TRANSITION RATE-DEPENDENT PROCESSING OF ONE-FORMANT SPEECH-LIKE STIMULI

Astrid van Wieringen and Louis C.W. Pols

ABSTRACT

Three listening experiments with carefully manipulated synthetic stimuli have been performed to examine the processing of dynamic information.
By means of comparing all possible combinations of triads of nine one-formant CVC-like stimuli, differing in target frequency and transition duration, an attempt is made at understanding the physical parameters cueing perceptual equivalences among these specific stimuli. From the first two experiments, the only difference in the stimuli being the direction of the transitions, it is concluded that two spectrally-complex sounds may sound alike -despite different target frequencies-, provided that the stimulus having the highest (or in exp. 2 lowest) target frequency is compensated by a longer/slower trajectory.
A third experiment showed that perceptual alikeness of two out of three spectrally-complex synthetic stimuli is for a major part dependent on transition rate. The results of the experiments are discussed in terms of psycho-acoustic findings and -albeit the peripheral nature of the stimuli- also in terms of speech perception.

INTRODUCTION

Production

The relative importance of stationary and dynamic signal parameters must be studied in detail to understand how listeners make use of the continuously varying information in the speech signal. Due to the relatively slow articulatory changes, as well as to the reorganisation of sounds during speaking, the physical parameters of the speech signal (i.e. frequency, time, and amplitude) are subject to much variation. Apart from this, context effects such as speaking rate, speaking style, and stress, may cause a further blending of speech sounds. The unsegmentable speech wave and the decreasing distinctiveness of the speech sounds have been topic of much discussion during decades (Lindblom, 1963; Lindblom and Studdert-Kennedy, 1967; Lindblom and Moon, 1988). Lindblom has argued that the more speaking rate increases, the less distinctive speech sounds become, due to the increasing coarticulation and increasing reduction of the quasi-stable vowel-like portions. Physically speaking this holds that formant frequencies, instead of reaching their hypothetical target values, reduce spectrally to a more neutral vowel. Moreover, the dynamic trajectories moving to and from the quasi-stationary portions may undershoot the target frequencies.
That an increase in speaking rate not necessarily results in more coarticulation or vowel reduction was shown by Van Son and Pols (1990). They measured the effect of speaking rate for one speaker and found that even at a relatively fast speaking rate the ideal targets (or nuclei) could be measured well. The effect of speaking rate on vowel reduction may be speaker dependent and may well be absent when pronouncing clearly.

In measuring the dynamic parts of the speech signal a small effect of shorter duration on formant trajectories was found for the vowels with the higher target frequencies (Van Son & Pols, submitted). By means of a 15 point to point comparison, as well as global measures of the formant track shape, based on a Legendre polynomial, the transitions of these vowels were shown to level-off somewhat.

Perception

In spite of the ongoing discussion on the dependency or correlation of coarticulation on speaking rate many speech sounds are found to interact under different conditions to such an extent that the speech signal is more dynamic than stationary. And, in spite of the mutual interaction of these speech sounds the perceptual mechanism is able to resolve the context-conditioned variation into a meaningful percept. Obviously, many contextual factors contribute to the identity of a sound, and various research is focused on understanding integration of cues at different levels of perception.
The undershoot issue has been dealt with perceptually by among others Lindblom and Studdert-Kennedy (1967), Fujisaki and Sekimoto (1975), Pols et al. (1984), and House et al. (1989) to examine whether listeners compensate perceptually for the spectral reduction and/or variation in the speech signal.
By asking listeners to label /u/-like and /i/-like stimuli in two different symmetrical synthesized contexts Lindblom et al. found that listeners indeed compensate for undershoot -at the faster rate- as the boundaries shifted in the direction of the context loci.
The results were somewhat different for the two contexts, which differed not only in transition rate, but also in transition direction, suggesting the importance of the dynamic information, not only that of stationary information to recognize vowels.
The same was acknowledged by Fujisaki et al. (1975), who in a discrimination and identification tests, in which a truncated formant transition was matched with a steady-state, found that complete transitions between locus and target frequency were not even necessary for the listener to recognize vowels: incomplete formant transitions tended to extrapolate to the target of the steady-state.
In a real-time matching experiment, however, with different types of stimuli Pols et al. (1984) did not find any clear indication of perceptual extrapolation or overshoot. Neither did House and Neuburg (1983) in matching tests with sweeps of different durations, complexity and formant patterns. Presumably, a major part of the results of all perception experiments dealing with highly manipulated or synthetic stimuli is stimulus and task dependent; therefore, interpretations should be considered carefully.
Although not clear to which extent, dynamic information does contribute to the perception of the more stable portions -the vowels- of the speech signal. The perceptual mechanism is flexible, in that the information delivered by many perceptual attributes, not that by one alone, brings about a certain percept. This notion of equivalence is basic in many ways. Consider, for example, speech perception: higher-level knowledge is necessary to discriminate between the words *two* or *too*. On a much more fundamental level, e.g. in constructing psychophysical scales, the measurements are often based on thresholds, on the capability of the mechanics of hearing. Between these two extremes different types of phonetic equivalences are measured, by trading spectral and temporal, or temporal and temporal parameters, whatever is appropriate for speech. For examples, Repp(1982), Best et al. (1981), Fitch et al.(1980) traded speech cues and found that the trade-offs between the correlates were mostly of a phonetic kind, even when using non-speech stimuli. That the phonetic contrasts are phonetically based was to be expected, as the trading mostly ocurred with phonetically based acoustic distinctions such as for examples, silence duration versus F1 onset frequency or

