IDENTIFICATION OF NASALS: AN ACOUSTICAL AND PERCEPTUAL ANALYSIS OF NASAL CONSONANTS[1)]

David House

INTRODUCTION

In general terms, the spectral cues governing perception and identification of nasal consonants are described as dichotomous. Information concerning the presence of a nasal[2)] is thought to reside in the nasal segment proper (the nasal murmur) while the formant transitions of the adjoining vowel segments are usually regarded as comprising the major spectral cues necessary for specific nasal phoneme identification. In an attempt to investigate the extent to which vowel formant transition information alone is sufficient for nasal phoneme identification in a VC-syllable in the context of the vowels /i/, /a/, and /u/, three experiments were carried out.

The first experiment took the form of a speech perception identification test using OVE III speech synthesis based on the Haskins tests of the 1950's where vowel formant transitions were progressively altered while keeping the nasal segment constant. The failure of this test to elicit labial responses after the /u/ vowel and a general non-conformity of /u/ vowel context results led to a spectrographic analysis of nasals following /u/ in natural speech. The appearance of a characteristic spectral peak at about 1100 Hz in the /m/ segment led to the third experiment which once again took the form of a synthetic speech identification test. This time, however, natural speech syllables were replicated as closely as possible using an OVE III synthesizer in an attempt to elicit labial responses after /u/ on the basis of the /m/ spectral peak keeping other information at a minimum. A parallel identification test using natural recorded speech syllables was also carried out in order to determine actual nasal phoneme identification difficulty in the /u/ vowel context in isolated syllables.

PROCEDURE

The first experiment comprised 60 stimuli in which a single nasal formant at 200 Hz followed Swedish vowel formants for /i/, /a/, and /u/. F_2 transitions in the vowel segments were altered in steps of 500 Hz ranging from a final value of 500 Hz up to 2500 Hz. The test was presented for four Swedish listeners, the task being to identify each stimulus as one of the following nine syllables: /im, in, iŋ, am, an, aŋ, um, un, uŋ/.

In the second experiment, five sets of the nine syllables were read in randomized sequences by five subjects. All but one of the subjects were native speakers of Swedish, the investigator representing American English. Spectrograms of the middle three sets of the recorded syllables were analysed regarding F_2 transitions and possible distinguishing spectral characteristics in the nasal segments. Particular attention was paid to the segments containing the vowel /u/. Sections from the middle of the nasal segments were also analysed.

The third experiment, designed to investigate perception of syllables in which the nasal follows the /u/ vowel, comprised listener identification tests using both natural and synthetic speech stimuli. In the natural speech portion of the test, the syllables /um/, /un/, and /uŋ/ were randomized and read by one of the Swedish speakers. Each syllable appeared 20 times. In the speech synthesis portion, an attempt was made to replicate the three syllables of natural speech using an OVE III synthesizer in such a way as to enable the testing of the following criteria of identification: a) an N2 at 1100 Hz in the nasal segment of /um/ with no formant transitions, b) an F_2 transition to 1800 Hz in /un/ with no N2 in the nasal segment, and c) the absence of both formant transitions and N2 in /uŋ/ (Fig. 1). The second nasal formant of /m/ was produced by extending the formants of the vowel branch from the vowel segment into the nasal segment, thereby using F2 to emphasize the nasal segment spectrum at 1100 Hz. This method was not completely satisfactory, however, as a substantial (-18 dB re: the vowel) amplitude decrease of the vowel formants in the nasal segment was necessary to maintain satisfactory nasal acceptability due to the effect of using vowel formants in a nasal segment, despite increased bandwidths. Nonetheless, given the single nasal formant nature of OVE III synthesis, the stimulus was considered to be the best possible means of testing the identification criteria given the equipment available.

The three synthetic speech stimuli were randomized and recorded, each stimulus appearing 20 times. Finally, the same stimuli were presented in the form of a pair test where the /uŋ/ stimulus was matched against /um/ stimuli having varying vowel branch amplitudes in the nasal segment in order to test the possibility of an optimum F2 (N2) - N1 amplitude relationship. The test was presented twice, the first time with instructions to identify the labial member of the pair /um/ and the second time with instructions to identify the velar /uŋ/. 13 listeners participated in the third experiment.

