Discrimination of single and complex consonant-vowel-like formant transitions

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

Published in:
The Journal of the Acoustical Society of America

DOI:
10.1121/1.413467

Citation for published version (APA):
Discrimination of single and complex consonant–vowel- and vowel–consonant-like formant transitions

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

(Received 16 May 1994; revised 4 April 1995; accepted 14 April 1995)

Two discrimination experiments were performed to determine auditory sensitivity for single and complex consonant–vowel (CV)–and vowel–consonant (VC)–like formant transitions. In experiment 1, difference limens in end-point frequency were determined by means of same/different paired comparison tasks for 20-, 30-, and 50-ms second formant (F2) speechlike transitions, followed or preceded by an 80-ms vowel-like steady state in initial or final position, respectively. The F2 transition was either single or part of a multiformal (complex) stimulus, also containing a fixed F1 transition with a steady state, a stationary third formant, and a 20-ms voice bar. Just-noticeable differences in end-point frequency decrease with increasing transition duration in all conditions and are smaller for single transitions than for transitions in a multiformal complex. Although difference limens in end-point frequency increase with increase in frequency extent, they are smaller in final than in initial position. As for relative rate-of-frequency change, the smaller the frequency extent, the larger the difference limens of decrementing transitions, possibly because the end points of the transitions approach the frequency of the steady state. Transitions varying in frequency and duration, in such a way that a constant rate of frequency change is maintained (experiment 2), yield considerably smaller difference limens than those varying in frequency extent at a constant transition duration (experiment 1). Discrimination is affected more by changes in duration than in frequency, even when the total duration of the stimulus remains constant. Both experiments show that transition rate is perceptually less important than end-point frequency or transition duration in discriminating short speechlike transitions. © 1995 Acoustical Society of America.

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

INTRODUCTION

In speech, distinctions in place of articulation of stop consonants can be conveyed by the physical properties of relatively short vocalic transitions (Delattre et al., 1955; Liberman et al., 1956, 1967). Detailed discrimination experiments with tone glides (Nábělek and Hirsh, 1969; Nábělek et al., 1970; Tsumura et al., 1973; Collins, 1984; Schouten, 1985, 1986; Dooley and Moore, 1988a,b; Cullen et al., 1992) and formant transitions (Elliott et al., 1989, 1991; Collins et al., 1990; Schouten and Pols, 1989; Porter et al., 1991; van Wieringen and Pols, 1992, 1994) simulating speech properties indicate how auditory sensitivity depends on the physical parameters such as transition direction, transition position, transition duration, etc. This is of importance to determine the psychoacoustic properties underlying place distinctions in speech and to model synthetic speech sounds.

In general, all studies show discrimination to benefit from an increase in duration. Due to the covariation of frequency extent, duration, and transition rate, however, it is very difficult to determine the relevant psychoacoustic cues. Just-noticeable differences in end-point frequency of isolated fixed-duration formant glides decrease from 70 to 63 and 57 Hz for 20-, 30-, and 50-ms stimuli, respectively (van Wieringen and Pols, 1992, 1994). The end point refers to the varying part of the transition, i.e., the onset of initial transitions converging to a similar frequency, or the offset of final transitions, diverging to different frequencies. Contrary to frequency difference limens, duration difference limens increase with increasing duration at a constant frequency extent, because the difference in time required for temporal discrimination increases as the duration of the standard increases (Creelman, 1962; Abel, 1972). On average, difference limens in transition duration are 2.7, 4.5, and 4.9 ms for 20, 30, and 50 ms, respectively (van Wieringen and Pols, 1992, 1994). Although discrimination of short transitions also depends on frequency region, direction (rising or falling), and position (initial or final) of the transitions (Cosgrove et al., 1989; Collins, 1984), sensitivity varies most with changes in signal duration and bandwidth characteristics. The latter depends on both the frequency extent and the duration of the stimulus (Porter et al., 1991; van Wieringen and Pols, 1992, 1994).

Following these discrimination experiments on short, isolated formant transitions varying in frequency and duration, the present experiments examine the psychoacoustic cues involved in the processing of short single and complex transitions, which are preceded (final) or followed (initial) by an 80-ms (vowel-like) steady state. Difference limens in end-point frequency of short transitions preceded or followed by a steady state are expected to be larger than isolated transitions, because more information has to be processed before the transitions can be distinguished. Also, complex stimuli are expected to yield larger difference limens than single
ones, because of masking effects of the acoustic environment and possibly because the more speechlike nature of the sound prevents adequate discrimination of similar phonemes.