duration of silence versus duration of fricative noise. Due to the nature of the stimuli and the kind of task, i.e. identification, there is much evidence for the phonetic origin of the equivalences, and not so much for auditory, psycho-acoustical factors. Discrimination, identification, and matching tasks are well suited to measure integration of cues accurately.
In the tests discussed in this paper we do not deal with the kind of trading relations or phonetic equivalences mentioned above. We address perceptual alikeness on an auditory level by using a different task and psycho-acoustical stimuli. On a more fundamental level, presumably void of higher level knowledge, we have tested the integration of different physical parameters in terms of perceptual alikeness. A possible mechanism compensating for any kind of articulatory undershoot may not only be present at a phonetic, but also at an auditory level. Such is tested by means of a triadic test, in which listeners are asked to judge the most and least similar pairs of stimuli. Perceptual equivalences in terms of identity are not measured, but insight can be gained into the cues underlying identification.

Although the variables tested in the following experiments may be overruled in speech perception by other parameters, or by a different context or more complex stimuli, we will discuss how two relatively simple speech-like sounds may sound alike, despite different physical parameters.

EXPERIMENT 1

The purpose of performing the triadic experiment described below is to determine if and to what extent there exists a trade-off between transition duration and target frequency. If the identity of a vowel is not only determined by the formant frequency at the target, but also by the rate and direction of the formant transitions, careful trading of the spectral and temporal cues should not change the perception of the stimulus. Testing which cues are responsible for the identity of a vowel or phoneme need not necessarily be measured by an identification test, as in such a test much of the information leading to a certain identity is lost in the course of processing. Instead of measuring boundary shifts a perceptual compensation process underlying e.g. articulatory undershoot may be measured by asking listeners which pair in a triad is most and which pair is least similar. In the case of overshoot a faster transition of one stimulus would be inclined to extend to a higher target frequency of another stimulus, and still found to belong together in the particular triad. To measure whether listeners already compensate for context-conditioned variability visible in the speech signal at a psycho-acoustical, peripheral level, very simple, one-formant stimuli were used.
By manipulating the transition duration -and therefore the transition rate-, together with the target frequency in a triadic experiment the trade-off between the dynamic and stationary parameters was tested.

Hypothesis

We hypothesized that the compensation for variation would be such that -in order for two sounds to be perceptually alike, a lower -undershooted- target frequency of one stimulus would be compensated by a faster transition of another one.

Stimuli (Convex)

Nine synthetic one-formant stimuli were generated by means of a digital formant synthesizer designed by Weenink (1988). With a pulse as source the fundamental frequency was kept constant at 110 Hz. The nine stimuli were constructed by varying three transition durations with three target frequencies. The different transition durations (10 ms, 20 ms, and 30 ms) rose linearly from 220 Hz to three different target frequencies (550 Hz, 625 Hz, and 700 Hz), the bandwidth always being 10% of the formant frequency. The target frequency remained constant for 15 ms, followed by a falling transition of identical duration as its rising one. No other parameters such as voice bar were added to make the stimulus more speech-like. Together there were nine different symmetrical CVC-like stimuli. Figure 1 illustrates the different stimuli schematically. Note also that the total durations of the stimuli vary from 35 to 75 msec.
The stimuli were sampled at 10 Hz via a 12-bit D/A converter; a 5-msec hamming window at the beginning and ending of the stimulus was used to avoid clicks.

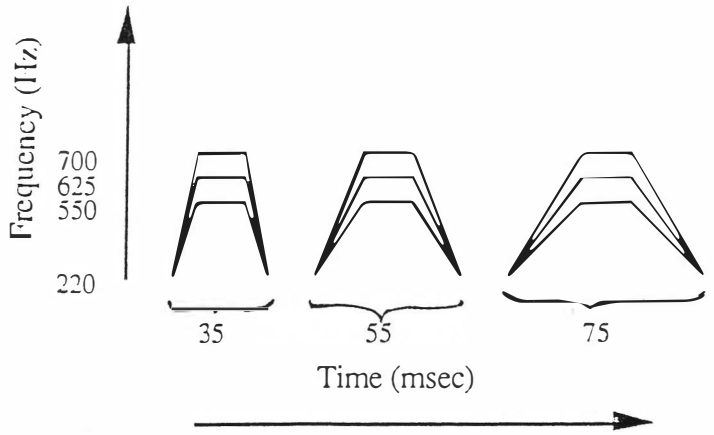

Figure 1. Illustration of the stimuli under test.

