Identification and discrimination of sweep tones

M. E. H. SCHOUTEN
Institutes of English and Phonetics, University of Utrecht, Utrecht, The Netherlands

In Experiment 1, subjects were asked to identify the direction (up or down) of sweep tones centered around frequencies of 400, 1300, and 2700 Hz. Durations were 20, 30, 40, or 50 msec, and the rates of the (unidirectional) sweeps were 0, 5, 10, 20, 40, and 60 octaves/sec. The main result was that, on the whole, stimuli with zero- or low-sweep rates were judged to move "down," irrespective of the actual direction. Experiment 2 was a discrimination experiment, in which subjects had to discriminate between falling, rising, and level sweep tones centered around 1300 Hz. It turned out that discrimination between rising and falling tones did not differ significantly from discrimination between rising and level tones, indicating again that level tones tend to be perceived as going down. The parameters were chosen in such a way that they resembled those of formant transitions in plosive consonants; some implications with regard to speech perception are discussed.

In speech, rapid frequency transitions carry a great deal of information, especially about plosive consonants. It is therefore of interest to find out how such rapid transitions are perceived in isolation, as long as it is kept in mind that any psychophysical experiments involving frequency transitions should later be extended to experiments involving increasingly speech-like stimuli; this consideration will determine the selection of the various parameter values to be used, such as duration, center frequency, and rate of frequency change. (I do not mean to say that it would be uninteresting to increase our knowledge about the perception of rapid frequency transitions as such, but merely that the ultimate aim of the experiments described in this paper is to discover the role of such transitions in speech perception.)

It was thought appropriate to start our investigation of frequency transitions with transitions involving pure tones—"sweep tones" from now on. Since these sweep tones had to be comparable to formant transitions in speech, their center frequencies were chosen at 400, 1300, and 2700 Hz, within the ranges of the first, second, and third formants, respectively. Note that, in order to make it possible to combine various sweep tones into one slightly more complex stimulus, the frequencies are not exact multiples of each other.

The possible sweep rates to be considered covered a wide range. Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967) give an extreme example of an F1 transition going from 120 to 720 Hz in 50 msec, that is, around 2.5 octaves (oct) in 50 msec, or at a rate of 50 oct/sec. In a similar way, transitions described by Howell (1980) can be recalculated as 4 or 8 oct/sec for

F2 and 15 oct/sec for F3, with all durations 40 msec. It therefore seemed prudent to have a range of sweep rates running from 0 oct/sec to over 50 oct/sec, while retaining the option to modify the ranges to be used in the later experiments on the basis of the results of the earlier experiments.

Sweep duration was at first fixed at 30 msec, since Nabelek and Hirsh (1969) had found that sweep discrimination was best at 30 msec and since this seemed to be an appropriate duration for formant transitions; after a pilot experiment, a more extended range of durations was adopted.

In speech, a formant transition is seldom, if ever, surrounded by two steady states, so in speech these transitions cannot be perceived as differences between two successive steady states: the formant transition is either preceded or followed by a steady state, in the form of a vowel. Initially, it was therefore decided to include position of the steady state as an experimental variable: a steady state could occur before or after the sweep or could be left out completely. As a result, the present experiments cannot be directly compared with those of Tsumura, Sone, and Nimura (1973) or Arlinger, Jerlvall, Ahrén, and Holmgren (1977), who used steady states at both ends of their sweeps, or with that of Horst (1982), whose bell-shaped bandsweeps started and ended at the same steady state. However, in a pilot experiment conducted at the start of the present series of experiments, it appeared that this variable (position of the steady state) had no effect at all on subjects' responses. Consequently, although one cannot entirely rule out any effect of steady-state position, this variable was dropped from the main experiments. In the pilot experiment, the three center frequencies were mixed, and it turned out that subjects judged sweeps around 400 Hz to be "falling" and sweeps around 2700 Hz to be "rising," regardless of the actual direction and of the steady-state position. The subjects were apparently judging pitch—rather than direction—and giv-

ing it the nearest available label (the same thing happened in the condition without a steady state). If they had been judging direction, position of the steady state might have produced an effect. This question awaits further investigation.

In a fairly theoretical paper (Schouten, 1980), I had taken the position that plosive consonants were probably identified on the basis of a psychophysical threshold for sweep tones: above a particular sweep rate, *sweep direction* should become detectable, resulting in a ternary classification of sweeps: upward, downward, or no change. Although this must be a fairly peripheral process, it seemed to be of interest to learn whether or not the peripheral classification could be maintained in subjects' responses: So, in addition to a *sweep discrimination* experiment (Experiment 2), other experiments were included in which subjects had to report whether a sweep went up or down (stimuli without any audible change would presumably result in a band of 50% responses). The initial hypothesis, then, was that there would be three clearly delimited categories.

