PERCEPTUAL ASYMMETRY BETWEEN INITIAL AND FINAL GLIDES: PSYCHO-ACOUSTICS AND COCHLEAR ENCODING

Astrid van Wieringen, Paul Boersma and Louis C.W. Pols

Abstract

Psycho-acoustic experiments show a perceptual asymmetry between tone and formant transitions in initial and final position: difference limens in end point frequency are significantly larger for transitions followed by a steady-state than preceded by one. Sensitivity of end point detection decreases as a result of auditory masking and/or a memory recency effect. The present paper describes how sensitivity varies with duration, stimulus complexity and position of the transition (section I) and tentatively examines whether peripheral indications can be found for the perceptual asymmetry, by means of a cochlear model (section II).

Introduction

Just noticeable differences (JND's) in end point frequency are often found to be smaller for transitions diverging to different frequencies than for those converging to a similar frequency (Elliott *et al.*, 1991; Cullen *et al.*, 1992). In addition to this, falling transitions yield smaller difference limens than rising ones and sensitivity is, on average, greater for transitions incrementing with respect to standard transitions rather than for decrementing transitions (Porter *et al.*, 1991; van Wieringen and Pols, 1992, submitted). The difference in sensitivity between converging and diverging transitions increases when a steady-state is added (Collins, 1984, 1990; van Wieringen *et al.*, 1993). Psycho-acoustical studies suggest that the *following* steady-state affects end point discrimination of the preceding transition more than the *preceding* steady-state does for discrimination of the following transition.

Apart from experimental differences, many different factors may account for the varying sensitivity: converging tone transitions, even without steady-state, could be perceived as having a more similar pitch than diverging transitions, higher perceptual weight could be assigned to the most recent information (Brady *et al.*, 1961), later occurring information could (physically) mask preceding input or discrimination could be affected by a memory recency effect. Results of psycho-acoustical experiments indicating the perceptual importance of the terminal frequency (e.g., Brady *et al.*, 1961) suggest that higher-level (central) factors influence the perceptual asymmetry. In the present paper the perceptual asymmetry is examined on a lower, peripheral, level, by comparing excitation patterns of rising and falling glides in initial and final position.

In section I, difference limens in end point frequency are determined for short tone (section 1a) and formant (section 1b) glides with trajectories similar to stop consonants in speech (Delattre *et al.,* 1955). Just noticeable differences were compared for 20 ms, 30 ms and 50 ms single and multi-formant speechlike stimuli preceded or followed by an 80 ms steady-state (section 1c) to examine how sensitivity varies with stimulus complexity (van Wieringen and Pols, submitted). However, as the observed asymmetry is more difficult to examine with broad-band stimuli than with tone glides, difference limens were determined of 40 ms tone glides, followed or preceded by a 200 ms steady-state, in initial and final position, respectively (section 1d). In both the tone and formant stimuli the transition rate was varied by changing the frequency extent of the transition at a constant transition duration.

In section II, excitation patterns of several kinds of tone glides are compared to determine whether any of the physical properties inducing perceptual asymmetry are reflected by the basilar membrane or by neural responses. Tone glides are preferred above formant transitions, as they are affected less by masking effects of broad bands of energy.

Section I : difference limens in end point frequency

I A: tone glides - stimuli and procedure

In this section the stimulus characteristics and the measuring procedure, used to determine end point sensitivity of tone glides, will be discussed. Difference limens in end point frequency were determined by different methods for the tone and formant glides as data of the two experiments were collected on two different locations[1] . We assume that the procedures are not of major influence on the results. Just noticeable differences in onset frequency were determined for 40 ms tone glides preceded or followed by a 200 ms steady-state. The end point frequencies of the initial-falling and the final-rising standard transitions swept from the 1850 Hz steady-state frequency to 2050 Hz, thus having a frequency range of 200 Hz (Figure 1, left hand CV and VC).
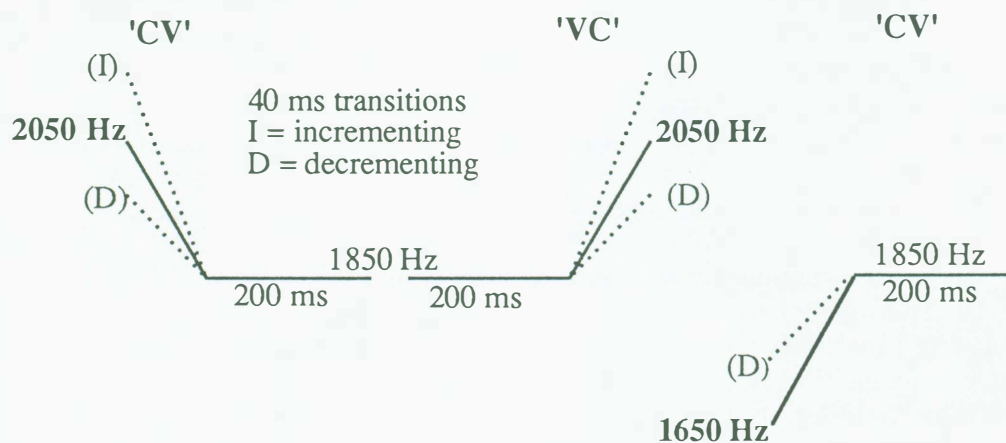

Figure 1. Schematic representation of the initial and final tone glides, incrementing (I) and decrementing (D) with respect to the standard transition. The standard transitions range between 1850 Hz and 2050 Hz (left part) and 1650 Hz and 1850 Hz (right part).

---

[1] The data on tone glides were collected while the first author worked with Jack Cullen at the Louisiana State University (U.S.A.) in 1992.

Transitions having higher rates of frequency change (than the standard) varied from 2055 Hz to 2250 Hz in 5 Hz steps, and those lower than the standard varied from 2045 Hz to 1855 Hz for the falling or rising transitions in initial and final position, respectively. There was an additional condition in which transitions rose from 1650 Hz to 1850 Hz in initial position (Figure 1, right part). In this condition, only the transitions decrementing in rate of frequency change were examined. All stimuli were generated at 20 kHz sampling frequency (12 bit resolution) and low-pass filtered at 4.9 kHz. JND'S in end point frequency between standard and comparison tone glides were determined using a four-interval, 2-alternative forced-choice paradigm. The subject was presented with four intervals each time, of which the first and the fourth intervals always contained the standard stimulus, and either the second or the third randomly contained the comparison stimulus. The subject's task was to decide whether the second or third interval contained the comparison signal.

Stimuli were presented monaurally at a comfortable loudness level of 70 SPL over a Sennheiser headphone in a quiet listening booth. Feedback was given after each response. Thresholds were measured adaptively by a modified PEST procedure (Taylor and Creelman, 1967). The threshold was defined at 75% correct and the JND was taken as the difference between the end point frequencies of the standard and comparison stimuli. Two experienced subjects participated in the task.

I B: formant glides - stimuli and procedure

Difference limens in end point frequency were determined of formant transitions with short and rapid trajectories similar to stop consonants in speech. Transition rate was varied by changing the frequency extent of the second formant transition at a fixed transition duration of either 20 ms, 30 ms, or 50 ms. The standard transitions began or ended at either /b/-like or /d/-like loci, i.e., at either 700 Hz or 1800 Hz, respectively. Difference limens in end point frequency were determined of single (Figure 2) and of complex F2-transitions (Figure 3) preceded or followed by an 80 ms steady-state. The complex stimulus consisted of a fixed F1-transition with steady-state, a stationary third formant, and a 20 ms voice bar to make the stimuli sound more speech-like.
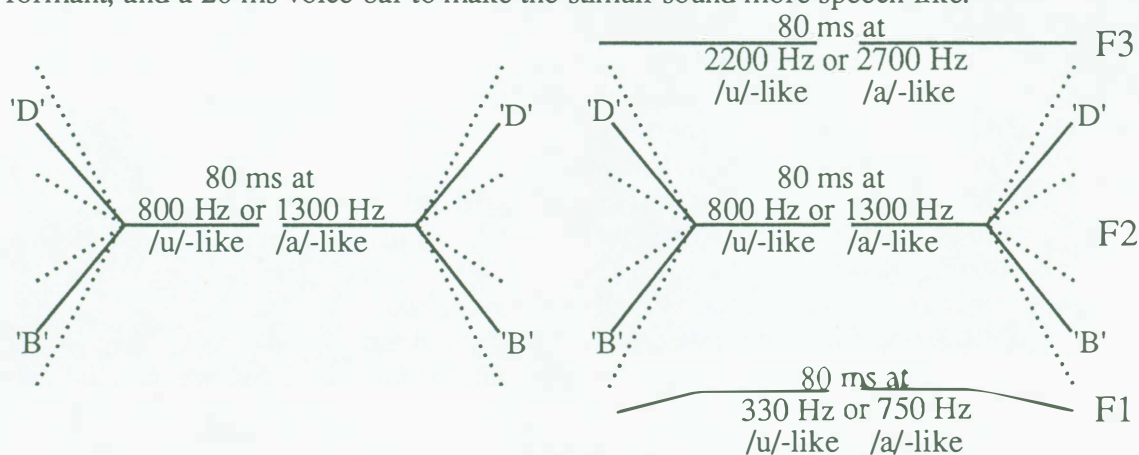