Figure 2 illustrates the hypothesized relationship between the stimuli, i.e. that a stimulus with a faster transition will extrapolate to a higher target frequency (and still sound alike). In other words, the groups of stimuli marked by arrows would be judged to be more alike than those placed diagonally in the reversed direction.

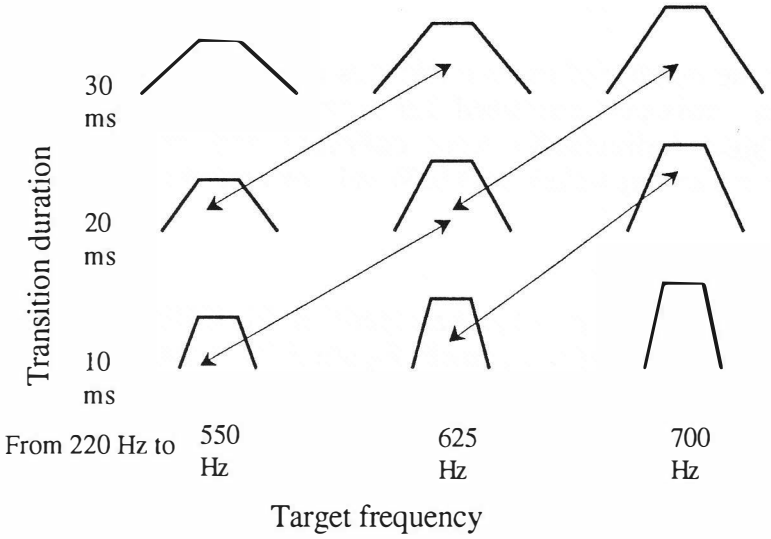

Figure 2. Illustration of the relationship among one-formant CVC-like stimuli. It is hypothesized that the stimuli joined by arrows will sound more alike than those placed diagonally in the opposite direction.

Procedure

The test was done by means of a real-time triadic comparison test.
The subject was seated in front of a terminal of the μVAX II. By pressing three keys, one for each stimulus of the triad, the subject was allowed to listen to the stimuli -over headphones- as often as he wished. After having decided on the most and least similar pairs the next triad of stimuli was offered. In order to have a balanced design in a triadic experiment and to prevent the test from being too long the set of stimuli is usually very limited. As we used nine stimuli a total of 84 triads were offered. There was a new randomisation of stimulus order for every subject. The test consisted of two presentations of one randomisation sequence, resulting in a total of 168 triads (no break between sequences).

Subjects

Fourteen subjects, -eight of whom were totally naive as to the kind of stimuli and the task- participated in the test. They were paid for listening. The other six subjects were phonetically trained listeners. None of the subjects were informed on the differences between the stimuli, only on the task. Although differences between the stimuli were difficult to hear at the beginning of the test, none of the subjects mentioned any problems. The test lasted on average five quarters of an hour (individually paced).

Results

By counting the number of times a stimulus is found to be most or least similar a (dis)similarity matrix can be obtained. The responses of the most and least similar pairs of each subject individually were collected and analysed by means of a multidimensional scaling technique (MDS) and a two-tailed t-test.

Kruskal

The stimuli were analysed by a Kruskal algorithm (MINISSA) to obtain an impression of the perceptual space of the stimuli. Figure 3 illustrates the object space of the cumulative data.
Examining the object space it will be noticed that the specific relationship, that of a shorter transition extending to a higher frequency is absent in the plot. This is because the MDS algorithm yields a rank-order of the interpoint distances between the stimuli, and is therefore not capable of showing a finer perceptual structure between the stimuli. The plot is interesting though in that it illustrates that the stimuli were all equally well distinguished. Should the steady-state have been 100 msec, there would have been three clusters, one for each target frequency, as listeners would have focussed on this parameter in judging the most and least similar pairs.

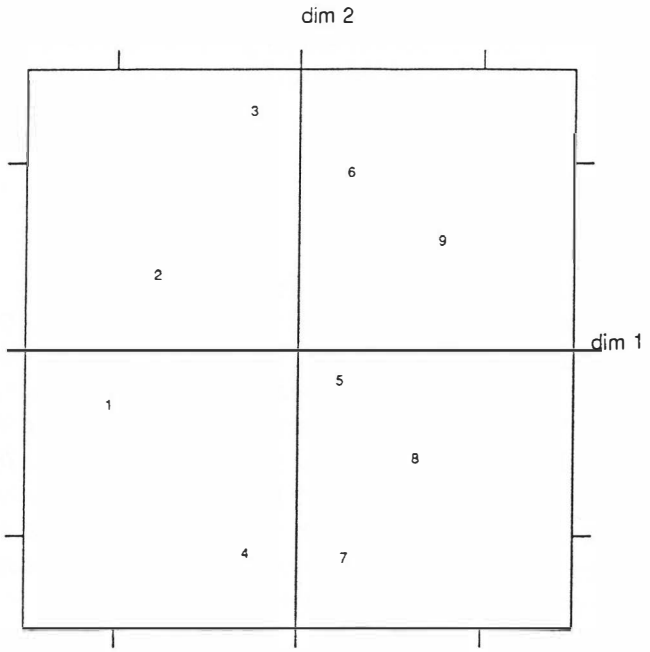

Figure 3. Two-dimensional Kruskal of the nine stimuli under test belonging to all the subjects. The three different target frequencies (vertically) and three different transition durations (horizontally) are placed at more or less equal distances from each other.