Since Sergeant and Harris (1962) and Pollack (1968) dealt only with slow rates of frequency change, there appear to be only a few studies in the literature that are relevant in the present context. The first one of these is Nabelek and Hirsh (1969). They asked subjects to discriminate among tone bursts in which an octave, a fifth, or a major second was covered in 10, 30, 100, or 300 msec, and found that discrimination was best at 30 msec. All stimuli ended at 250, 1000, or 4000 Hz and remained steady for a while at the terminal frequency. The results seem to suggest that there is no optimum rate of change, although rate of change was directly related to duration, which did have an optimum value. Another finding was that low rates of change were discriminated less well than high ones, a finding that seems to run counter to Cullen and Collins (1982), who found lower detection thresholds in noise for lower sweep rates. There is a measure of agreement on another aspect of sweep-tone stimuli: Cullen and Collins (1982), Gardner and Wilson (1979), and Nabelek (1978) all report thresholds for rising tones that were lower than those for falling tones.

Although all these findings suggest a clear perceptual imbalance between upward and downward sweeps, there are considerable differences between the experimental paradigms used. Nabelek (1978) compared detection thresholds for a number of sweep tones with the threshold for steady tones at the terminal frequencies of the sweep tones; his figures suggest, at least for durations of between 10 and 100 msec, that downward-sweep thresholds were closer to those of the steady tones than were the upward-sweep thresholds. The reverse was true of the results presented by Cullen and Collins (1982): the steady-tone thresholds they obtained were lower than the thresholds for rising sweep tones, which in turn were lower than the falling thresholds; in their experiment, the steady tones always had a frequency of 2000 Hz, which was the center frequency of all their sweep tones. Gardner and Wilson

(1979) performed a discrimination experiment in which rising tones turned out to be easier to discriminate from steady tones at the center frequency of 1000 Hz than falling tones. In an identification experiment involving isolated F2 transitions which subjects had to label as speech sounds, Pisoni (1976) found two categories: one consisted of the clearly rising transitions and the other contained the falling transitions, the steady formants, and the slightly rising transitions. Smoorenburg and Coninx (1980) found that the masking effect on steady tones diminished as the tones went from rising to falling to steady.

We thus have three studies indicating that falling tones are perceptually more like steady tones than are rising tones, and one study (Cullen & Collins, 1982) indicating the reverse. It therefore seemed useful to include steady tones in the present identification experiments, forcing subjects to label them as either rising or falling: this should show whether subjects tend to perceive them as rising, falling, or neither (in the latter case, responses should be distributed randomly over the two response categories). With respect to the three categories I expected to find (rising, steady, falling), it was predicted that the steady category would *not* be centered at 0 oct/sec: the results of three papers had suggested that the center of the steady category was somewhere in the falling half of the range.

There is one important respect in which the experiments described in the present paper differed from the vast majority of psychophysical experiments, and that is in the sophistication of the subjects. In psychoacoustics, the aim is usually to determine to what limits of performance the ear can be driven, whereas in speech perception research, the aim is, or should be, to find out how various sound stimuli function in human speech perception. This difference is one reason why psychoacoustics and speech perception research have so rarely met over common problems, a situation greatly to be regretted. The present research was an attempt to bridge the gap between the two: in addition to giving the parameters of psychoacoustic stimuli values appropriate for speech, it was also decided not to select experienced subjects and to give them no more than a minimal amount of training.

It is my strongly held conviction that, although the training of subjects has many advantages from a purely experimental point of view, its use severely restricts the wider validity of the experimental results. Consequently, psychoacoustics has generally remained a self-contained area with little overt relevance to what goes on in the perception of everyday sounds, such as speech sounds. Another reason for this is, of course, the choice of stimuli in psychoacoustic experiments. In this research, we used stimuli that were dynamic and had many characteristics derived from speech. Nevertheless, they were still a long way from normal speech sounds. For one thing, they occurred in isolation; for another, they were pure tones that changed in frequency rather than band filters that swept through the harmonics of a fundamental frequency. In a sense, therefore, they were purely psychoacoustic stimuli and could not be perceived and identified in the same way

as speech sounds. Even here, however, it was felt to be better not to train subjects extensively, but to see how they would classify these stimuli in their own way. This was expected to increase the relevance of the results to later, increasingly speech-like experiments.

The main disadvantage of doing without training was expected to be a rather wide spread in levels of performance among the subjects; in many cases, therefore, in addition to the average results, those for the best and poorest subjects will also be presented; it is felt that this will be more informative than just providing standard deviations.

Finally, a word about the identification task. The subjects were forced to classify stimuli as going up or going down; this seemed the only meaningful classification that could be used, although it has little to do with speech perception. The labels are not completely arbitrary, however: they made sense to the subjects, even to those who claimed to be unable to use them consistently.

