Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants

PETER W. JUSCZYK and ELIZABETH THOMPSON
Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

In the present study, three aspects of the 2-month-old's perception of multisyllabic utterances were explored. First, do infants perceive phonetic contrasts occurring either in the initial (Bada-Gada) or medial (Daba-Daga) positions of multisyllabic utterances? Second, are infants more likely to perceive these contrasts in stressed as opposed to unstressed syllables? Third, will infants detect a difference between two stress patterns? Our results indicate the following: (1) Two-month-olds are sensitive to place-of-articulation differences occurring in either the initial or medial positions of multisyllabic stimuli. (2) Whether the contrast occurred between stressed or unstressed syllables had little or no effect on the infant's ability to detect it. (3) Infants as young as 2-months old respond to differences between stress patterns.

For most of its brief history, the study of speech perception in human infants has been focused on the *variety* of contrasts that infants are capable of perceiving (e.g., Eimas, 1975a; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Morse, 1972; Trehub, 1976). While this line of research has been fruitful, it has dealt only with one aspect of the development of speech perception, namely, the ability to distinguish *single syllables* on the basis of *initial* segment differences. However, these kinds of distinctions represent only a small subset of those which the infant will have to make if he is to become a successful perceiver of speech. In fact, the speech sounds which infants hear from adults often differ in their medial or final segments, as well as their initial ones. Moreover, many of the words which adults use have more than one syllable, and each syllable may not be enunciated to the same degree (i.e., some syllables are stressed and others are unstressed). At the present time, we know very little about the extent to which the infant's perceptual system is capable of coping with the information present in *multisyllabic* utterances.

Our previous research (Jusczyk, 1977) has demonstrated that infants are sensitive to differences between initial segments (**dab** vs. **gab**) and final seg-

ments (**bad** vs. **bag**) of single syllables. Thus, given that infants do appear to process more than just the initial segment of a syllable, we might expect that they would be sensitive to contrasts between segments of different multisyllables. However, the perception of phonetic contrasts in multisyllables is complicated by a number of factors. One such factor is the effect that embedding a phonetic contrast in the middle of an utterance may have on the infant's ability to perceive it. A second consideration is the presence in multisyllabic words of suprasegmental features of speech such as stress patterns. Stress is produced by increasing, either singly or in combination, a syllable's pitch, duration, or intensity (Fry, 1958; Lehiste, 1970; Lieberman, 1960). While the factors which give rise to the adult's perception of stress differences have been studied (e.g., Fry, 1958; McClean & Tiffany, 1973; Morton & Jassem, 1965), little attention has been given to the question of how stress differences might affect the adult's (or the infant's) perception of phonetic contrasts.

It is not unreasonable to suppose that one effect of stress may be to enhance the perception of a syllable by calling more attention to it. Indeed, we often act as if this were the case when we attempt to clarify a statement by stressing the misunderstood element. Developmental psychologists also have implied that stress directs attention to an item by suggesting that stress plays an important role in what words a child learns first (Blasdell & Jensen, 1970; Brown & Fraser, 1963; Miller & Ervin, 1964). If stress really does serve to highlight a particular item, then we might expect that infants would be more likely to perceive differences between stressed syllables than between unstressed ones.