Difference limens of transitions varying in transition rate and frequency extent at a constant transition duration (experiment 1) are compared to those varying at a constant rate of frequency change and thus varying in frequency extent and transition duration (experiment 2), in order to examine the underlying psychoacoustic cues, i.e., in particular, the perceptual importance of transition rate. If discrimination is based on the extent of the transition, discrimination of shorter transitions should be better than that of longer ones, provided the rate of frequency change is constant (experiment 2), as the extent of frequency change is smaller for the shorter transitions. However, discrimination could also be based on the difference in end points between the standard and the comparison transitions; this would imply that longer transitions would yield smaller difference limens, due to an increase in processing time. With stationary stimuli an increase in processing time does not necessarily result in improved discriminability (frequency discrimination). However, with dynamic stimuli a certain time interval is necessary to be able to perceive a change in frequency. Given that this interval is the same for 20-, 30-, and 50-ms stimuli, the spectral change of a longer stimulus is less than that of a shorter one. We expect that the less the formant frequency changes, the easier it is to determine the end-point frequency of the stimulus.

Transitions maintaining a constant rate-of-frequency change (experiment 2) should yield smaller difference limens in end-point frequency than those of the same duration varying in transition rate (experiment 1), due to the extent of frequency change being smaller and the frequency uncertainty (within a certain unit of time) being smaller as a result of a lower transition rate. In addition to these spectral changes, discrimination can also be affected by the temporal variation, which is necessary to maintain a constant rate-of-frequency change. Temporal cues strongly affected discrimination of isolated formant transitions (van Wieringen and Pols, 1992, 1994), where frequency extent was held constant and transition duration was varied. In the present experiments auditory sensitivity was not measured for transitions varying in transition duration at a constant frequency extent because the auditory system is highly sensitive to changes in duration. Because we did not examine this condition separately, end-point frequency and frequency extent can be used as similar terms in this study. In order to reduce temporal sensitivity, the durations of the transition and the steady state are both varied in experiment 2 to maintain a constant total duration. The potential perceptual importance of transition rate as a cue to the discrimination of short formant transitions is examined by determining difference limens in end-point frequency of two different rates of frequency change, i.e., 5 and 10 Hz/ms, for equally long transitions. However, discrimination could also be affected by the temporal extent, which is necessary to maintain a constant rate-of-frequency change (experiment 2) should yield smaller difference limens, due to an increase in processing time. With stationary stimuli an increase in processing time does not necessarily result in improved discriminability (frequency discrimination). However, with dynamic stimuli a certain time interval is necessary to be able to perceive a change in frequency. Given that this interval is the same for 20-, 30-, and 50-ms stimuli, the spectral change of a longer stimulus is less than that of a shorter one. We expect that the less the formant frequency changes, the easier it is to determine the end-point frequency of the stimulus.

In experiment 1, transition rate was varied by changing the frequency extent of the second formant transition at a fixed transition duration of either 20, 30, or 50 ms. The standard transitions began or ended at either /b/-like or /d/-like loci, i.e., at either 700 or 1800 Hz, respectively. Difference limens in end-point frequency were determined for single (Fig. 1) and for complex F2 transitions (Fig. 2) preceded or followed by an 80-ms steady state. The complex stimulus consisted of a fixed F1 transition with a steady state, a sta-
tionary third formant, and a 20-ms voice bar to make the stimuli sound more speechlike. The transition duration of the $F1$ transition was the same as that of the $F2$ transition. The transition rates of the standard transitions tested in experiment 1 are listed in Table I.

The transitions preceded or followed a steady-state signal with either an /a/-like or /u/-like formant pattern in order to examine whether discrimination is affected by the different vowel-like timbres. For the /a/-like and /u/-like stimuli the frequency of the steady state was 1300 Hz; for the /u/-like and /a/-like stimuli it was 800 Hz. The $F1$ transition of the complex /a/-like and /u/-like stimuli always rose or fell from 220 to 750 Hz, and the frequency of the third formant was fixed at 2200 Hz. The $F1$ transition of the complex /u/-like and /a/-like stimuli always rose or fell from 220 to 330 Hz, and the third formant frequency was fixed at 2200 Hz (Fig. 2).

In experiment 2, the frequency extent and the transition duration of the first and second formant transitions were both varied in such a way that a constant rate-of-frequency change of 5 or 10 Hz/ms was maintained (Fig. 3). The second formant frequency was 1300 Hz and the frequency extents of the 5-Hz/ms transitions were 150 Hz for the 30-ms standard transitions and 250 Hz for the 50-ms standard transitions.

The frequency extent of the 10-Hz/ms transitions was 500 Hz for the 50-ms standard transitions. The total duration of the stimuli remained constant (110 or 130 ms) by changing the durations of the first and second formant frequencies of the standard transition (30 or 50 ms) and of the standard steady state (80 ms, $F2=1300$ Hz). The durations of the transitions and the steady state varied in steps of 1 ms for each 5- or 10-Hz/m change in end-point frequency. Discriminability was tested for 30- and 50-ms transitions only.