In Experiment 1, two groups of subjects, under different experimental conditions, were asked to indicate whether a sweep tone they heard went up or down. In Experiment 1a, the subjects judged stimuli with different center frequencies in separate sessions; in Experiment 1b, pairs of sweep tones with different center frequencies, moving in parallel or opposite directions, were combined into single stimuli in an attempt to more closely approximate real speech.

EXPERIMENT 1a

Method

Stimuli. The stimulus parameters were: two directions (up and down), six sweep rates (0, 5, 10, 20, 40, and 60 oct/sec), four durations (20, 30, 40, and 50 msec), and three center frequencies (400, 1300, and 2700 Hz). The center frequencies were reached at the time centers of the sweep tones; the sweep tones were calculated using the formula

$$L(t) = \cos\left[\frac{2}{S \cdot \ln 2} \cdot Fc \cdot (2^{st} - 1)\right],$$

in which t is time in seconds (samples were 40 μsec and 10 bits), Fc is center frequency (400, 1300, 2700 Hz), and S is sweep rate in octaves/second. This is the same method as the one used by Smoorenburg and Coninx (1980), and it leads, of course, to exponentially rising and falling sweep tones. The lowest frequency reached was 141 Hz (50 msec, 60 oct/sec, around 400 Hz); the highest was 7637 Hz (50 msec, 60 oct/sec, around 2700 Hz).

Given the stimulus parameters, stimuli could not be made both to start and to end at zero crossings, but any time window used to counteract the inevitable wideband onset and offset spectra would have affected especially the shorter stimuli more than was thought permissible; for this reason, the unwanted transients were countered by adding 100 msec of low-level white noise at both ends of the sweeps. This noise was renewed after every 24 stimuli by sampling the output of a white-noise generator, with sampling frequency again being 25 kHz. Pols and Schouten (1981) have shown that low-level noise and time windows have the same smoothing effect, at least for the identification of plosive consonants. Some masking was thus unavoidable, but it was probably very small. It was, of course, no problem to smooth the beginning of an initial noise burst

and the end of a final noise burst by means of a 5-msec cosine window. Both noise bursts had durations of 100 msec.

Figure 1 presents an example of a stimulus (taken from a tape recording) with 100-msec bursts of white noise at both ends; it is a 100-msec sweep around 250 Hz, with a sweep rate of 20 oct/sec, going from 125 to 500 Hz. This sweep was not used in the experiment, but is shown here because its scale and parameters are such that they can demonstrate what goes on. The second noise burst is not shown in its entirety.

The stimuli were low-pass filtered at 12 kHz and recorded on audiotape. There was a separate tape for each of the three center frequencies. There were five different random orders of four duration blocks per frequency; sweep duration was constant in each block, but all six sweep rates occurred four times, twice up and twice down, resulting in 24 randomly ordered stimuli per block. Thus, each subject responded 10 times to each of the 2 (directions) × 4 (durations) × 6 (rates) = 48 different stimuli per frequency.

Procedure. A stimulus was presented every 2 sec over headphones in an anechoic room. The response sheets consisted of five pages of four columns each; each column constituted a constant-duration block of 24 stimuli. The end of a column was signaled by means of a 5-sec steady-state tone; the end of a page was signaled by two such tones. A response was given by ticking one of two boxes, the box on the left containing an upward-moving diagonal line and the one on the right, a downward one. The three center frequencies (400, 1300, 2700 Hz) were presented in counterbalanced orders with at least 1 week between them. A session took less than 1 h, including a brief training period that consisted of two stimulus blocks with durations of 50 and 40 msec. At no time was any feedback given, and training was deliberately kept short.

Subjects. Since Experiment 1a was run at a time when no funds were available for paid subjects, we had to fall back on colleagues at the Institute for Perception TNO. This is not such a random selection as one might have liked, since the majority of the subjects had quite often participated in psychophysical experiments. There were 12 subjects. Three quite experienced subjects said that they found it impossible to perform the task and demanded training. They had to be replaced by three uncomplaining subjects.

Results

The results are depicted in Figures 2, 3, and 4 for the average results, for the upper quartile of subjects, and for the lower quartile of subjects, respectively. Quartiles were determined on the basis of the overall numbers of correct responses; poor subjects were usually poor throughout all conditions. The ordinates of the figures do not represent correct responses, however, but percentages of "down" responses; the abscissas show the sweep rates
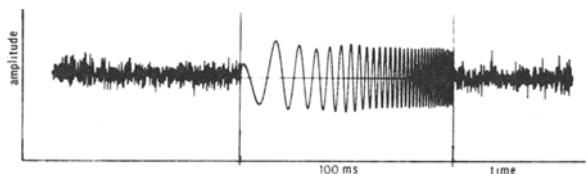