Figure 2   Single   Figure 3   Complex

Figures 2 & 3. Schematic representation of 20 ms, 30 ms, and 50 ms single (Figure 2) and complex (Figure 3) formant stimuli followed (initial) or preceded (final) by an 80 ms steady-state. The F2 transition varies in end point frequency from /b/-like (700 Hz) or /d/-like (1800 Hz) loci. The dashed lines indicate the transitions with a higher and lower rate-of-frequency change than the standard transition.

The transitions preceded or followed a steady-state signal with either an /a/-like or /u/-like formant pattern. For the C/a/-like and /a/C-like stimuli the F2-frequency of the steady-state was 1300 Hz, for the C/u/-like and /u/C-like stimuli it was 800 Hz. The F1-transition of the complex C/a/-like and /a/C-like stimuli always rose or fell from 220 Hz to 750 Hz, and the frequency of the third formant was fixed at 2700 Hz. The F1-transition of the complex C/u/-like and /u/C-like stimuli always rose or fell from 220 Hz to 330 Hz, and the third formant frequency was fixed at 2200 Hz (Figure 3).

The stimuli were software-generated by a digital formant synthesizer on a μVax II (Klinkers, Weenink, 1988). The pulse voice 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 formant frequency values were updated every 1 millisecond at a sample frequency of 1.2 megaHz. After low pass filtering, they were downsampled to 20 kHz (16 bit resolution). 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 50 dB) to minimize external auditory factors. The 16 bit stimuli were transferred from the μVAX II to an IBM-compatible PC in order to have real-time capability for training and testing. The stimuli were generated real-time by means of an OROS-AU22 D/A converter.

Difference limens were collected 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. 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. Every set of randomized stimuli was well balanced in that each stimulus was offered twice, once before and once after the standard transition, resulting in eighty responses per standard stimulus, per testing session. Approximately fifty percent of each testing series consisted of catch trials, i.e., physically identical pairs of stimuli, to reduce response bias towards responding 'different' for all pairs of stimuli. All conditions were tested separately. Each 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, resulting, on average, in 40 responses per standard transition.

The subject was 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 he/she 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 on the fixed proportion (50 %) of catch trials.

Difference limens were determined by the theory of signal detectability (TSD, Egan *et al.*, 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)max (Macmillan and Creelman, 1991), can be calculated that separates sensitivity from response bias. Thresholds corresponded to frequency extents yielding a P(C)max of at least 0.75. P(C)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.

Four normal hearing subjects (aged between 23 and 36 years), of which three had served as subjects in previous experiments (van Wieringen and Pols, 1992, 1993),

participated individually in the experiments. Difference limens in end point frequency were determined for the C/a/-like and /a/C-like (two subjects), and for the C/u/-like and /u/C-like(two subjects) stimuli, during two or three 10-minute session a day with short breaks in between, for many days. They received 2 hours of practice before actual data collection began. Two subjects were paid for their participation.

I C: average JND's of tone and formant glides

The average just noticeable differences in end point frequency of initial and final, tone, single and complex transitions with steady-state, are plotted in Figure 4 in terms of differences in end point frequency (Hz) as a function of transition duration. The data, which are averaged over 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 isolated transitions (van Wieringen and Pols, 1992, 1993), i.e., they decrease with increasing transition duration due to an increase in processing time (Elliott *et al.*, 1991; Porter *et al.*, 1991). Due to the presence of a steady-state, thresholds in end point frequency of single transitions with steady-state are larger than those of isolated transitions, which were 70 Hz, 63 Hz, and 58 Hz for 20 ms, 30 ms and 50 ms, respectively (see crosses in Figure 4). Discriminability in end point frequency is better for single F2-transitions with steady-state than for F2-transitions which are part of complex stimuli, presumably because of the increasing complexity of psychoacoustical cues involved in the processing of more speech-like stimuli (many more frequencies are excited by broad-band stimuli).
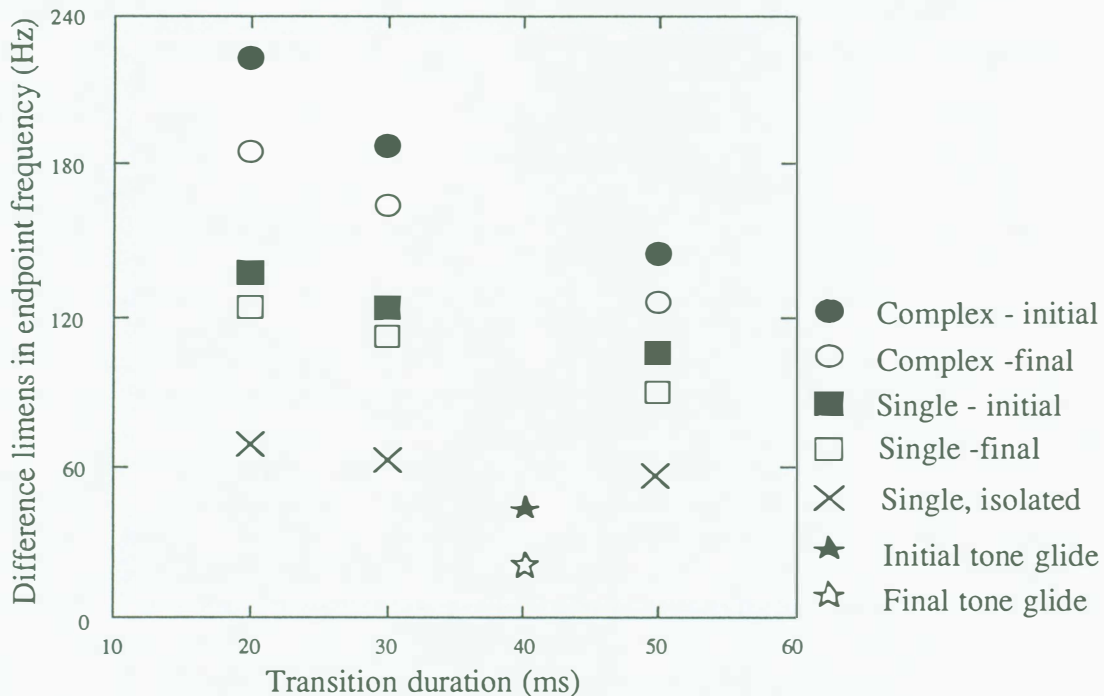

Figure 4. Average difference limens in end point frequency (Hz) of tone sweeps (stars), and single (squares) and complex (circles) formant glides in initial and final position (closed versus open symbols). Difference limens of isolated formant glides (van Wieringen and Pols, 1992, 1993) are also included in the plot (crosses).

Detailed analyses show how sensitivity varies as a result of the direction, the position and the rate-of-frequency change. Data suggest that temporal backward masking affects discrimination more than forward masking.

I D: detailed results - JND-formant glides

Difference limens in end point frequency of incrementing and decrementing initial and final formant transitions are listed in Table I (single) and Table II (complex). The tables illustrate the difference in sensitivity between the different kinds of stimuli. A split-plot factorial ANOVA (Kirk, 1982) with two formant patterns (/u/ vs. /a/, between-blocks), and two subjects, three transition durations (20 ms, 30 ms, 50 ms), two positions (CV and VC), two frequency extents per position (/b/ and /d/ =direction), and higher or lower rate of frequency change per formant pattern (within blocks) was conducted on the difference limens expressed in end point frequency (Hz) or in transition rate (Hz/ms) for the single and complex transitions with steady-state separately. The subjects factor was random, the others were fixed.

