Acoustic Correlates of Stress in Young Children’s Speech

Margaret Kehoe
Carol Stoel-Gammon
Eugene H. Buder
University of Washington
Seattle

This study examined the acoustic correlates of stress in children’s productions of familiar words. Previous research has employed experimental words rather than familiar words to examine children’s phonetic marking of stress, or has not adequately controlled for phonetic environment. Subjects in this study included 22 children, aged 18–30 months, and 6 adults. Fundamental frequency, duration, and amplitude measures were extracted from stressed and unstressed syllables in two types of comparisons: one that controlled phonetic environment and syllable position (interword) and one that measured the relative effects of stress within the same word (intraword). When the tokens were analyzed on the basis of target stress pattern, results revealed no differences between adults and children in their acoustic marking of stress. Listener judgments showed that approximately 30% of children’s two-syllable productions were coded unreliably or were perceived as inaccurately stressed. Overall findings indicate that children control fundamental frequency, amplitude, and duration to derive perceptually identifiable stress contrasts in the majority of their productions but they are not completely adult-like in their marking of stress.

KEY WORDS: phonological acquisition, word stress, acoustic correlates

When acquiring language, children must learn not only the segmental but the suprasegmental features of their language, including linguistic stress. Recent literature has explored children’s perception and production of stress, yet controlled acoustic phonetic data on children’s use of word stress remain limited. Perception studies show that prelinguistic infants are sensitive to aspects of stress in the acoustic signal (Jusczyk & Thompson, 1978; Kuhl & Miller, 1975; Spring & Dale, 1977) and show a selective preference for the stress pattern of their native language by 9 months of age (Jusczyk, Cutler, & Redanz, 1993). Production studies suggest that children’s early utterances are influenced by a rhythmic constraint. Children have difficulty producing utterances that do not conform to a strong-weak, or trochaic, metrical pattern (Allen & Hawkins, 1980; Gerken, 1990; Gerken, Landau, & Remez, 1990).

One area of stress development that has been relatively unexplored is children’s acquisition of the phonetic features of stress. In order to study stress development, it is necessary to define clearly what is meant by “stress.” Typically, stress refers to the prominence given to a syllable in a word or to a word in a sentence. It is considered a mental phenomenon (Hayes, 1991) but is strongly associated with a number of acoustic features. The results of acoustic and perceptual experimentation reveal that the stressed syllable in adult speech is marked by increased magnitude of fundamental frequency (f0), increased syllable length, increased amplitude, and changes in the quality of the vowel (Fry, 1955, 1958; Lehiste, 1970; Lieberman, 1960, 1967). These acoustic features give rise to the perceptual parameters of pitch, duration, and loudness, respectively. Studies that have examined the relative strength of these correlates show that pitch change is the most important factor in...
stress perception and outweighs duration and loudness effects (Fry, 1955, 1958; Morton & Jassem, 1965).

Despite an array of studies with adults, there are few reports on the acoustic correlates of stress in young children. Yet it is possible that children employ phonetic properties that are different from those of adults in marking stress. Maturational changes in speech production abilities may result in earlier control over some phonetic parameters before others. Conversely, children's use of stress may be adult-like from the earliest stages of acquisition and imply all acoustic features. Most studies of children's acquisition of stress have employed perceptual judgments without acoustic analysis. Two notable exceptions are the works of Allen and Hawkins (1980) and, more recently, Pollock, Brammer, and Hagerman (1993). One of the main claims of Allen and Hawkins (1980) is that children's early multisyllabic utterances contain a high proportion of unreduced syllables, giving the impression that the speech is syllable-timed. Learning to stress in English, according to Allen & Hawkins (1980), involves learning to reduce or shorten weak syllables. This suggests that duration as a parameter is not adequately controlled in children's early marking of stress.

Allen and Hawkins (1980) conducted acoustic analyses of 50 utterances of three children, approximately 3 years of age. The utterances were all two or more words in length and were extracted from natural speech samples. They measured duration and fo values from accented and unaccented syllables in different phrase positions and found that the marking of stress in nonfinal position was cued primarily by pitch change, whereas the marking of stress in final position was cued primarily by duration (see also Allen & Hawkins, 1978).