In the present study, we explored aspects of the

2-month-old's perception of stop consonant differences in multisyllabic utterances. Three issues were of primary concern to us. First of all, does position within an utterance affect the infant's ability to detect a phonetic contrast between two multisyllables? Consequently, we tested the perception of the same place-of-articulation difference ([b] vs. [g]) both in the initial position (e.g., [**B**ada] vs. [**G**ada] and the medial position (e.g., [Da**b**a] vs. [Da**g**a]) of a two-syllable utterance. Second, are infants more likely to perceive a given contrast between stressed syllables or unstressed ones? Thus, we compared the infant's ability to detect the same contrast in both stressed (e.g., [Dabá] vs. [Dagá]) and unstressed (e.g., [Dába] vs. [Dága] syllables. Third, are infants as young as 2 months old sensitive to change in stress patterns? Previous research on the perception of suprasegmental information by infants has been limited to the study of intonation contours for sentences (Kaplan, 1969) and the detection of pitch changes in single syllables (Kuhl, 1976; Morse, 1972). In her study, Kaplan found that 8-month-olds, but not 4-month-olds, are able to perceive intonation contour differences. By contrast, Morse (1972) found that 6-week-olds discriminated a rising pitch from a falling one. Recently Kuhl and Miller (1975, see also Kuhl, 1976; Spring & Dale, 1977) replicated and extended Morse's findings by showing that infants do not appear to recognize pitch changes in vowel stimuli when the vowel is varied randomly throughout the experimental session. For this reason, we also investigated the 2-month-old's capacity to respond to segmental information such as a difference in stress patterns (e.g., [Báda] vs. [Badá]).

METHOD

Procedure

Each infant was tested individually in a small laboratory room. The infant was placed in a reclining seat which faced a rear-projection screen approximately 2 ft away. An image of a man was displayed on the screen for the entire test session. The projection screen was situated just above a loudspeaker through which the test stimuli were played. Each infant sucked on a blind nipple which was held in place by one of the experimenters who wore headphones and listened to recorded music throughout the test session. A second experimenter in an adjacent room monitored the apparatus.

The experimental procedure was a modification of the high-amplitude sucking technique developed by Siqueland and DeLucia (1969). For each infant, the high-amplitude sucking criterion and the baseline rate of high-amplitude sucking were established before the presentation of any stimuli. The criterion for high-amplitude sucking was adjusted to produce rates of 15 to 30 sucks/min. After a baseline rate was established, the presentation of stimuli was made contingent upon the rate of sucking. Since the stimuli were approximately 670 msec in duration and a 750-msec interstimulus interval was used, a maximum of one stimulus was presented every 1.4 sec. If the infant produced a burst of sucking responses with interresponse times less than 1.4 sec, then each response did not produce one presentation of the stimulus. Rather, the timing apparatus was reset so as to

provide continuous auditory feedback until 1.4 sec after the last response of the sucking burst.

The criterion for satiation to the first stimulus was a decrement in sucking rate of 25% or more over 2 consecutive minutes compared to the rate in the immediately preceding minute. At this point, the auditory stimulation was changed without interruption by switching channels on the tape recorder. For infants in the experimental conditions, the change resulted in the presentation of a second acoustically distinct stimulus. For infants in the control condition, the channels on the tape recorder were switched but no acoustic change was made. The postshift period lasted for 4 min. The infant's sensitivity to the change in auditory stimulation was inferred from comparisons of the response rates of subjects in the experimental and control conditions during the postshift period.

Stimuli

So that we might closely approximate those features of naturally produced bisyllabic utterances such as stress patterns and spacing between syllables, our speech stimuli were produced at the Haskins Laboratories in the following manner. Naturally produced speech samples were recorded on auditory tape. Next, using the CONVERT program at Haskins, we used the natural speech spectrum data to provide values for the parameters of the OVE III serial synthesizer. The resulting synthetic utterances were then digitized, using the Haskins PCM system (Cooper & Mattingly, 1969). The output from the PCM system was then used to prepare the audio tapes employed in this experiment.

Eight different synthetic bisyllabic stimuli were created (Dába, Dabá, Dága, Dagá, Báda, Badá, Gáda, Gadá). Each stimulus consisted of one stressed and one unstressed syllable. For half of the stimuli, the stressed syllable was first (Dába, Dága, Bága, Gáda). For the other half the unstressed syllable was first (Dabá, Dagá, Badá, Gadá). In order to simulate the closure of the second stop consonant; a pause of 78 msec occurred in all the stimuli between the termination of the first syllable and the onset of the second one.