Statistical analyses of the single transitions with steady-state showed three significant main effects, i.e., transition duration, transition position, and higher or lower rate of frequency change. The factor duration is highly significant for the data in terms of Hz, i.e., $[F(2,4) = 214.3, p < 0.001]$ and in terms of Hz/ms, i.e., $[F(2,4) = 2099.7, p < 0.001]$. Other significant effects are position, i.e., $[F(1,2) = 93.9, p < 0.01]$ for the difference limens in end point frequency (Hz), and $[F(1,2) = 131.2, p < 0.01]$ for the transition rate thresholds (Hz/ms), and relative rate-of-frequency change, i.e., $[F(1,2) = 217.8, p < 0.005]$ for the difference limens in end point frequency (Hz), and $[F(1,2) = 395.5, p < 0.005]$ for the difference limens in transition rate (Hz/ms). The factor frequency extent was not significant, and there were no significant first or second order interactions.

Statistical analysis of the complex stimuli revealed two significant main effects. The transition duration factor was highly significant, i.e., $[F(2,4) = 54345.2, p < 0.001]$ for the transition rate thresholds (Hz/ms), and $[F(2,4) = 4599.9, p < 0.001]$ for the difference limens in end point frequency (Hz). The position of the transition was also highly significant, i.e., $[F(1, 2) = 3070.9, p < 0.001]$ for the difference limens expressed in Hz/ms, and $[F(1, 2) = 656.1, p < 0.005]$ for the difference limens expressed in Hz. The factors frequency extent and increment/decrement were not significant. For the difference limens in transition rate (Hz/ms) the interaction position (CV and VC) by duration was significant, i.e., $[F(2, 4) = 81.22, p < 0.005]$.

Difference limens in end point frequency of single and complex transitions with steady-state in initial and final position increase with increasing frequency extent. Therefore, transitions starting at the /b/-like locus of C/u/-like and /u/C-like stimuli yield smaller difference limens in end point frequency than their C/a/-like and /a/C-like counterparts, as the frequency extent of the first is 100 Hz and that of the second is 600 Hz. However, the opposite holds for the transitions with end points at the /d/-like locus: these yield larger thresholds 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. In general, difference limens increase with increasing frequency extent in initial position.

The present experiments have shown that discrimination becomes more difficult with increasing stimulus complexity. The difference between incrementing and decrementing transitions is not significant for the complex stimuli, presumably as a result of the surrounding formants masking discrimination. To gain insight into the perceptual

IFA Proceedings 17, 1993

asymmetry on a peripheral level, discrimination of initial and final tone glides is examined in detail as they yield the most subtle differences.

TABLE I. Difference limens (DL) in end point frequency (Hz) of incrementing (I) and decrementing (D) initial and final single transitions with steady-state for the two formant patterns separately.

	Single C/a/-like (initial transitions)				Single /a/C-like (final transitions)			
	Rising /ba/-like 700 Hz-1300 Hz		Falling /da/-like 1800 Hz-1300 Hz		Rising /ad/-like 1300 Hz-1800 Hz		Falling /ab/-like 1300 Hz-700 Hz	
	DL-D	DL-I	DL-D	Dl-I	DL-D	DL-I	DL-D	DL-I
20 ms	145 Hz	135 Hz	140 Hz	130 Hz	125 Hz	120 Hz	130 Hz	120 Hz
30 ms	130 Hz	125 Hz	125 Hz	125 Hz	115 Hz	110 Hz	115 Hz	110 Hz
50 ms	110 Hz	105 Hz	105 Hz	105 Hz	95 Hz	90 Hz	90 Hz	85 Hz
	Single C/u/-like (initial transitions)				Single /u/C-like (final transitions)			
	Rising /bu/-like 700 Hz-800 Hz		Falling /du/-like 1800 Hz-800 Hz		Rising /ud/-like 800 Hz-1800 Hz		Falling /ub/-like 800 Hz-700 Hz	
20 ms	140 Hz	120 Hz	145 Hz	140 Hz	130 Hz	130 Hz	125 Hz	115 Hz
30 ms	125 Hz	105 Hz	135 Hz	125 Hz	120 Hz	115 Hz	115 Hz	100 Hz
50 ms	100 Hz	90 Hz	115 Hz	110 Hz	100 Hz	90 Hz	85 Hz	80 Hz

TABLE II. Difference limens (DL) in end point frequency (Hz) of incrementing (I) and decrementing (D) initial and final complex transitions with steady-state for the two formant patterns separately.

	Complex C/a/-like (initial transitions)				Complex /a/C-like (final transitions)			
	Rising /ba/-like 700 Hz-1300 Hz		Falling /da/-like 1800 Hz-1300 Hz		Rising /ad/-like 1300 Hz-1800 Hz		Falling /ab/-like 1300 Hz-700 Hz	
	DL-D	DL-I	DL-D	DL-I	DL-D	DL-I	DL-D	DL-I
20 ms	230 Hz	230 Hz	220 Hz	215 Hz	190 Hz	185 Hz	190 Hz	185 Hz
30 ms	195 Hz	195 Hz	185 Hz	175 Hz	170 Hz	165 Hz	165 Hz	160 Hz
50 ms	150 Hz	145 Hz	145 Hz	140 Hz	130 Hz	125 Hz	125 Hz	120 Hz
	Complex C/u/-like (initial transitions)				Complex /u/C-like (final transitions)			
	Rising /bu/-like 700 Hz-800 Hz		Falling /du/-like 1800 Hz-800 Hz		Rising /ud/-like 800 Hz-1800 Hz		Falling /ub/-like 800 Hz- 700 Hz	
20 ms	215 Hz	200 Hz	235 Hz	240 Hz	195 Hz	190 Hz	175 Hz	170 Hz
30 ms	185 Hz	170 Hz	200 Hz	195 Hz	180 Hz	175 Hz	155 Hz	145 Hz
50 ms	145 Hz	130 Hz	155 Hz	150 Hz	135 Hz	135 Hz	125 Hz	110 Hz

asymmetry on a peripheral level, discrimination of initial and final tone glides is examined in detail as they yield the most subtle differences.

TABLE I. Difference limens (DL) in end point frequency (Hz) of incrementing (I) and decrementing (D) initial and final single transitions with steady-state for the two formant patterns separately.

	Single C/a/-like (initial transitions)				Single /a/C-like (final transitions)			
	Rising /ba/-like 700 Hz-1300 Hz		Falling /da/-like 1800 Hz-1300 Hz		Rising /ad/-like 1300 Hz-1800 Hz		Falling /ab/-like 1300 Hz-700 Hz	
	DL-D	DL-I	DL-D	Dl-I	DL-D	DL-I	DL-D	DL-I
20 ms	145 Hz	135 Hz	140 Hz	130 Hz	125 Hz	120 Hz	130 Hz	120 Hz
30 ms	130 Hz	125 Hz	125 Hz	125 Hz	115 Hz	110 Hz	115 Hz	110 Hz
50 ms	110 Hz	105 Hz	105 Hz	105 Hz	95 Hz	90 Hz	90 Hz	85 Hz
	Single C/u/-like (initial transitions)				Single /u/C-like (final transitions)			
	Rising /bu/-like 700 Hz-800 Hz		Falling /du/-like 1800 Hz-800 Hz		Rising /ud/-like 800 Hz-1800 Hz		Falling /ub/-like 800 Hz-700 Hz	
20 ms	140 Hz	120 Hz	145 Hz	140 Hz	130 Hz	130 Hz	125 Hz	115 Hz
30 ms	125 Hz	105 Hz	135 Hz	125 Hz	120 Hz	115 Hz	115 Hz	100 Hz
50 ms	100 Hz	90 Hz	115 Hz	110 Hz	100 Hz	90 Hz	85 Hz	80 Hz

TABLE II. Difference limens (DL) in end point frequency (Hz) of incrementing (I) and decrementing (D) initial and final complex transitions with steady-state for the two formant patterns separately.

