DISCRIMINATION OF COARTICULATED GERMAN VOWELS IN THE SILENT-CENTER PARADIGM: "TARGET" SPECTRAL INFORMATION NOT NEEDED

Ocke-Schwen Bohn and Winifred Strange***
**English Department, Kiel University, Germany*
*** Dept. Communication Sciences and Disorders, U. South Florida, Tampa, USA*

ABSTRACT

German listeners were tested on their ability to discriminate naturally produced German /dVt/-syllables which were modified to manipulate the availability of formant target information (traditionally considered the primary information for vowel identity) and of dynamic spectral information. The results support Strange's [1] Dynamic Specification Theory, which states that vowels are specified by dynamic information defined over syllable onsets and offsets.

INTRODUCTION

In the traditional view of vowel perception, the target frequencies of the first two formants constitute the primary acoustic information for the perceptual identity of vowels. This Simple Target Model of vowel perception is inadequate because it fails to account for how listeners perceive speakers' intended messages in the face of various sources of variation in the acoustic signal. One kind of variation in vowel targets comes from coarticulation of vowels with consonants in consonant-vowel-consonant (CVC) syllables. Research by Strange and her collaborators has shown that vowels produced in CVC-syllables are identified with far greater accuracy than vowels produced in isolation, even though targets are often not reached in coarticulated vowels [1]. This led Strange to hypothesize that important information for vowel identity must be contained in the dynamic contour of the formants within the syllable.

Strange and her collaborators developed the Silent Center paradigm to test their hypotheses on the role of dynamic sources of information in vowel perception. Their methodology involves the systematic modification of CVC-syllables to explore the perceptual relevance of various sources of potential information contained in CVC-syllables. The methodologically most important modification in this paradigm is the generation of silent-center (SC) syllables, which are created by attenuating to silence the entire syllable nucleus, leaving only the initial and final transitions in their appropriate temporal relationship. The converse of SCs are vowel-centers (VCs), which are created by deleting the initial and final transitions, so that the syllable nuclei with target information are retained. Experiments employing SCs and VCs as stimuli allow one to test the perceptual importance of acoustic information associated either with the opening and closing gestures of the vocal tract in the production of CVC-syllables (i.e., SCs), or with the (approximate) target configuration of the vocal tract in vowel production (i.e., VCs). Experiments employing the SC paradigm typically also test the perception of initials (INIs) and finals (FINs), which are, respectively, the initial and final transitions alone. This is done to test whether SCs in their entirety, or their initial or final part alone, contribute to perceived vowel identity.

Several studies have examined the perception of AE vowels by AE listeners in the SC paradigm and found high levels of identifiability for SCs. Strange [1] concluded a review of these studies by stating that no single spectral cross-section adequately captures the perceptually relevant information; rather, the acoustic information for vowel identity resides in the changing spectral structure. Because most previous SC studies examined the perception of AE vowels, some ambiguity as to the nature of the dynamic information remained. According to Nearey's Compound Target Theory (CTT), the more or less diphthongized English vowels can be differentiated by contrasting patterns of vowel-inherent spectral change [2, 3]). Strange's Dynamic Specification Theory (DST), on the other hand, states that vowels are specified by dynamic information defined over syllable onsets and offsets [1]. The dynamic information

reflects each vowels' characteristic opening and closing phases in their appropriate temporal relationship and style of movement of the vocal tract.

In order to examine the generality of previous findings on the identification of AE vowels in the SC paradigm, and to assess alternative hypotheses about the nature of dynamic information, Strange & Bohn [4] examined the identification of coarticulated German vowels in /dVt/-syllables by native German listeners in the SC paradigm. Like English, German has a large vowel inventory whose monophthongs differ in tenseness and/or length, but which have little or no diphthongization. Strange's DST predicts that dynamic spectral information plays an important role in the perception of vowels which have little or no diphthongization. Nearey's CTT, on the other hand, predicts that the importance of dynamic spectral information is restricted to the perception of diphthongized vowels.

Strange & Bohn [4] reported that the availability of dynamic spectral vs. target spectral information affected German listeners' identification of coarticulated German vowels in much the same way as AE listeners' identification of coarticulated AE vowels. Vowel identity was maintained very well in German SC-syllables even though the vocalic nucleus with information on formant targets was not presented in that condition. Strange & Bohn also reported extremely high error rates in both INI and FIN conditions, indicating that neither onsets alone nor offstes alone were sufficient to maintain vowel identity for German listeners.