Stressed syllables had durations of 364 msec, and fundamental frequency (pitch) contours which began at 118 Hz, peaked 120 msec later at 130 Hz, and then declined to 106 Hz. By contrast, unstressed syllables had durations of 208 msec, and fundamental frequency contours which began at 106 Hz peaked 65 msec later at 117 Hz and declined to 106 Hz. In addition, the maximum amplitude of the stressed syllables was 5 dB greater than that of the unstressed ones.

The first 65 msec of each syllable represented a transitional period during which all three formants moved from their starting frequencies to their terminal steady-state values of 777, 1,139, and 2,150 Hz for the first, second, and third formant, respectively. All syllables had the same first-formant starting value of 388 Hz. The starting values for the second and third formants of each [da] and 1,744 and 2,557 Hz, respectively; while the comparable values for [ba] were 847 and 1,902 Hz and those for [ga] were 1,707 and 1,808 Hz.

Design

Table 1 shows the within-subject design employed in the present experiment.

All subjects were seen for two sessions. (Mean interval between sessions was 7.1 days; the range was 1 to 19 days). Subjects in Group I were presented with stimuli that contrasted in the medial position for both sessions. During one of the sessions, this contrast occurred in a stressed syllable (e.g., Dabá vs. Dagá), and during the other session it occurred in the unstressed syllable (Dába vs. Dága). Subjects in Group II were treated exactly like Group I except that they were presented with stimuli that contrasted in the initial position for both sessions. By comparison, subjects in Group III received stimuli that were *phonetically identical* for both sessions. However, during one of these sessions (STRESS CHANGE), the stress pattern was altered following satiation

Table 1
Design

	Condition A	Condition B
Group I (n = 12)	DÁBA-DÁGA	DABÁ-DAGÁ
Group II (n = 12)	BÁDA-GÁDA	BADÁ-GADÁ
Group III	Stress Change (e.g., BÁDA-BADÁ)	No Change (e.g., BADÁ-BADÁ)

to the first stimulus (e.g., Báda vs. Badá, Gáda, vs. Gadá, Dába vs. Dabá, or Dága vs. Dagá). During the other session (NO CHANGE), the subjects were randomly assigned one of the eight bisyllable stimuli for the entire session (e.g., Badá-Badá). The order of test conditions and the order of stimuli within a session were each counterbalanced across subjects.

Apparatus
A blind nipple was connected to a Grass PT5 volumetric pressure transducer, which was coupled in turn to a Type DMP-4A Physiograph. A Schmitt trigger provided a digital output of critical high-amplitude sucking responses. Additional equipment included a Sanyo four-track stereo tape recorder (Model #MR-939) with a remote speaker, a Lafayette 100-sec timer, a power supply, two relays, a counter, and a Physiograph dc preamplifier. Each criterion response activated the 100-sec timer for a 1.4-sec period or restarted the period. Auditory stimulation at a level of 79 ± 2 dB (A) SPL (approximately 15 dB above the background noise level caused by the ventilation system) was available to the infant whenever the timer was in an "active" state.[1]

Subjects
The subjects were 36 infants, 15 males and 21 females. Mean age was 7.4 weeks (range: 6 to 9 weeks). In order to obtain data for two complete sessions with 36 infants, it was necessary to test 91. Of the subjects excluded from the study, approximately 69% of them failed to complete the first session, while the remaining 31% failed to complete the second session. Subjects were excluded form the study for the following reasons: falling asleep (27%) or crying (36%) prior to shift, failure to maintain a minimum criterion sucking rate of 15 responses/min during the satiation period (13%), ceasing to suck during the course of the experiment (i.e., 2 consecutive minutes with less than 2 sucks/min) (9%), failure to keep the second appointment (6%), and experimenter error (9%).