In contrast to Allen and Hawkins (1980), who analyzed tokens from spontaneous speech, Pollock and colleagues (1993) examined the acoustic parameters of stress in 2-, 3-, and 4-year-old children's imitated productions of nonsense words. Their stimuli were two-syllable words that differed in accent. The researchers found that stress pattern of the adult model was presumed to be the pattern attempted by the child. The researchers found that 3- and 4-year-old children use higher fo, greater intensity, and longer duration to mark stressed syllables. In contrast, 2-year-old children use a single cue, longer duration. Examination of absolute duration across age showed that duration of the stressed vowel remained relatively constant, but that duration of the unstressed vowel decreased significantly over time. This supports Allen and Hawkins' (1980) claim that stress development involves learning to reduce the unstressed syllable.

Pollock et al. (1993) also conducted a second analysis based only on those tokens for which at least four out of five judges agreed on stress placement. This resulted in 67% of 2-year-old tokens being used in the second analysis. The perceptual judgments showed that only 55% (of the 67%) of the 2-year-old productions conformed to the target stress placement. Acoustic measures in this subsample revealed that the 2-year-olds also used increased fo and intensity, as well as duration, on syllables perceived as stressed. On the 33% of tokens for which high listener agreement was not reached, acoustic analysis indicated minimal differences in fo and intensity between target stressed and unstressed syllables. The authors conclude that 2-year-olds do have the production ability to use all acoustic parameters but they often display inaccurate stress placement.

The results of these two sets of investigators offer somewhat conflicting impressions on children's stress development. The claim of Allen and Hawkins (1980) that the speech of children is syllable-timed suggests that duration is not used to mark the difference between stressed and unstressed syllables, although their acoustic findings suggest that duration is a significant cue in phrase-final position. Pollock et al. (1993) find that 2-year-olds do not differentiate stressed and unstressed syllables by fo and intensity, but by duration only, whereas 3- and 4-year-olds use fo, duration, and intensity to mark stressed syllables. In sum, the literature findings remain equivocal on the acoustic correlates of stress in young children.

Furthermore, the research to date suffers from additional factors such as lack of control over the phonetic environment and the use of experimental words rather than naturalistic tokens to measure stress. Allen and Hawkins (1980) compared the accented syllable in final position with the accented syllable in nonfinal position but did not take into account the fact that syllable duration in final position will be naturally longer due to phrase-final lengthening effects. Final lengthening has been observed in the speech of English-speaking children by 2 years of age (Smith, 1978; Snow, 1992, 1994). In addition, Allen and Hawkins (1980) provide no information on the phonetic content of their accented and unaccented syllables. Yet the works of a number of researchers suggest that it is necessary to separate phonetic effects due to stress from other phonological phenomena (see Klatt, 1975, 1976; Lea, 1977; Lehiste, 1970; Oiller, 1973, in adult speech; Smith, 1978, in children's speech).

Pollock et al. (1993) control the effects of phonetic environment by employing nonsense words to measure phonetic correlates (e.g., [b5fi] and [bafi]). Although experimental paradigms offer clear advantages in terms of stimulus control and reduction of confounding variables, they are more questionable in terms of generalizability. Children may control the phonetic correlates of stress sooner in productions of familiar words than in imitated productions of nonsense words. For example, Hochberg (1988b) found that Spanish-speaking children acquire the rules of stress earlier in their spontaneous productions than in imitated productions of nonsense words. The same effect may be present in the acoustic marking of stress.

---

1Allen and Hawkins (1980) hint at the possibility of a difference in developmental schedule between control of fo and duration, citing a study by Kirk (1973) that suggests tonal contrasts are acquired earlier than rhythmic (durational) contrasts.

2Their conclusions are somewhat circular. Since the perception of stress is cued by the parameters of fo change, duration, and amplitude, it is not surprising that these factors are implicated in the second analysis based on listener judgment.
The present study fills a gap in current research by examining the acoustic correlates of stress in children's familiar word productions while controlling for the effects of phonetic environment. In order to adequately measure stress in young children's productions, two additional issues must be considered: (a) stimulus selection and (b) variability of stress placement in production.

**Stimulus Selection**

The optimal selection of stimuli to measure stress contrast in young children creates some methodological difficulties. Noun-verb pairs that differ only in stress placement (e.g., *Record* [noun] and *reCORD* [verb]) are not typically within the vocabularies of young children. (Throughout the manuscript, stressed syllables will be indicated in uppercase italics and unstressed syllables in lowercase italics.) Disyllabic words in the vocabularies of young children (e.g., *BABy, MONkey, shamPOO, giRAFFE*) do not have minimal pairs differing in stress pattern. That is, there are no word pairs, such as *BABy and baBY*, that are identical except for stress pattern.