	Complex C/a/-like (initial transitions)				Complex /a/C-like (final transitions)			
	Rising /ba/-like 700 Hz-1300 Hz		Falling /da/-like 1800 Hz-1300 Hz		Rising /ad/-like 1300 Hz-1800 Hz		Falling /ab/-like 1300 Hz-700 Hz	
	DL-D	DL-I	DL-D	DL-I	DL-D	DL-I	DL-D	DL-I
20 ms	230 Hz	230 Hz	220 Hz	215 Hz	190 Hz	185 Hz	190 Hz	185 Hz
30 ms	195 Hz	195 Hz	185 Hz	175 Hz	170 Hz	165 Hz	165 Hz	160 Hz
50 ms	150 Hz	145 Hz	145 Hz	140 Hz	130 Hz	125 Hz	125 Hz	120 Hz
	Complex C/u/-like (initial transitions)				Complex /u/C-like (final transitions)			
	Rising /bu/-like 700 Hz-800 Hz		Falling /du/-like 1800 Hz-800 Hz		Rising /ud/-like 800 Hz-1800 Hz		Falling /ub/-like 800 Hz- 700 Hz	
20 ms	215 Hz	200 Hz	235 Hz	240 Hz	195 Hz	190 Hz	175 Hz	170 Hz
30 ms	185 Hz	170 Hz	200 Hz	195 Hz	180 Hz	175 Hz	155 Hz	145 Hz
50 ms	145 Hz	130 Hz	155 Hz	150 Hz	135 Hz	135 Hz	125 Hz	110 Hz

I E: detailed results - JND - tone glides

In this section the difference limens of the tone glides (section 1a) will be presented. Just noticeable differences in end point frequency for transitions incrementing in transition rate with respect to the standard are, on average, 34.3 Hz in initial and 15.8 Hz in final position. Thresholds of 40 ms *isolated* (rising) transitions incrementing in transition rate are larger, i.e., 46.0 Hz and 35.3 Hz for converging and diverging glides, respectively (Cullen *et al.*, 1992), possibly due to subject differences (Table III).

The difference in discrimination between transitions incrementing and decrementing in transition rate is relatively large for initial transitions followed by a steady-state. Moreover, the results display individual differences in this condition. The average thresholds of initial falling ('/du/'-like) transitions decrementing in transition rate are 74.3 Hz. In final position, the difference between transitions incrementing and decrementing in transition rate is smaller: thresholds of rising transitions decrementing in transition rate are 40.0 Hz for one subject and 19.0 Hz for the other subject (29.5 Hz, on average). Average difference limens of isolated transitions decrementing in transition rate are 43.0 Hz and 27.0 Hz in initial and final position, respectively (Table III).

Table III. Difference limens (DL) in end point frequency (Hz) of rising and falling, incrementing (I) and decrementing (D) tone sweeps in initial and final position followed or preceded by a steady-state. The data of the isolated transitions (Cullen *et al.*, 1992) are also included in the table.

| | Initial | | | | Final | | | |
| | Rising | | Falling | | Rising | | Falling | |
	DL - I	DL - D	DL - I	DL - D	DL - I	DL - D	DL-I	DL-D
with steady -state	-	39.8 Hz	34.3 Hz	74.3 Hz	15.8 Hz	29.5 Hz	-	-
isolated	46.0 Hz	43.0 Hz			35.3 Hz	27.0 Hz		

Thresholds of the rising transition in initial position decrementing in transition rate are, on average, 39.8 Hz. The incrementing condition was not tested, see Figure 1. Due to the transition sweeping from low to high frequencies, discrimination of rising transitions is not as much affected by the frequencies of the steady-state when the end point approaches the steady-state as the falling ones. Isolated transitions also display a rise/fall effect; this is a frequency region effect with rising transitions being smaller than falling ones in initial position, and falling transitions being smaller than rising ones in terminal position (Cullen *et al.*, 1992). The data of the present experiments allow an across-frequency-region comparison, rather than a within-frequency-region comparison, as the frequencies of the steady-state are 1850 Hz (instead of 2050 Hz).

The difference limens of final transitions are significantly smaller than those of initial ones, as the final transitions are clearly perceived as glides, whereas the initial ones are not. This perceptual asymmetry may result from the absence of a gradual energy onset and offset of the tone glides. Transitions with a gradual onset in energy are perceived less like a click, especially at the beginning of the stimuli, than those which have maximum energy from the beginning of the stimuli. However, a pilot study (i.e., discrimination of a glide versus a steady-state) showed that the difference limens of

initial and final transitions with a gradual on-/offset are comparable to the ones determined in the present study and that the perceptual asymmetry between initial and final transitions is equally large (personal communication, Egbert de Boer).

In conclusion, just noticeable differences in end point frequency are significantly smaller for final than for initial transitions, as was also clear from Figure 4. Compared to thresholds of isolated transitions (Cullen *et al.*, 1992) more backward masking occurs with falling transitions followed by a steady-state. Moreover, initial transitions incrementing in rate of frequency change yield smaller thresholds than those decrementing in rate of frequency change in the presence of a steady-state (in agreement with Porter *et al.*, 1991): masking affects discrimination increasingly as the end points of the transition approach the frequencies of the steady-state. Discrimination is better for initial rising than for initial falling transitions, suggesting that later occurring lower frequencies mask the earlier high frequencies for the falling glides.

I F: discussion

Although the difference limens in end point frequency are larger for complex formant glides than for single formant transitions, and larger for formant transitions than for tone sweeps, discrimination functions of tone and formant glides are comparable in their general behaviour. Difference limens decrease with increasing transition duration, and are larger in initial than in final position. Sensitivity decreases with increasing stimulus complexity as a result of masking of surrounding formants. The tone glides show much sharper perceptual asymmetries than the formant glides with respect to direction, position, and rate-of frequency change of the transition.

The perceptual asymmetry between initial and final transitions for both isolated transitions and transitions preceded or followed by a steady-state, suggests that discrimination is, to a large extent, based on the difference between the end point frequencies of the standard and comparison stimuli. Transitions that converge to a common frequency in a steady-state are discriminated more poorly due to temporal backward masking affecting discrimination of the preceding end point frequencies (probably a memory recency effect). Backward masking affects falling transitions more than rising transitions due to the subsequent lower frequencies sweeping across the basal regions of the cochlea. Low frequency glide components increasingly affect discrimination when the difference between the end point of the transition and the steady-state decreases, i.e., when transitions decrement in transition rate. In final position, discrimination is not affected by the masking phenomena, due to the end points occurring last.

If discrimination were based on rate of frequency change rather than on extent of frequency change, the difference between increments and decrements would be equally large in initial or final position. Our data argues against rate of frequency change as a cue in the processing of these transitions or the critical band range affecting discrimination, as thresholds in final position were (nearly) identical for transitions incrementing and decrementing in transition rate.

The difference between the thresholds in initial and final positions and between those incrementing and decrementing in transition rate indicates that discrimination is based on the difference in end point frequency between the standard and comparison transitions and on the frequency difference between end point and steady-state. Due to transitions diverging to different frequencies in final position, discrimination is not affected by

masking phenomena of the preceding steady-state, although thresholds are somewhat poorer for transitions decrementing in transition rate than those incrementing in transition rate in final position. The relative magnitude of the thresholds in final position, given that they are the same for higher and lower rates of frequency change, indicate that discrimination is based on differences in end point frequency, rather than on some rate cue.

The *similarity* of thresholds of the transitions incrementing and decrementing in transition rate in *final* position argues against rate of frequency change as a cue in processing short dynamic stimuli. Although thresholds of *isolated* transitions in initial and final position were similar with respect to higher and lower rates of frequency change, the difference in performance with regard to the *position* of the transition also indicated that discrimination of 40 ms transitions is based on differences in end point frequency, rather than on a rate cue (Cullen *et al.*, 1992). As a result of the difference in *frequency region* (/u/ vs. /a/), thresholds of *isolated* rising transitions are smaller than those of falling ones in initial position, whereas they are larger than those of falling transitions in final position (Cullen *et al.*, 1992).