RESULTS

Figure 1 displays the mean number of high-amplitude sucking responses as a function of minutes and experimental groups. For purposes of statistical comparisons, we examined each subject's rate of sucking during five intervals: baseline minute, third minute before shift, average of Minutes 1 and 2 before shift, average of Minutes 1 and 2 after shift, and average of all 4 min after shift. Difference scores were calculated for each subject for the following rate comparisons: (1) acquisition of the sucking response—third minute before shift less baseline; (2) satiation—third minute before shift less average of last 2 min before shift; (3) release from satiation— average of first 2 min after shift less average of last 2 min before shift; and (4) release from satiation for full 4 min after shift—average of 4 min after shift less average of last 2 min before shift.

As is usually the case in studies using this procedure, subjects in all sessions acquired the conditioned high-amplitude sucking response and satiated to the first stimulus prior to shift. An indication of subjects' *postshift* performance is provided in Table 2. Randomization tests (Siegel, 1956) were used to make comparisons between performance in each of the five experimental sessions and that of the NO CHANGE session for both the first 2-min and the full 4-min periods after shift. The tests indicated reliable (p < .05 or better) difference between each of the experimental sessions and that of the NO CHANGE session for the first 2 min after shift. Comparisons of performance for the full 4-min period yielded similar results, with one exception. Group I subjects in the Dabá-Dagá session, although in the right direction, did not differ reliably (p < .20) from the NO CHANGE session for this period.
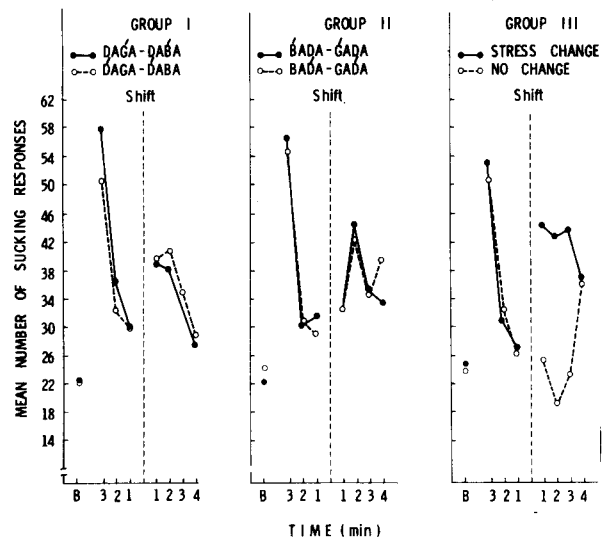
In order to evaluate the effect of stress on the

Figure 1. Mean number of high-amplitude sucking responses as a function of time and experimental group. Time is measured with reference to the moment of the stimulus shift, marked by the vertical dashed line. The baseline rate of sucking is indicated by the letter "B."

Table 2
Mean Change in Response Rate After Shift

		Release From Satiation (Minutes After Shift)	
		First 2	Full 4
Group I	Dabá-Dagá	5.50*	1.31
	Dába-Dága	8.88*	4.65*
Group II	Báda-Gáda	8.17*	6.02*
	Badá-Gadá	7.08*	6.92*
Group III	STRESS CHANGE	14.21*	12.60*
	NO CHANGE	-6.75	-3.08

*Indicates a reliable difference (p < .05 or better) according to randomization tests when compared against performance in the NO CHANGE session.

perception of the phonetic contrast, two kinds of comparisons were made. For the first kind of comparison, the data from Groups I and II were collapsed across groups. Friedman two-way analyses of variance (Siegel, 1956) were then used to compare performance for the session in which the phonetic contrast was stressed to that in which it was unstressed. The results indicated no difference in performance on the stressed vs. the unstressed phonetic contrasts for either the first 2-min or the full 4-min periods after shift. For the second kind of comparison, separate randomization tests for matched samples (Siegel, 1956) were used to assess performance across type of session for subjects in Groups I and II. In neither group was there any indication that subjects performed reliably better for contrasts between stressed as opposed to unstressed syllables.