RESULTS

For vowels /i/ and /a/ in the first experiment, results concerning F_2 transition patterns and nasal phoneme identification conform very well with the Haskins results. Low F_2 transitions elicited labial responses, transitions ending at 1500 Hz elicited dental responses, and higher F_2 transitions elicited velar responses. When the vowel /u/ preceded the nasal segment, however, the results changed drastically (Fig. 2). Low F_2 transitions elicited velar responses and high transitions dental responses while there was only a single labial preference.

In the second experiment, spectrogram analysis of F_2 transitions in the vowels /i/ and /a/ preceding the three nasal phonemes conforms well with the results obtained in the first experiment. Further, two of the five speakers, representing English and Swedish, produced /m/ segments having an extremely well defined second formant at 1100 Hz which was lacking in the other nasal segments. Although this /m/ resonance at 1100 Hz was not as

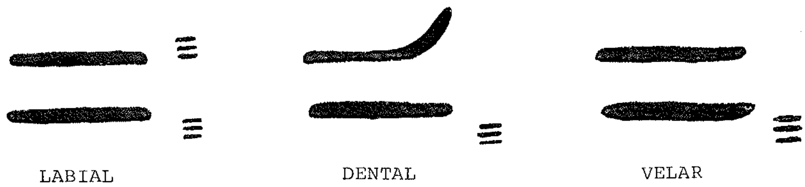

Figure 1. Schematic synthetic stimuli used for synthesis.

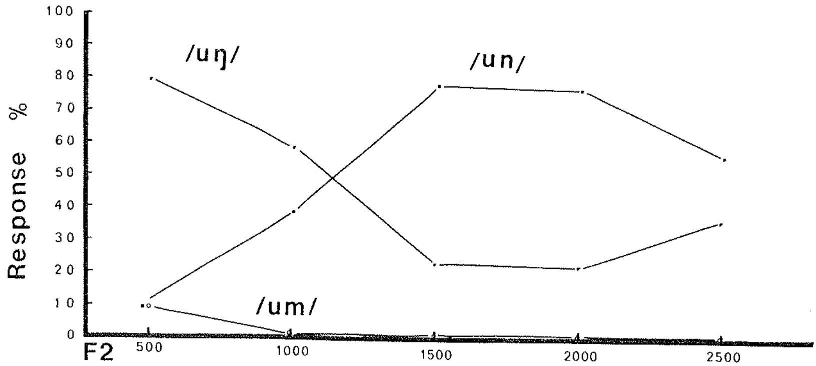

Figure 2. Frequency (Hz) of F2, final value along the abscissa, percent response for each stimulus along the ordinate.

sharply defined in the spectrograms of the other subjects, a spectral peak at this frequency could easily be discerned. Spectrograms of the /u/ vowel syllables differed from the other vowels in that no F_2 transitions were present in the labial or velar context (Fig. 3).

Identification of the natural speech syllables /um/, /un/, and /uŋ/ in the third experiment proved to be a very easy task. Correct identification of the three syllables was nearly 100%, slight confusion (less than 10%) occuring, as could be expected, between /um/ and /uŋ/ (Fig. 4). Identification of the synthetic stimuli where dental and velar responses were expected (Fig. 5) corresponded acceptably with the natural speech test (greater than 75%). Response where labials were expected did not correspond well with the natural speech results. Responses for this stimulus were distributed fairly evenly among the three possible phonemes in the identification test. Labial identification was obtained, however, in more than one-third of the responses (34.3%), a result that compares favorably with that of the first experiment (Fig. 2) where practically no labials were elicited. Results of the pair test where the /uŋ/ stimulus was paired against different versions of /um/ did not differ substantially from the results of the identification test. While a clear preference for the /uŋ/ stimulus was obtained when velar identification was requested (over 80%), no clear preference for any of the /um/ stimuli was obtained when labial identification was requested. Thus there was no optimum stimulus for eliciting a labial response.