Vertically, the stimuli having the same target frequency, either 550 Hz, 625 Hz, or 700 Hz are separated by approximately the same distance. Horizontally, the stimuli are distinguished equally well by the 10ms (left), 20 ms (middle) and 30 ms (right) transition durations
The cumulative data naturally obfuscates the individual effects. The individual analyses show the distances between the two lower target frequencies often to be somewhat smaller than that between 625 Hz and 700 Hz. An explanation for these differences is

not at hand, but the data may reflect some kind of linguistic influence, due to the nature of the stimuli. Although simple, the stimuli are perceived by some as being speech-like. Especially the relatively fast transition, rising to a 700 Hz target frequency was said to be perceived as /bab/-like. Therefore, despite the 'phonetic categorization' seen in figure 3, the perceptual cue value of the parameters may be different for the stimuli with the 625 Hz target frequency, due to a different linguistic value. In the second experiment the effect of integrating target frequency and transition duration will be analysed further in terms of transition direction. In spite of the results varying individually, there seem to be no differences between the naive and phonetically trained subjects. Therefore, this distinction will not be made in further experiments of this kind.

T-test

The existence of a specific relationship, which cannot be concluded from the MDS Kruskal algorithm, can be tested by means of a two-tailed t test on two different groups, each consisting of four pairs of stimuli, the one group consisting of the pairs of stimuli marked by arrows in figure 2, the second being those placed in the reversed direction. The t test was performed on the two presentations of each subject. With 28 presentations (2 * 14) the difference in the means of the two groups can be tested with $\nu = 222$ (28 * 4 * 2 -2) and amounts to $t_{222} = -3.26$. The probability that the two groups are the same is less than 0.0005.

Conclusion

The results of the triadic tests are indicative of a specific relationship between the dynamic and stationary portions of the signal. Given that the transition is rising, two physically different synthetic stimuli are judged to be similar if a higher value of the steady-state of the signal is compensated by a longer/slower dynamic trajectory. Note that the measured percept is of a peripheral kind. From this type of test it is not possible to determine to what extent the phonetic percept remains unchanged.

The standing question is why the pairs marked by arrows sound more similar than those in opposite direction. The results suggest that not one cue in particular, i.e. formant frequency or transition duration/rate but the interaction between the two cues is responsible for a certain perceptual alikeness of two stimuli. In the following test the transition direction will be reversed to examine whether the same relationship between the stimuli also occurs in opposite direction.

EXPERIMENT 2

The second experiment is very much like the first one, the only difference in the stimuli being the direction of the transitions, which was falling instead of rising. The test was performed to investigate whether the relationship between the transition duration and the target frequency of the first experiment was also present with a reversed transition direction. Linguistic effects -if present at the psycho-acoustical level- would be different from possible linguistic effects in the first experiment. The stimulus percept was more VCV-like as the initial frequency was higher and the target frequency lower.

Hypothesis

It is expected that two physically different stimuli will sound alike if the stimulus with the lower target frequency has a longer/slower transition. The relatively fast transition of one stimulus could extrapolate to a lower target frequency of another stimulus. Two of such stimuli would sound alike in the context of three physically different stimuli. Note that with these stimuli, in which the initial target frequency is high, the excursion of the transition increases with decreasing target frequency, and a lower target frequency does not imply a reduced percept.

Stimuli

The stimuli differed by those used in the first experiment only by the direction of the transitions, which went down instead of up. The initial frequency was reversed from 220 Hz to 700 Hz. The target frequencies were 220 Hz, 295 Hz, 370 Hz (differing by 75 Hz). The transition durations were once again 10ms, 20ms, and 30ms. All other parameters remained the same as in experiment 1.

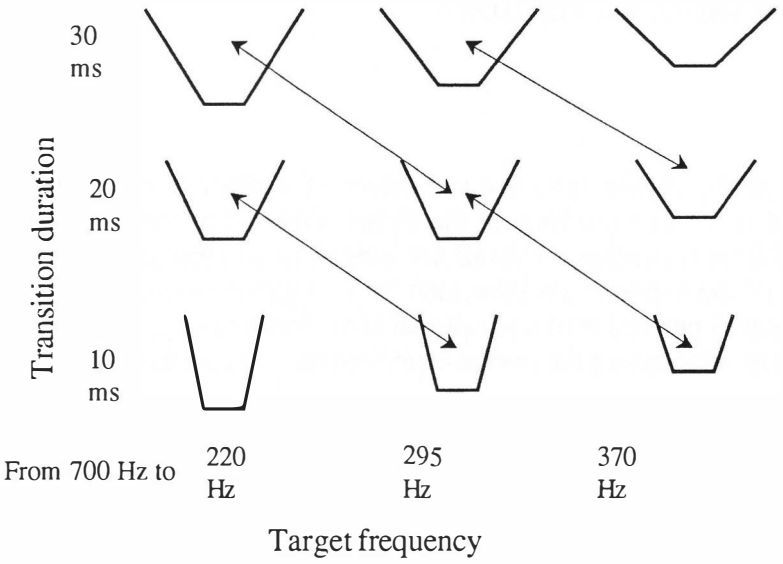

Figure 4. illustrating the stimuli of experiment 2. Arrows indicate pairs of stimuli which will sound more alike.

