Frequency and duration discrimination of short first-formant speechlike transitions

van Wieringen, A.; Pols, L.C.W.

DOI
10.1121/1.408344

Publication date
1994

Published in
The Journal of the Acoustical Society of America

Frequency and duration discrimination of short first-formant speechlike transitions

Astrid van Wieringen and Louis C. W. Pols
Institute of Phonetic Sciences, University of Amsterdam, Herengracht 338, 1016 CG Amsterdam, The Netherlands

(Received 2 July 1992; revised 29 April 1993; accepted 30 August 1993)

Frequency and duration discrimination thresholds of short rising and falling one-formant speechlike transitions without a steady state were determined by means of same/different paired comparison tasks in two experiments. When frequency extent is varied (experiment 1), just noticeable differences *decrease* with increasing transition duration. Expressed in Hz, thresholds are, on average, 70, 63, and 58 Hz for 20, 30, and 50 ms, respectively. However, when transition duration is varied at a constant frequency extent (experiment 2), difference limens *increase* with increasing duration and are, on average, 2.7, 4.5, and 4.9 ms for standard transitions of 20, 30, and 50 ms, respectively. The thresholds determined in the two experiments indicate that different psychoacoustical cues are used depending on whether final frequency (experiment 1) or transition duration (experiment 2) are varied. Both experiments were performed at two different frequency regions (between 200 and 700 Hz and between 500 and 1000 Hz), but the results did not differ per region. In addition, no significant differences were found between rising and falling transitions. Particular attention was paid to a methodological issue, viz., the extent to which sensitivity changes as a result of different proportions of catch trials. It was found that the listeners maintained the same response strategies throughout the tests, as their performance is similar, irrespective of the number of catch trials included in the testing sessions.

PACS numbers: 43.71.Es, 43.66.Fe, 43.66.Mk

INTRODUCTION

Following an acoustical description, speech formant frequencies characterize the time-varying changes of the vocal tract. Of the continuously changing formant frequencies the relatively slowly changing ones are perceived as vowel-like, whereas relatively rapid changes in frequency are important acoustic cues for the perception of consonants.

In order to understand the auditory principles underlying speech perception it is necessary to determine the sensitivity of the hearing system for the relatively slow, as well as for the faster transients. The sensitivity to possible cues for vowel perception has been closely examined in the literature by determining the just noticeable differences in frequency, duration, loudness, and timbre of long and short *stationary* stimuli, ranging from sinusoids (Moore, 1973; Feth, 1974; Jesteadt and Bilger, 1974; Jesteadt and Sims, 1975; Jesteadt *et al.*, 1977; Wier *et al.*, 1977; Fastl, 1978; Fastl and Hesse, 1984; Tyler *et al.*, 1983; Nelson *et al.*, 1983) to speech formant stimuli (Flanagan, 1955a,b, 1957; Fairbanks and Grubb, 1961; Danaher *et al.*, 1973). The discriminability of cues for consonant perception is a more complex issue, not only because it involves an additional factor, i.e., transition rate, but also because the short, rapid, consonantlike glides (between 20 and 50 ms) have a broader bandwidth than the longer ones. In addition to this, as one dimension (e.g., frequency) of a glide always covaries with another one (e.g., duration), it is very difficult to isolate the auditory properties on which a subject's response is based. Discrimination can be based on transition rate, (end point) frequency, average frequency, duration, bandwidth, or a combination of such parameters. Interpretation of results becomes even more complicated with more speechlike stimuli, when fundamental frequency and additional varying resonances are involved.

To avoid some of the complexities associated with speechlike stimuli, many studies have measured the just noticeable differences in frequency and duration of *tone* glides (Sergeant and Harris, 1962; Pollack, 1968; Nábělek and Hirsh, 1969; Nábělek *et al.*, 1970; Nábělek, 1978; Tsumura *et al.*, 1973; Fujisaki and Sekimoto, 1975; Arlinger *et al.*, 1977; Gardner and Wilson, 1979; Collins and Cullen, 1978; Collins, 1984; Cullen and Collins, 1978, 1982; Tyler *et al.*, 1983; Schouten, 1985, 1986; Dooley and Moore, 1988a,b; Aslin, 1989; Cullen *et al.*, 1992). Comparison of these difference limens with those determined from stationary stimuli should give insight into the perceptual cues used in the discrimination of long and short dynamic stimuli. The issue is more complex, however, as difference limens determined from glides of the same duration show that transition rate affects discrimination differently, depending on whether frequency extent is varied at a constant duration (frequency discrimination) or duration is varied at a constant frequency extent (duration discrimination). In the present study we are interested in the discrimination of short, rapid transitions with more speechlike properties. This may improve our understanding of the perception of synthetic speech, and also that of some consonants in natural speech. In the next two sections, frequency (Sec. A) and duration (Sec. B) discrimination of short and long tone and formant glides will be reviewed.

A. Rate-of-change discrimination and frequency discrimination

Discrimination of fixed-duration glides changing in frequency may depend on various physical properties, such as (center) frequency, frequency extent, glide duration, the presence or absence of steady state(s), transition direction, and task. Because of the large number of the potential sources of information involved, it is impossible to draw firm conclusions about relevant auditory constraints. Furthermore, because of the interdependence of frequency, duration, and transition rate, it remains unclear whether discrimination is based on the sampling of end points, or on the differences in rate (sampling of different points in a glide), or on the difference between the initial and final frequencies.

Although frequency discrimination improves with increase of duration due to an increase in processing time (Nábělek *et al.*, 1969, 1970; Tsumura *et al.*, 1973; Dooley and Moore, 1988b; Elliott *et al.*, 1989, 1991; Porter *et al.*, 1991), the perceptual cues underlying discrimination may vary with transition duration. Whereas discrimination of relatively long glides, containing "sufficient" spectral information, may be based on pitch or timbre cues, that of short (< 50 ms) glides may rely more on rate of frequency change, with difference limens improving as rate of frequency change increases (Nábělek *et al.*, 1969, 1970; Porter *et al.*, 1991). Difference limens in frequency of short formantlike transitions in which rate of frequency change varies, range between 7.5 and 8.5 Hz/ms for 30-ms transitions and between 1.0 and 3.75 Hz/ms for 60-ms ones (Elliott *et al.*, 1989, 1991; Porter *et al.*, 1991).

Several studies have examined the discrimination of complex stimuli to determine whether, and to what extent, psychoacoustical properties found with tone glides also apply to more complex stimuli. Due to the large number of variables, firm conclusions are difficult to draw. Some findings are discussed below.