The present study was designed to test the robustness of the findings of Strange & Bohn by using a new set of stimuli and employing a different experimental paradigm. Four vowel contrasts which German listeners confused most frequently in Strange & Bohn's [4] identification experiments were examined for their discriminability in experimental conditions which selectively presented different types of acoustic information contained in CVC-syllables. Compared to previous identification studies, the discrimination paradigm used in the present study presents listeners with a very simple task in terms of memory load and cognitive demands. One might expect, therefore, that performance levels in INI and FIN conditions will be much higher in a task involving simple discrimination rather than identification with, e.g., 14 response alternatives as in the Strange & Bohn study. If this were so, an important assumption of the DST would be questioned, namely, that vowel identity is specified not by onsets alone or offsets alone, but by trajectory information defined over syllables onsets **and** offsets in their appropriate temporal relationship. On the other hand, the DST would be strengthened if the pattern of results obtained from a simple discrimination experiment were much the same as the patterns obtained previously from identification experiments (i.e., low performance levels for INIs and FINs as opposed to the SC condition).

METHODS

Stimuli

Six tokens each of the vowels /i/, /I/, /e/, /ɛ/, /U/, /o/ were produced in isolated /dVt/-syllables by a male native speaker and recorded onto DAT. The vowels were selected for discrimination because the pairs /i/-/e/, /e/-/I/, /I/-/ɛ/ and /o/-/U/ were most frequently confused in the identification experiments of Strange & Bohn [4]. From the digitized waveforms, measurements of target syllable duration, voice onset time and fundamental frequency were used to make the final selection of four instances each of the six vowels.

SCs were generated by attenuating to silence the center portion of each of the target syllables, leaving onset and offset portions in their original temporal position. The onset and offset portions included the major part of the transitions associated with opening and closing gestures for all vowels. - VCs were the converse of SCs. They were generated by attenuating to silence the onset and offset portions. INIs were generated by silencing both center and offset portions in each syllable. FINs were generated by silencing both onset and center portions of each syllable.

Subjects

50 subjects who met the selection criteria (no history of hearing loss according to self-report, native speaker of North German, limited exposure to

languages other than German) were recruited mainly from linguistics courses at Kiel University and participated as unpaid volunteers. The mean age of the 33 female and 17 male subjects was 26 years (SD = 5.5).

Procedure

The stimulus syllables (unmodified syllables, SCs, VCs, INIs, and FINs) were redigitized and stored on the hard disk of a 486-PC. Groups of 10 subjects each were assigned to one of the five listening conditions (defined by stimulus type) and tested individually in a sound treated chamber, where stimuli were presented from a loudspeaker. Each subject was tested for discrimination of the contrasts /i/-/e/, /e/-/I/, /I/-/ɛ/, /o/-/U/ in a pseudo-randomized order. Because the results of the discrimination experiment were to provide baseline data for a study of infant vowel perception that employed the change/no change procedure [5], the adult subjects were tested in an age-appropriate version of that procedure. The subjects listened to presentations of the background stimuli (e.g., the four randomly presented tokens of /dit/ for the /i/-/e/ contrast) while being engaged in a simple distractor game, and were instructed to raise their hand when they detected a change to the foreground (e.g., /det/ for the /i/-/e/ contrast). The ISI was 1.5 sec; a change trial consisted of the presentation of three foreground stimuli. The change from background to foreground stimuli was initiated by an assistant, who could not hear whether she initiated a change or a control trial, and who observed the subject through a one-way mirror. The assistant pressed a button when she observed a hand signal by the subject during trials. Custom software controlled feedback for correct signals (illumination of a display above the loudspeaker.) The software also kept track of correct responses, false alarms, correct rejections, and misses throughout the 25 trials (15 changes and 10 controls) for each contrast.

RESULTS

Figure 1 gives the overall results for the five stimulus conditions for the vowel contrasts /i/-/e/, /e/-/I/, /I/-/ɛ/, /o/-/U/, expressed as percentages of correct responses averaged across subjects within each group. Discrimination levels for unmodified syllables (mean % correct: 97.5), SCs (mean % correct: 96.1), and VCs (mean % correct: 99.1) did not differ significantly. Vowel identity was well maintained in the SC condition, even though the vocalic nucleus with information on formant targets was not presented in that condition. All four vowel contrasts were discriminated highly accurately in the SC condition (/i/-/e/: 96.5%, /e/-/I/: 95.6 %, /I/-/ɛ/: 95.6, /o/-/U/: 96.8 %).