Figure 1. Example of the type of stimulus used in Experiment 1. For reasons of clarity, this is a 100-msec, 20-oct/sec sweep around a center frequency, going from 125 to 500 Hz, embedded on both sides in 100 msec of white noise; it was not used in the actual experiment. Not all the noise is shown. The apparent level fluctuations are due to the fact that only one of every three 40-μsec samples is displayed.
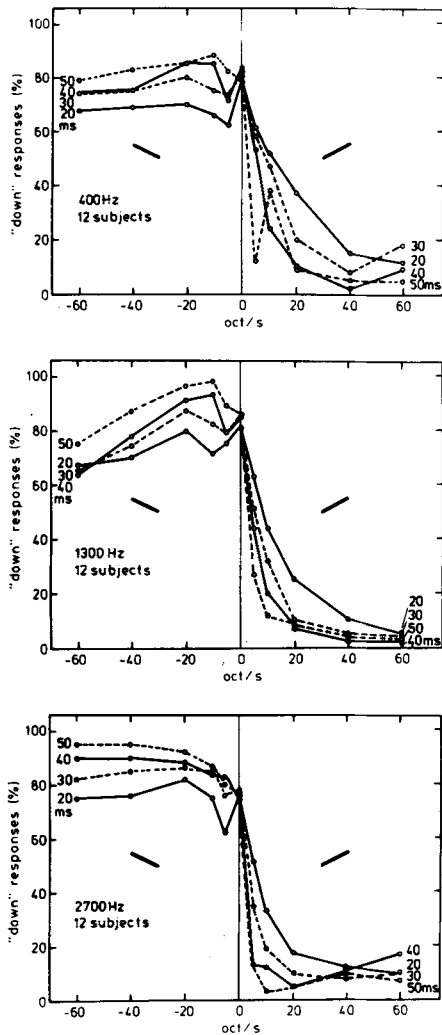
Figure 2. Average percentages of "down" responses in Experiment 1a. Negative values along the abscissa indicate rates of falling frequency; responses to steady tones (0 oct/sec) are situated on the thin vertical line in the center of the panels. The parameter within the panels is duration. Stimulus direction is indicated in each half of a panel by means of a rising or falling line.

in octaves/second, from steeply falling, through steady state, to steeply rising. Each figure consists of three panels for the center frequencies; the graph parameter is duration in milliseconds.

All the main effects were significant at the 1% level, as were all the two-way interactions involving subjects. Separate analyses of variance on the upper and lower quartiles showed that even in these cases the subjects factor had a significant effect (p < .01), as did the other factors, except frequency in both cases and duration in the case of the lower quartile. All analyses of variance were four-way (sweep rate × duration × center frequency × subjects, with only subjects regarded as a random factor); the steady-state tones were excluded from these analyses. A five-way analysis, which included direction as

a factor in its own right, revealed that this factor had no significant effect. All analyses were, of course, performed on the correct-score percentages and not on the percentages of "down" responses shown, for the sake of clarity, in the figures.

Discussion

Figure 2 shows that, on average, longer durations produce higher correct scores (in the left half of each panel, correct scores are the same as the "down" scores indicated, but in the right half, the curves stand for incorrect response percentages). The effect of duration is certainly not dramatic. The effect of sweep rate is more complicated: it is positive for the rising stimuli, but appears to depend on center frequency for the falling stimuli. Figures 3 and 4 show why this is so: although the upper quartiles behave as one would expect (a positive effect
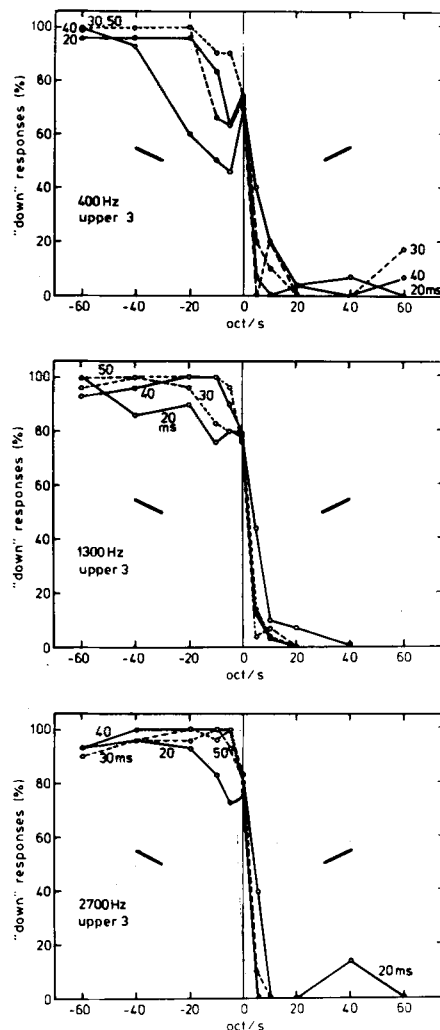