The stimuli were generated by a digital formant synthesizer on a μVax II (Klinkers, Weenink, 1988). The pulse source had a fundamental frequency of 110 Hz. The formant bandwidth was always 10% of the (changing) formant frequency. To ensure a precise generation of these formant transitions, the stimuli were sampled at 1.2 MHz. After low-pass filtering, they were downsampled to 20 kHz (16-bit resolution). The formant frequency values were updated every 1 ms. Although the first period started at a zero crossing, a 2-ms cosine window was used at both end points to avoid on/offset clicks. Stimuli were centrally added to 300 ms of low level white noise (signal-to-noise-ratio of approximately 70 dB) to minimize external auditory factors. The 16-bit stimuli were transferred from the μVAX II to an IBM-PC in order to have real-time capability for training and testing. The stimuli were generated by means of an OROS-AU22 DSP board with D/A converter.

B. Procedure

Difference limens were determined by means of the method of constant stimuli in a same/different procedure. 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. Randomized sets of discrete stimuli differing in steps of 10 Hz served as testing series in experiment 1. The testing series covered a frequency range of 200 Hz (or 250 Hz for the 20-ms complex transitions) above and below the standard transition. In experiment 2 the testing series included 15 stimuli above and below the standard transition. The order of presentation of every set of randomized stimuli was balanced in that each stimulus was presented twice, once before and once after the standard transition, resulting in 80 responses per standard stimulus per testing session.

Approximately 50% of each testing series consisted of catch trials, i.e., physically identical pairs of stimuli, to reduce response bias toward responding "different" for all pairs of stimuli. All conditions were tested separately. Each

| TABLE I. Stimulus characteristics of the standard 20-, 30-, and 50-ms $F2$ transitions of experiment 1, in terms of end-point frequencies (Hz), frequency extents (Hz), and transition rates (Hz/ms). |
|---|---|---|---|---|---|
| **End-point frequencies** | **Frequency extents** | **Transition duration** | **Transition rates** |
| | | | 20 ms | 30 ms | 50 ms |
| /a/-like and /u/-like | 700–800 | 100 | 5.0 | 3.3 | 2.0 |
| /a/-like and /u/-like | 700–1300 | 600 | 30.0 | 20.0 | 12.0 |
| /u/-like and /a/-like | 1800–1300 | 500 | 25.0 | 16.7 | 10.0 |
| /u/-like and /a/-like | 1800–800 | 1000 | 50.0 | 33.3 | 20.0 |

FIG. 3. Schematic representation of stimuli in experiment 2. An example is given of a 50-ms $F2$ transition varying in end-point frequency and transition duration at a constant rate-of-frequency change of 5 Hz/ms in initial and final position. The end-point frequency of the standard transition is 1550 Hz. The dashed lines mark the transitions that are shorter and longer than the standard transition. The eped duration of the transition and the steady state are both varied to maintain a constant total duration of 130 ms. Discriminability is also measured for transitions varying at a constant rate-of-frequency change of 10 Hz/ms in initial and final position. The standard durations of these transitions were 30 and 50 ms. A fixed first formant transition with a steady state, a stationary third formant, and a voice bar are added to make the stimulus sound more speechlike. The frequency extent of the 10-Hz/ms transitions was 500 Hz for the 50-ms standard transitions. The total duration of the stimuli remained constant (110 or 130 ms) by changing the durations of the first and second formant frequencies of the standard transition (30 or 50 ms) and of the standard steady state (80 ms, $F2=1300$ Hz). The durations of the transitions and the steady state varied in steps of 1 ms for each 5- or 10-Hz/m change in end-point frequency. Discriminability was tested for 30- and 50-ms transitions only. The stimuli were generated by a digital formant synthesizer on a μVAX II (Klinkers, Weenink, 1988). The pulse source had a fundamental frequency of 110 Hz. The formant bandwidth was always 10% of the (changing) formant frequency. To ensure a precise generation of these formant transitions, the stimuli were sampled at 1.2 MHz. After low-pass filtering, they were downsampled to 20 kHz (16-bit resolution). The formant frequency values were updated every 1 ms. Although the first period started at a zero crossing, a 2-ms cosine window was used at both end points to avoid on/offset clicks. Stimuli were centrally added to 300 ms of low level white noise (signal-to-noise-ratio of approximately 70 dB) to minimize external auditory factors. The 16-bit stimuli were transferred from the μVAX II to an IBM-PC in order to have real-time capability for training and testing. The stimuli were generated by means of an OROS-AU22 DSP board with D/A converter.
testing session included a new order of randomization. At least 20 responses were collected per subject for each standard-comparison pair of transitions above and below the standard transition. Averaged over higher and lower rates of frequency change, each difference limen is, therefore, based on, on average, 40 responses.