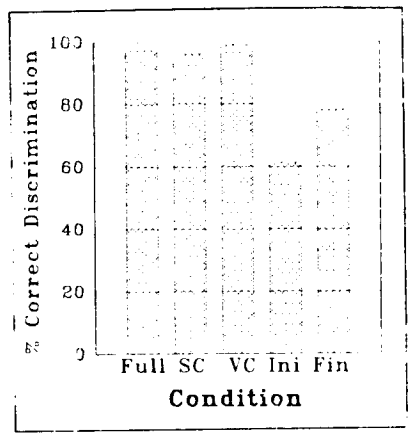

Figure 1: Overall per cent correct responses in vowel discrimination in unmodified syllables (Full), silent-center syllables (SC), vowel center (VC), initials (INI), and finals (FIN) conditions.

Both FINs and INIs on the one hand were discriminated siginificantly less accurately than unmodified syllables, SCs, and VCs on the other. This shows that neither onsets alone nor offsets alone were sufficient to maintain vowel identity. Overall discrimination of FINs (78.3 % correct) was significantly better than of INIs (61.2 % correct), but more detailed analyses revealed that this was true only for contrasts differing in tenseness (/e/-/I/, /o/-/U/). Performance on FINs and INIs was not significantly different for the tense /i/-/e/ or the lax /I/-/ɛ/ contrast.

CONCLUSIONS

The most important finding was that German listeners discriminated German vowels highly accurately when only dynamic spectral information specified over onsets and offsets together was presented. The pattern of results for the discrimination of confusable German vowel contrasts was very similar to that reported previously for the identification of German and AE vowels by native listeners. For an adequate perceptual representation of vowels in these two languages with large vowel inventories, target spectral information is not necessary; rather, trajectory information specified over syllable onsets and offsets is a very good source of information for vowel identity in both German in AE.

The finding that INIs and FINs were discriminated less accurately than SCs provides important support for Strange's DST [1], which states that vowel identity is specified by dynamic information defined over syllable onsets **and** offsets. Even when listeners were confronted with the very simple task of detecting a change from foreground to background stimuli, onsets alone (INIs) or offsets alone (FINs) did not specify vowel identity as well as trajectory information defined over syllable onsets **and** offsets in their appropriate temporal relationships (SCs).

The finding that German vowels presented as SCs were identified and discriminated highly accurately is incompatible with Nearey's ([2, 3]) Compound Target Theory (CTT). The CTT predicts that distinct patterns of vowel-inherent spectral change (VISC) contribute importantly to perceived vowel identity, at least when the more or less diphthongized vowels of AE are presented as SCs. However, acoustic analyses of the German vowels presented in the experiments reported here revealed very little or no VISC. Formant movement, particularly for F2, was observed in many of the vowels, but this movement was associated with coarticulatory influences from the preceding and following alveolar consonants (see also [6]). This means that the dynamic sources of information which German listeners used so successfully in the SC conditions were associated with the opening and closing gestures at the margins of CVC-syllables, as predicted by Strange's [1] DST.

Further studies are underway to establish the generality of these perceptual results with German vowels and adult German listeners. One series of studies in progress examines whether German vowels produced in multiple consonant contexts by multiple speakers are identified accurately when listeners are presented with SCs. Another series examines how the three types of acoustic information present in coarticulated German vowels (dynamic spectral, target spectral, and temporal) contribute to perceived vowel identity in prelingual German infants [5]. Preliminary results from these experiments suggest that almost all infants who can discriminate vowel contrasts with unmodified syllables can also discriminate the same contrast when vowels are presented as SCs.

ACKNOWLEDGEMENTS

Research supported by a grant from the *Deutsche Forschungsgemeinschaft* (DFG grant Bo-1055/3-1) to O. Bohn, and by a grant from the National Institutes of Health (NIDCD 00323) to W. Strange. The authors wish to thank Desiderio Saludes and Sonja Trent for their assistance in generating stimulus materials, and Kirsten Schriever and Anja Steinlen for assistance in testing subjects. Very special thanks go to Linda Polka for her help in setting up the speech perception lab at Kiel University.

REFERENCES

[1] Strange, W. (1989), "Evolving theories of vowel perception." *J. Acoustical Soc. America* 89, 2081-2087.

[2] Nearey, T. (1989), "Static, dynamic and relational properties in vowel perception." *J. Acoustical Soc. America* 85, 2088-2113.

[3] Andruski, J. E. & Nearey, T. M. (1992), "On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables." *J. Acoustical Soc. America* 91, 390-410.

[4] Strange, W. & Bohn, O.-S. (1995), "Dynamic specification of coarticulated German vowels: I. Perceptual studies." To appear in *J. Acoustical Soc. America*.

[5] Bohn, O.-S. & Polka, L. (1995), "What defines vowel identity in prelingual infants?" *Proceedings 13th International Congress of Phonetics*.

[6] Strange, W. & Bohn, O.-S. (in prep.), "Dynamic specification of coarticulated German vowels: II. Acoustical studies."