WHAT DEFINES VOWEL IDENTITY IN PRELINGUAL INFANTS?

Ocke-Schwen Bohn and Linda Polka***
**English Department, Kiel University, Germany*
*** School of Human Communication Disorders, McGill University, Montreal, Canada*

ABSTRACT
This is the first report on experiments examining which acoustic properties of coarticulated vowels (target spectral, dynamic spectral, temporal) define vowel identity in prelingual infants. German-learning infants were tested for discrimination of German vowel contrasts in the Silent Center paradigm. Results indicate that infants derive vowel identity from dynamic spectral information, and that target information is not needed for perceived vowel identity.

INTRODUCTION
Previous studies of vowel perception by adult American English (AE) listeners and by adult German listeners have shown that three types of acoustic information contained in consonant-vowel-consonant (CVC) syllables contribute to vowel identity: target spectral information, dynamic spectral information of the syllable onsets and offsets, and temporal information. Strange and her collaborators established the relative importance of these three types of information [1], [2], [3]. The most important finding was that adult AE listeners and adult German listeners identify coarticulated native vowels highly accurately if target spectral information has been electronically removed from the CVC stimuli. This indicates that vowel identity does not depend on acoustic information on spectral targets. Instead, the perceptually relevant information for vowel identity seems to reside in the changing spectral structure of coarticulated syllables.

Several studies by Strange and her collaborators examined in detail the sources of dynamic information in the perception of native coarticulated vowels by adult AE listeners ([1], [2]) and adult German listeners ([3], [4]). These studies found that Vowel Centers (VCs), which consist only of the syllabic nuclei with target information, are not perceived more accurately than Silent Center (SC) syllables, which consist only of the dynamic portions of the syllable onsets and offsets in their appropriate temporal relationship. Vowel identity is maintained very well in SCs even though the vocalic nucleus with information on formant targets is silenced in SCs. These studies also showed that syllable onsets alone (INIs) or syllable offsets alone (FINs) do not provide sufficient information on vowel identity for adult listeners.

These findings and others on the insufficiency of target information for vowel perception are accounted for by Strange's Dynamic Specification Theory (DST), which 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. DST provides an elegant solution to an important problem in vowel perception research, viz., perceptual constancy. Unlike DST, Target Models of vowel perception have to account for the target undershoot problem (formant targets are often not reached in coarticulated vowels) and for speaker normalization (formant targets for the same vowel category differ greatly across men, women, and children). Context- and speaker-dependent variation pose problems only for those theories of vowel perception that view formant (or gestural) targets as objects of perception. Research motivated by the DST, however, strongly suggests that coarticulated vowels are specified by styles of movement that are invariant across consonant contexts [2] and across different speakers [5].

The experiments reported here are the first to examine the role of the three types of acoustic information of CVC syllables in infant vowel perception. This was done by testing German-learning infants' discrimination of naturally produced German /dVt/-syllables which were modified to manipulate the availability of the three types of acoustic information. No study has ever examined how target spectral information, dynamic

spectral information, and temporal information contribute to perceived vowel identity in prelingual infants. This is somewhat surprising given the fact that many studies of infant vowel perception implicitly assume that spectral targets alone specify vowel identity (e.g. [6]).

METHOD
Stimuli
Six tokens each of the German vowels /i/, /I/, /e/, /ɛ/, /U/, /o/, /ɑ/ were produced in /dVt/-syllables by a male native German speaker and recorded onto DAT. The vowels were selected to be presented in the contrasts /i/-/e/, /e/-/i/, /I/-/ɛ/, and /o/-/U/ because these contrasts were confusable in the identification experiments reported by Strange & Bohn [3]. The maximal /i/-/ɑ/ contrast was selected as a control contrast. Measurements of syllable duration, voice onset time and fundamental frequency of multiple tokens of the seven vowels were used to make the final selection of four instances each of the six vowels.

To test the role of target vs. dynamic spectral information, the original syllables were modified as follows. SCs were generated by attenuating to silence the center portion of each of the original syllables, leaving onset and offset portions in their original temporal position. The onset and offset portions included the major part of the transitions. VCs were generated by silencing the onset and offset portions. INIs were generated by silencing both center and offset portions, and FINs were generated by silencing both onset and center portions.

To test the role of temporal information, all eight tokens for a given contrast were electronically edited so that they had the same duration. This was done by iterating or deleting full pitch periods (for full syllables with neutral duration - FNDs, and for VCs with neutral durations - CNDs), or by adding or deleting silence (for SCs with neutral durations - SCNDs).