Procedure

The procedure was exactly the same as in the first experiment.

Subjects

Seven subjects, all of whom were familiar with the task, but again not with the differences in the stimuli, participated in the test. Even with this small number of subjects it was already clear that reversing the transition direction resulted in the same differences between the two groups of two pairs of stimuli.

Results

As in the previous experiment the difference in the means of both groups of pairs of stimuli were compared by means of a two-tailed t test (n=56). The t ratio t_{110}=-4.86 ($p<0.0005$) indicated that at each level of significance the two groups were totally different.

Conclusion

The results of this experiment are the same as that of the first experiment, i.e. that the longer the excursion of the transition the further away the target frequency must be -in this case lower- to be judged as a similar pair.

Discussion

In this test it is once again found that listeners make use of both the transition duration and the target frequency in judging (dis)similarity between sounds. Contrary to Lindblom et al. we are not investigating the role of target frequency in context of transitions, but the integration of both cues. However, both the first and second experiment indicate that the effect of target frequency (the vowel in identification tasks) is not stronger, and perhaps even less than that of transition duration/rate.

Prompted by the findings of the first and second experiment, i.e. the longer transition allowing a more extreme target frequency, a third experiment was performed in which the rate of the transitions were adjusted to be more or less the same. Consider the four upper left stimuli of figure 5. In spite of the total duration of the stimuli being the same for the stimuli joined by arrows the specific relationship between target frequency and transition duration only exists for the stimuli which are marked by a solid arrow, not for those marked by a dashed one.
The perceptual differences between the two diagonally placed pairs of stimuli may also be explained by the difference in transition rate. In table 1a the transition rates (in msec/100 Hz) are listed similarly to the placement of the nine stimuli in figure 5. Note that the transition rates of the stimuli marked by arrows are more or less the same, whereas those of stimuli placed in opposite direction differ markedly.
Instead of a compensation effect the listener may focus specifically on the transition rate, especially with one-formant stimuli. The following test is performed to test this.

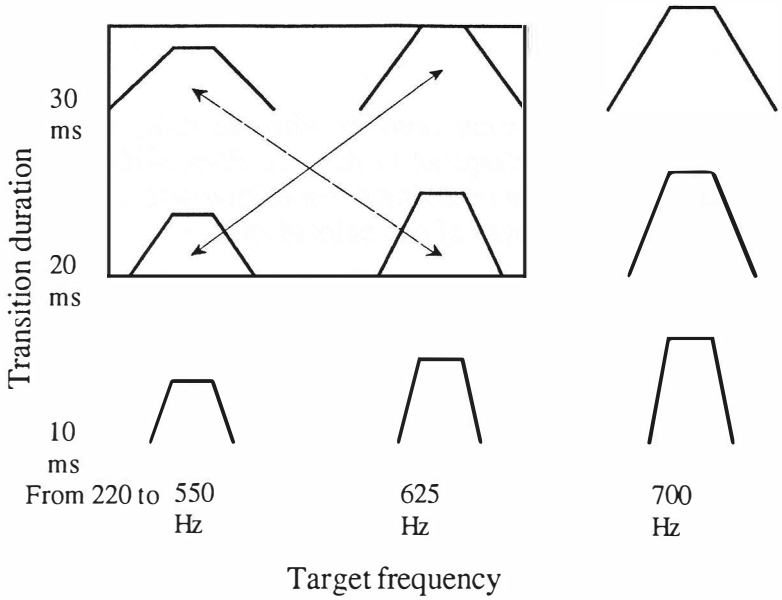

Figure 5. The total durations of both pairs of stimuli joined by arrows are identical. Yet, the pair joined by an arrow slanting to the right is perceived as sounding more similar than the one placed in opposite direction.

EXPERIMENT 3

To test whether transition rate was indeed the main cue in the two previous experiments the transition durations of some of the stimuli had to be changed such that the unmarked pairs would be perceived as being similar. This can be realized either by adjusting the target frequency, or by enlarging the difference in transition rate of the marked pairs, or by changing the durations of the unmarked ones so as to make all the stimuli more alike. The first alternative was not chosen to avoid introducing an extra variable in the judgements on the most and least similar pairs For the second alternative the same results as above were to be expected, whereas for the third there would be no difference between the two groups. The third alternative was chosen, in order to change as few rates as possible. As one cannot demand that the transition rates be identical -only one transition duration per target frequency would be left over- the rates were changed arbitrarily and the new transition durations were calculated and rounded off per decimal. The four new transition rates (in msec/100 Hz) and transition durations together with the unchanged ones are listed in table 1b.