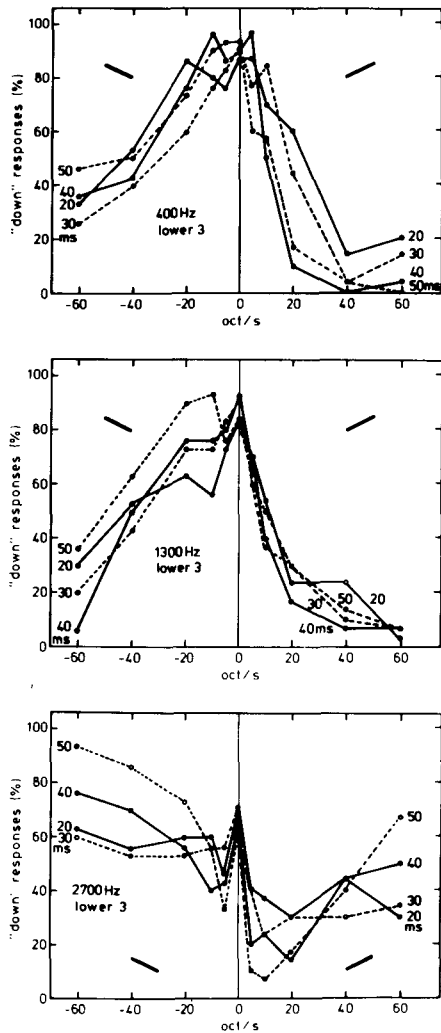Figure 3. Same as Figure 2, but averaged over the "upper 3" subjects.

Figure 4. Same as Figure 2, but averaged over the "lower 3" subjects.

of sweep rate), the lower quartiles give "up" responses to most of the steeply rising and falling stimuli and "down" responses to the slowly rising and falling stimuli, at least around the center frequencies of 400 and 1300 Hz. A curious pattern emerges from the lower quartile at 2700 Hz: here the steep stimuli receive more "down" responses and the shallow stimuli more "up" responses. A possible explanation might be that these poorer subjects could not follow the higher ends of the fast-sweep tones, so that the perceived average pitch of these fast sweeps was lower than that of the slow sweeps.

One thing that emerges clearly from all three figures is that steady tones (0 oct/sec) are regarded by all subjects as predominantly falling. The better listeners (Figure 3) do not allow this bias to influence their responses to even the slowest sweeps, but the poorer listeners (Figure 4) have a fairly wide range of sweep rates which they are apparently unable to distinguish from

steady tones and which, as a result, receive a majority of "down" responses.

The question about the perceptual imbalance between rising and falling sweeps appears to be getting an unambiguous answer: perceptually, there is a greater difference between rising and steady tones than between falling and steady tones, so in any experiment in which rising and falling tones are compared with steady ones, an imbalance must occur. We have found no evidence for a ternary categorization into rising, steady (or apparently steady), and falling tones.

EXPERIMENT 1b

Method

Stimuli. Two frequency combinations were used, namely 400 + 1300 Hz and 1300 + 2700 Hz: in every stimulus, two sweep tones of equal duration and equal sweep rate were simply added up, taking into account a −6-dB/oct slope, which reduced the level of the higher of the two sweep tones. The two tones in a stimulus moved either in parallel ("parallel" stimuli) or in opposite directions ("opposite" stimuli). Because of the results from Experiment 1a, some of the stimulus parameters were changed. There were now six sweep rates (0, 5, 10, 20, 30, and 40 oct/sec), and three durations (20, 30, and 40 msec). The stimuli were again either upward or downward; in the case of the opposite stimuli, these terms always refer to the *lower* of the two tones. The stimuli were embedded in noise and recorded in the same way as in Experiment 1a.

Again, there were blocks of 24 stimuli, during which duration and frequency combination were held constant; within a block, each of the six sweep rates occurred four times, namely in a parallel upward, a parallel downward, an opposite upward, and an opposite downward stimulus. Each block occurred five times with different internal random orders; this resulted in 5 × 2 (frequency combinations) × 3 (durations) = 30 blocks, recorded on one tape in five different orders. Each stimulus thus occurred five times.

Procedure. The procedure was the same as that in Experiment 1a, except that now the response sheets consisted of 7½ pages of four blocks, there was only one session covering both frequency combinations, and the training period used 40-msec stimuli in two blocks with different frequency combinations. The experiment took 35 min.

Subjects. There were again 12 subjects; all of them were first- and second-year students of English at Utrecht University. No selection was made—the subjects simply entered their names on a list on a notice board. They were rewarded in terms of course credits. No subjects were discarded for any reason other than defective hearing.

Results

The presentation of the results is now restricted to one figure, namely Figure 5, in which the scores averaged over all 12 subjects are shown, with the four stimulus types and two frequency combinations in four separate panels. The scores are again in percentages of "down" responses. Analyses of variance, performed separately on the parallel and the opposite data and leaving out the steady tones, revealed that only the subjects and rate factors had a significant effect in both.