DISCUSSION

The present study demonstrates that 2-month-olds are sensitive to subtle differences between multisyllabic utterances. First, infants are capable of distinguishing, on the basis of a single phonetic feature difference, multisyllables which contrast in either their initial (e.g., **B**ada vs. Gada) or medial (e.g., Daba vs. Daga) segments. Second, 2-month-olds will respond to a change in the stress pattern of a multisyllabic utterance (e.g., Dága vs. Dagá). Moreover, there was no indication in the present study that stress pattern interacts with the perception of a stop consonant contrast (i.e., subjects were no more likely to detect a contrast in a stressed as opposed to an unstressed syllable).

The present results extend previous developmental findings in two directions. First, they indicate that infants can detect place-of-articulation differences not only for different positions in single syllable stimuli (Eimas, 1974; Jusczyk, 1977; Miller, Morse, & Dorman, 1977; Morse, 1972), but for the initial and medial positions of *multisyllabic* stimuli as well.[2] Second, in contrast to Kaplan's (1969) findings for the perception of intonation contours, it was shown that infants as young as 2 months of age are sensitive to suprasegmental speech information such as stress patterns. -

In view of the claims made by a number of developmental psycholinguists (Blasdell & Jensen, 1970; Brown & Fraser, 1963; Miller & Ervin, 1964), the finding that stress position had little or no effect on the infant's ability to perceive the phonetic contrast was somewhat surprising. One might have expected that the increase in pitch, intensity, and duration associated with the stressed syllables might have made them more salient, and thus differences between them would have been more detectable. There are several possible explanations for our results. It

may be the case that as long as a syllable is at an audible level, stress does not affect the perception of its elements. Thus, we would have to look for another explanation as to why the language learning child often omits unstressed syllables in his productions. Alternatively, stress may interact with the perception of some contrasts but not others. In the present study, we looked only at stop consonants. These phones are all characterized by a period of complete closure of the vocal tract, and this alone may make them quite prominent to a listener. Perhaps stress interacts with the perception of other phones, such as those characterized by less abrupt transitions (e.g., glides).

One additional result which deserves comment is the failure to find a reliable difference between the NO CHANGE sessions and one of our experimental sessions (i.e., the Dagá-Dagá session) for the *full 4-min period* after shift. In fact, an examination of the mean recovery in response rates for the full 4-min period after shift for Groups I and II (see Table 2) shows them to be generally lower than those observed using single-syllable stimuli differing in place-of-articulation (Eimas, 1974; Jusczyk, 1977). One possible explanation for the reduced level of responding for the full 4-min period after shift with multisyllabic stimuli has to do with the novelty of the second stimulus. Proportionally, there is less that is novel about two multisyllabic stimuli differing by the same feature. For this reason, subjects may resatiate to the second stimulus more quickly when multisyllabic stimuli are employed.

In conclusion, the present study has added new information concerning the remarkable capacities that the young infant has available for speech perception. Not only are infants able to perceive a wide variety of speech contrasts (e.g., Eimas, 1975a, 1975b; Trehub, 1976), but they are capable of detecting information about such contrasts across different positions in a syllable (Jusczyk, 1977) and, as we have shown, in strings longer than a single syllable.

REFERENCE NOTE

1. Trehub, S. E. *Infants discrimination of multisyllabic stimuli: The role of temporal factors.* Paper presented at the annual convention of the American Speech and Hearing Association, Houston, Texas, November 21, 1976.

REFERENCES

BLASDELL, R., & JENSEN, P. Stress and word position as determinants of imitation in first-language learners. *Journal of Speech and Hearing Research*, 1970, 13, 193-202.

BROWN, R., & FRASER, C. The acquisition of syntax. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963.

COOPER, F. S., & MATTINGLY, I. G. A computer-controlled PCM system for the investigation of dichotic speech perception.

Journal of the Acoustical Society of America, 1969, **46**, 115 (A).