Subjects were seated in front of a terminal and heard two formant glides in 300-ms low level noise binaurally at a comfortable listening level. By pressing the appropriate mouse key they indicated whether both stimuli in a pair were the same or different. Feedback was given after each response. Before each testing series subjects were informed that the proportion of catch trials was fixed at 50%, so that they could adjust their criterion accordingly.

Difference limens were calculated according to the theory of signal detectability (TSD, Egan and Clarke, 1966). Assuming that two assumptions are met, i.e., that the data are normally distributed and of equal variance, a measure of performance, such as \( d' \) or \( P(C)_{\text{max}} \) (Macmillan and Creelman, 1991), can be calculated that separates sensitivity from response bias. Difference limens in frequency corresponded to frequency extents yielding a \( P(C)_{\text{max}} \) of at least 0.75. \( P(C)_{\text{max}} \) was determined for each of the higher and lower rate transitions separately. Difference limens are measured relative to the end-point frequency of the standard transition.

C. Subjects

Four normal hearing subjects (aged between 23 and 36 years), three of whom had served as subjects in previous experiments (van Wieringen and Pols, 1992, 1994), participated in experiment 1. Difference limens in end-point frequency were determined for the Clu/-like and iu/C-like (two subjects) and for the iu/-like and iu/C-like (two subjects) stimuli, during two or three 10-min sessions a day with short breaks in between. Two of these subjects participated in experiment 2. They received 2 h of practice before actual data collection began. Two subjects were paid for their participation.

II. RESULTS

A. Experiment 1: Discrimination of fixed duration transitions

The average just-noticeable differences in end-point frequency of initial and final, single and complex transitions with steady states are plotted in Fig. 4 in terms of the differences in end-point frequency (Hz) as a function of transition duration. The data, which are averaged over four subjects, formant patterns, frequency extents (transition directions), and higher/lower rate of frequency change, display the same tendency as the just-noticeable differences in frequency of transitions without a steady state (van Wieringen and Pols, 1992, 1994), i.e., they decrease with increasing transition duration. Due to the presence of a steady state and possibly also the difference in frequency region, difference limens in end-point frequency of these second formant single transitions with steady states are larger than those of the first formant, isolated transitions, which were 70, 63, and 58 Hz for 20, 30, and 50 ms, respectively (included in Fig. 4). Discriminability in end-point frequency is better for single F2 transitions with steady states than for F2 transitions which are part of complex stimuli, presumably because the surrounding formant frequencies partially mask discrimination of the varying F2 transition.

A fully factorial ANOVA (Kirk, 1982) including stimulus complexity (single and complex), subjects (four), transition durations (three), transition position (two), and relative rate-of-frequency change showed that all the factors except that of subjects (\( p > 0.5 \)) were highly significant (\( p < 0.001 \)). The statistical analyses were repeated for the single and complex stimuli separately due to the large number of factors and since the difference between the difference limens of single and complex stimuli was substantial.

Two split-plot factorial ANOVAs with two formant patterns (between blocks), two subjects, three transition durations, two positions, two directions per position, and relative rate-of-frequency change per formant pattern (within blocks) was conducted on the difference limens in end-point frequency, expressed either in end-point frequency (Hz) or in transition rate (Hz/ms). The results of the statistical analyses of the single and complex stimuli are listed in Table II. The factor subjects was not significant, neither within nor across blocks (formant patterns).

Statistical analyses of the single transitions with steady states showed three significant main effects, i.e., transition duration, transition position, and higher or lower rate of frequency change. The factors transition direction and formant patterns were not significant and there were no significant first- or second-order interactions.

Statistical analysis of the complex stimuli revealed two highly significant main effects, i.e., transition duration and transition position. The factors formant patterns, transition direction, and increment/decrement were not significant (\( p > 0.02 \)). For the difference limens in transition rate (Hz/ms) the interaction of position by duration was significant.

\[ \text{FIG. 4. Difference limens in end-point frequency (Hz) of single and complex transitions with steady states, averaged over four subjects, two formant patterns, two directions (two frequency loci), and higher/lower rates of frequency change for initial and final position separately. Difference limens in end-point frequency of isolated transitions (van Wieringen and Pols, 1992, 1994) are also included in the plot (crosses).} \]
TABLE II. Summary table for the ANOVAs conducted on the single and complex formant transitions in terms of Hz and Hz/ms.