There are two potential solutions to this problem, both of which have advantages and disadvantages. One is to conduct an interword comparison using words such as *MONkey and KEY*. In this situation, the syllable /ki/ constitutes the unstressed syllable of the disyllable *MONkey* and the stressed syllable of the monosyllable *KEY*. The advantage of this comparison is that it achieves close phonetic matching and control over phrase position, since both syllables can occur phrase finally. The disadvantage is that syllable number is uncontrolled; the comparison consists of a monosyllable and a disyllable. It is well-established in adult speech that syllables in longer words tend to be shorter than the same syllables in shorter words, an effect known as "compression" (Lehiste, 1970). Therefore, in this comparison the unstressed syllable may be naturally shorter because it occurs within a disyllable.

Another solution is to conduct an intraword comparison using the stressed and unstressed syllable within the same word, such as the syllables *MON* and *key* in the word *MONkey*. The advantage of an intraword comparison is that it captures the "relative notion" of stress; the stressed and unstressed syllable create the relationship of prominence. However, an intraword comparison cannot achieve control over the effects of phrase position and inherent segmental factors. In single-word utterances and at the end of phrases, the unstressed syllable *key* will occur phrase-finally and thus be subject to phrase-final lengthening effects, whereas the stressed syllable *MON* will be less subject to these effects. In addition, control over segmental factors will be compromised.

The preceding discussion highlights the difficulties associated with the selection of stimuli to analyze stress in children's productions. Because it is not possible to control for all confounding factors, both an interword and intraword comparison are adopted in this study. The combined results will provide useful information on what acoustic features children use to mark stress.

**Variability of Stress Placement**

Accuracy of production also needs to be considered when analyzing stress in children. In adult studies, we can generally assume that the stress pattern produced conforms to the target stress pattern. We cannot necessarily assume this with children's speech in which stress placement is more variable (Hochberg, 1988a; Klein, 1984; Pollock et al., 1993). Inaccurate or variable stress placement may reflect lack of control over the phonetic features of stress, or incorrect application of the lexical stress rule. For example, the word *MONkey* may be perceived as *monKEY* or *MONkey* because the child hasn't sufficiently reduced the duration of the unstressed syllable to create a stress contrast, suggesting lack of phonetic control. Conversely, the child may not have learned the lexical stress rule of English, namely that two-syllable words have initial stress.

Teasing out these alternatives requires coordinated perception and production studies using nonsense syllables (of the *TAki vs. taKI* variety, see Allen & Hawkins, 1980) or languages other than English, where minimal pairs distinguishable only by stress exist. Such studies should determine whether children can appropriately use linguistically meaningful stress pattern differences. In the case of the present study, the apparent occurrence of stress errors in children's early productions suggests that listener judgment should also be included in an acoustic study of stress acquisition. Variability and changes in phonetic correlates with age may be associated with changes in stress accuracy.

The present study was designed to address some of the unresolved issues in the acquisition of stress by examining the acoustic correlates of stress in young children's spontaneous productions. Because previous research suggests that by age 3–4 years, children mark differences in stress with all acoustic features, a younger age group was targeted. The age range selected, 18–30 months, represents a period when stress contrast is first being produced. Analysis was also conducted on adult productions to provide information on the differences between adults and children in the marking of stress. The basic questions examined are: (a) whether children at their earliest stages of word production differentiate stress on the basis of f0, duration, and amplitude factors; and (b) whether there are age-related differences in the marking of stress.

**Method**

**Subjects**

Subjects were 22 children, six 18-month-olds, eight 24-month-olds, and eight 30-month-olds, and six female adults. Children were recruited through an experimental subject pool associated with the University of Washington and...
formed a subgroup of a larger research program, the main focus of which was to study phonological development. An attempt was made to analyze the speech samples of eight children at each age range; however, only six 18-month-old children produced enough suitable tokens to be included in the study. All children were within 1 week of their age range. At each age range, an equal number of male and female subjects was tested. Communication development was judged to be normal for all children on the basis of the MacArthur Communicative Developmental Inventory: Toddlers (Fenson et al., 1991). The 6 adult subjects were graduate students and faculty members at the University of Washington (age 22–40 years).

