dB change, and the effect is similar for both sweep directions. Second, end level has an effect on judgments of change, and that, too, is the same—about 20% greater for each 15-dB increase—for both sweep directions. Third, the vertical placement of the functions is about the same for both sweep directions, indicating a similarity in the judgments for any given end level.

Figure 7 shows judgments of upsweeps and downsweeps for the same sweep locations, separately for 15- and 30-dB ranges. The small difference in judged change favors the upsweep in each case. Assuming that these small but consistent differences are in fact reliable, Figure 7 thus presents a puzzle: If judgments of change for upsweeps and downsweeps are affected in the same way by end level and by sweep size, and have the same absolute level, how can upsweeps be judged larger than downsweeps for the same sweep ranges? The answer lies in an observation we made in the discussion of the results for tones: Whereas the sweep *ranges* are the same, the upsweeps *end at a higher decibel level* than the downsweeps; since, as end level increases, so do judgments of change, the upsweep will be assigned a larger value.

Finally, it should be noted that this small advantage for upsweeps of noise was not statistically reliable in Neuhoff's (1998) initial study. Possibly the method of magnitude estimation or our method of analysis (or both) is more sensitive to small differences than is the visual analogue scale and analysis used by Neuhoff.

The judgments were subjected to analyses parallel to those described above.

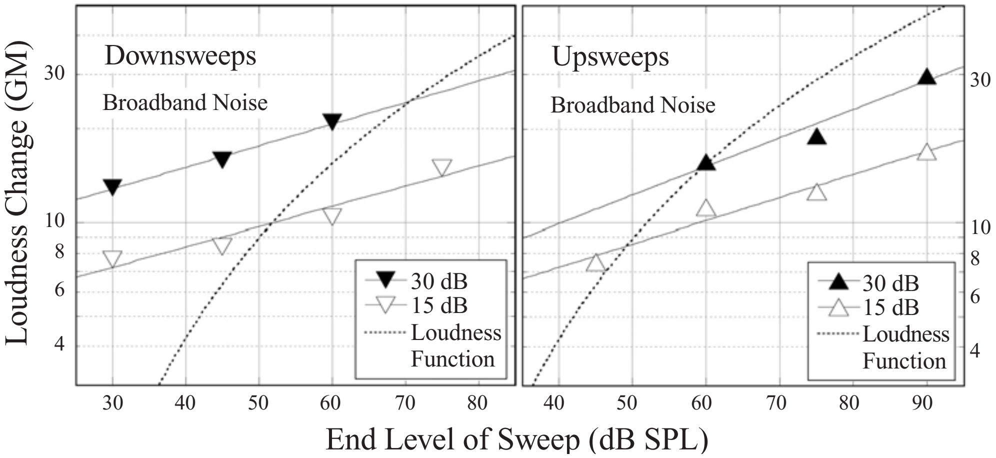

Figure 6. Loudness change (geometric mean magnitude estimates, $N = 13$) of a broadband noise that swept down 15 or 30 dB in 1.8 sec (left panel) or swept up 15 or 30 dB in 1.8 sec (right panel), in Experiment 2. The dotted curve represents the magnitude estimation function for loudness of a broadband noise, as shown in the right panel of Figure 3.

1. The effects of sweep size (15 vs. 30 dB) and end level of sweep (30, 45, and 60 dB for downsweeps; 60, 75, and 90 dB for upsweeps) were evaluated with separate ANOVAs of log judgments of change for downsweep and upsweep. Just as for the 1-kHz tone, both sweep size and end level have a reliable effect, whereas their interaction is not reliable. For downsweeps: $F(1,12) = 30.60$, $MS_e = 0.047$, for sweep size; $F(2,12) = 7.33$, $MS_e = 0.027$, for end level; and $F(2,24) < 1$ for their interaction. For upsweeps: $F(1,12) = 45.32$, $MS_e = 0.016$, for sweep size; $F(2,12) = 24.72$, $MS_e = 0.014$, for end level; and $F(2,24) = 2.56$, $MS_e = 0.0060$, for their interaction.

2. The judgment ratios for 30 versus 15 dB are 1.85 for downsweeps and 1.55 for upsweeps, a difference in the same direction as for the tone, but not a reliable one [$t(12) = 1.14$].

3. The mean individual exponents (for log judged change regressed on end level of sweep) are 0.11 for downsweeps and 0.15 for upsweeps, not reliably different by a *t* test [$t(12) = 1.08$]. Since the exponents do not differ, we also tested for differences in mean individual intercepts (0.82 for downsweeps and 0.64 for upsweeps); the difference is not reliable by a *t* test [$t(12) = 1.36$].