The "upper 3" and "lower 3" quartiles are not shown graphically here; suffice it to say that the "upper 3" in this experiment were not as good as the "upper 3" in Experiment 1a, but that the pattern was the same: very lit-
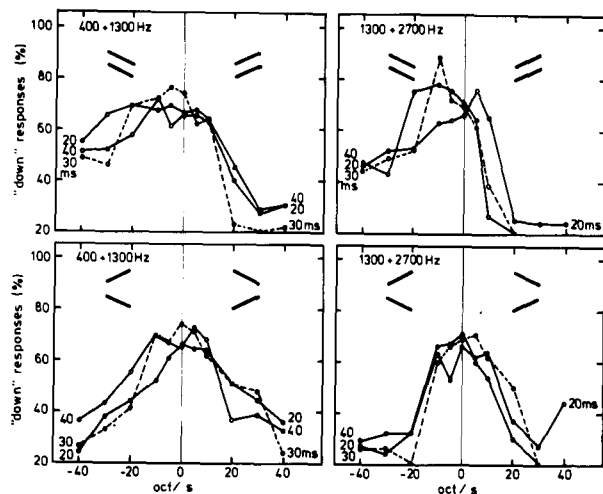
Figure 5. Average percentages on "down" responses in Experiment 1b. Stimulus direction is indicated by means of two rising or falling lines within each half of each of the four panels.

tle bias was evident in their responses to non-steady-state stimuli. The "lower 3" subjects here again showed the same bias effects as their counterparts in Experiment 1a. All this applies, of course, only to the parallel stimuli, with which a "correct" response was possible.

Discussion

The average subject in Experiment 1b was not as good a listener as the average subject in Experiment 1a; if the same processes are at work in both experiments, we should expect the overall score in Experiment 1b to be relatively more like that of the "lower 3" in Experiment 1a. If we look at the scores for the parallel stimuli, we see that that is indeed the case: downward stimuli show diminishing scores for increasing sweep rate. The opposite stimuli confirm this impression: low sweep rates are interpreted as steady-state stimuli; the greater the sweep rate, the more the average listener is inclined to regard it as going up.

The effect, then, is the same as that found in Experiment 1a: a very strong bias influence.

EXPERIMENT 2

The question subjects had to answer throughout Experiment 1 ("does this sweep tone go up or down?") was not a very natural or realistic one, since it is unlikely that the ability to recognize sweep direction plays any part in speech perception. An ability to *discriminate* rising and falling sweeps and steady tones is much more likely to be useful in speech perception: such an ability seems to be necessary for discrimination between rising, falling, and steady formants. From a speech-perception point of view, then, a discrimination experiment is probably more important than an identification experiment. In the light of the findings from Experiment 1, the stimulus param-

eters were changed slightly for Experiment 2; in particular, the range of durations was moved down the scale. It was also decided to restrict this experiment to one single center frequency, namely 1300 Hz, because Experiment 1a had shown little difference between frequencies.

Method

Stimuli. The stimuli were very similar to those of Experiment 1: exponentially rising, falling, and steady tones were surrounded by 100 msec of low-level white noise; the center frequency was 1300 Hz in all cases; the durations were 10, 15, 20, 25, and 30 msec; and the sweep rates were 0, 5, 10, 20, 30, and 40 oct/sec. A trial consisted of two pairs of stimuli, with three of the stimuli being the same and one stimulus being different from the other three stimuli; the subjects had to indicate which of the two pairs contained the different stimulus. There were 15 blocks of 40 trials each; within a block, all stimuli of all trials had the same duration. There were three conditions: (1) falling vs. steady, (2) falling vs. rising, and (3) rising vs. steady stimuli. Within a 40-trial block, only one condition obtained. In Condition 2 (rising vs. falling), only equal sweep rates were compared. This resulted in five different trials per block, one for each of the five sweep rates; within a block, there were eight replications, which differed in the position of the odd one out among the two stimuli in a trial: each of the two different stimuli was the odd one out half the time, and each occurred in all four positions in a trial.

The stimuli were generated in the same way as in Experiment 1, and were recorded on audio tape.

Procedure. The 15 blocks of trials were presented to subjects in one 45-min session, with short breaks between blocks; this was preceded by a brief training session in which feedback on correct responses was given. The subjects, who heard the stimuli over Sennheiser HD 414 reproducers mounted inside a pair of Peltor ear pads, were asked to indicate, by pressing a button marked "1" or "2," which of the two pairs of stimuli contained the odd one out. There were 2 sec between trials; the interval between the two pairs of stimuli was 500 msec, and the interval within a pair was 250 msec. Responses were stored automatically; occasional failures to respond were stored as "1" or "2" in a random way (the subjects had been told that no response was an incorrect response, and that guessing resulted in a 50% chance of being correct, and thus increased their chances of winning a bonus).