Data Collection

All recordings were conducted in acoustically treated rooms. Children were recorded in two 30–40 minute sessions while they interacted with a parent and occasionally, an experimenter. Toys, books, and stimulus pictures were placed in the experimental room to aid elicitation of certain target words. Parents were told that spontaneous productions were preferred over imitated productions; however, if children did not produce words spontaneously, productions could be obtained through elicitation or imitation techniques.

All sessions were videotaped using a camera situated in an adjacent test room. Both the parent and child wore vests containing small microphones (Countryman electret condenser microphones with flat frequency responses) connected to FM transmitters (HME TX822 transmitter linked to HME RX722 receiver for the child; Telex WT-50 transmitter linked to Telex FMR-50 receiver for the parent). The audio signals picked up from these receivers were recorded onto separate HiFi audio channels of the videotape simultaneously with the video signal.

Adult subjects were tested with the same recording equipment as the child subjects. They were asked to read 12 sentences in which target words were embedded in sentence-final position. They were asked to say the sentences as naturally as possible and to produce the last word in the sentence with the greatest amount of prominence. If the target word was not produced with sufficient prominence, subjects were requested to read individual sentences again. Sentences were chosen over isolated words because speakers often adopt a rehearsed manner when reciting lists of words. In addition, sentences were considered the most natural context to record stress, second to conversational speech, thus trying to simulate the same conditions as for the children. The sentence list read by all the adults is presented in the Appendix.

Data Base

Words selected for analysis included phonetically matched pairs of disyllables and monosyllables from the children's speech samples. The majority of words were spontaneous productions (73%), although a proportion of words were also spontaneous or elicited imitations (27%). The proportions of spontaneous and imitated productions were similar across age ranges.

All target disyllabic words contained a trochaic stress pattern (a stressed syllable followed by an unstressed syllable), the most frequently occurring stress pattern of two-syllable words in English (Cutler & Carter, 1987) and the dominant rhythmic pattern of children's early vocabularies (Allen & Hawkins, 1979). Examples of monosyllabic and disyllabic pairs include MONkey and KEY, PUPpy and PEa, BUunny and KNEE, and Tissue and SHOE. When possible, eight word-pairs were selected for acoustic analysis from each child. The monosyllable and the unstressed syllable of the disyllable were matched according to the following criteria:

1. Emotional level: All "potential" tokens were screened by the first author in terms of emotional level. Three categories were employed to code lexical items: (a) low, (b) normal, (c) high (see Snow, 1992). "Low" items were typically produced with reduced volume, often whispered, and conveyed a sense of shyness or hesitation. "High" items were produced with a great deal of emphasis, often shouted, and conveyed a high arousal level. Words from "low" and "high" categories were not considered representative items for normal stress parameters and were excluded from analysis. These constituted only a small proportion of the total lexical items.

2. Phrase position: Only words said in isolation or in phrase-final position were included in the analysis. These two utterance positions should be similar in terms of duration effects (Kubaska & Keating, 1981; Snow, 1992).

3. Segmental effects: Where possible, tokens were matched for segmental content (e.g., the /ki/ of MONkey matched with KEY). Matching was based on the phonetic form used by the child rather than the target adult form. Therefore, the production of [bi] for BEE could be matched with the production of [pabj] for PUPpy since both onset consonants were voiceless. If phonetic matching was not possible, the following intrinsic and extrinsic phonetic factors were considered:

- Intrinsic vowel quality. High vowels constituted the main vowel type used in the analysis. Their predominance was the result of the high proportion of two-syllable forms such as [phAph i] and [khith ij] observed for puppy and kitty respectively; (c) Mean f0 values, found in stressed syllables, reflect effects due to aspiration only.

*Although the term "phonetic matching" is used, the matching was essentially phonemic since there will be differences in the degree of aspiration for voiceless stop consonants due to the effects of stress. That is, there will typically be greater aspiration in the /k/ of KEY than MONkey, thus resulting in a lack of phonetic equivalence f0 values. Although aspiration may influence f0 values, there are a number of reasons to suggest that the effect of aspiration did not result in the f0 differences observed between stressed and unstressed syllables in this study: (a) Only 25% of the data set contained pairs with unvoiced-initial stops subject to aspiration; (b) A re-examination of this subset suggested that not all "unstressed" syllables were unaspirated. For example, forms such as [phAph i] and [khith ij] were commonly observed for puppy and kitty respectively; (c) Mean f0, which would be less subject to aspiration effects than maximum f0, displayed similar trends to maximum f0. In sum, it is unlikely that the increased f0 values, found in stressed syllables, reflect effects due to aspiration only.