<table>
<thead>
<tr>
<th></th>
<th>Single</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formant patterns (two)</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Subjects (two per form. p.)</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Transition duration</td>
<td>Hz [F(2,4)=214.3, p&lt;0.001]</td>
<td>Hz [F(2,4)=4599.9, p&lt;0.001]</td>
</tr>
<tr>
<td></td>
<td>Hz/ms [F(2.4)=2099.7, p&lt;0.001]</td>
<td>Hz/ms [F(2,4)=54345.2, p&lt;0.001]</td>
</tr>
<tr>
<td>Transition position</td>
<td>Hz [F(1,2)=93.9, p&lt;0.01]</td>
<td>Hz [F(1,2)=656.1, p&lt;0.005]</td>
</tr>
<tr>
<td></td>
<td>Hz/ms [F(1,2)=131.2, p&lt;0.01]</td>
<td>Hz/ms [F(1,2)=3070.9, p&lt;0.001]</td>
</tr>
<tr>
<td>Transition direction</td>
<td>n.s</td>
<td>n.s</td>
</tr>
<tr>
<td>Rel. rate of freq. change</td>
<td>[F(1,2)=217.8, p&lt;0.005]</td>
<td>n.s</td>
</tr>
<tr>
<td></td>
<td>[F(1,2)=395.5, p&lt;0.005]</td>
<td>n.s</td>
</tr>
<tr>
<td>Position × duration</td>
<td>Hz/ms [F(2,4)=81.22, p&lt;0.005]</td>
<td></td>
</tr>
</tbody>
</table>

The significant effects related to factors of transition position and relative rate-of-frequency change will be discussed separately in the next sections, followed by a description of some relevant psychoacoustic cues.

1. Transition position (initial versus final)

The just-noticeable differences in end-point frequency of single and complex transitions with steady states, averaged over the four subjects, are listed in Tables III and IV, respectively. The data are listed separately for transitions incrementing and decrementing in relative rate-of-frequency change (and end-point frequency) with respect to the standard transition. Difference limens in end-point frequency of both single and complex transitions with steady states are consistently larger in initial than in final position. This is in agreement with difference limens of isolated tone and formant transitions (Elliott et al., 1989; Cullen et al., 1992) as well as transitions with a steady state (Collins, 1984; Elliott et al., 1989; Porter et al., 1991). Our data on short isolated first formant transitions were similar (van Wieringen and Pols, 1992, 1994) with respect to position (also direction) of the transitions, although we did not examine the discriminability of the isolated converging and diverging transitions systematically.

2. Relative rate-of-frequency change for the single transitions (incrementing versus decrementing)

Relative rate-of-frequency change is discussed only for the single transitions, as this was not a significant effect for the complex transitions. Just-noticeable differences in end-point frequency of single transitions incrementing in transition rate with respect to the standard transition are significantly smaller than those decrementing with respect to the standard, especially when the frequency extent is relatively small (see Table III). Previous studies with tone and formant glides have also found that transitions increasing in rate-of-frequency change are more discriminable than those decrementing in transition rate, especially for frequency extents varying between 100 and 300 Hz (Collins, 1984; Porter et al., 1991; Cullen et al., 1992). As the frequency extents become larger, the difference between incrementing and dec-
TABLE IV. Difference limens in end-point frequency (Hz) of rising and falling, initial and final complex transitions with steady states for relative rate-of-frequency change and formant patterns (/a/-like and /u/-like) separately. The standard deviations varied between 0.0 and 12.6 Hz.

<table>
<thead>
<tr>
<th>Complex /a/-like (initial transitions)</th>
<th>Complex /u/-like (final transitions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rising /a/-like (700–1300 Hz)</td>
<td>falling /u/-like (1800–1300 Hz)</td>
</tr>
<tr>
<td>20 ms</td>
<td>D.L.-D</td>
</tr>
<tr>
<td></td>
<td>230 Hz</td>
</tr>
<tr>
<td></td>
<td>D.L.-I</td>
</tr>
<tr>
<td></td>
<td>230 Hz</td>
</tr>
<tr>
<td>30 ms</td>
<td>D.L.-D</td>
</tr>
<tr>
<td></td>
<td>195 Hz</td>
</tr>
<tr>
<td></td>
<td>D.L.-I</td>
</tr>
<tr>
<td></td>
<td>195 Hz</td>
</tr>
<tr>
<td>50 ms</td>
<td>D.L.-D</td>
</tr>
<tr>
<td></td>
<td>150 Hz</td>
</tr>
<tr>
<td></td>
<td>D.L.-I</td>
</tr>
<tr>
<td></td>
<td>145 Hz</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Complex /a/-like (final transitions)</th>
<th>Complex /u/-like (initial transitions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rising /u/-like (1300–1800 Hz)</td>
<td>falling /a/-like (800–1300 Hz)</td>
</tr>
<tr>
<td>20 ms</td>
<td>D.L.-D</td>
</tr>
<tr>
<td></td>
<td>190 Hz</td>
</tr>
<tr>
<td></td>
<td>D.L.-I</td>
</tr>
<tr>
<td></td>
<td>185 Hz</td>
</tr>
<tr>
<td>30 ms</td>
<td>D.L.-D</td>
</tr>
<tr>
<td></td>
<td>170 Hz</td>
</tr>
<tr>
<td></td>
<td>D.L.-I</td>
</tr>
<tr>
<td></td>
<td>165 Hz</td>
</tr>
<tr>
<td>50 ms</td>
<td>D.L.-D</td>
</tr>
<tr>
<td></td>
<td>130 Hz</td>
</tr>
<tr>
<td></td>
<td>D.L.-I</td>
</tr>
<tr>
<td></td>
<td>125 Hz</td>
</tr>
</tbody>
</table>