Subjects. There were 12 subjects, all of them students of English at Utrecht University, chosen at random. They were paid for their services; moreover, a bonus of 25 guilders was given to the subject giving the highest number of correct responses.

Results and Discussion

A five-way analysis of variance [sweep rates × durations × position of odd stimulus (in first or second pair) × conditions × subjects, with only subjects regarded as a random factor] showed that four of the five main effects were significant at the 1% level, but that the position of the odd one out did not have any influence on the results. The conditions factor was also significant at the 1% level, but accounted for much less variation in the data than did the other factors.

The results are shown graphically in Figure 6 as a function of condition, sweep rate, and duration. The upper and lower quartiles are not shown here, since variation within these quartiles was almost as great as that between them, if we disregard for the moment differences in the overall level of correct responses.
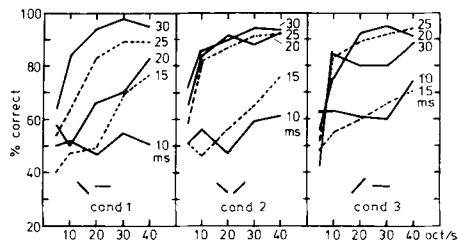
Figure 6. Correct sweep discrimination scores in Experiment 2 of, from left to right, falling vs. steady tones (Condition 1), falling vs. rising tones (Condition 2), and rising vs. steady tones (Condition 3).

Figure 6 makes it clear that, as we saw from the analysis of variance a moment ago, there is not a great deal of difference between the three conditions, although Condition 1 (falling vs. steady) does stand out a little. Since this could mean that most of the variance accounted for by the conditions factor is caused by the difference between Condition 1, on the one hand, and Conditions 2 and 3, on the other, two further analyses of variance were carried out on two subsets of the data: in one, the conditions factor was limited to Conditions 1 and 2; in the other, it was limited to Conditions 2 and 3. When only Conditions 1 (falling vs. steady) and 2 (falling vs. rising) were included, the conditions factor remained significant at the 1% level and the percentage of variance explained by this factor increased slightly but not to much higher than 1% (the same applied, incidentally, to Conditions 1 and 3). When the analysis consisted only of the Conditions 2 (falling vs. rising) and 3 (rising vs. steady), however, the conditions factor turned out to have no effect at all (p > .59). This indicates that Conditions 2 and 3 are essentially the same, and that discriminating between falling and rising sweeps does not differ much from discriminating between steady tones and rising sweeps, and hence that falling sweeps and steady tones have more in common than rising sweeps and steady tones. In absolute terms, the effect here (a significant difference between Conditions 1 and 2 and no difference between Conditions 2 and 3) is rather small, but it is in line with the most consistent finding from Experiment 1: There is a perceptual imbalance between rising and falling sweep tones, and this is due to the tendency on the part of subjects to hear steady tones as falling sweeps.

Figure 6 also shows that 10 msec is not enough for sweep discrimination and that, with a duration of 15 msec, well over 30 oct/sec of sweep rate is needed in order to get anywhere near a score of 75%, which in this type of experiment is often regarded as a meaningful threshold. If we do take 75% as our criterion, it can be said that for successful discrimination between falling sweeps and steady tones, the threshold is given approximately by the following combinations of duration and sweep rate: 20 msec and 35 oct/sec; 25 msec and 15 oct/sec. For discrimination of rising sweep tones, 10 oct/sec is enough, provided that duration is 20 msec or more.

GENERAL DISCUSSION

The conclusions to be drawn from the four experiments in this paper can best be summarized as follows: (1) Discrimination of rising sweeps requires sweep rates of at least 10 oct/sec and durations of at least 20 msec; discrimination of falling sweeps requires longer durations and/or higher sweep rates. (2) There is a perceptual imbalance between rising and falling sweeps, and it is caused by the fact that steady-state tones and sweeps with an imperceptible rise are somehow perceived as falling sweeps.[1] This is in agreement with findings by Gardner and Wilson (1979), Nabelek (1978), Pisoni (1976), and Smoorenburg and Coninx (1980), but runs counter to findings by Cullen and Collins (1982). Cullen and Collins used an adjustment method, in which subjects were given ample time to reach a "just masked" level. Theirs was an experiment with low stimulus uncertainty, whereas all the other experiments had high stimulus uncertainty. Although it is not quite clear how this difference in stimulus uncertainty influenced the results, it is true to say that it may change perception completely: according to Cutting (1982), for example, categorical perception of both speech sounds and nonspeech sounds is a function of stimulus uncertainty. It makes a great deal of difference whether one asks subjects to reach an instantaneous decision about the category a stimulus belongs to, or whether one allows them all the time they need to examine the stimulus in all its detail.