*
children’s words in English that end in the phoneme /i/ (e.g., MONkey, COOkie, BAby, and KItty). If vowel matches could not be obtained, matching of the vocalic nucleus was possible with any of the high or mid tense vowel set /i, e, u, o/. Therefore the word TWO could be matched with the word KItty since the vowel in TWO and the vowel in the unstressed syllable of KItty are high tense vowels.

- Manner and voicing of the prevocalic consonant. The prevocalic consonant was matched in terms of voicing and general manner of articulation categories because both of these factors may influence intensity and f0 measures (Lea, 1977). Sonorant-initial syllables (nasals, liquids) were always matched and obstruent-initial syllables were always matched together.

- Presence of the postvocalic consonant. All open syllables were matched with open syllables and closed syllables were matched with closed syllables. The voicing and manner of articulation of the final consonant were also considered since these features influence the duration of the preceding vowel (Klatt, 1976). In the majority of cases, children’s use of the plural morpheme allowed for identical phonetic matches for closed syllables (e.g., BAbies and BEES, PUppies and PEAS).

Other requirements of stimulus selection were: (a) Tokens had to be free of background noise that may prevent accurate acoustic measurement; (b) Tokens had to receive the nuclear accent (the greatest prominence in the phrase) if part of a multiword utterance; (c) Tokens had to be part of an utterance that functioned pragmatically as a statement since a rising intonational contour, typical of a question or request, may result in the unstressed syllable having a greater maximum f0 than the stressed syllable; and (d) Tokens of the words MOmmy and DAddy were included only when used in a referential sense because productions that functioned as vocatives were often shrill and exaggerated, making them inappropriate for stress analysis.

As a result of the word selection procedure, eight word-pairs were selected for each child matched in terms of utterance position, emotional level and phonetic content. Because word selection was dependent on an individual child’s speech sample, different sets of word-pairs were selected for each child. Nevertheless, the selection of stimuli and assignment to word-pairs was based on strict criteria and performed similarly across children.

The stimulus set for the adults consisted of six rather than eight word pairs because less variability was anticipated with the adult measures than the child measures. The word-pairs included BAbies-BEES, DAddy-DEE, MOmmy-ME, MONkeys-KEYS, Tissue-SHOE, and MOney-KNEE. The stimulus set for the adults remained constant across subjects.

Missing data included 12 pairs out of a total possible of 212 pairs (5.6% of the data). Where possible, disyllables were included in the analysis even when a suitable monosyllabic match could not be found, thus resulting in fewer missing items for the intraword comparison.

Acoustic Analysis

Acoustic analysis of stimuli was performed using CSpeech Version 4.0 (Milenkovic & Read, 1992), a PC-based speech signal processing package. Target words were digitized at a sampling rate of 22,000 Hz with 15 bit resolution and automatic anti-aliasing performed by the CSpeech system. The total duration, f0 contour, and RMS (root mean square) amplitude contour of each stressed and unstressed syllable were measured. The domain of measurement included the “vocalic nucleus” in which voicing was detectable. This included the vowel and any preceding sonorant; it did not include a preceding obstruent, but would typically include voicing into a following obstruent.

The onset and offset of each syllable was considered to be the first or last detectable periodic cycle in the time waveform. Inspection of the time waveform combined with spectrographic display and auditory judgment usually were sufficient for boundary identification. The CSpeech display of the time waveform, spectrogram and pitch contour of a representative token is shown in Figure 1. Criteria for duration measurement included very low amplitude periodicity and double period pulses seen in fry register. These phonatory waveforms are often encountered in children’s voices, especially in open syllable codas. Damped, relatively aperiodic cycles or isolated pulses following clear breaks in periodicity were not included. When voicing was continuous across syllable boundaries, as in the case of sonorant-initial syllables (e.g., MOney), the intersyllabic boundary was placed at the completion of oral cavity closure (e.g., as in the reduced energy and altered formant structure occurring at the point of oral tract closure for nasals).