To summarize, these results resemble those for a 1-kHz tone in showing that both sweep size and end level influence judged change and that there is no interaction between them; they differ from those for a 1-kHz tone, in that the relative balance between the two factors is the same in downsweeps as in upsweeps. The influence of end level is not as strong for the noise in upsweeps: For tone, the exponent of the power function relating judged change to end level of sweep is 0.31, whereas for noise it is 0.16. And, for noise, the influence of sweep size is greater: For tone, the ratio of judgments for 30 versus 15 dB was 1.2, whereas here it is 1.55. The details of the way in which end level of a sweep determines differences in judgments of change for upsweeps and downsweeps differs between a 1-kHz tone and a broadband noise, but the larger picture is the same in both cases. Rising tones have an apparent advantage because they end at a higher level than falling tones spanning the same interval and because judgments of change are susceptible to the influence of the sweep's end level.

Why should the end level control judgments of change? One possibility is that listeners are uncertain about the start level when the sweep begins without a prior stimulus at the same level—either a continuous plateau or a short burst followed by silence. Comparing judgments of the start level of a 2-sec sweep with no plateau with those with a 20-sec plateau, we saw in Experiment 1 that, on an upsweep, judgments were larger and, on a downsweep, smaller when there was no plateau. However, the no-plateau judgments were no more variable than the plateau judgments, suggesting a constant error in locating the start point, rather than uncertainty about its location.

REANALYSIS OF NEUHOFF'S DATA

Do the effects we demonstrate here also occur in Neuhoff's data? In his 1998 experiment, it is impossible to tell, because a single sweep size of 15 dB was used in all the conditions. Thus, although listeners were instructed to judge the amount of change, there was no independent assessment of their ability to conform to the instructions. However, we can look at the relation between judgments of change and the end point of the sweep for the pure tone; this is shown in Figure 8. It is suggestive that judgments for both up- and downsweeps are influenced by the end level and that the influence is greater for up-

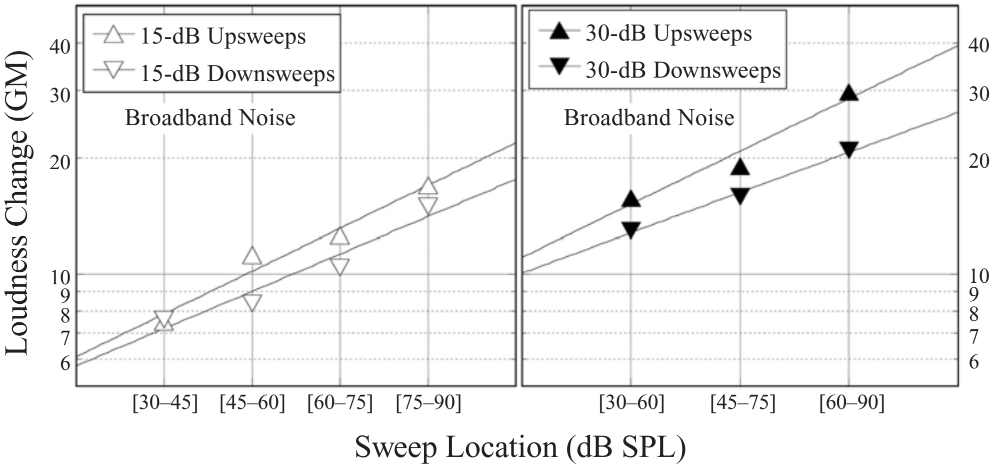

Figure 7. Comparison of loudness changes at the same sweep range for upsweeps and downsweeps of a broadband noise, in Experiment 2. The data are the same as those shown in Figure 6, with data for 15-dB sweeps shown in the left panel and those for 30-dB sweeps in the right panel. The fitted lines are the same as those shown in Figure 6.

sweeps. This is just what would be predicted by the outcome of the present experiment.

Neuhoff (2001) reported a similar study, in which listeners were presented on each trial with two 30-dB sweeps, one up and one down, in 1.8 sec; the stimulus was either white noise, for one group, or a synthetic vowel sound, for another. They were instructed to say which sweep exhibited the greater amount of change and to indicate how much change by placing a cursor on an unlabeled visual analogue scale. His data are so presented that they cannot be plotted as in Figure 8. However, there were significantly more choices for the upsweep when the end level was 90 dB than when it was 70 dB, as would be expected from the results of Experiment 2. This outcome suggests that the influence of end level on judged change that we have demonstrated here operated in Neuhoff's (2001) study as well.