First, frequency difference limens of formant transitions (Danaher *et al.*, 1973; Mermelstein, 1978; Horst, 1982, 1989; Schouten and Pols, 1989; Cosgrove *et al.*, 1989; Elliott *et al.*, 1989, 1991; Porter *et al.*, 1991) are, in general, similar to those of tone glides, i.e., they also decrease with increase of transition duration. The modulation discrimination thresholds determined by Horst (1982, 1989) are comparable to those of stationary pure tones (1%–3.5%). However, relatively large differences are found for those transitions varying in a multiformant complex: The just noticeable differences of the second formant in 250-ms two-formant consonant-vowel combinations (Danaher, 1973) appeared to be 3%–4% of the reference frequencies (1100 or 2200 Hz), whereas those determined by Mermelstein (1978) in five formant synthetic stimuli ranged from 9%–14% of the formant frequency.

Second, dynamic cues influence discrimination of formant glides, as the frequency thresholds for 120-ms *transitions* were smaller than those of 300-ms *stationary* sounds (Elliott *et al.*, 1989). This difference between stationary stimuli and glides is larger than the one determined with experiments using tone glides (Dooley and Moore, 1988b), either as a result of the complexity of the stimulus or because the 750-ms glides of Dooley and Moore (1988a) were much longer than the 120 ms ones of Elliott *et al.* (1989).

Third, a rise/fall effect may disappear with more complex stimuli. Schouten and Pols (1989) performed discrimination experiments between exponentially rising and falling, rising and level, and falling and level band sweeps (with and without a steady state) of various rates and durations in 4-IFC paradigms. Contrary to the experiments with tone glides (Schouten, 1985), there was much less tendency to perceive level and slowly rising band sweeps, centered at 1300 Hz, as falling. When the band sweeps started or ended at 1300 Hz (with or without steady states) the results show no evidence of the asymmetry between falling and rising sweeps. Discrimination improves with increase in duration for stimuli moving to and from 1300 Hz, and, on the whole, also with increase in rate.

Despite this difference between tone and formant glides, the relationship between the complexity of the stimulus and the rise/fall effect is not clear (Nábělek and Hirsh, 1978; Collins and Cullen, 1978; Cullen and Collins, 1982; Schouten, 1985, 1986; Pols and Schouten, 1987; Moore *et al.*, 1989; Porter *et al.*, 1991; Cullen *et al.*, 1992). No significant differences were found by Nábělek and Hirsh (1969), nor by Arlinger *et al.* (1977) when performing experiments with sinusoids. Neither did Elliott *et al.* (1989, 1991) find different thresholds in frequency for rising or falling second-formant transitions, with and without a plateau. Porter *et al.* (1991), however, obtained smaller difference limens in frequency for falling second-formant transitions than for rising ones. Their data suggest that the relatively high-frequency region affects discrimination of transitions.

In conclusion, discrimination of transitions varying in frequency and duration cannot be explained solely as a function of transition duration nor of frequency extent. Together with transition duration, discrimination depends on the changing dimension, namely transition duration or frequency extent, the kind of stimulus, the presence or absence of a steady state, and the direction of the transitions. Although difference limens increase with stimulus complexity, discrimination of tone and formant glides appears to be based on similar psychoacoustical cues. A key question, therefore, concerns the extent to which discrimination is affected by the changing dimension, i.e., frequency extent, transition duration, and especially transition rate, as the latter varies in both experiments. In experiment 1, discrimination of fixed duration transitions is examined. In the next section discrimination of tone glides varying in duration at a constant frequency extent will be reviewed.

B. Rate-of-change discrimination and duration discrimination

The ability to discriminate differences in the rate of change of glides by varying the duration of stimuli with a constant frequency extent, (Pollack, 1968; Náběek and Hirsh, 1969; Dooley and Moore, 1988a) involves a covariation of rate of change and duration. As the difference in time required for temporal discrimination increases as the standard time increases (Creelman, 1962; Abel, 1972), it is difficult to determine whether discrimination is based on temporal or on spectral cues, or on a combination of both. Different psychoacoustical cues are involved in the discrimination task depending on whether frequency extent or transition duration are varied together with transition rate. Thresholds for this duration condition also increase with increasing duration. Relatively few studies have compared frequency and duration conditions. With infants, Aslin (1989) found larger difference limens for glides varying in transition duration than in frequency. However, the experiments differ in task and stimulus type (a steady state was added in the duration condition).

Dooley and Moore (1988a) compared discrimination in duration for relatively long steady and gliding sinusoids (standard: 750 ms). By asking listeners to indicate the longest stimulus of a pair, they took advantage of the covariation between duration and rate of change. Smaller difference limens were found for dynamic than for stationary stimuli, suggesting that glide rate was used as an extra cue in their experiments.

In order to prevent listeners from using the total duration of the transition as a cue, some studies allow the transition to be preceded or followed by a steady state (Pollack, 1968; Náběek and Hirsh, 1969; Náběek *et al.*, 1970; Tsumura *et al.*, 1973; Arlinger *et al.*, 1977; Collins, 1984; Aslin, 1989; Schouten and Pols, 1989; Elliott *et al.*, 1989; Porter *et al.*, 1991). However, listeners may still be able to detect differences in the durations of the glide, thereby making use of other cues than glide rate. In addition to this, the steady state may also interfere with the discrimination of glides. Tsumura *et al.* (1973) measured frequency discrimination as a function of transition duration and steady-state duration and found thresholds, especially those of short transition durations, to increase with decreasing plateaus. In experiment 2 difference limens of isolated single formant transitions are determined to examine psychoacoustic cues for glides varying in transition duration.

C. Present study

The present experiments are part of a study dealing with the perception of dynamic formant glides, in particular short stop-consonantlike transitions. The first-formant transitions of voiced plosives in natural speech rise and fall between approximately 220 and 700 Hz at a relatively high rate of change. In this study, the perceptual importance of transition rate as a cue to the discrimination of transitions is examined. As transition rate always covaries with frequency extent or transition duration, just noticeable differences for these short transitions will be determined in this study by varying the frequency extent, at a constant duration (experiment 1), as well as by varying the transition duration, at a constant frequency extent (experiment 2), in two separate experiments.