DISCUSSION

Where the identification of /un/ and /uŋ/ are concerned, the results of this investigation demonstrate that a perceptually significant spectral cue for the identification of dental nasals following the vowel /u/ can be said to lie in the second formant transition of the vowel, while a spectral cue for identification of velars in the same context lies in a more general downward shift of energy from the vowel segment to the nasal segment.

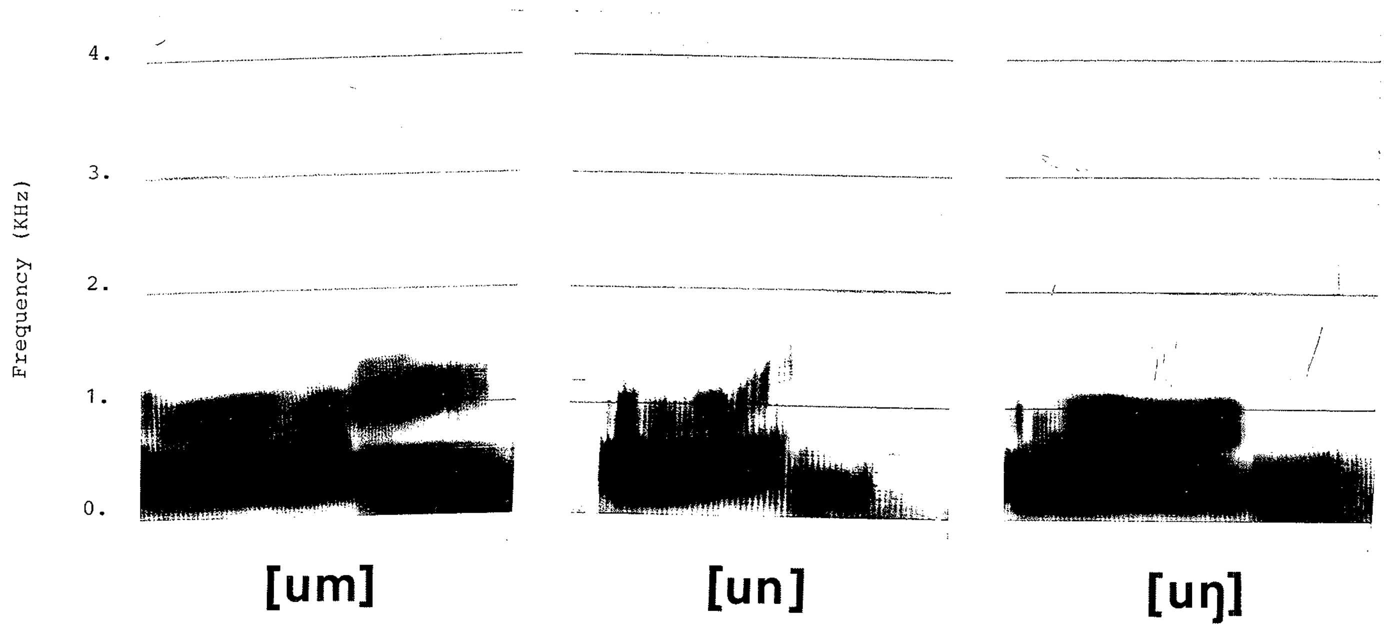

Figure 3. Representative spectrograms, Swedish Subject.