In conclusion, the presence of a steady-state affects discrimination of a preceding transition, especially when the difference between the end point frequency and the steady-state decreases. In addition to a higher memory demand in initial than in final position (to remember the discriminative end points while information is added to memory), the additional steady-state physically masks the end points. In final position the vibrations gradually die away, while they are continuously followed by new waves in initial position (temporal overlap of cochlear responses in masking). In the following section we will examine excitation patterns of several kinds of tone sweeps to determine whether the perceptual asymmetry partly originates at a peripheral level. Excitation patterns are determined of <u>tone sweeps,</u> rather than of formant glides, as the former yield the clearest perceptual differences.

Section II. Excitation patterns of tone glides

Introduction

In this section we use a model of the inner ear to compare excitation patterns of initial and final tone glides. Several attempts have been made to construct models of the auditory periphery (e.g., Seneff, 1986; Jenison *et al.*, 1993;), most of which emphasize the modelling of frequency selectivity (masking in the frequency domain instead of in the time domain). A prerequisite of our auditory model is asymmetry in time, rather than asymmetry in frequency, as we want to gain insight into the relatively large perceptual difference between initial and final glides, rather than between the difference between rising and falling ones. The *gammatone*, which is used to filter the input stimulus, meets this requirement (see Section II A).

Roughly, our model consists of 1) derivation of the excitation patterns from a series of linear filters, modelled after De Boer (1975, 1978, 1979, 1990) and 2) calculation of the firing probability of action potentials of the auditory nerve. Calculation of the neural firing probability is adapted from Meddis (1986a) and Meddis and Hewitt (1988, 1990). We first discuss the two stages of this cochlear model in short (A), and then compare the neural responses for different types of tone sweeps (B).

II A: Cochlear modelling

To examine the perceptual asymmetry peripherally, tone glides are tested to a model yielding the pattern of activity of a set of auditory nerve fibres. These auditory nerve fibres are represented by a set of neural units, obtained by convolution of the input with a series of gammatones (adapted from De Boer, 1975):

$$h_i(t) = \left[(t-\alpha_i)/\beta_i\right]^{\gamma-1} \cdot e^{-[(t-\alpha_i)/\beta_i]} \cdot \cos(\omega_i t + \psi_i)$$

In general, the gammatones prove to be a good approximation of the peripheral coding by the basilar membrane. The bandwidth of the filters varies as a function of β, the decay constant, which is made to vary as the -0.663 power of the resonance frequency. The degree of the filter, γ, is 4. The latency, the intrinsic delay time α, depends on the resonance frequency as well as on the amplitude of the stimulus. Figure 5 shows the impulse response of the basilar membrane (BM), modelled as a series of 26 linear filters (details of the computation are found in De Boer, 1975, p. 1034).
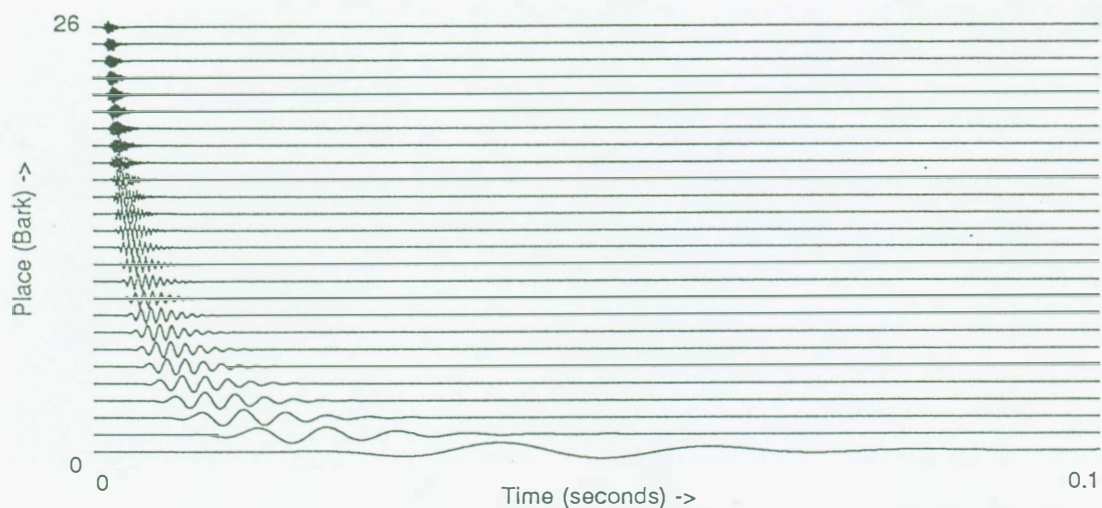

Figure 5. The impulse response of 26 filters (1 Bark steps) as a function of time. The series of gammatones corresponds to the length of the basilar membrane (between 0 and 12 kHz) with low-frequency filters showing longer latencies than high-frequency ones.

Two important properties of the neural units, i.e., frequency selectivity (or firing probability) and latency, were derived from the reverse correlation technique by De Boer (1975, 1979). Figure 5 shows how the latency varies as a function of the resonance frequency: the lower the frequency the longer the latency. The asymptotic decay of excitation shows that forward masking is greater (longer decay time) than backward masking (short rise time).

The second stage of the cochlear model consists of determining the firing probabilities for the different neural units. Meddis (1986a) and Meddis and Hewitt (1988, 1990) have published the implementation of a model which simulates a number of important properties of auditory nerve activity. Briefly, their model assumes that the hair-cell/auditory-nerve junction consists of three reservoirs of neurotransmitter

substance and that the rate in which the substance is released from the hair cell into the presynaptic cleft depends directly on the instantaneous displacement of the neighbouring cochlear partition. The model shows rapid as well as short-term adaptation and phase-locking for the low-frequency stimuli (see Meddis and Hewitt, 1990, p.1813-1814 for parameters and implementation of the model).

II B: Neural responses of initial and final tone glides

Neural responses of several 40 ms tone glides, preceded (VC) and followed (CV) by a 200 ms stationary part were examined. In order to compare excitation patterns of the trailing part, all stimuli were followed by 100 ms of silence. The initial tone glides started on the zero-crossing to reduce click effects at the onset (they do not build up gradually). Figure 6 shows the pattern of the presynaptic vibrations after gammatone filtering of the CV (6a) and VC (6b) stimuli.
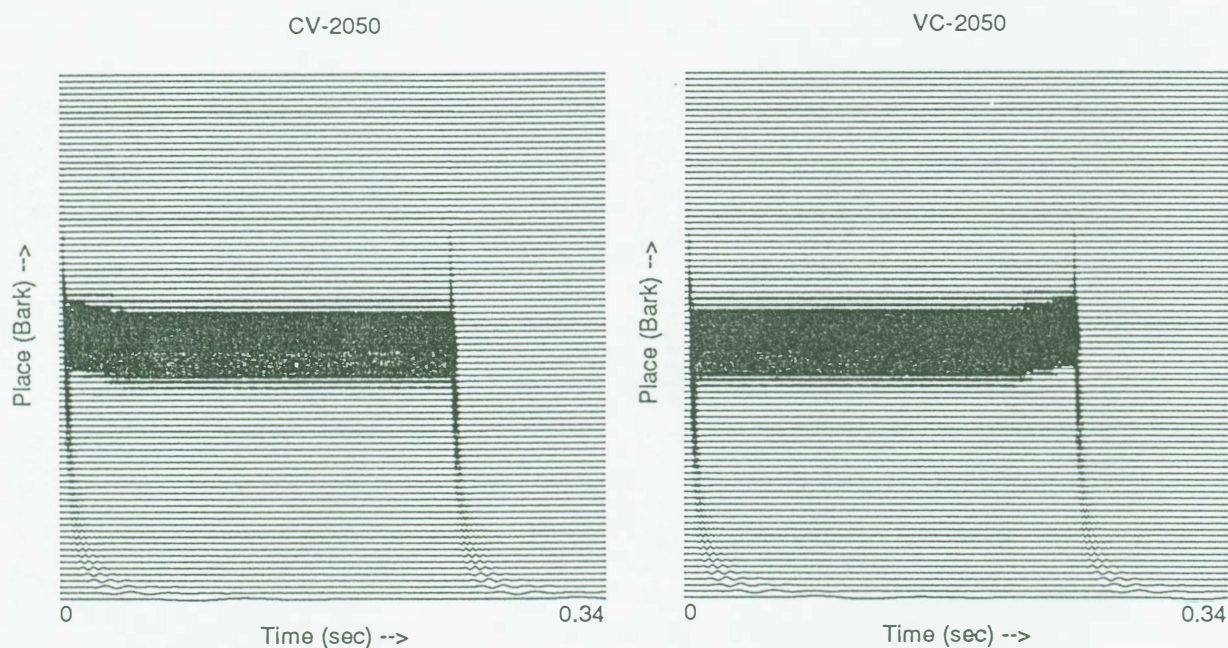

Figure 6. Patterns of vibration of the initial and final tone glides filtered by a series of gammatones (the first stage of the model). The response patterns of the CV and VC transitions are plotted between 0 and 26 Bark in steps of 0.3 Bark (place on the BM) as a function of time. Transients at the beginning at the end of the stimuli are visible for both CV and VC transitions.