Although difference limens of various durations and frequency extents have already been determined (see Sec. A), those of short and rapid (plosivelike) formant transitions not bounded by plateaus have not been examined extensively. Whereas most studies include short and long transitions, we focus on short durations only (20–50 ms). In addition to this, difference limens of short formant transitions, varying in transition duration at a constant frequency extent, have not yet been reported in the literature. In speech, end point frequency typically signals place of articulation, whereas transition duration signals manner of articulation. However, within a short durational range the difference limens in both conditions can be compared: The first formant transition is similar for stop consonants. In an attempt to get more insight into the psychoacoustical cues used in the processing of short transitions we have used stimuli that differ from those of Dooley and Moore (1988) in that they are shorter (20–50 ms), and more complex. Unlike Náběek and Hirsh (1969), Elliott *et al.* (1989, 1991) and Porter *et al.* (1991) we examined the first formant region. In speech, the first formant transitions of the stop consonants /b/, /d/, and /g/ are generally similar when they precede or follow the same vowel. Two frequency regions were chosen, in one the first formant transitions are typical of stop consonants (between 200 and 700 Hz), in the other (between 500 and 1000 Hz) the pattern does not correspond to any linguistic category. Both frequency regions are relatively low, so that thresholds would not be expected to differ on the basis of the difference in frequency. However, it is possible that discrimination is influenced by a linguistic reference. Also, three different frequency extents (330, 430, and 530 Hz) are examined per testing region to determine whether transition rate affects discrimination of transitions of the same duration. Thresholds of both rising and falling transitions were determined in our study to examine whether the presence or absence of a rise/fall effect is correlated with transition duration (20, 30, and 50 ms), stimulus complexity, frequency region, or absence of a steady state.

In summary, discrimination in frequency (experiment 1) and duration (experiment 2) of relatively short, formant transitions are examined, of which, unavoidably, either the endpoint frequency or the transition duration covary with rate. This should reveal whether similar psychoacoustical cues are used in the two conditions and whether thresholds per condition are comparable with respect to direction, rate of frequency change, and frequency region. Furthermore, the extent to which sensitivity and response bias are separated, is examined by comparing the subjects' performance for different proportions of catch trials (18.75%–81.25%) included in the testing sessions.

I. GENERAL METHOD

The just noticeable differences in transition rate were measured by varying the frequency extent at a constant duration (experiment 1), and by varying the transition duration at a constant frequency extent (experiment 2). Thresholds of rising and falling transitions were measured as a function of transition duration, frequency extent, and frequency region. Difference limens were collected by means of the method of constant stimuli in a same/different procedure. A same/different task does not require the listener to recognize a specific feature of the signal, such as transition direction. This would be a very difficult task at very short durations, where listeners hardly even perceive the glide. In the method of constant stimuli the number of stimuli is predetermined by the experimenter, and the sequence of trials does not depend on the subject's response. Although time-consuming, this method yields an accurate estimation of the listener's sensitivity when taking the responses to catch trials (physically identical pairs of stimuli, which are intermixed in the testing session; expected response is "same") into account. If analyzed by the theory of signal detectability (TSD, Egan *et al.*, 1966, 1975), the subject's *bias* (which is present in all psychophysical procedures) can be separated from his *sensitivity*. In order for bias and sensitivity to be separated, two assumptions have to be met, i.e., that the data are normally distributed and of equal variance. The extent to which these assumptions are satisfied is examined by including varying proportions of catch trials in each testing series.

A. Stimuli

One-formant linear rising and falling transitions were generated by a digital formant synthesizer on a μVAX II (KLINKERS, Weenink, 1988). The glottal function (a pulse train) was held constant at a (fundamental) frequency of 110 Hz. The formant bandwidth was always 10% of the changing formant frequency. Two frequency regions were examined. The onset frequency of the rising transitions (or the offset frequency of the falling ones) was either 220 or 520 Hz.

Six different standard trajectories, three per frequency region (extents of 330, 430, and 530 Hz), and three different transition durations, 20, 30, and 50 ms were tested. These transition durations are typical of stop-consonant transitions in speech. Depending on the condition, either the frequency or the transition duration of the standard trajectory was changed. The rates of frequency change of rising and falling standard transitions is given in Hz/ms in Table I. Compared to Porter *et al.* (1991), whose rates of frequency change were 10, 5, and 2.5 Hz/ms for 30, 60, and 120 ms, respectively, our rates of frequency change are deliberately somewhat higher, because of the shorter transition durations.

To ensure a precise generation of these formant transitions, the formant frequency values were updated every 1 ms at a sample frequency of 1.2 MHz. After low-pass filtering, they were downsampled to 20 kHz (at a 16-bit resolution). A 2-ms Hamming window at both end points (1 ms for the 20-ms transitions) was used to avoid clicks,

TABLE I. Transition rates (in Hz/ms) of standard stimuli in the frequency and duration conditions for rising and falling transitions.

	20 ms	30 ms	50 ms
220–550 Hz 520–850 Hz	16.5 Hz/ms	11.0 Hz/ms	6.6 Hz/ms
220–650 Hz 520–950 Hz	21.5 Hz/ms	14.3 Hz/ms	8.6 Hz/ms
220–750 Hz 520–1050 Hz	26.5 Hz/ms	17.6 Hz/ms	10.6 Hz/ms

even though the first period of all the stimuli started on a zero crossing. The stimuli were then added to the center of 150 ms of low-level white noise (signal-to-noise ratio of approximately 50 dB) to minimize external auditory factors. Due to the different transition durations the amount of noise preceding and following the transition varied. The 16-bit stimuli were transferred from the μVAXII to an IBM-PC. They were presented using an OROS-AU22 D/A converter (the measured signal-to-noise ratio of the total system is 92 dB).

B. Stimulus characteristics

1. Frequency condition: Varying transition rate by end point frequency

In experiment 1 discriminability in transition rate of rising and falling transitions was determined by varying the frequency extent (at a fixed initial frequency). The transitions were not bounded by plateaus. We investigated whether detection of differences in rate of frequency transitions was frequency dependent by determining difference limens in two different frequency regions. Transitions in the lower frequency region are similar to the first-formant transitions of voiced plosives (Liberman *et al.*, 1956), between 220 and 750 Hz, whereas those in the higher frequency region, between 520 and 1050 Hz, have no linguistic reference. Frequency extents and rates of frequency change were identical across series. Each series consisted of 18 standard transitions, i.e., fixed transitions, which were compared to comparison transitions of the same duration, but varying frequency extent. Figure 1 illustrates some of the stimuli in one of the transition durations, i.e., 30 ms, for both series, schematically.

Although the rate of frequency change is faster for shorter than for longer transitions, the difference in step size in Hz/ms is the same for each of the three frequency

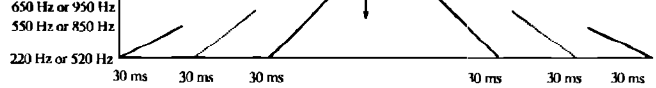

FIG. 1. Illustration of stimuli in the frequency condition for a standard transition of 30 ms (dashed) with a varying final frequency.

extents per transition duration. Each standard transition was compared to stimuli of a higher and lower rate of frequency change.