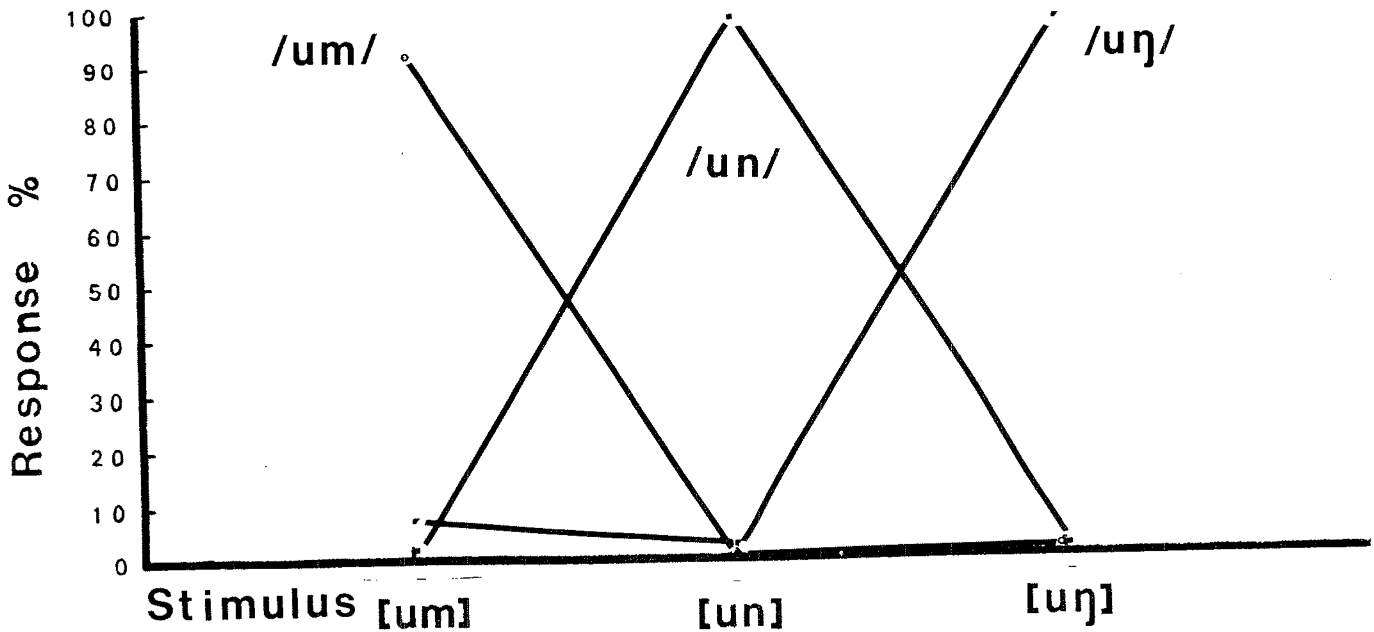

Figure 4. NATURAL SPEECH. Stimulus syllables along the abscissa, percent response for each stimulus along the ordinate.