The figures illustrate the response patterns of the initial (CV) and final (VC) transitions filtered by 85 gammatones (Barks) against time (s). The CV transition glides from 2050 Hz to 1850 Hz in the first 40 ms (labeled CV-2050), while it rises from 1850 Hz to 2050 Hz for the VC stimulus (labeled VC-2050). The stimuli are sampled at 48 kHz and filtered in 0.3 Bark steps. The plots show that the successive filters are capable of processing the relatively fast transients. Although maximum excitation occurs over a limited part of the cochlea (frequency selectivity of the neural units) the low and high frequency regions show some excitation due to onset and offset transients. For initial as well as for (the mirror image) final sweeps maximum excitation occurs in the 42nd neural unit (12.45 Bark or 1870 Hz).

The auditory-neural transduction is simulated for the above-mentioned stimuli as well as for two other stimuli which have a 30 Hz lower end point frequency. That CV

transition glides from 2020 Hz tot 1850 Hz (the steady-state frequency) and the VC transition glides from the 1850 Hz stationary frequency to 2020 Hz (CV-2020 and VC-2020, respectively). The psycho-acoustic experiments have shown that the initial transitions are not discriminable, whereas the final transitions are (see Table III). The responses, i.e., the firing rates, show the neural adaptation effects at the beginning and the end of the signals.

Figures 7-10 compare the firing frequency of one of the nerve fibres, i.e., the 42nd, of the two CV and two VC stimuli. Only a limited time span of 65 ms is shown in each plot, namely the part containing the transition. The peak at the beginning of the neural response, here only visible in the CV stimuli, illustrates adaptation; it is followed by a rapid decrease in firing rate, which reaches a plateau during the larger, stationary, part of the stimulus. Latency precedes the peak at the beginning of all the signals. The two VC plots show the transient from stimulus to silence (the stimulus ends at 0.24 s).
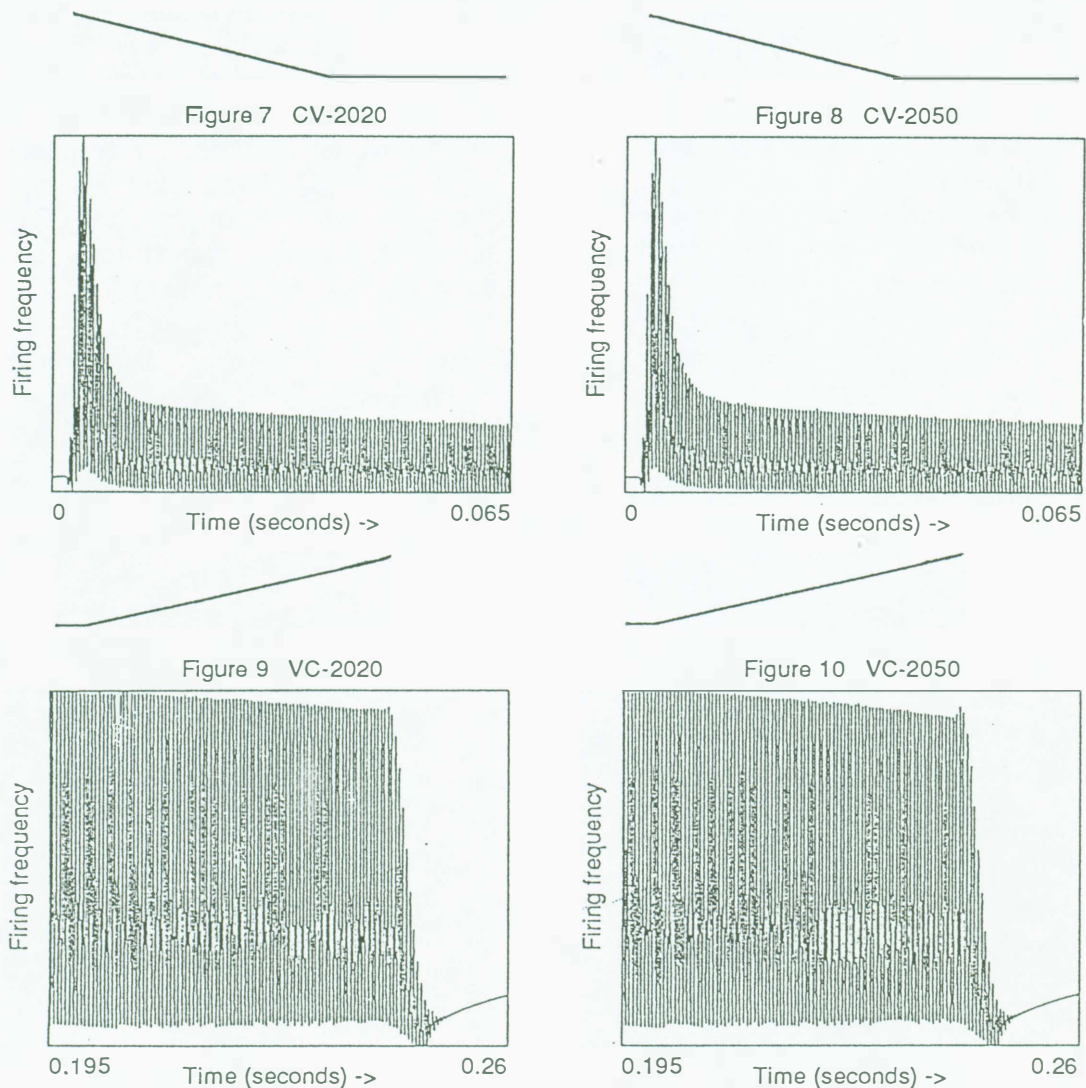
Figures 7-10. The firing frequency of one auditory nerve fibre as a function of limited time span (65 ms at the beginning and at the end of the CV- and VC-stimuli, respectively). Firing frequency is simulated for four different stimuli: the non-discriminable initial (CV) and the discriminable final (VC) ones. CV 2050 and VC 2050 are the standard stimuli. A schematic representation of the (length of the) corresponding tones is included at the top of the figures. Compared to figures 7 and 8, a smaller part of the steady-state is shown in figures 9 and 10 to illustrate the transient from stimulus to silence of the final transitions.

The model incorporates both the rapid adaptation which may last 2-5 ms, and (subsequent) short-term adaptation, which may last 40-50 ms (the nerve fibres are modelled as independent of each other). Psycho-acoustical experiments have shown that discrimination is based on differences in end point frequency, rather than on differences in transition rate (Porter *et al.*, 1991; van Wieringen and Pols, submitted). In modelling the auditory-neural responses the discriminable information of the CV lies in the first few milliseconds, during neural adaptation, while that of the VC transition appears at the end of the stimulus, after neural adaptation. Analysis of individual nerve fibres does not yield clear insight into why two initial glides differing in 30 Hz are not discriminable (CV-2020 vs. CV-2050) whereas comparable final glides are. This is not surprising, as many fibres contribute to the perception of a stimulus. However, the present analyses have shown that it may be possible that the initial transitions, contrary to the final ones, are driven into saturation and that this affects discrimination in end point frequency.