Randomized sets of discrete stimuli covering a range of 200 Hz below and above the standard transition in steps of 10 Hz, served as test blocks. A 200-Hz frequency range (40 stimuli in total) was chosen to include a number of readily discriminable stimuli in the testing sessions. Testing sessions containing the longest transitions were usually considered more difficult, presumably because the difference in rate is much smaller with the longer than with the shorter transitions. Every block of randomized stimuli was balanced in that each stimulus was presented twice, once before and once after the standard transition, resulting, in total, in 80 responses per standard stimulus per testing session.

All the transition duration and direction conditions were tested separately. However, to avoid listeners focusing on one standard transition only, two standard transitions, differing only in frequency extent, were included in one testing session. This was possible as some standard transitions had similar testing ranges. For instance, a 220–550 Hz and 220–650 Hz were tested together, with a testing range varying from 220–350 Hz to 220–850 Hz. Each testing session included a new order of randomization.

At least 20 responses were collected for each standard-comparison pair of transitions. When the comparison transitions higher and lower than the standard transition are added together, there are at least 40 responses for each frequency deviation, i.e., for each difference between the end point frequencies of the standard and comparison transitions. These were collected for three frequency extents, three transition durations, and two transition directions, from each of four subjects. In total, nearly 3000 responses were collected for each standard-comparison pair in the 200-Hz range (40 responses × 3 frequency extents × 3 transition durations × 2 transition directions × 4 subjects).

In addition to this, the sets of randomized stimuli included a proportion of catch trials. The number of these physically identical pairs of stimuli, which were added to the standard-comparison pairs, varied per testing session. With regard to the total number of physically different trials, there were five different percentages of catch trials, i.e., 18.75%, 25.0%, 37.5%, 50%, and 81.25%. These were randomly interspersed in the testing session.

2. Duration condition: Varying transition rate by duration

Transition rate was also varied by changing the transition duration for a constant frequency extent. As in the frequency condition, thresholds in experiment 2 were once again determined as a function of direction, fixed frequency extent, frequency range, and transition duration (20, 30, and 50 ms).

The same final frequencies were examined as in the frequency condition, i.e., 550, 650, and 750 Hz in the first series, and 850, 950, and 1050 Hz in the second series. While the frequency extent remained the same, transitions now differed in steps of 1 ms each. The testing range was

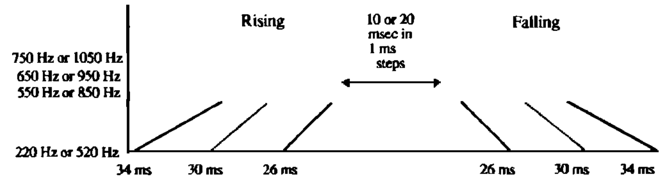

FIG. 2. Illustration of stimuli in the duration condition for a standard transition of 30 ms (dashed) at either 550 or 850 Hz.

10-ms above and below the standard transition for the 20-ms stimuli, and 20 ms for the 30- and 50-ms transitions. Some of the stimuli used in this condition are illustrated schematically in Fig. 2. The testing procedure was exactly the same as in the frequency condition. Transition durations were tested separately: Two different standard transitions, differing only in transition duration, not in final frequency, were included in one testing session. Catch trials were added and intermixed with the physically different pairs of stimuli, but the number of physically identical pairs of stimuli was not varied systematically. One fixed percentage, i.e., 37.5%, was used throughout the test.

The number of trials varied per testing session, depending on the testing range. As in the frequency condition, at least 20 responses were collected for each trial having higher and lower rates of frequency change than the standard transition. Once again these were collected for three frequency extents, three transitions durations, and two directions.

As glide rate does not remain constant, but decreases with increasing duration, and temporal properties strongly influence discrimination, we also determined duration difference limens of 50-ms *stationary* one-formant stimuli (whose frequency was endpoint frequency of the transition). If glide rate is used as a perceptual cue, thresholds of the dynamic stimuli should be smaller than those of the stationary ones. As the differences in glide rate, decrease with increasing transition duration, 50-ms thresholds were compared to stationary ones. These stationary thresholds were collected from three subjects (who also participated in the frequency condition). They were determined in the same way as the difference limens of the dynamic stimuli.

C. Procedure

Blocks of trials were presented real time on a PC (16-bit resolution) in two or three 10-min sessions a day with short breaks in between. Subjects were seated in front of a terminal and heard two formant glides in noise successively at a comfortable listening level. By pressing the appropriate mouse key they were required to indicate for each pair whether it was the same or different. Following a response, two new stimuli were offered immediately. There was no feedback. Before the testing session they were told the proportion of catch trials, so that they could adjust their criterion accordingly.

The $P(C)_{max}$ (Egan, 1965) was calculated for each frequency extent.[1] Thresholds correspond to frequency de-

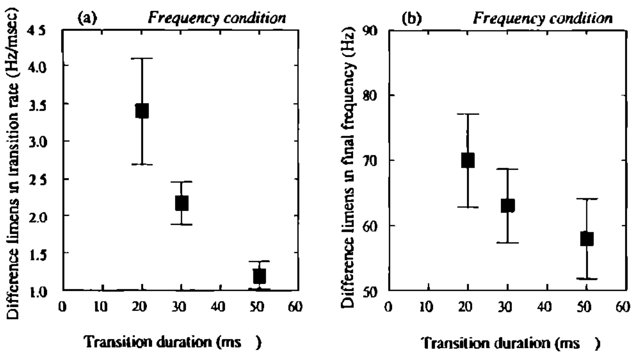

FIG. 3. Difference limens and standard deviations in the frequency condition expressed in transition rate (left panel) and final frequency (right panel). Data are averaged over four subjects, two frequency regions, two directions, and three frequency extents.

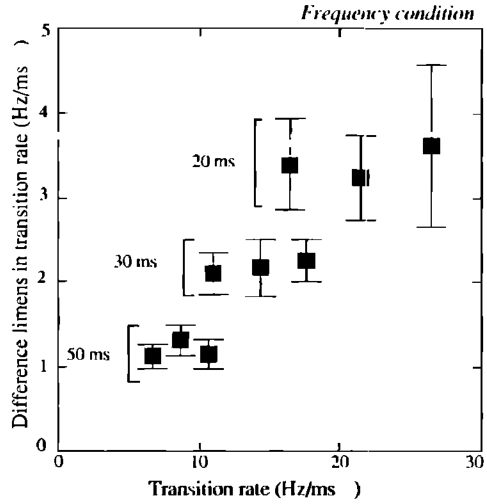

FIG. 4. Difference limens and standard deviations in the frequency condition in terms of transition rate. The three points per transition duration reflect the three frequency extents (three transition rates, see Table I). Data are averaged over subjects, frequency regions, and transition direction, but the three different frequency extents per transition duration are displayed separately.

viations yielding a $P(C)_{max}$ of at least 0.75. All conditions were analyzed separately.