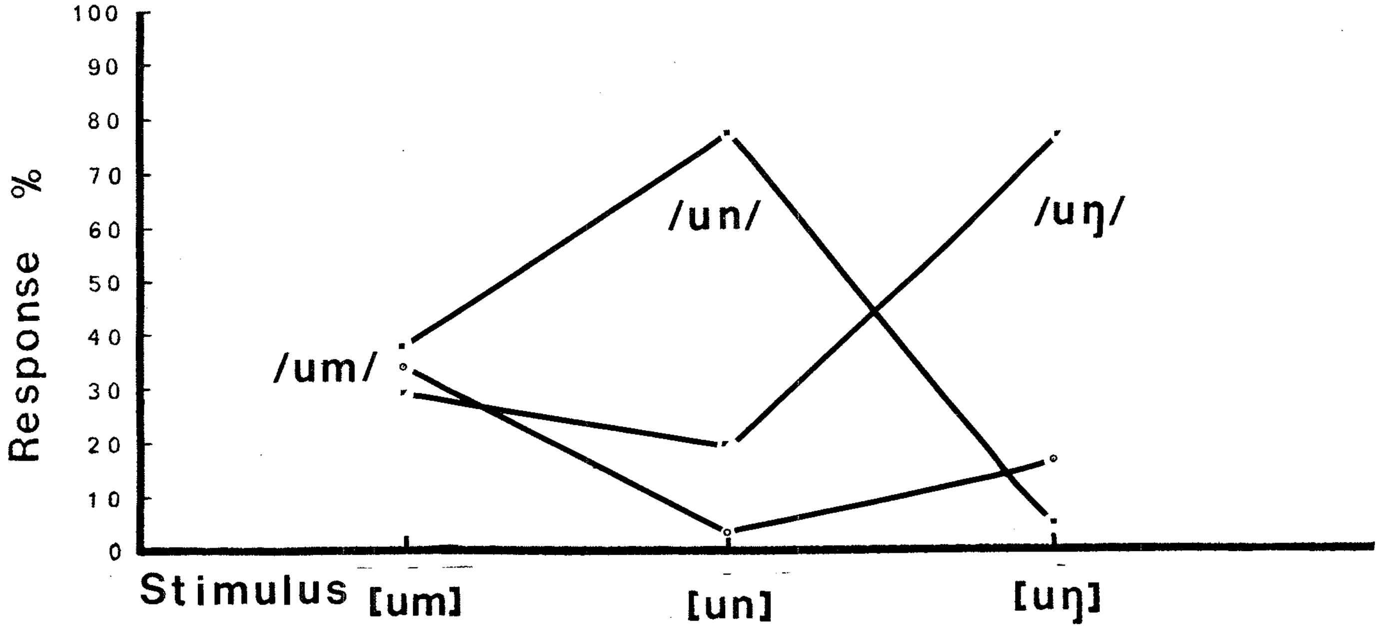

Figure 5. SYNTHETIC SPEECH. Stimuli along the abscissa, percent response for each stimulus along the ordinate.

Regarding identification of labials, results of this test (although possibly attributable to guessing and far from conclusive) do indicate a possibility that identification could be based on a spectral cue (perhaps a pole-zero pair emphasizing the envelope at 1100 Hz) located within the nasal segment. The OVE III synthesis available for this investigation did not enable this point to be sufficiently tested. A number of labial responses were obtained when expected, however, and very few where not expected showing that it is possible to elicit labials by virtue of a more complex nasal segment spectrum. As labial responses were not obtained in the first experiment, the addition of the second /m/ resonance can at least be seen as a step in the right direction.

This investigation further demonstrates the complicated nature of the perceptual mechanism and once again questions the concept of perceptually invariant spectral cues. Evidence points toward a double mechanism of spectral cues having no dependence upon phonemic units. Thus, the same phoneme /ŋ/ in different contexts relies upon different cue mechanisms. Following /i/, the formant transitions can provide the information. Following /u/, a general downward shift of energy from the vowel to the nasal segment can provide the information. This also indicates a possible difference in the point in time at which the perceptual mechanism determines phonemic identity.

SUMMARY

The purpose of this investigation was to determine the extent to which nasal phonemic identification is governed by formant transitions in the vowel, following the transition patterns for place of articulation as described in the Haskins perception experiments. While these patterns seem to hold true for nasal consonants following the vowels /i/ and /a/, a different type of cue is needed for non-dental nasals following /u/. The identification of /uŋ/ can be made on the basis of a downward energy shift from the vowel segment into the nasal segment. It is hypothesized that identification of /um/ is related to a pole-zero pair creating a spectral

peak at about 1100 Hz.

ACKNOWLEDGEMENTS

I am greatly endebted to Sidney Wood for his extensive help with the speech synthesis used in this investigation. Thanks also to Anne-Christine Bredvad-Jensen, Gösta Bruce, and Lennart Nord for their interest and encouragement.

REFERENCES

Fant, G. (1960) Acoustic Theory of Speech Production. The Hague: Mouton & Co.

Fujimura, O. (1962) Analysis of Nasal Consonants. Journal of the Acoustical Society of America 34: 1865-1875

House, A.S. (1975) Analog studies of nasal consonants. Journal of Speech and Hearing Disorders 22: 190-204

Liberman, A.M. et al. (1952) Perception of Synthetic Speech Sounds. Journal of the Acoustical Society of America 24: 597-606

Malécot, A. (1956) Acoustic Cues for Nasal Consonants. Language 32: 274-284

Nord, L. (1976) Perceptual Experiments with Nasals. Quarterly Progress and Status Report of the Speech Transmission Laboratory, Royal Institute of Technology, Stockholm 2/3: 5-8

Nord, L. (1976) Experiments with Nasal Synthesis. Quarterly Progress and Status Report of the Speech Transmission Laboratory, Royal Institute of Technology, Stockholm 2/3: 14-19

FOOTNOTES

1) This is a condensed version of a longer paper, copies of which are available from the author.

2) The term nasal will be used here to pertain exclusively to nasal consonants.