As indicated in the introduction, the choice of parameter values in the experiments was determined by what seemed to be relevant to the perception of speech and particularly to the perception of plosive consonants. Plosive consonants are at least partly cued by formant transitions; the results of Experiment 2 indicate that a rising transition needs a duration of at least 20 msec and a sweep rate of 10 oct/sec to be detected, and that a falling transition needs 20 msec and 35 oct/sec or 30 msec and 10 oct/sec. Anything between these is indeterminate; the results of Experiment 1 indicate that intermediate values will be regarded as falling transitions. Our tentative conclusion with regard to formant transitions is therefore that they have to rise at a rate of at least 10 oct/sec for at least 20 msec, or they will be "processed" as falling transitions. We cannot entirely exclude the possibility that the intermediate transitions form a separate category, but it seems more than likely that the left-hand panel of Figure 6 shows the detectability of differences within one category of sweeps rather than between two different categories.

The remarks just made with respect to speech should be treated with a great deal of caution. For one thing, formant transitions are not tonal sweeps, and they never occur on their own, but are always accompanied by other formants, which are bound to have some masking effect and thus to raise the threshold. In dealing with speech, we shall, as a result, probably have to mark up the threshold values we have found for sweep tones. More

importantly, unlike sweep tones, speech formant transitions are not characterized by a rising or falling pitch; they are more like band filters being swept through the harmonics of a fundamental frequency. It is still far from certain that perception of the outputs of such sweeping band filters will parallel perception of sweep tones. However, as a result of the experiments described in this paper, we may now know enough about the perception of sweep tones with speech-like parameters to make the jump to experiments on the perception of swept bands.

REFERENCES

ARLINGER, S. D., JERLVALL, L. B., AHRÉN, T., & HOLMGREN, E. C. (1977). Thresholds for linear frequency ramps of a continuous pure tone. *Acta Otolaryngologica*, **83**, 317-327.

CULLEN, J. K., & COLLINS, M. J. (1982). Audibility of short-duration tone glides as a function of rate of frequency change. *Hearing Research*, **7**, 115-125.

CUTTING, J. E. (1982). Plucks and bows are categorically perceived, sometimes. *Perception & Psychophysics*, **31**, 462-476.

GARDNER, R. B., & WILSON, J. P. (1979). Evidence for direction-specific channels in the processing of frequency modulation. *Journal of the Acoustical Society of America*, **66**, 704-709.

HORST, J. W. (1982). *Discrimination of complex signals in hearing*. Doctoral thesis, University of Groningen.

HOWELL, P. (1980). Pitch as a phonemic cue. *Memory & Cognition*, **8**, 285-296.

LIBERMAN, A. W., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.

NABELEK, I. V. (1978). Temporal summation of constant and gliding tones at masked auditory threshold. *Journal of the Acoustical Society of America*, **64**, 751-763.

NABELEK, I. V., & HIRSH, I. J. (1969). On the discrimination of frequency transitions. *Journal of the Acoustical Society of America*, **45**, 1510-1519.

PISONI, D. B. (1976). *Some effects of discrimination training on the identification and discrimination of rapid spectral changes* (Research on Speech Perception Progress Report No. 3, pp. 121-141). Bloomington: Indiana University.

POLLACK, I. (1968). Detection of rate of change of auditory frequency. *Journal of Experimental Psychology*, **77**, 535-541.

POLS, L. C. W., & SCHOUTEN, M. E. H. (1981). Identification of deleted plosives: The effect of adding noise or applying a time window. A reply to Ohde and Sharf. *Journal of the Acoustical Society of America*, **69**, 301-303 (L).

SCHOUTEN, M. E. H. (1980). The case against a speech mode of perception. *Acta Psychologica*, **44**, 71-98.

SERGEANT, R. L., & HARRIS, J. D. (1962). Sensitivity to non-directional frequency modulation. *Journal of the Acoustical Society of America*, **34**, 1625-1628.

SHORE, S. E., & CULLEN, I. K. (1984). Cochlear microphonic responses of the peripheral auditory system to frequency-varying signals. *American Journal of Otolaryngology*, **5**, 34-42.

SMOORENBURG, G. F., & CONINX, F. (1980). Masking of short probe sounds by tone bursts with a sweeping frequency. *Hearing Research*, **3**, 301-316.

TSUMURA, T., SONE, T., & NIMURA, T. (1973). Auditory detection of frequency transition. *Journal of the Acoustical Society of America*, **53**, 17-25.

NOTE

1. A recent paper by Shore and Cullen (1984) may provide part of the answer: They found that displacements in the cochlear partition in response to rising sweeps occur closely together in time, whereas those in response to falling sweeps are dispersed in time and appear to follow the course of the sweep. At any point in time, therefore, the excitation pattern of falling sweeps is like that of steady tones in that it shows a local maximum, whereas that of rising sweeps is vastly different. However, Shore and Cullen used durations (2, 4, 8, and 16 msec) and frequencies (6 to 12 kHz) that were very different from the ones used in the present research.