Instead of examining individual fibres, which all show similar neural adaptation effects, the resonance frequency of one instant in time can be determined for all the fibres. The stimuli are now sampled with a step size of 0.01 ms and filtered in 1 Bark steps. Figure 11 shows the firing frequency of all fibres, log-scaled, as a function of the neural units (Bark) at two instants in time: the solid line illustrates the response at exactly halfway the transition, i.e., at 0.02 s for the CV and 0.22 s for the VC stimuli. The dashed line shows the response at 0.001 s from the beginning of the initial (CV) or at the end of the final (VC) transitions, i.e., at 0.239 s. The initial transitions show a low frequency response due to the transient at the beginning of the stimuli.
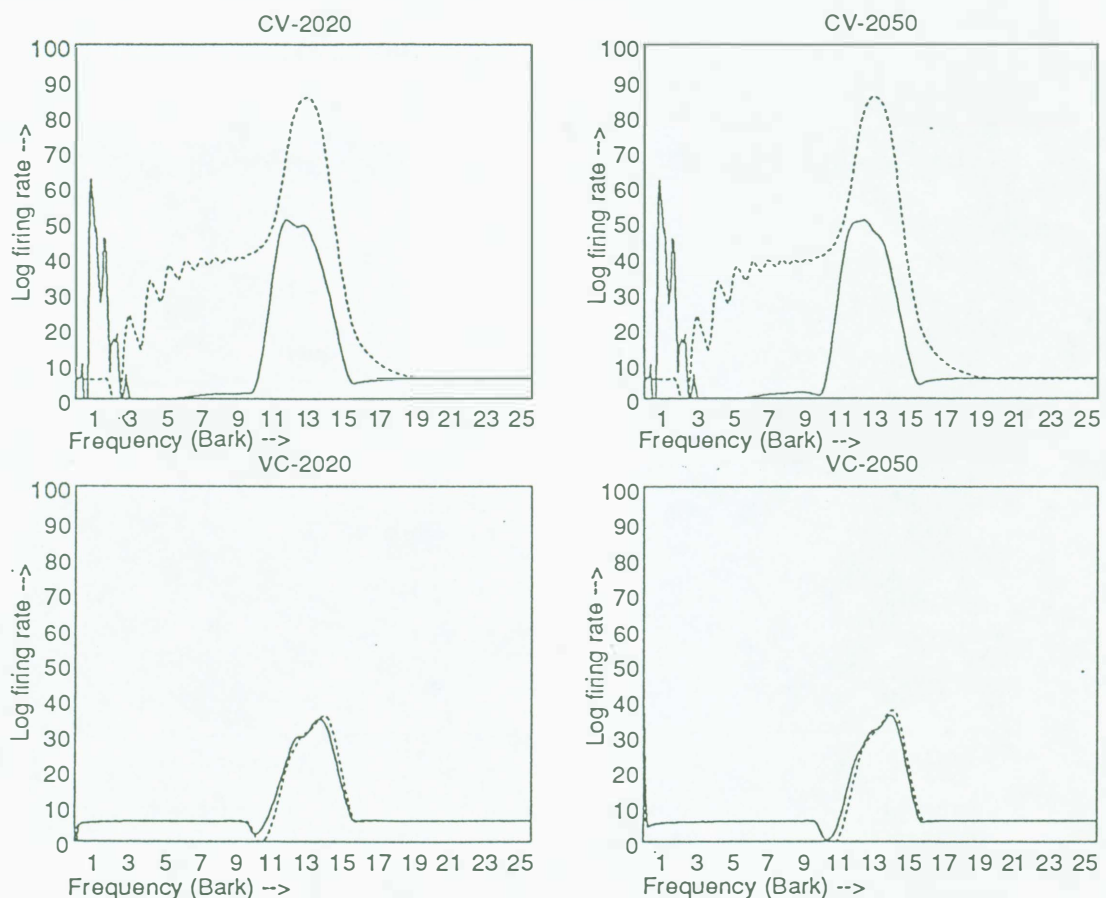

Figure 11. Firing frequency of all the nerve fibres as a function of frequency (Bark) at two instants of time, i.e., at 0.001 s and 0.02 s for the two initial stimuli, and at 0.22 s and 0.239 s for the two final stimuli.

Although the resonance frequency changes in the course of the glide, the plots do not clearly show the width of the resonance frequency to change: their width remains approximately 2 Bark (they are somewhat narrower at the beginning than at the end, for CV transitions). The plot illustrates the response patterns for the initial (CV) transitions. Those of the final (VC) transitions are similar with respect to the width of the resonance frequencies. However, they are not preceded by low frequency transients.

Although psycho-acoustical experiments show that the two VC stimuli are discriminable, the present analysis shows no evident underlying peripheral mechanism. Figure 11 once again shows that the resolution is too small with respect to the small psycho-acoustical differences determined for initial and final glides. We suggest that discrimination in end point frequency of the initial transition is negatively affected by the relatively high discharge rate at the beginning of the stimulus, while discrimination in end point frequency of final transitions is not affected by saturation, as adaptation has already taken place. However, this cannot be ascertained by the present model.

If neural adaptation proves to be one of the effects influencing discrimination in end point frequency, the perceptual asymmetry between initial and final transitions should decrease, i.e., discrimination of the initial transition should improve, if the initial transition is preceded by a noise burst. As all fibres are then stimulated by the broadband noise, the initial transition is processed after neural adaptation, and discrimination of the CV transition should be comparable to that of the VC transition. A pilot perception experiment showed that discrimination in end point frequency did not improve for the initial transition preceded by a noise burst. Rather, performance decreased, as a result of forward masking of the noise burst on the transition! Possibly, discrimination between two transitions varying in end point frequency is affected by a memory recency effect, especially if the transitions are followed by a steady-state. However, a memory recency effect does not explain why a transition preceding a steady-state sounds so much like a stationary sound. Discrimination in end point frequency of initial transitions followed by a steady-state is likely to be affected by several factors, including transients and a decay in short term memory.

Discussion neural coding of tone glides

The present study shows that the perceptual asymmetry found in the psycho-acoustic experiments is not clearly reflected in the responses of the auditory nerve fibres (as modelled so far). Ideally, the properties cueing discrimination of initial and final transition, should be observed on a peripheral level. However, such a conception is difficult to pursue when the discriminable portions of the transitions, i.e., the end points, are exposed to neural adaptation. Psycho-acoustical experiments have shown that discrimination is based on the end points or the difference between the initial and final frequencies of the transition, rather than on transition rate (e.g., Porter *et al.*, 1991; van Wieringen and Pols, 1992, submitted; van Wieringen *et al.*, 1993). As the frequency resolution decreases when the fibre is driven into saturation, shown by a deterioration of the mean firing rate (Evans, 1975a), discrimination in end point frequency is significantly more difficult for initial transitions: the fibres are firing at maximum, and a change in firing rate does not occur over a wide frequency range. Adaptation is represented by a linear filtering process applied to the neural activity function (the weighted sum of all firing probabilities). Also, the rise time of the filters of the initial CV tones, which are continuously changing in frequency is shorter than the decay time of the final VC sweeps. This may also affect discrimination in end point frequency of CV-tones (negatively, in terms of temporal overlap of cochlear responses) or VC-tones (positively, due to a longer decay time), as the rise and decay times include

the discriminable parts of the stimuli. Psychoacoustically, the change in pitch of the final tone sweep is perceived very clearly, whereas the timbre of the transient seems to be the dominant cue in the CV-transition. Physically, the gammatones prove to be a good approximation of the peripheral coding by the basilar membrane; the impulse response of the peripheral filters is provided well by the wave forms of the gammatones. As these are able to trace the relatively fast sweeps they therefore serve as a valid input for the simulation of neural coding.

It must be kept in mind that the cochlear model in the present study is a relatively simple approximation of the coding by the auditory system. The neural units are modelled as independent from each other, each with their own latency and frequency selectivity properties (the description of the pattern of excitation is based on a tone burst as stimulus (De Boer, 1975)). Also, nonlinearities are not included and the model holds specific assumptions about the way in which firing probability depends on the wave form of excitation. A more realistic model would probably incorporate coupling of the essential properties of the hair cells. However, the present model is sophisticated enough to explain the most prominent properties of neural coding, including the properties cueing discrimination in end point frequency.