EIMAS, P. D. Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, 1974, **16**, 513-521.

EIMAS, P. D. Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics*, 1975, **18**, 341-347. (a)

EIMAS, P. D. Speech perception in early infancy. In L. B. Cohen & P. Salapatek (Eds.), *Infant perception*. New York: Academic Press, 1975. (b)

EIMAS, P. D., SIQUELAND, E. R., JUSCZYK, P. W., & VIGORITO, J. Speech perception in infants. *Science*, 1971, **171**, 303-306.

FRY, D. B. Experiments in the perception of stress. *Language and Speech*, 1958, **1**, 126-152.

JUSCZYK, P. W. Perception of syllable-final stop consonants by 2-month-old infants. *Perception & Psychophysics*, 1977, **21**, 450-454.

KAPLAN, E. L. The role of intonation in the acquisition of language. (Unpublished PhD dissertation, Cornell University, 1969.) *Dissertation Abstracts International*, 1970, **30**, 3407B. (University Microfilms No. 69-20, 977)

KUHL, P. K. Speech perception in early infancy: The acquisition of speech-sound categories. In S. K. Hirsh, D. H. Eldredge, I. J. Hirsh, & S. R. Silverman (Eds.), *Hearing and Davis: Essays honoring Hallowell Davis*. St. Louis, Mo: Washington University Press, 1976.

KUHL, P. K., & MILLER, J. D. Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 1975, **190**, 69-72.

LEHISTE, I. *Suprasegmentals*. Cambridge Mass: MIT Press, 1970.

LIEBERMAN, P. Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, 1960, **32**, 451-454.

McCLEAN, M. D., & TIFFANY, W. R. The acoustic parameters of stress in relation to syllable position, speech loudness and rate. *Language and Speech*, 1973, **16**, 283-290.

MILLER, C., MORSE, P. A., & DORMAN, M. F. Cardiac indices of infant speech perception: Orienting and burst discrimination. *Quarterly Journal of Experimental Psychology*, 1977, **29**, 533-545.

MILLER, W. R., & ERVIN, S. The development of grammar in child language. In U. Bellugi & R. Brown (Eds.), The acquisition of language. *Monographs of the Society for Research in Child Development*, 1964, **29**, 9-35.

MORSE, P. A. The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology*, 1972, **14**, 477-492.

MORTON, J., & JASSEM, W. Acoustic correlates of stress. *Language and Speech*, 1965, **8**, 159-181.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

SIQUELAND, E. R., & DELUCIA, C. A. Visual reinforcement of non-nutritive sucking in human infants. *Science*, 1969, **165**, 1144-1146.

SPRING, D. R., & DALE, P. S. Discrimination of linguistic stress in early infancy. *Journal of Speech and Hearing Research*, 1977, **20**, 224-232.

TREHUB, S. E. The discrimination of foreign speech contrasts by infants and adults. *Child Development*, 1976, **47**, 466-472.

NOTES

1. It should be noted that due to the nature of our equipment it was sometimes possible for an infant to receive a stimulus in the middle of a bisyllable. However, given the nature of the infants' typical sucking patterns, which usually consist of short bursts of responses, the vast majority of stimulus presentations were uninterrupted. In any event, the interrupted presentations were equally likely to occur in control and experimental sessions.

2. Sandra Trehub (Note 1) has also recently reported data on infants' perception of multisyllabic contrasts. She found that infants perceive a voicing contrast in the medial position of a bisyllabic utterance provided that each syllable has a duration of 500 msec. When the syllable durations were shortened to 300 msec, Trehub failed to find reliable evidence for the discrimination of either bisyllables or trisyllables differing in voicing. She has explained her results in terms of temporal constraints on the information processing capacities of young infants. However, given that infants in the present study did discriminate between unstressed second syllables with durations of only 208 msec, such temporal constraints, if they do exist, would not appear to be a general limitation upon the infant's capacity to process all speech sounds.