DISCUSSION AND CONCLUSIONS

One major conclusion is that listeners asked to judge directly the size of short continuous sweeps of an auditory signal are heavily influenced by the level reached at the end of the sweep. In fact, the judgment of perceived change in an auditory sweep can be modeled as a power function of the decibel level of the stimulus at the end of the sweep. Its exponent reflects the influence of the terminal level of the sweep, and its intercept, in part, the influence of the sweep range. The difference in judgments between upsweep and downsweep will depend on the re-

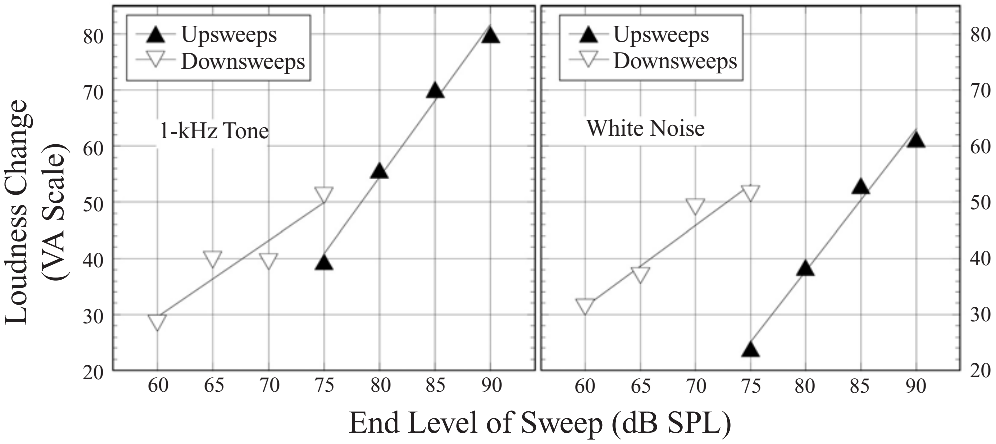

Figure 8. Loudness change (mean visual analogue scale values, $N = 12$) of 15-dB sweeps, as reported by Neuhoff (1998), for a 1-kHz tone (left panel) and a white noise (right panel). Loudness change on a linear scale is plotted as a function of the end level of a sweep in decibels.

lation of the two functions. If, as for a 1-kHz tone, the functions intersect within the experimental range of sound pressure levels and the upsweep function has a higher exponent, the upsweep will be assigned a higher judgment than the downsweep above the intersection. If, as for a broadband noise, the functions are parallel, the relation between judgments assigned to the upsweep and downsweep throughout the range of sound pressure levels will depend on the intercepts—that is, on the degree to which sweep size controls judgments in each condition.

A second major conclusion bears on the choice of method for evaluating perceived sweep size. Neuhoff (1999) argued for the hypothesized superiority of direct judgments of change, as compared with "static judgements of loudness at discrete points in time" (p. 673), for studying dynamic signals. But we are now able to say that this method is afflicted by a susceptibility to the biasing effect of the sweep's end level and that the magnitude of this effect depends on the nature of the stimulus and the direction of change. It is questionable whether such a procedure can form the basis for valid conclusions about the perception of dynamic signals, much less about the survival value of sweeps in one direction as opposed to the other.

Third, we can ask whether Neuhoff (1998) is correct in asserting the existence of a *bias for rising intensities*. It seems to us that the answer depends on how the question is asked. If listeners report the loudness at the start and end levels, we have clear evidence that the range of loudnesses experienced is never greater for upsweeps than for downsweeps. If, instead, listeners report only on the apparent size of the sweep, they do indeed say that short (under 2 sec) sweeps at levels over 60 dB SPL seem bigger on the upsweep than on the corresponding downsweep. However, we have shown that, when listeners make such judgments, they are more influenced by the level at the end of the upsweep than by the size of the sweep itself. For that reason, it may be more appropriate to speak of a *bias for end levels* when describing Neuhoff's (1998) finding.

Neuhoff (1998, 1999) has argued persuasively for the importance of studying dynamic signals, those whose intensity changes over time, in order to understand the information they convey in a way that static ones cannot. We agree that this is an important problem, one that deserves careful attention to the methods used in studying it. Neuhoff's approach contains three implicit assumptions that merit explicit statement and evaluation. First, he assumes that perceived change can be more accurately measured by asking observers to judge dynamic signals directly, rather than by inferring it from loudness judgments of start and end points. The assumption is plausible and, indeed, appealing in its face validity. However, we have presented evidence here that such direct judgments of change differ greatly in their sensitivity to real change—that in some conditions they are highly insensitive to change, and that in some conditions they may correspond more closely to the end level of the sweep than to the actual size of that sweep.