General discussion

In conclusion, the presence of a steady-state affects discrimination of a preceding transition, especially when the difference between the end point frequency and the steady-state decreases. Several studies suggest that several sources contribute to the perceptual asymmetry. From a psychoacoustical point of view, backward masking occurs as a result of a memory recency effect (Massaro, 1975; Moore, 1989): it is more difficult to remember two end points when the transitions are followed by a steady-state, than when they occur at the end of the stimulus. Therefore, a decrease in sensitivity as a result of the presence of a following steady-state may, for the larger part, be explained by a memory effect (Collins, 1984). Several psycho-acoustical experiments, among which those of Brady *et al.* (1961), indicate that the terminal frequency is more important than the initial frequency and therefore also questions the equal importance of the locus in CV and VC formant transitions. Although several speech experiments, among which those of Sharf and Hemeyer (1972) have found the VC transition to be more consonant specific than the CV one, others favour the perceptual prominence of the CV transition (e.g., Kashino, 1990; Kashino *et al.*, 1992), arguing that the relative importance of the CV and VC transitions depends on many factors, such as context and language background. A third set of studies (e.g., Pols and Schouten, 1978) finds the CV and VC transitions to be perceptually equally prominent and also concludes that the perceptual asymmetry between initial and final tone and formant glides is not reflected in the speech transitions, possibly as a result of many available redundant cues (van Wieringen and Pols, in preparation).

It is suggested that the perceptual asymmetry between CV and VC transitions is caused by temporal backward masking of the stationary part on the preceding transition rather than by forward masking of the final transition. Discrimination of isolated, diverging glides and glides preceded by a steady-state is more comparable than that of isolated, converging glides and those which are followed by a steady-state. The VC speech transition is therefore not expected to be more consonant specific than the CV one, in an experimental design. Rather, the CV transition may become less consonant specific as a result of the following steady-state. However, as complex formant stimuli yield a less clear perceptual asymmetry than single formant transitions or tone glides, natural speech stimuli may not yield an initial-final effect due to stimulus complexity. In

addition to a memory recency effect, which is a crucial factor in every 'uncertainty' paradigm, such as a discrimination task, the discriminable portion may well be affected on a peripheral level by the extent of neural adaptation. The data cannot solely be explained by a (cognitive) memory effect: in that case practice would diminish the perceptual asymmetry. Not only were the psycho-acoustic listeners highly trained, they also continued to hear the final transition more as a glide than the initial one. The difference in sensitivity with respect to the direction of the transition may be explained by high frequency components masking lower ones, rather than the reverse: falling transitions generally yield lower difference limens as parts of the cochlea are stimulated in the course of the transition which have not been stimulated before. Contributions to the perceptual asymmetry may come from different levels of the auditory system and it is still a matter of debate of how small variations are encoded, such that they are 'eventually' perceived differently. Further research is necessary, in the interest of speech recognition systems, and in the interest of our knowledge on peripheral coding, such as neural spike activity.

Acknowledgement

References

Brady, P.T., House, A.S. & Stevens, K.N. (1961): "Perception of sounds characterized by a rapidly changing resonant frequency", *J. Acoust. Soc. Am.* **33**: 1357-1362.

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

Collins, M.J., Cullen, J.K., Jackson, D.F. & Porter, R.J. (1990): "Discrimination of formant-like frequency transitions preceded or followed by steady-states", *Abstracts of the Thirteenth Midwinter Meeting of the Association for Research in Otolaryngology*: 181-182.

Cullen, J.K., jr., Houtsma, A.J.M. & 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.

de Boer, E. (1975): "Synthetic whole-nerve action potentials for the cat", *J. Acoust. Soc. Am.* **58**: 1030-1045.

de Boer, E. & H.R. de Jongh (1978): "On cochlear encoding: Potentialities and limitation of the reverse-correlation technique", *J. Acoust. Soc. Am.* **63**: 115-135.

de Boer, E. (1979): "Travelling waves and cochlear resonance", in *Models of the Auditory System and Related Signal Processing Techniques* (eds. M. Hoke & E. de Boer). *Scand. Audiol. Suppl.* **9**: 17-33.

de Boer, E. (1990): "On ringing limits of the auditory periphery", *Biol. Cybern.* **63**: 433-442.

Delattre, P.C., Liberman, A.M. & Cooper, F.S. (1955): "Acoustic loci and transitional cues for consonants", *J. Acoust. Soc. Am.* **27**: 769-773.

Egan, J.P. & Clarke, F.R. (1966): "Psychophysics and signal detection", In J.B. Sidowski (Ed.) *Experimental methods and instrumentation in psychology,* New York: McGraw-Hill: 211-246.

Elliott, L.L., Hammer, M.A. & Carrell, Th. (1991): "Discrimination of second-formant-like frequency transitions", *Percept. Psychophys.* **50**: 1-6.

Evans, E.F. (1975a): "Cochlear nerve and cochlear nucleus", In *Handbook of sensory physiology* (eds. W.D. Keidel & W.D. Neff), Springer, Berlin, **5/2**: 1-108.

Jenison, R.L., Greenberg, S., Kluender, K. R. & Rhode, W.S. (1991). "A composite model of the auditory periphery for the processing of speech based on the filter response functions of single auditory-nerve fibers", *J. Acoust. Soc. Am.* **90**: 773-786.

Kashino, M. (1990): "Distribution of perceptual cues for Japanese intervocalic stop consonants", *Proceedings ICSLP 90*, Kobe: 557-560.

Kashino, M., Wieringen, A. van & Pols, L.C.W. (1992): "Cross-language differences in the identification of intervocalic stop consonants by Japanese and Dutch listeners", *Proceedings ICSLP 92*, Banff, **2**: 1079-1082.

Kirk, R.E. (1982): *Experimental design: procedures for the behavioral sciences* (Brooks/Cole).

Macmillan, N.A. & Creelman, C.D. (1991): *Detection theory: a user's guide* (Cambridge University Press).

Massaro, D.W (1975): "Preperceptual images, processing time, and perceptual units in speech perception", In: *Understanding language. An information-processing analysis of speeech perception, reading, and psycholinguistics.* (D.W. Massaro, ed.) New York: Academic Press: 125-150.

Meddis, R. (1986a): "Simulation of mechanical to neural transduction in the auditory receptor", J. *Acoust. Soc. Am.* **79**: 702-711.

Meddis, R. (1988): "Simulation of auditory-neural transduction. Further studies", *J. Acoust. Soc. Am.* **83**: 1056-1063.

Meddis, R. & Hewitt, M.J. (1990): "Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse", *J. Acoust. Soc. Am.* **87**: 1813-1816.

Moore, B.C.J. (1989): *An introduction to the psychology of hearing*, 3rd edition, Academic Press, 350 pages.

Pols, L.C.W. & Schouten, M.E.H. (1978): "Identification of deleted consonants", *J. Acoust. Soc. Am.* **64**: 1333-1337.

Porter, R.J., Cullen, J.K., Collins, M.J. & 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.

Seneff, S. (1988). "A joint synchrony/mean rate model of auditory speech processing", *J. Phon.* **16**: 55-76.

Sharf, D.J. & Hemeyer, Th. (1972): "Identification of place of consonant articulation from vowel formant transitions", *J. Acoust. Soc. Am.* **51**: 652-658.

Taylor, M.M. & Creelman, C.D. (1967): "PEST: efficient estimates on probability functions", *J. Acoust. Soc. Am.* **41**: 782-787.

Weenink, D.J.M. (1988): "Klinkers: een computerprogramma voor het genereren van klinkerachtige stimuli", *IFA-report nr. 100*.

Wieringen, A. van & Pols, L.C.W. (1992): "Discrimination of short speech-like formant transitions", *J. Acoust. Soc. Am.* **92**, **4** (2): 2298 (A).

Wieringen, A. van, Cullen, J.K. & Pols, L.C.W. (1993): "The perceptual relevance of CV- and VC transitions in identifying stop-consonants: cross-language results", *Proceedings Eurospeech'93*, Berlin, **2**: 1499-1502.

Wieringen, A. van & Pols, L.C.W. (1994): "Frequency and duration discrimination of short first formant frequency transitions", *submitted to J. Acoust. Soc. Am.*

Wieringen, A. van & Pols, L.C.W. (submitted): "Discrimination of single and complex CV- and VC-like formant transitions", *submitted to J. Acoust. Soc. Am.*

Wieringen, A. van & Pols, L.C.W. (in preparation): "Plosive identification of original and time-reversed initial and final Dutch vocalic transitions: cross-language results.