D. Subjects

Thresholds of four normal hearing subjects were determined, two in the first and two in the second frequency region. All four subjects (aged between 22 and 34 years) participated in both the frequency and the transition duration conditions. They received about 2 h of practice before data collection began. The subjects have absolute thresholds less than 10 dB HL at all audiometric frequencies. Two subjects were paid for their participation.

E. Results

Thresholds and standard deviations of rising and falling transitions of the frequency and duration conditions (averaged over four subjects, two frequency regions, and transition durations) are plotted in Figs. 3–6. Thresholds in the frequency condition are plotted in transition rate as a function of transition duration [Fig. 3(a)], in terms of the difference in final frequency (Hz) as a function of transition duration [Fig. 3(b)], and in transition rate (Hz/ms) as a function of transition rate as specified in Table I (Fig. 4).

In Fig. 3(a), difference limens in transition rate are, on average, 3.4, 2.2, and 1.2 Hz/ms, and in Fig. 3(b) difference limens in final frequency are 70, 63, and 58 Hz for 20, 30, and 50 ms, respectively. A split-plot factorial ANOVA (Kirk, 1982) with frequency region (between blocks), subjects, transition duration, transition direction, frequency extent, and higher or lower rate of frequency change (within blocks) was conducted on the thresholds in final frequency (Hz) and transition rate (Hz/ms). The subjects factor was random, the others were fixed. There was a significant main effect of transition duration, for the transition rate thresholds expressed in Hz/ms [$F(2,4)=98.3$ $p<0.001$] and the final frequency thresholds in Hz [$F(2,4)=8.85$, $p<0.005$]. None of the other factors nor any of the interactions were statistically significant.[2] Except for the 20-ms condition, thresholds in final frequency are comparable with Elliott *et al.* (1989), who also measured transitions without a steady state. The thresholds of the shortest transitions are lower than those of Elliott *et al.* (1989: 110 Hz, on average), possibly because we varied the final frequency and Elliott *et al.* (1989) the initial frequency. Difference limens of 20-ms transitions could be determined reliably, despite their short durations. The thresholds are also in agreement with the difference limens in onset frequencies (Porter *et al.*, 1991), although not with respect to the rise/fall and higher/lower rate effects reported in that study (for the transitions longer than 45 ms). The results of the experiment are generally in agreement with a sampling theory of glide perception (Pollack, 1968; Dooley and Moore, 1988b), which states that if sampling the end points underlies perception of glide discrimination, the thresholds should become poorer as stimulus duration decreases. However, if transition duration is very short, listeners may not be sampling end points. Discrimination might be based on the differences in bandwidth, which is directly related to transition duration.

To examine the perceptual importance of transition rate, thresholds of the 20-, 30-, and 50-ms transitions are plotted as a function of their transition rates in Fig. 4. As one can see from the three points per transition duration (indicating three different transition rates, see Table I) the differences in transition rate per transition duration did not influence discrimination. If discrimination were solely based on transition rate, difference limens of the 30-ms transition with the smallest frequency extent would be similar to a 50-ms one with the largest frequency extent (see transition rates of 11.0 and 10.6 Hz/ms, respectively, in Table I). Also, it is expected that if discrimination were based on transition rate, lower rate transitions would yield smaller difference limens than higher rate ones per transition duration. If discrimination were based on frequency extent only, difference limens would increase with increase of frequency extent per transition duration. If discrimina-

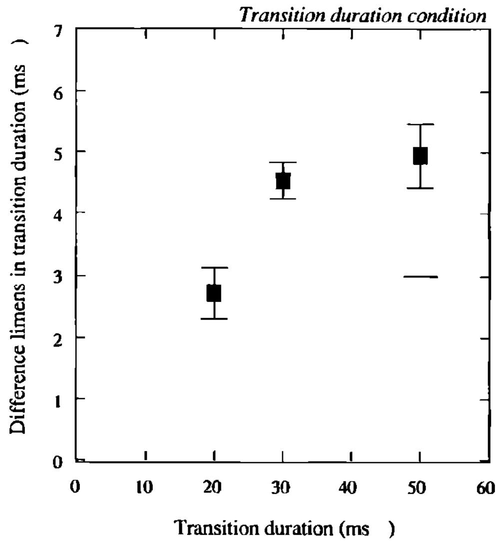

FIG. 5. Difference limens and standard deviations in the transition duration condition, averaged over frequency region, direction, and frequency extent. The difference limen of the 50-ms stationary stimuli is marked by a solid line.

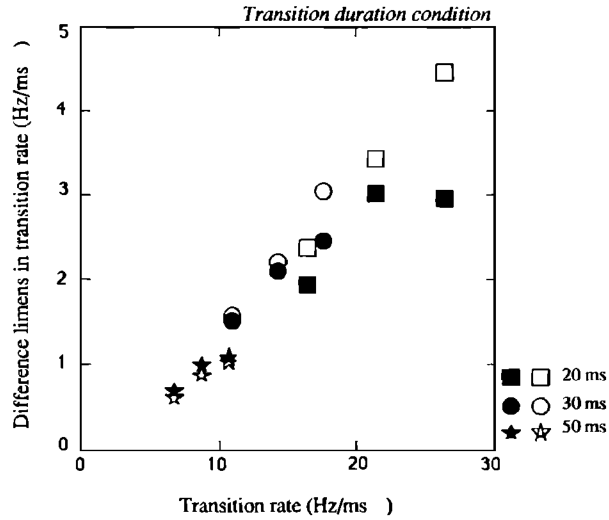

FIG. 6. Thresholds determined in the transition duration condition, expressed in transition rate, for transitions shorter (unfilled symbols) and longer (filled symbols) than the standard. Data are averaged over subjects, direction, and frequency region, yet plotted separately for the three different frequency extents (or transition rates) per transition duration.

tion were based on final frequency, difference limens would be similar for the three frequency extents per transition duration (as the difference limen is defined as the difference between the end points of the standard and comparison stimuli). For most of the transition durations, thresholds are nearly identical, suggesting that discrimination is based on the difference between the end points of the standard and comparison signals, rather than on transition rate.

By analogy with Weber's law, the ability to discriminate temporal changes between intervals should decrease with increasing duration (experiment 2). Contrary to the difference limens in the frequency condition, thresholds in the transition duration condition (Fig. 5) increase with increase of transition duration. Averaged over the two frequency regions and three frequency extents, thresholds are 2.7, 4.5, and 4.9 ms for the 20-, 30-, and 50-ms standard transitions, respectively (Fig. 5). A split-plot factorial ANOVA with frequency region (between blocks) and subject, transition direction, transition duration, and frequency extent (within blocks) was conducted on the difference limens in milliseconds. Transition duration was a significant main effect, i.e., for the average difference limens [$F(2,4)=113.3$, $p<0.001$], for those shorter than the standard [$F(2,4)=22.3$, $p<0.001$], and for those longer than the standard [$F(2,4)=163.5$, $p<0.001$]. There were no other significant factors nor significant interactions.