Subjects
80 infants served as subjects. All were healthy, full term infants with no history of ear infections (by parental report). The infants aged between 7 and 11 months were being raised in monolingual German-speaking families in Kiel, Germany.

Procedure
Infants were tested using the headturn procedure (for details of our implementation, s. [7]). In this procedure a syllable is played from a loudspeaker every 1.5 sec and at random intervals this background syllable changes to a target syllable for a brief interval. Discrimination is assessed by first conditioning the infant to turn his/her head in the direction of a visual reinforcer above the loudspeaker when they detect a change in the background syllable. Correct headturns are reinforced by the activation of a visual reinforcer (an electronic animal that moves) accompanied by verbal praise. We implemented this procedure as a category change paradigm in which the background and the target consist of multiple tokens of each syllable type.

Discrimination of a vowel contrast in a given condition (e.g., /i/-/e/ as SC) was tested in a single session. The infant was seated on a parent's lap across a small table from an experimenter (E1). The loudspeaker and an array of visual reinforcers, located behind a smoked plexiglass panel, were arranged to one side of the parent and infant. The parent and E1 listened to music over headphones to prevent them from hearing the stimuli and influencing the infant. A second experimenter (E2), located ouside the test room, observed the infant through a one-way window and operated the computer.

The session begins with a conditioning stage in which the infant is given an opportunity to learn the contingency between the vowel change and availablity of the visual reinforcer (s. [7]). During the testing stage, E2 initiates trials when the infant is in a "state of readiness" (not fussing, facing E1 etc.). E2 is blind to the trial type and pushes a button when she observes a headturn during the trial interval. The visual reinforcer is activated automatically for a change trial when E2 records a headturn by pushing a response button. Twenty-five trials were presented during testing stage.

Infant testing was conducted in an sound-treated chamber. Custom software controlled stimulus delivery, activation of the reinforcers, and trial selection (i.e. presentation of change vs. no-change trial), and also recorded the number of trials, hits, misses, correct rejections and false alarms.

Design

Groups of 10 subjects each were assigned to one of the eight listening conditions, which were defined by the 4 vowel contrasts (/i/-/e/, /e/-/i/, /I/-/ε/, /o/-/U/) and by availability of temporal information (in unmodified syllables, SCs, and VCs) vs. its neutralization (in FNDs, SCNDs, and CNDs). Each subject was first tested for discrimination of the unmodified test contrast. Only infants who discriminated full syllables (criterion: 7/8 consecutive correct trials and > 60 % correct responses) were then tested on separate days for discrimination of the contrast tested initially in the edited conditions. Infants were randomly assigned to the two series of experiments. In the first series, infants were tested for discrimination of SCs, VCs, and INIs or FINs. In the second series, infants were tested for discrimination of FNDs, SCNDs, and CNDs. The vowel category which served as the background was counterbalanced within each group.

RESULTS

Figure 1 gives the overall results for the eight stimulus conditions for two vowel contrasts (/i/-/e/, /e/-/I/), expressed as percentage of correct responses averaged across subjects. Overall discrimination levels for unmodified syllables (mean % correct: 69.3), SCs (mean % correct: 67.0), and VCs (mean % correct: 63.7) did not differ significantly. Compared to the Full condition, 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. The mean per cent correct values for INIs (47.5) and FINs (51.9) suggest that vowel onsets or vowel offsets alone do not preserve vowel identity.

The two vowel contrasts did not differ significantly in discriminability in the Full condition (/i/-/e/: 69.7 % correct, /e/-/I/: 68.9 % correct) and in the SC condition (/i/-/e/: 66.8 % correct, /e/-/I/: 67.1 % correct). In the VC condition, the /i/-/e/ contrast was significantly less discriminable (52.9 % correct) than the /e/-/I/ contrast (74.3% correct).

Neutralization of the temporal contrast reduced the discriminability of both contrasts in the Full conditions (Full syllables: 69.3 % correct; FNDs: 55.2 % correct), in the SC conditions (SC: 67.0 % correct; SCND: 56.3 % correct), and in the VC conditions for the /e/-/I/ contrast (VC: 74.3 % correct; CND: 53.0 % correct), but it did not affect the discriminability of the /i/-/e/ contrast in the VC conditions (VC: 52.9 correct; CND: 53.0 % correct).