Table 1a and 1b. Transition rates and transition durations of stimuli used in experiments 1, 2, and 3.

Tabel 1a

Experiment 1 and 2

30ms	9.09	7.4	6.25	|
20ms	6.06	4.93	4.16	|
10ms	3.03	2.46	2.08	|
	550 Hz	625 Hz	700 Hz	
	370 Hz	295 Hz	220 Hz	

Table 1b

Experiment 3

7.57	25ms	7.40	30ms	6.25	30ms
6.06	20ms	4.93	20ms	7.29	35ms
3.03	10ms	6.17	25ms	5.20	25ms
550 Hz		625 Hz		700 Hz	

Hypothesis

It is expected that with transition rate being more or less equal there would be no difference between the perceptual judgments of the pairs of stimuli placed diagonally in either direction.

Stimuli

Apart from four adjusted transition rates all the other parameters, as well as the generation of the stimuli, were the same as in experiment 1.

Procedure

The procedure was the same as in the two previous tests, i.e. a real time triadic comparison test consisting of two presentations of a randomisation sequence of 84 triads.

Subjects

Eight subjects, four of which had participated in one of the previous tests, judged the most and least similar pairs of stimuli. Due to the stimuli being more alike now the test lasted on average a quarter of an hour longer.

Results

Contrary to the t ratios of the two previous experiments a two-tailed t test now indicated that both groups of four pairs of stimuli were the same ($n=64$; $\nu = 126$). With $t = 0.37$ the null hypothesis that the groups are the same could not be rejected at any level of significance.

Discussion

The results of the third experiment suggest that perceptual equivalence depends to a large extent on transition rate (peripheral). That is to say, for the one-formant stimuli used in the three experiments. It is not unthinkable that with more complex stimuli the effect of transition rate is less. It may not necessarily be important for speech perception.

INDSCAL

Indscal, a simple weighted Euclidean model (Caroll and Chang, 1970), is used here to test the extent to which listeners remained consistent during the two presentations of a sequence of stimuli. Indscal considers perceptual differences across individual subjects and determines the relative weight of each dimension for each subject. As dimensions are uniquely defined in Indscal -due to the orientation of axes of the resulting solution being fixed, it is possible to determine which physical aspects of the stimuli play a major role in the perceptual judgments. The physical difference between the sounds

often serve as the guideline for the choice axes and the subsequent interpretations. Here the dimensions relating to acoustic information would be target frequency and transition duration The choice of attributes represents a theoretical decision, because it argues that listeners use these attributes in perceiving and jugding sounds.

By considering the two presentations of each subject separately there are twice as many subjects than actually participated. A three-way Indscal was used to examine how consistent subjects were during the course of judging.

Figure 6 a, b, and c illustrates the weighing of the subjects in two dimensions for exp. 1, exp. 2, and exp. 3, respectively. In all three plots the two presentations of each subject are indicated by the upper and lower cases of the same alphabetical letter. The placing of the subjects in the perceptual space for all three plots are fairly similar: the subjects seem to make use of both the dimensions transition duration and target frequency in judging the stimuli. Or, in other words, it may be possible, that, as listeners integrate these factors, they make their judgements on the basis of one difference in the sounds. This could be transition rate in our task.

In all three plots the stimuli are placed fairly close to the unity circle, indicating a well fitted solution. The two-dimensional solutions of the first, second and third experiments account for 78.2%, 81.3%, and 81.1% of the variance, respectively. These data correspond to a correlation of 0.882, 0.90, and 0.90 between data and spatial model, for exp. 1, exp.2, and exp.3, respectively.