Although the difference limens vary with transition duration, it remains unclear whether discrimination is based on temporal or spectral properties. Due to the temporal properties involved in this condition we also measured difference limens of *stationary* 50-ms stimuli, of three subjects, to compare psychoacoustical cues in both conditions. It is expected that if discrimination benefits from transition rate, difference limens of dynamic stimuli would be smaller. If not, one would expect them to be the same. The most striking finding, however, was that the difference limens are larger in the dynamic condition than in the stationary one. Difference limens of stationary complex stimuli were measured at three frequencies (the end point frequencies of the transitions) for 50-ms standard stimuli. On average, thresholds were 3 ms in all conditions (Fig. 5).

To determine the perceptual importance of transition rate difference limens are expressed in transition rate of the comparison transitions as a function of the transition rate of the standard transitions in Fig. 6 (to be compared with Fig. 4 in the frequency condition). Data are averaged over four subjects, direction, and frequency region. Following the difference limens in the frequency condition, sensitivity increases with increase of transition duration. Compare the difference limens of 20-ms (squares), 30-ms (circles), and 50-ms (stars) transitions. However, interpretation of the results in this condition is more complicated and therefore hardly comparable to those of the frequency condition, as the difference in slope does not remain constant per transition duration, but decreases as the transition becomes longer. In other words, 20- vs 22-ms stimuli will be more difficult to discriminate than 18- vs 20-ms ones, not only because the temporal difference limen is larger (also for steady states), but also because the relative difference in slope is smaller. Therefore, the difference limens of the transitions shorter (unfilled symbols) and longer (filled symbols) than the standard, were analyzed separately. The difference between the thresholds of the transitions longer and shorter than the standard increases with decrease of transition duration, due to the increase of transition rate and, subsequently, the increasing difference in the asymmetry of the slope (Fig. 6). It is difficult to determine the extent to which discrimination is based on temporal or on spectral properties in this condition. Transition rate may be an important psychoacoustical cue, as difference limens

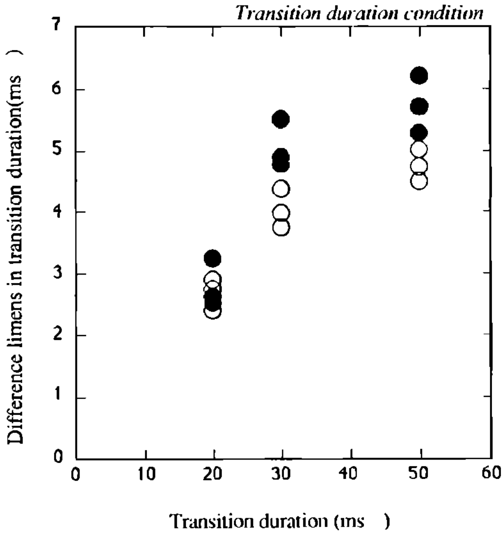

FIG. 7. Thresholds determined in the transition duration condition, expressed in transition duration, for transitions shorter (unfilled symbols) and longer (filled symbols) than the standard. Data are averaged over subjects, direction, and frequency region.

increase with increase of transition rate. Also, some thresholds, for which the standards have similar transition rates but different durations (see Table I), are very similar (due to the temporal thresholds, these difference limens will never be identical across transition durations). Compare, for instance, the thresholds of the 20- and 30-ms transitions, with transition rates of 16.5 and 14.3 Hz/ms, respectively. On average, their difference limens are 2 Hz/ms. In comparison with the thresholds determined in the frequency condition, the difference limens of the 50-ms transitions are very small, i.e., on average, 1 Hz/ms. This corresponds to a temporal difference of, on average, 5 ms (see Figs. 5 and 7), due to the small differences in transition rate for the longer transitions. It is not very surprising that difference limens increase with increase of transition rate, because the difference in transition rate increases with increase of frequency extent for the same standard duration. Compare the thresholds of the 30-ms transition with the smallest frequency extent and the 50-ms one with the largest frequency extent. They are relatively close to each other, between 1.0 and 1.5 Hz/ms, suggesting that spectral properties influence discrimination. However, as the difference limen in transition rate of the 30-ms transition is much smaller than the 30-ms difference limens in final frequency (Fig. 4), it is more likely that discrimination is governed by temporal cues. It must be kept in mind that the temporal, as well the spectral differences are small, and that the frequency extent factor was not significant.

Figure 7 illustrates the difference limens in transition duration for transitions shorter and longer than the standard. Thresholds of the transitions shorter than the standard (unfilled symbols) are smaller than those longer than the standard (filled symbols), especially at 30 and 50 ms. In temporal differences, thresholds for the three frequency extents appear to be fairly similar (see Fig. 6). No differences were found between rising and falling transitions.

Apart from transition duration, no significant differences were found statistically for transition direction, frequency extent or frequency region. Also, thresholds for stimuli having higher rates of frequency change than the standard are comparable to those having lower rates of frequency change than the standard (despite some individual differences) in the frequency condition. Discrimination experiments with formant transitions followed by a steady state (Porter *et al.*, 1991) have shown difference limens to improve as rate of frequency change increases, possibly as a result of the relatively large extents of frequency change. By contrast, increasing or decreasing the extent of frequency change in our experiments affected discrimination similarly.

F. Discussion

The thresholds determined in the two experiments indicate that different psychoacoustical cues are used depending on whether final frequency (experiment 1) or transition duration (experiment 2) are varied. The thresholds determined for formant sweeps in our frequency condition are in agreement with Elliott *et al.*, (1989,1991), and Porter *et al.* (1991) and for tone glides in agreement with Sergeant and Harris (1962), Pollack (1968), and Arlinger *et al.* (1977). Discrimination of tone and formant glides clearly benefits from an increase of processing time, i.e., when the transitions become longer (Dooley and Moore, 1988a,b; Elliott *et al.*, 1989, 1991; Porter *et al.* 1991). With shorter transition durations, between 20 and 50 ms, discrimination may be based on rate of frequency change (Porter *et al.*, 1991). However, no support for rate cues was found in our experiments, where similar thresholds were determined for transitions incrementing and decrementing in transition rate.