3. Effective cues

If the data for both formant patterns are organized in terms of the frequency extent, difference limens in end-point frequency of single and complex transitions with steady states in initial and final position increase as the frequency extents become larger (experiment 1). Transitions starting at the /b/-like end point of /C/a/-like and /a/C-like stimuli yield smaller difference limens in end-point frequency than their /C/u/-like and /u/C-like counterparts, as the frequency extent of the first is 100 Hz and that of the second is 600 Hz (see Table I for frequency extents and transition rates). However, the opposite holds for the /d/-like stimuli: these yield larger difference limens for the /C/u/-like and /u/C-like stimuli than for the /C/a/-like and /a/C-like ones, due to the frequency extent of the first being twice as large. As the formant bandwidth varies with center frequency, the spectral peaks of the shorter frequency extents are sharper than those of the larger frequency extents. Such bandwidth characteristics may partly explain the difference in discriminability for the transitions tested in the present study. Discrimination also seems to be affected by the direction of the transition in final position. Due to masking of subsequent lower frequencies, difference limens in frequency of the falling /a/C-like transitions are similar to or smaller than the rising ones in final position, despite a 100-Hz larger frequency extent of the /a/ C-like stimuli (Tables III and IV).

B. Experiment 2: Discrimination of constant rate transitions

Just-noticeable differences in end-point frequency of complex stimuli, of which the second formant has a constant rate of frequency change (Fig. 5), are much smaller than those which vary in frequency extent at a constant transition.
duration (Fig. 4). The difference limens in end-point frequency for the transitions varying at 10 Hz/ms are 51 and 52 Hz for the 30-ms standard transition, and 50 and 47 Hz for the 50-ms standard transition in initial and final position, respectively. They are lower for the 50-ms standard transitions with a rate-of-frequency change of 5 Hz/ms, i.e., 42 and 38 Hz, in initial and final position, respectively (Fig. 5). Two ANOVAs were performed, one comparing the two different durations and one comparing the two different rate-of-frequency changes, as the statistical design was not completely balanced. A three-way factorial design with subjects as a random factor, and duration, position, and relative rate-of-frequency change as fixed factors revealed a significant effect of duration \( F(2,7) = 23.4, p < 0.001 \).

A three-way factorial design with subjects as a random factor and transition rate, position, and relative rate-of-frequency change as fixed factors showed highly significant effects of rate-of-frequency change, i.e., \( F(2,7) = 188.2, p < 0.001 \), and position of the transition, i.e., \( F(2,7) = 31.9, p < 0.001 \). Decreasing the rate of frequency change from 10 to 5 Hz/ms for the same transition duration improves discrimination, as is illustrated in Fig. 5 for the 50-ms transitions. This improvement in discrimination is partly explained by a larger temporal difference and partly by a decrease of frequency uncertainty with decrease of transition rate.

The smaller difference limens in the constant rate condition can be explained by the excellent temporal resolution of the auditory system, rather than by the lower rate of frequency change (compared to the difference limens varying in transition rate). Small temporal changes affect discrimination markedly (see also van Wieringen and Pols, 1992, 1994), even when the total duration of transition and steady state remains constant. Discriminability can, to some extent, also be cued by changes in loudness, because a change in duration is associated with a change in energy (e.g., Watson and Gengel, 1969). We have tried to prevent loudness from being a cue by keeping the total duration of the stimulus constant. Nevertheless, loudness may still cue discrimination if the subject listens to changes in the duration of the transition. The difference limens of the 5- and 10-Hz/ms transitions involve temporal changes of approximately 8 and 5 ms, respectively. These changes, which correspond to a 0.5- to 1-dB change in loudness (Watson and Gengel, 1969 for a 1000-Hz tone), are somewhat smaller than the difference limens in intensity for vowels (Flanagan, 1955).