Fig. 6a

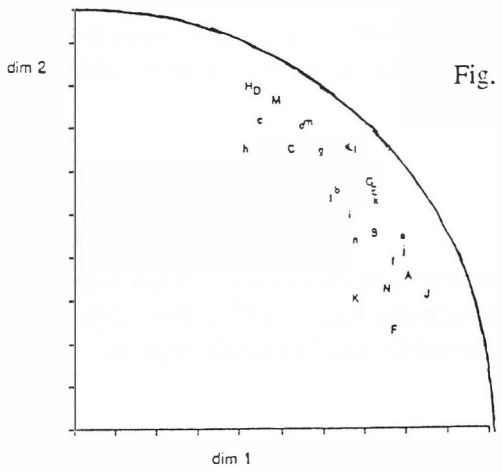

Fig. 6c

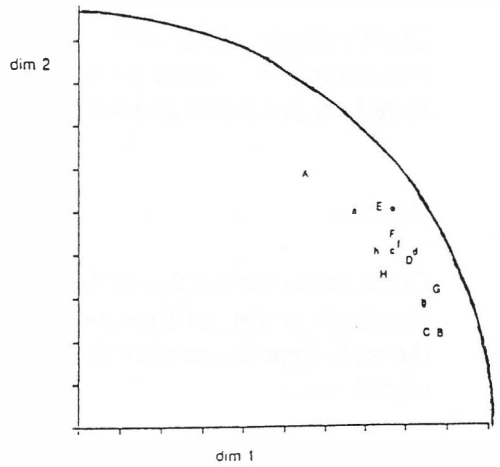

Fig. 6b

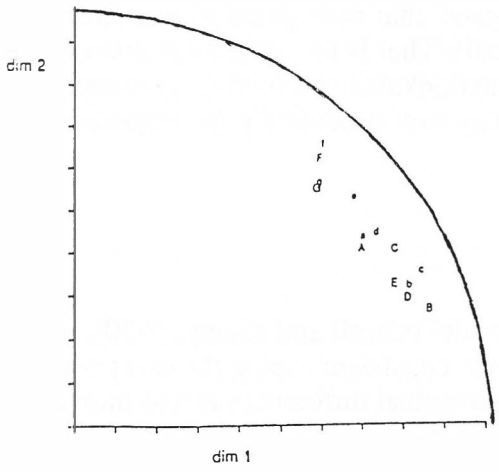

Figure 6 a,b,c. Indscal of experiments 1, 2, and 3, respectively.

GENERAL DISCUSSION

The three experiments discussed above indicate that much of the perceptual 'alikeness' in this specific task between the one-formant synthetic stimuli is based on transition rate. This implies that, although both the stationary and the dynamic portions of the sound integrate to one percept, there is a dominance of the time-varying trajectory over the target frequency. The first two experiments showed a specific type of perceptual compensation between target frequency and transition duration when either of the variables was changed: while presumably maintaining a similar percept a faster transition of one stimulus was allowed to extend to a more *extreme* target frequency of another one, dependent on the direction of the transition. With a rising transition a *higher* target frequency would be compensated by a faster transition; with a falling transition a *lower* target frequency would be compensated by a faster transition.
Without jumping to major conclusions some kind of perceptual overshoot seems to occur in both tasks. A lower frequency is chosen for falling transitions, a higher one for rising transitions. The typical relation between target frequency and trajectory will be studied in detail in future. This will also be done with respect to the manner in which the ear processes the incoming information. Before integration of the cues the listener may hear-out the individual components and respond selectively to the criterial properties (transition rate here). Should the target frequencies have differed by 200 Hz or the steady-states have been 200 msec instead of 15 msec, perceptual alikeness may not be based on transition rate anymore, as the relative importance of one cue depends on that of others. In pilot tests the salience of the target frequency at relatively long durations was assessed. We chose to use very short -15 ms- steady states as these can occur in rapidly spoken natural speech as well.
Although not the topic of this paper, the reversal of transition direction in our tasks is worth considering for a moment. First, listeners may not be able to identify falling tones or formant sweeps as well as rising ones (Pols and Shouten, 1982; Schouten, 1985; Pols and Schouten, 1987; Schouten en Pols, 1989), the triadic tests show the changes in rate to be processed similarly for both for rising and falling formant transitions. Second, in the second experiment listeners may have been less inclined to use Dutch linguistic categories due to the nature of the stimuli; however, as listener always make use of categories to mark out the different stimuli it is not clear from our few stimuli whether these are linguistically based or not.

Despite the auditory nature of the stimuli it is interesting to consider the importance of the transition rate in these experiments in terms of speech perception. Due to the non-linguistic nature of the stimuli and that of the task (that does not demand labelling), we were not examining the perception of the vowel in symmetrical CVC syllables, but rather the interaction of the transition and vowel. However, as communication depends on the the recognition of the sounds, the role of the transition should also be interpreted in terms of identifying the vowel and/or syllable

Speech perception

There is an ongoing debate on whether either the dynamic or the stationary portions of the signal are responsible for the identification of speech sounds.
Those who believe that the transition between the consonant and the vowel are a better source of phonetic recognition than the steady-state, as for examples Strange et al. (1976,1983), Gottfried and Strange (1980) ascribe the contextual advantage to vowel information being distributed throughout the entire syllable. This means that the

formant transitions -usually considered as specifying consonantal place and manner -, are also responsible for the recognition of vowels. Even after deleting the entire vocalic portion of /bVb/ syllables, listeners were capable of identifying vowels accurately. In the experiments by Benguerel (1989), the formant transitions were perhaps not sufficient to cue vowel identity, but definitely contributed to the recognition of the vowels in different contexts.
According to Furui (1986), who also supports the notion that vowel information does not only depend on spectral targets, the auditory system can predict the spectral targets on the basis of transitional information, the endpoints of the syllable being especially important.

The dependency of recognition on dynamic information is not acknowledged so strongly by, among others, Diehl et al. (1981) and Bladon (1985). Diehl failed to find vowels to be identified more accurately in the context of formant transitions, while Bladon, in examining the dynamic auditory processing of dipthongs, concluded that identification was brought about by the endpoints of the transition (the target frequency).