Apart from transition rate, discrimination can also be based on frequency or bandwidth cues. Consider the spectrum of a signal, which is a function of duration. The signal bandwidth increases with decreasing transition duration, and this can be an important factor limiting frequency discrimination in short duration stimuli. In other words, the spectral bandwidth of a 20-ms signal is approximately 50 Hz. The formant bandwidth of the varying formant was always 10% of the formant frequency. With short transition durations, it is unlikely that initial or final frequencies, for falling and rising transitions, respectively, serve as perceptual cues. The 20-ms transitions hardly seem to be perceived as glides. Discrimination is more likely to be based on the differences between initial and final frequencies of the two stimuli or on the differences in bandwidth, both of which are related to each other.

When transitions vary in transition duration, as in experiment 2, temporal, as well as spectral properties are used. The shorter the transition, the better the temporal resolution of the auditory system, as difference limens increase with increasing transition duration (Fig. 7). Discrimination is, to some extent, also based on transition rate, as some transitions having similar rates but different durations yield similar thresholds (Fig. 6). However, it is unlikely that spectral differences of less than 1 Hz/ms are perceived at 50 ms. In our study it seems as if spectral properties hamper discrimination, as the thresholds for the stationary stimuli are smaller than the dynamic ones.

Difference limens were similar for the three frequency extents per transition duration, the two frequency regions, and for rising and falling transitions. Results of Nábělek and Hirsh (1969) indicate that the extent of frequency change influences discriminability. Best thresholds are achieved at small glide rates for small frequency differences, and at high rates for large frequency differences. Frequency extent did not influence discrimination in our tests, possibly because all our glide rates were high (see Table I), and differences among frequency extents and transition rates were relatively small. This may also explain why the different frequency extents in the transition duration condition do not influence discrimination. For a small frequency interval Nábělek *et al.* (1969) found discriminability to be better at 30 than at 10 ms. In terms of milliseconds, our data show smaller thresholds at the shorter durations (but this involves a relatively large step in transition rate).

The lack of a consistent difference in results between rising and falling transitions agrees with Nábělek and Hirsh (1969), Arlinger *et al.* (1977), Dooley and Moore (1988), Elliott *et al.* (1989, 1991), and Schouten and Pols (1989) who also failed to find differences between rising and falling frequency changes. Porter *et al.* (1991) argue that that the rise/fall effect is due to the difference in relative temporal sensitivity of critical bands in the lower or higher frequency regions. Critical band units in the higher frequency region (above 2000 Hz) have better temporal resolution than those in the lower frequency region. In this line of thought, a rise/fall effect would not be expected in our study: The frequencies of our transitions are well below 2000 Hz. Discrimination is similar in both frequency regions, although the transitions in the lower region correspond to the first-formant transitions of voiced plosives and those in the higher frequency region have no linguistic reference.

The durations used in the present experiments resemble stop consonants in speech. Assuming that our data are correct and that there are no rise/fall, or higher rate/lower rate effects at very short durations, then the perception of stop consonants rising to and falling from targets should be similar. The discrepancy between CV and VC transitions (Collins, 1984) could be brought about by (masking of) the surrounding vowel context, not by the transition itself. Our future research will focus on the perceptual relevance of physical properties of one-formant transitions in a suprathreshold paradigm. Subsequently, more complex stimuli, i.e., those involving two or more formant transitions, will be tested in psychoacoustical and speech-perceptual experiments.

ACKNOWLEDGMENTS

We are grateful to Bert Schouten for helpful criticisms on an earlier draft of this paper. This research was supported by the University of Amsterdam.

[1]The maximum proportion correct, $P(C)_{max}$, is obtained through a transformation of the d' associated with the stimulus–response matrix (Macmillan and Creelman, 1991). This bias-free measure of sensitivity is preferred to d', as small differences in probability values are easier to compare than small differences in d'.

[2]The same/different psychophysical procedure is analyzed directly within the context of a decision-theory model. It is assumed that the listener's criterion to determine whether a pair is the same or different remains constant during a test, but may be changed by informing the listener on the proportion of catch trials included in the testing series. We have explicitly examined the extent to which discrimination and response bias are separated by comparing the values of $P(C)_{max}$ of the different proportions of catch trials per testing session. The listener's behavior consists of two aspects, sensitivity and response bias. As the same stimuli are used, sensitivity should not change. If listeners adjust their criterion according to the proportion of catch trials, the criterion will shift but the difference between the two distributions, i.e., sensitivity, should remain the same. Five different proportions of catch trials were included and systematically tested in the testing series of the frequency condition, i.e., 18.75%, 25.0%, 37.5%, 50%, and 81.25%. The $P(C)_{max}$'s for the five different proportions of catch trials (in different testing sessions), averaged over the three frequency extents and three transition durations, are nearly the same, indicating the proportion of catch trials does not influence the sensitivity. It is concluded that the listeners maintained the same response strategies throughout the tests, as their performance is similar, irrespective of the proportion of catch trials included in the testing sessions.

Abel, S. M. (1972). "Duration discrimination of noise and tone bursts," J. Acoust. Soc. Am. 51, 1219–1224.

Arlinger, S. D., Jerlvall, L. B., Ahren, L. B., and Holmgren, E. C. (1977). "Thresholds for linear frequency ramps of a continuous pure tone," Acta Otolaryngol. 83, 317–327.

Aslin, R. N. (1989). "Discrimination of frequency transitions by human infants," J. Acoust. Soc. Am. 86, 582–590.

Collins, M. J., and Cullen, J. K. Jr. (1978). "Temporal integration of tone glides," J. Acoust. Soc. Am. 63, 469–473.

Collins, M. J. (1984). "Tone-glide discrimination: normal and hearing-impaired listeners," J. Speech Hear. Res. 27, 403–412.

Cosgrove, P., Wilson, J. P., and Patterson, R. D. (1989). "Formant transition detection in isolated vowels with transitions in initial and final position," Proc. IEEE ICASSP, 278–281.

Creelman, C. D. (1962). "Human discrimination of auditory duration," J. Acoust. Soc. Am. 34, 582–593.

Cullen, J. K. Jr., and Collins, M. J. (1978). "Temporal integration of two-component tone glides," J. Acoust. Soc. Am. 64, 1526–1527.

Cullen, J. K., and Collins, M. J. (1982). "Audibility of short-duration tone glides as a function of rate of frequency change," Hear. Res. 7, 115–125.

Cullen, J. K., Jr., Houtsma, A. J. M., and Collier, R. (1992). "Discrimination of brief tone-glides with high rate of frequency-change," Abstracts of the Fifteenth Midwinter Meeting of the Association for Research in Otolaryngology, 67.

Danaher, E. M., Osberger, M. J., and Pickett, J. M. (1973). "Discrimination of formant frequency transitions in synthetic vowels," J. Speech Hear. Res. 16, 439–451.