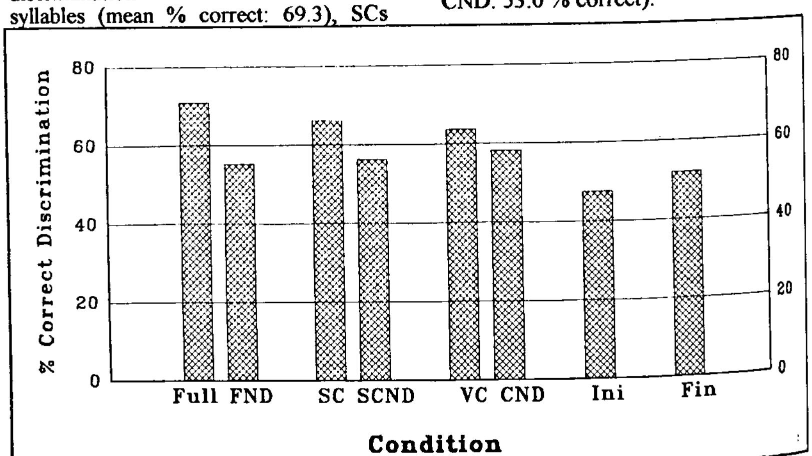

Figure 1: Overall per cent correct responses in vowel discrimination in unmodified syllables (Full), full syllables with neutral duration (FND) silent-center syllables (SC), silent-center syllables with neutral duration (SCND), vowel centers (VC), vowel centers with neutral duration (CND), initials (INI), and finals (FIN) conditions for the vowel contrasts /i/-/e/ and /e/-/I/.

CONCLUSIONS

The most important finding was that German-learning infants discriminated two German vowel contrasts equally well when these contrasts were presented either as unmodified full syllables or as SCs, which preserve only the dynamic spectral information of the syllable onsets and offsets in their appropriate temporal relationships. This suggests that infants do not need target spectral information to differentiate the spectrally similar high front vowel contrasts /i/-/e/ and /e/-/I/. Rather, trajectory information specified over syllable onsets and offsets is a good source of information for vowel identity in prelingual infants, as it is in adult native speakers of AE and of German.

The overall pattern of results for the German infants is quite similar to that for German adults, who discriminated the same contrasts in a related study [4]. Within each age group of infants and adults, discriminabilty of Full, SC, and VC syllables did not differ significantly, but discrimination levels for INIs and FINs were lower than for SCs in both the adult and the infant study.

The first results from our experiments on the acoustic specification of vowels in infants support Strange's Dynamic Specification Theory, which states that vowels are specified by dynamic information defined over syllable onsets and offsets together. German infants discriminated two German vowel contrasts by making use of the dynamic sources of information associated with the opening and closing gestures at the margins of the CVC syllables. This indicates that infants perceive coarticulated vowels in terms of their characteristic styles of movement. We suggest that perceptual representations of these styles, which seem to be invariant across consonant contexts [2] and across different speakers [5], contribute importantly to perceptual constancy for vowel categories in infants. Further research is underway to establish the generality of our first results by examining how accurately infants discriminate vowels produced in varying contexts and by multiple speakers when presented only with dynamic information specified over syllable onsets and offsets.

One interesting aspect of our study is that infants' discrimination abilities suffer considerably if contrastive temporal information is not available. Unlike adult German listeners, for whom the neutralization of duration contrasts had only a very selective effect for individual vowel contrasts in specific experimental conditions (unpublished data from the study reported in [4]), German infants discriminated both contrasts at lower levels of performance when the contrasts were temporally neutralized. Further research will have to show whether German infants' sensitivity to temporal manipulations reflects L1 experience with the German vowel system, or whether duration differences have a universally important function in learning to differentiate vowel contrasts.

ACKNOWLEDGMENTS

Research supported by a grant from the *Deutsche Forschungsgemeinschaft* (DFG grant Bo-1055/3-1) to O. Bohn. We thank Desiderio Saludes and Sonja Trent for assistance in generating stimulus materials, and Kirsten Schriever, Tatjana Soldat, Anja Steinlen, and Amira Yassine for assistance in testing infants. Special thanks go to Winifred Strange for her essential support, advice, and encouragement.

REFERENCES

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

[2] Jenkins J. J. (1987), "A selective history of issues in vowel perception." *J. Memory and Language* 26, 542-549.

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

[4] Bohn, O.-S. & Strange, W. (1995) "Discrimination of coarticulated German vowels in the Silent-Center Paradigm". *Proc. 13th Int. Congr. Phon. Sciences.*

[5] Jenkins, J. J. et al. (1994), "Vowel identification in mixed-speaker silent-center syllables" *J. Acoust. Soc. America* 95, 1030-1043.

[6] Kuhl, P. K. et al. (1992), "Linguistic experience alters phonetic perception in infants by 6 months of age." *Science* 255, 606-608.

[7] Polka, L. & Werker, J. F. (1994), "Developmental changes in the perception of nonnative vowel contrasts". *J. Experim. Psychology: Human Perception & Performance* 20, 421-435.