### III. DISCUSSION

Discrimination experiments were performed to determine the extent to which general psychoacoustic properties underlie the perception of speech sounds, specifically that of plosives in the present study. The transitions examined in the present study have trajectories similar to stop consonants in speech, i.e., they are short and rapid, and have the appropriate end-point frequencies. Despite the increase in difference limens with increase of stimulus complexity, the global pattern of the discrimination functions for tone sweeps and single or multiformalt stimuli is similar: all difference limens decrease with increasing transition duration (van Wieringen et al., 1993a). Several studies have determined the effect of transition duration (Pollack, 1968; Dooley and Moore, 1988b; Elliott et al., 1989, 1991; Porter et al., 1991). We interpret the improvement in discriminability in terms of an increase in processing time. The spectral change of a longer stimulus is less than that of a shorter one. It is expected that the less the formant frequency changes, the easier it is to determine the end-point frequency of the stimulus. Furthermore, which psychoacoustic cues are used by the listeners depend on whether frequency extent or transition duration are varied experimentally. Relatively small temporal changes strongly affect discrimination, not only when the duration of isolated transitions is varied (van Wieringen and Pols, 1992, 1994), but also when the durations of the transition and the stationary part are varied at a constant total duration (experiment 2). Difference limens are, therefore, smaller for the transitions that maintain a constant rate-of-frequency change (and thus vary in duration), than those that vary in transition rate (at a fixed duration). The improvement in discrimination with decrease of transition rate (10 vs 5 Hz/ms) is explained by a decrease in frequency uncertainty per unit of time (due to a decrease in spectral variation).

### A. Stimulus complexity

Difference limens determined for varying second formant transitions in multiformalt complexes with steady states are larger than those for single transitions with steady states, possibly because the first and third formant partially mask the varying second formant frequency. Compare tone and formant glides: tone glides yield smaller difference limens than formant glides, because only one harmonic varies in frequency, while the amplitudes of several harmonics change when a formant transition varies in frequency. Compared to a single transition it is even more difficult to detect a change in the harmonics of a complex stimulus when the changing formant is surrounded by other (fixed) formant frequencies. This is in agreement with Danaher et al. (1973), who also examined how the level of the formant frequencies affects discrimination. It is also possible that the speechlike quality of the stimuli causes difference limens in end-point frequency to increase with increasing stimulus complexity, because of the increasing difficulty of attending selectively to varying parameters in the presence of additional acoustical cues.

### B. Backward masking

With the exception of the 30-ms transition in the constant rate condition, all difference limens for transitions of similar extents and durations were smaller in final than in initial position (Figs. 4 and 5). This is in agreement with previous experiments using formant glides (Elliott et al., 1989; Porter et al., 1991) and even with isolated tone glides (Cullen et al., 1992). As the difference in sensitivity between converging and diverging isolated transitions increases when a steady state follows a transition, rather than when it precedes one, discrimination seems to be affected more by backward than by forward masking. In other words, the following steady state affects end-point discrimination of the preceding transition more than the preceding steady state does for end-point discrimination of the following transition. The data is
supported by previous discrimination experiments with tone and formant glides (Collins, 1984; Elliott et al., 1991; Lacerda, 1987) indicating better difference limens for transitions diverging from a common frequency (final position) than for those converging to a common frequency (initial position). As the end point of the transition is an important cue in differentiating between standard and comparison stimuli, the results could be interpreted in terms of a memory recency effect, especially when the transitions are followed by a steady state. For instance, discrimination experiments with tonal patterns (Watson et al., 1976) show how performance is influenced by memory and attention, rather than by the limitations of the auditory system. The data cannot, however, solely be explained by a (cognitive) memory effect: in that case practice would diminish the perceptual asymmetry. Not only were the listeners highly trained, they also continued to hear the final transition more as a glide than the initial one. In addition to a higher memory demand in initial than in final position, the initial transition could be partially masked by the following steady state. In final position the vibrations gradually die away, while they are continuously followed by new waves in initial position. By means of a model of the inner ear, excitation patterns of rising and falling tone glides in initial and final position were compared to determine whether backward masking can be explained by a peripheral mechanism (van Wieringen et al., 1993b). That study showed some evidence of a perceptual asymmetry in the responses of the auditory-nerve fibers (as modeled so far). Discrimination in end-point frequency of initial transitions followed by a steady state is therefore likely to be affected by several factors, including transients and a decay in short-term memory.