The second assumption is that the *dynamic* nature of the signal creates a difference between up- and downsweeps, yet we know of no evidence supporting such a view. The early studies of SIF (e.g., Canévet & Scharf, 1990) included a control condition in which the start and end levels of the downsweep were presented discretely, with an intervening silent period corresponding to the duration of the sweep. The difference between the judgments of the end level with and without a sweep was taken as a measure of SIF. Neuhoff has reported no similar control to justify the attribution of greater perceived change in the upsweep to the dynamic character of the signal.

The third assumption is that a phenomenon observed for sweeps 1.8 sec in duration can be generalized to real-world situations in which dynamic signals occur over a wide range of durations. In the study of SIF, we found that its magnitude was highly dependent on sweep duration (Teghtsoonian et al., 2000), and it would not be surprising to find that the same was true of any bias for rising intensities.

In other research, Neuhoff has pursued somewhat different approaches to the study of dynamic signals. Ghazanfar, Neuhoff, and Logothetis (2002) reported that rhesus monkeys orient longer to a rising-intensity complex tone than to one of falling intensity. Seifritz, Neuhoff, et al. (2002) presented human listeners with rising- and falling-intensity 1-kHz pulsed tones and identified brain areas that were specifically active during stimulation; they concluded, "Rising compared to falling intensity activated a distributed neural network subserving space recognition, auditory motion perception and attention" (p. 2147). We note here that the same comments made above apply to these later studies: (1) The size of the sweep was not varied, (2) there was no control condition presenting discrete stimulus levels sampled from the entire sweep, and (3) the duration of the sweep was not varied.

We suggest that the foundation for the claims that dynamic stimuli command unique perceptions and that upsweeps in intensity have greater perceptual salience than downsweeps is not yet firmly established and that it is thus premature to speculate about an evolutionary basis for the phenomenon.

REFERENCES

Canévet, G. (1986). Estimation de sonie pour des sons purs à variation monotone de niveau. *Acustica*, **61**, 256-264.

Canévet, G., & Scharf, B. (1990). The loudness of sounds that increase and decrease continuously in level. *Journal of the Acoustical Society of America*, **88**, 2136-2142.

Canévet, G., Scharf, B., Schlauch, R. S., Teghtsoonian, M., & Teghtsoonian, R. (1999). Perception of changes in loudness. *Nature*, **398**, 673.

Canévet, G., Teghtsoonian, R., & Teghtsoonian, M. (2003). A comparison of loudness change in signals that continuously rise or fall in amplitude. *ACUSTICA/Acta Acustica*, **89**, 339-345.

Ghazanfar, A. A., Neuhoff, J. G., & Logothetis, N. K. (2002). Auditory looming perception in rhesus monkeys. *Proceedings of the National Academy of Sciences*, **99**, 15755-15757.

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature*, **395**, 123-124.

Neuhoff, J. G. (1999). "Perception of changes in loudness": Reply. *Nature*, **398**, 673-674.

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology*, **13**, 87-110.

Scharf, B., & Fishken, D. (1970). Binaural summation of loudness: Reconsidered. *Journal of Experimental Psychology*, **86**, 374-379.

Schlauch, R. S. (1992). A cognitive influence on the loudness of tones that change continuously in level. *Journal of the Acoustical Society of America*, **92**, 758-765.

Seifritz, E., Neuhoff, J. G., Bilecen, D., Scheffler, K., Mustovic, H., Schächinger, H., Elefante, R., & Di Salle, F. (2002). Neural processing of auditory looming in the human brain. *Current Biology*, **12**, 2147-2151.

Stecker, G. C., & Hafter, E. R. (2000). An effect of temporal asymmetry on loudness. *Journal of the Acoustical Society of America*, **107**, 3358-3368.

Teghtsoonian, R., Teghtsoonian, M., & Canévet, G. (2000). The perception of waning signals: Decruitment in loudness and perceived size. *Perception & Psychophysics*, **62**, 637-646.

NOTES

1. After the original manuscript was submitted, a referee pointed out a fifth procedural difference: Neuhoff (1998, 1999) had used binaural presentation, whereas the studies reported here, like earlier research on SIF, used monaural presentation. The significance of this disparity was explored in replications that are described in the Appendix.