Notwithstanding the discussion whether the formant transitions subserve the perceptual extraction of the target frequency, and are then sufficient in cueing vowel identity, it is clear that different cues are prominent at various levels of processing. Whether identification is effectuated by the speed of the transition, on the endpoints, or on the target frequency, depends on the dynamic information present. As all parameters have perceptual cue value, the perceptual mechanism will compensate for undershoot or coarticulation by using prominent local dynamic cues, such as transition rate, as well as widespread contextual information. It is important to keep in mind that the conditions under which variables interact or compensate for each other are very stimulus, task, and context dependent. In extending their research on pure tones to that of more speech-like stimili, Schouten and Pols (1985, 1989) found that the tendency of listeners to perceive level and slightly rising tones in sweep tones as falling (1985) was less clearly present with formant sweeps. As the hearing mechanism does not change for different stimuli, other parameters are active in the processing of the direction of the sweep. The more cues present, the more they will interact resulting in enhancement of some, masking of others.

In this paper we have addressed potential cues in one experimental context. As has been said before listeners may not necessarily make use of this cue in natural speech. Future research, with increasingly complex stimuli, will focus on the perception of the vowel cued by dynamic information, not only in terms of the amount of information they carry, but also in terms of the conditions under which they are active.

REFERENCES

Benguerel, A.-P., Ukrainetz McFadden, T. (1989). "The effect of coarticulation on the role of transitions in vowel perception", Phonetica 46, 80-96.

Best, C., Morrongiello, B., and Robson, R. (1981). "Perceptual equivalence of acoustic cues in speech and nonspeech signals", Percep. Psychophys. 29, 191-211.

Bladon, A. (1985). "Diphthongs: a case study of dynamic auditory processing", Sp. Comm. 4, 145-154.

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.

Carroll, J.D., and Chang, J.J. (1975). Indscal

Diehl, R.L., Buchwald McCusker, S., and Chapman, L.S. (1981). "Perceiving vowels in isolation and in consonantal context", J. Acoust. Soc. Am. 69 (1), 239-248.

Fitch, H.L., Halwes, T., Erickson, D.M., and Liberman, A.M. (1980). "Perceptual equivalence of two acoustic cues for stop-consonant manner", Percep. Psychophys. vol 27(4), 343-350.

Fujisaki, H., and Sekimoto, S. (1975). "Perception of time-varying resonance frequencies in speech and non-speech stimuli", in: Structure and process in speech perception (eds. A. Cohen & S.G. Nooteboom), 269-282.

Furui, S. (1986). "On the role of spectral transition for speech perception", J. Acoust. Soc. Am., 80(4), 1016-1025.

Gottfried, T.L., and Strange, W (1980). "Identification of coarticulated vowels", J. Acoust. Soc. Am., 68(6), 1626-1635.

House, A.S., and Neuburg, E.P. (1983). Unpublished data made available by personal communication in 1983, (1970).

Lindblom, B. (1963). "Spectrographic study of vowel reduction", J. Acoust. Soc. Am. 35, 1773-1781.

Lindblom, B.E.F., and Studdert-Kennedy, M. (1967). "On the role of formant transitions in vowel recognition", J. Acoust. Soc. Am. 42(4), 830-843.

Lindblom, B., and Moon, S.-J. (1988). "Formant undershoot in clear and citation-form speech", Phonetic Experimental Research Institue of Linguistic University of Stockholm (PERILUS), VIII, 21-31.

Pols, L.C.W., and Schouten, M.E.H. (1982). "Perceptual relevance of coarticulation", in: The representation of speech in the peripheral auditory system, (eds. R. Carlson & b. Granström), 201-208.

Pols, L.C.W., Boxelaar, G.W., and Koopmans-van Beinum, F.J. (1984). "Study of the role of formant transitions in vowel recognition using the matching paradigm", Proceedings of the Institute of Acoustics, 6(4), 371-378.

Pols, L.C.W., and Schouten, M.E.H. (1987). "Perception of tone, band, and formant sweeps", in: The psychophysics of speech perception (ed. M.E.H. Schouten), 231-240.

Repp, B.H. (1983). "Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization", Speech Comm. 2, 341-361.

Schouten, M.E.H. (1985). "Identification and discrimination of sweep tones", Percep. Psychophys. 37(4), 369-376.

Schouten, M.E.H. (1986). "Three-way identification of sweep tones", Percep. Psychophys. 40(5), 359-361.

Schouten, M.E.H., Pols, L.C.W. (1989). "Identification and discrimination of sweep formants", Percep. Psychophys. 46(3), 235-244.

Son, R.J.J.H. van, and Pols, L.C.W. (1990). "Formant frequencies of Dutch vowels in a text, read at normal and fast rate", J. Acoust. Soc. Am. 88(4), 1683-1693.

Son, R.J.J.H. van, and Pols, L.C.W. (1990). "Formant movements of Dutch vowels in a text, read at normal and fast rate", Submitted to the J. Acoust. Soc. Am.

Strange, W., Jenkins, J.J., and Johnson, Th.L. (1983). "Dynamic specification of coarticulated vowels", J. Acoust. Soc. Am. 74, 695-705.

Strange, W., Verbrugge, R.R., Shankweiler, D.P., and Edman, T.R. (1976). "What enables a listener to map a talker's vowel space?", J. Acoust. Soc. Am., 60, 198-212.

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