Dooley, G. J., and Moore, B. C. J. (1988a). "Duration discrimination of steady and gliding tones: A new method for estimating sensitivity to rate of change," J. Acoust. Soc. Am. 84, 1332–1337.

Dooley, G. J., and Moore, B. C. J. (1988b). "Detection of linear frequency glides as a function of frequency and duration," J. Acoust. Soc. Am. 84, 2045–2057.

Egan, J. P. (1965). "Masking-level differences as a function of interaural disparities in intensity of signal and of noise," J. Acoust. Soc. Am. 38, 1043–1049.

Egan, J. P., and Clarke, F. R. (1966). "Psychophysics and signal detection," in *Experimental Methods and Instrumentation in Psychology*, edited by J. B. Sidowski (McGraw-Hill, New York), pp. 211–246.

Egan, J. P. (1975). *Signal Detection Theory and ROC Analysis* (Academic, New York).

Elliott, L. L., Hammer, M. A., Scholl, M. E., Carrell, T. D., and Wasowicz, J. M. (1989). "Discrimination of rising and falling simulated single-formant frequency transitions: Practice and duration effects," J. Acoust. Soc. Am. 86, 945–953.

Elliott, L. L., Hammer, M. A., and Carrell, T. (1991). "Discrimination of second-formant-like frequency transitions," Percept. Psychophys. 50, 1–6.
Fairbanks, G., and Grubb, P. (1961). "A psychophysical investigation of vowel formants," J. Speech Hear. Res. 4, 203–219.
Fastl, H. (1978). "Frequency discrimination for pulsed versus modulated tones," J. Acoust. Soc. Am. 63, 275–276.
Fastl, H., and Hesse, A. (1984). "Frequency discrimination for pure tones at short durations," Acustica 56, 41–47.
Feth, L. L. (1974). "Frequency discrimination of complex periodic tones," Percept. Psychophys. 15, 375–378.
Flanagan, J. L. (1955a). "Difference limen for vowel formant frequency," J. Acoust. Soc. Am. 27, 613–617.
Flanagan, J. L. (1955b). "Difference limen for the intensity of a vowel sound," J. Acoust. Soc. Am. 27, 1223–1225.
Flanagan, J. L. (1957). "Estimates of the maximum precision necessary in quantizing certain 'dimensions' of vowel sounds," J. Acoust. Soc. Am. 29, 533–534.
Fujisaki, H., and Sekimoto, S. (1975). "Perception of time-varying resonance frequencies in speech and non-speech stimuli," in *Structure and Process in Speech Perception*, edited by A. Cohen and S. G. Nooteboom (Springer-Verlag, New York), pp. 269–282.
Gardner, R. B., and Wilson, J. P. (1979). "Evidence for direction specific channels in the processing of frequency modulation," J. Acoust. Soc. Am. 66, 704–709.
Horst, J. W. (1982). "Discrimination of complex signals in hearing," Ph.D. thesis, University of Groningen.
Horst, J. W. (1989). "Detection and discrimination of frequency modulation of complex signals," J. Acoust. Soc. Am. 85, 2022–2030.
Jesteadt, W., and Bilger, R. C. (1974). "Intensity and frequency discrimination. I. One- and two-interval paradigms," J. Acoust. Soc. Am. 55, 1266–1279.
Jesteadt, W., and Sims, S. L. (1975). "Decision processes in frequency discrimination," J. Acoust. Soc. Am. 57, 1161–1168.
Jesteadt, W., Wier, C. C., and Green, D. M. (1977). "Intensity discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. 61, 169–177.
Kirk, R. E. (1982). *Experimental design: Procedures for the Behavioral Sciences* (Brooks/Cole, Belmont, MA).
Liberman, A. M., Delattre, P. C., Gerstman, L. J., and Cooper, F. S. (1956). "Tempo of frequency change as a cue for distinguishing classes of speech sounds," J. Exp. Psychol. 52, 127–137.
Macmillan, N. A., and Creelman, C. D. (1991). *Detection Theory: A User's Guide* (Cambridge U.P., Cambridge).
Mermelstein, P. (1978). "Difference limens for formant frequencies of steady-state and consonant-bound vowels," J. Acoust. Soc. Am. 63, 572–580.
Moore, B. C. J. (1973). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. 54, 610–619.
Moore, B. C. J., Oldfield, S. R., and Dooley, G. J. (1989). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," J. Acoust. Soc. Am. 85, 820–836.
Nábělek, I. V., and Hirsh, I. J. (1969). "On the discrimination of frequency transitions," J. Acoust. Soc. Am. 45, 1510–1519.
Nábělek, I. V., Nábělek, A. K., and Hirsh, I. J. (1970). "Pitch of tone bursts of changing frequency," J. Acoust. Soc. Am. 48, 536–553.
Nábělek, I. V. (1978). "Temporal summation of constant and gliding tones at masked auditory threshold," J. Acoust. Soc. Am. 64, 751–763.
Nelson D. A., Stanton, M. E., and Freyman, R. L. (1983). "A general equation describing frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. 73, 2117–2123.
Pollack, I. (1968). "Detection of rate of change of auditory frequency," J. Exp. Psychol. 77, 535–541.
Pols, L. C. W., and Schouten, M. E. H. (1987). "Perception of tone, band, and formant sweeps," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Dordrecht, The Netherlands), pp. 231–240.
Porter, R. J., Cullen, J. K., Collins, M. J., and Jackson, D. F. (1991). "Discrimination of formant transition onset frequency: Psychoacoustic cues at short, moderate, and long durations," J. Acoust. Soc. Am. 90, 1298–1308.
Schouten, M. E. H. (1985). "Identification and discrimination of sweep tones," Percept. Psychophys. 37, 369–376.
Schouten, M. E. H. (1986). "Three-way identification of sweep tones," Percept. Psychophys. 40, 359–361.
Schouten, M. E. H., and Pols, L. C. W. (1989). "Identification and discrimination of sweep formants," Percept. Psychophys. 46, 235–244.
Seargent, R. J., and Harris, J. D. (1962). "Sensitivity to unidirectional frequency modulation," J. Acoust. Soc. Am. 34, 1625–1628.
Tsumura, T., Sone, T., and Nimura, T. (1973). "Auditory detection of frequency transition," J. Acoust. Soc. Am. 53, 17–25.
Tyler, R. S., Wood, E. J., and Fernandes, M. (1983). "Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners," J. Acoust. Soc. Am. 74, 1190–1199.
Weenink, D. J. M. (1988). "Klinkers: een computerprogramma voor het genereren van klinkerachtige stimuli," IFA-Rep. No. 100.
Wier, C. C., Jesteadt, W., and Green, D. M. (1977). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. 61, 178–184.