C. Discriminative cues

Discrimination experiments with both isolated as well as single and complex formant transitions with steady states generally show the end-point frequency or the difference between the onset and offset frequencies of the transition to be perceptually important cues (Dooley and Moore, 1988b; Pols and Schouten, 1987; Elliott et al., 1989, 1991; Collins et al., 1990; van Wieringen and Pols, 1992, 1994). The perceptual importance of the end-point frequency was also illustrated in matching experiments where listeners matched the frequency of a steady state near the end-point frequency of the transition (Brady et al., 1961). If discrimination were based on rate-of-frequency change, difference limens in end-point frequency of isolated transitions (Cullen et al., 1992) and of initial and final transitions would be more similar, and the difference between incrementing and decrementing transitions would be larger in final position. Due to the covariation of frequency extent with transition rate or transition duration, none of the physical cues are disclaimed: transitions with a smaller frequency extent also have a lower rate-of-frequency change, and both cues may contribute to an increase in sensitivity. Transition rate is perceptually important in both experiments, with lower rates yielding smaller difference limens, and incrementing transitions being more discriminable than decrementing transitions, possibly because frequency perception is more accurate per unit of time for the lower rate transitions. However, the similarity of difference limens for incrementing and decrementing transitions in final position, and for transitions with relatively larger frequency extents, indicates that frequency cues, such as end-point frequency or frequency extent, are perceptually probably more important than rate cues (and temporal cues in experiment 2). Although longer transitions have lower rates of frequency change than shorter transitions, smaller difference limens occur as a result of the increasing frequency discrimination, due to an increase in processing time, rather than to a decrease in transition rate. This is illustrated by the 30-ms /bu/-like and /ub/-like (3.3 Hz/ms) and /da/-like and /ad/-like (16.7 Hz/ms) transitions, which have a lower rate-of-frequency change than the 50-ms /a/C-like and /C/a/-like ones (of 10 and 12 Hz/ms) and the /d/-like and /l/-like ones (20 Hz/ms), respectively. Despite the lower transition rate, auditory sensitivity is, on average, better for the longer, 50-ms stimuli. Identification experiments have also shown that transition rate is perceptually less salient than frequency extent (Schwab et al., 1981).

The difference in difference limens for transitions in initial and final position suggests that consonants preceding a vowel are more poorly discriminated than those following a vowel. The difference between discrimination functions of speechlike CV and VC syllables (Lacerda, 1987) was explained by the temporal sequence of auditory excitations, which is reversed for the two kinds of stimuli. The excitation of auditory channels produced by VC syllables is not disturbed by new auditory patterns. Some identification experiments with natural speech tokens have also shown the VC transition to be more consonant specific than the CV transition (Sharf and Hemeyer, 1972), whereas others show both transitions to be perceptually equal in identifying consonants (Pols and Schouten, 1978; Pols, 1979). Contrary to synthetic syllables, natural CV and VC syllables are not mirror images of each other, and it is not easy to relate the perception of nonspeech sounds to that of speech sounds. However, it is of interest to examine the sensory smearing of auditory information further with natural speech, especially with respect to backward masking of the vowel on the preceding consonant. Subsequent experiments with more speechlike and (manipulated) natural speech tokens will examine whether the perceptual asymmetries found in psychoacoustic and speech perceptual experiments are related to each other.

ACKNOWLEDGMENTS

We thank Bert Schouten for numerous comments on this paper. The first author was financially supported in a Ph.D. program by the University of Amsterdam.

of brief tone-glides with high rate of frequency-change," Abstracts of the
Fifteenth Midwinter Meeting of the Association for Research in Otolaryn-
gology, p. 67.
of formant transitions in synthetic vowels," J. Speech Hear. Res. 16, 439–
451.
and transitional cues for consonants," J. Acoust. Soc. Am. 27, 769–773
steady and gliding tones: A new method for estimating sensitivity to rate
glides as a function of frequency and duration," J. Acoust. Soc. Am. 84,
2045–2057.
Experimental Methods and Instrumentation in Psychology, edited by J.
Elliott, L. L., Hammer, M. A., Scholl, M. E., Carrell, Th.D., and Wasowicz,
J. M. (1989). “Discrimination of rising and falling simulated single-
Soc. Am. 86, 945–953.
Sciences (Brooks/Cole, Belmont, MA).
Lacerda, F. (1987). “Effects of peripheral auditory adaptation on the
discrimination of speech sounds," Phonetic Experimental Research Institute
at the Institute of Linguistics, University of Stockholm (PERILUS) VI.
“Tempo of frequency change as a cue for distinguishing classes of speech
sounds," J. Exp. Psychol. 52, 127–137.
Liberman, A., Cooper, F., Shankweiler, D., and Studdert-Kennedy, M.
Exp. Psychol. 77, 535–541.