2. A reviewer asked why we calculated change in loudness as ratios, rather than as differences. We believe that the logic of ME procedures dictates that change between two values should be stated as a ratio. Observers are asked to make their judgments reflect ratio relations in their subjective impressions. At a purely practical level, we note that, since our results for upsweeps and downsweeps are so close together in absolute values, the results obtained using differences would not lead to any change in our conclusions.

APPENDIX
Partial Replications of Experiments 1 and 2, Using Binaural Presentation

Monaural presentation is commonly used in research on loudness, so, since the method of presentation was not specified by Neuhoff (1998), we mistakenly assumed that it was monaural. A referee pointed out our error and asked whether our Experiment 1 might have failed to support Neuhoff's results because the loudness of his signals was augmented by binaural summation. To answer this query, we undertook a partial replication of Experiment 1 with binaural stimulation, and the results are reported here. (Throughout this Appendix, we use the term *binaural* to refer to the case in which the same level is presented to both ears; this is sometimes termed *diotic* presentation.)

The critical condition in our Experiment 1 is the one most similar to Neuhoff's (1998, 1999) conditions, with a 2-sec signal duration, no plateau preceding the signal, and a sweep between 75 and 90 dB. Our original finding was that the ratio defined by the loudness judgments for the start and end levels was unaffected by direction of sweep. However, the loudness levels of our signals were lower than those of Neuhoff: The loudness of a tone presented binaurally is up to 1.7 times louder than the same tone presented monaurally (Scharf & Fishken, 1970). Accordingly, we undertook replications of that condition, using both binaural and monaural presentations, so that we would have a direct comparison within the same study. Eight listeners served in the binaural condition, but only 6 were available for retesting in the monaural condition.

The results are shown here in Figure A1 and should be compared with the original finding reported in the main text as Figure 2. The left panel shows the results for the monaural condition. It is clear by visual inspection that the original finding was confirmed: The ratios of the loudness judgments for end level are unaffected by sweep direction. It is noteworthy that the judgments in the replication are about twice as great as those in the original study. Since the replication was done in a different laboratory, with different equipment, and drew from a different subject population, it is not possible to specify the source of this difference. It should be remembered that the scale factor was not set by the experimenter and could, therefore, vary across individuals. Despite the disparity in the absolute level of judgments, the main conclusion is clear: As before, direction of change has no influence on the inferred loudness range.

The results for binaural presentation are shown in the right panel of Figure A1. Comparison with the left panel shows the expected increase in loudness; on average, the loudness was about 1.5 times greater with two ears than with one. However, the ratios of end-level loudnesses are approximately the same as those for monaural presentation, and, once again, there is no difference due to sweep direction. Thus, despite an unavoidable change in the laboratory in which the study was run and in the equipment used, the replication confirms our original finding and shows that it holds even when the loudness level is increased by a factor of 50%. We conclude that the findings reported for Experiment 1 were not artifacts of our monaural signals.

It was not obvious to us whether a replication of Experiment 2, using binaural presentation, was needed. Experiment 2 confirmed Neuhoff's (1998) finding that direct judgments of loudness change are bigger for upsweeps than for the corresponding downsweeps. The question at hand was not whether such a result occurs but whether such judgments are influenced by factors other than sweep size. Nevertheless, we carried out a replication of Experiment 2, using binaural presentation of a 1-kHz tone with a sample of 11 observers. The results are shown in Figure A2: The left panel is for downsweeps; the right panel, for upsweeps.

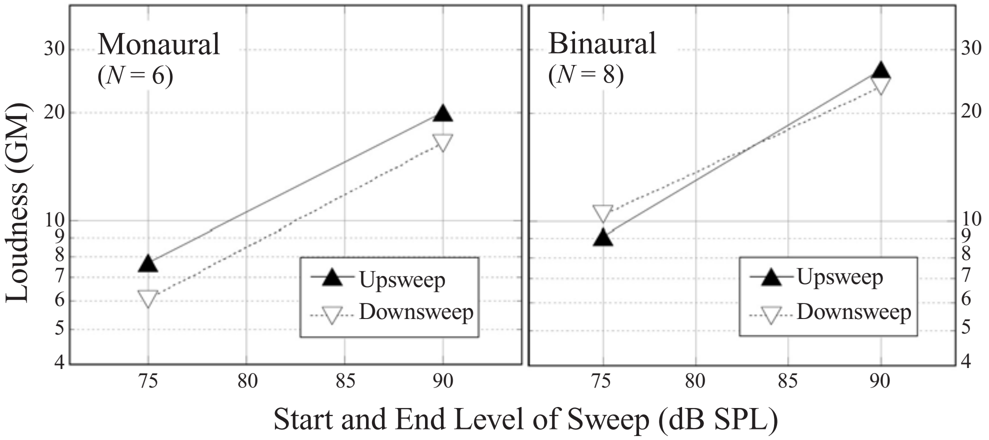

Figure A1. Loudness (geometric mean magnitude estimates) of a 1-kHz tone obtained for the start and end levels of a 2-sec sweep with no plateau, up from 75 to 90 dB and down from 90 to 75 dB. The left panel shows the results for monaural presentation (*N* = 6); the right panel, for binaural presentation (*N* = 8). See the Appendix.

APPENDIX (Continued)

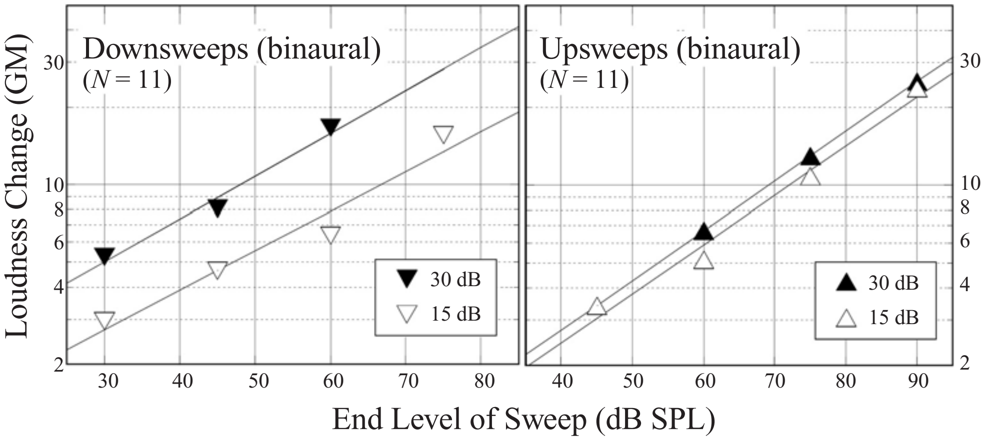

Figure A2. Loudness change (geometric mean magnitude estimates, $N = 11$) of a 1-kHz tone increasing or decreasing over a range of 15 or 30 dB SPL with binaural presentation, at different locations on the scale of tone intensity. The left panel shows the results for downsweeps; the right panel, for upsweeps. For example, in the right panel, the three filled triangles (going from left to right) show average judged loudness change for upsweeps from 30 to 60, 45 to 75, and 60 to 90 dB, respectively. See the Appendix.

The analysis is the same as that used in preparing Figure 4 in the main text; it shows judgments of loudness range as a function of a sweep's end level. In each panel, separate functions are shown for the two sweep sizes, 15 and 30 dB. Just as for the results with monaural presentation, there are two noteworthy features. First, for both sweep directions, end level plays an important role. If judgments were purely a function of sweep size, the functions shown in Figure A2 (and in Figure 4 in the main text) would be horizontal, rather than the increasing power functions apparent in both figures. Second, there is a marked difference in the effect of sweep size as a function of sweep direction. As was found for monaural presentation, in the case of downsweeps, judgments of sweep range are substantially larger for a range of 30 dB than for one of 15 dB. However, for upsweeps, these judgments are remarkably insensitive to actual sweep size; the judgment is determined almost entirely by end level. Whether the signal is presented to one ear or two has little if any effect on the basic form of our finding.

As in the case of our partial replication of Experiment 1, there are differences in the average scale factor employed by this sample of listeners, so a direct comparison of the vertical location of these functions cannot be made with those shown in Figure 4 in the main text. However, 5 of the 11 observers, along with 7 new observers, were retested under both binaural and monaural conditions and showed that binaural presentation results in judgments that are about 20% greater than those obtained with monaural presentation. Why should a fixed change sound larger when presented to two ears rather than to one? In this case, it is not loudness per se that is being judged, an attribute for which there is a large body of evidence showing some degree of summation in binaural presentation. When the same sweep is presented at the same level but to two ears, there is no a priori reason for any summation to occur. The fact that it *is* judged to be somewhat larger in size points once again to the fact that, in judging sweep size, end-level loudness plays an important role.

(Manuscript received January 20, 2004;
revision accepted for publication September 8, 2004.)