Pitch contours were tracked by the CSpeech “pitch” command that invokes a center-clipped autocorrelation algorithm. The analysis set-up allows manipulation of analysis range, window, update (determining the amount of time shift between respective data windows) and parameters related to voicing detection. In this study, the analysis parameters were varied on an individual basis for each token to produce an accurate pitch contour as validated against visual inspection and measurement of glottal periods in the waveform. In general, the analysis parameters that served the majority of tokens consisted of Analysis Range: 100–500 Hz and Analysis Window: 10/15 msec. In all cases, the analysis update was set to 10 msec so that pitch contours for each token were tracked with the same f0 data point intervals.

At the analysis stage, some tokens, characterized by significant f0 breaks or unusual vocal registers (e.g., excessive fry), were excluded because they posed difficulty for acoustic analysis. The f0 changes that result from these vocal phenomena were not considered linguistic manifestations of stress, justifying the exclusion of these tokens. These tokens were replaced with additional tokens using the selection criteria described above. CSpeech also offers

The majority of disyllabic words ended in the phoneme /i/. The exceptions included the word TISSUE, which was spoken by a number of children on a number of occasions, and the word HIPPO, which was spoken by one child on two occasions.
FIGURE 1. CSpeech display of the word *PUuppy* spoken by 24-month-old female child. The figure shows time waveform, spectrogram (Frequency Range: 7.5 KHz; Analysis Bandwidth: 450 Hz), and pitch contour.

Editing facilities for the $f_0$ waveform. These facilities (zeroing and interpolating functions) were used judiciously when minor voice perturbations interfered with the $f_0$ contour.

An RMS amplitude envelope was obtained for the stressed and unstressed syllable of each disyllable, using the syllable boundaries determined through previous duration measures. Amplitude values were subject only to intra-word analysis since amplitude variations unrelated to stress (e.g., recording artifacts, emotional level) may influence interword comparisons.

The Analysis Measure function of the CSpeech software computed summary statistics of $f_0$ and amplitude contours for each syllable measured. Mean $f_0$, maximum $f_0$, and $f_0$ range were extracted from each pitch contour. A single average dB value was extracted from the RMS amplitude contour. These values, along with the onset and offset time measures for each syllable, were entered into a statistical analysis program (SYSTAT for Windows, Version 5, 1992).

Because criteria for duration measurement can sometimes be somewhat subjective, and because durations formed the basis for $f_0$ and amplitude calculations, intra- and inter-judge reliabilities were measured for subsets of the duration measurements. Intra-judge remeasures for 40 tokens produced a Pearson $r$ of .99 and a mean difference of 2.8 msec (SD = 9.0 msec). Inter-judge remeasures for 36 tokens produced a Pearson $r$ of .98 and a mean difference of 5.5 msec (SD = 25.7 msec). The duration differences associated with the reliability statistics were much smaller than the subsequent duration differences of interest.

**Listening Test**

In order to obtain perceptual judgments of stress placement, all children's productions of two-syllable words ($N = 173$) were transferred to an audio cassette tape and randomized with respect to age and subject. Each of the 173 words was presented three times consecutively on the tape with a 2-second interval between repetitions, to allow the listeners sufficient time to identify the stress contour.

Five monolingual American students who were currently taking or had taken a course in advanced phonetic transcription and had experience in the transcription of stress participated as listeners in the listening task. All listeners performed a short practice run prior to the main listening test to receive exposure to both the stimuli and the procedure. In the listening task, they were given a response form that contained a gloss of the 173 words in the order of presentation on the tape. Judges were asked to indicate whether they heard the first syllable as stressed, the second syllable as stressed, or both syllables as equally stressed. After hearing the three repetitions of each word, they placed a stress diacritic above the appropriate syllable(s) on the response form. The listening task lasted approximately 30 minutes. Subjects were not permitted to rewind the tape during the experiment or re-listen to any token, but they were encouraged to take a short break if they experienced any fatigue as a result of the listening task.

Tokens were coded into four main categories on the basis of listener judgments: correct, misplaced stress, level stress,
and unreliable. Following Pollock et al. (1993), tokens judged to have the adult stress pattern (i.e., stress on the first syllable) by four out of the five judges were coded as correct. Tokens judged to have stress on the second syllable or stress equally on both syllables by four out of the five judges were coded as mismatch or level stress, respectively. Tokens for which fewer than four out of five judges agreed on stress placement were coded as unreliable.

The three authors listened to a tape of the adult disyllabic words. They were unanimous in their judgment that all tokens were produced with correct stress placement.

**Results**

The dependent variables extracted for analysis were maximum f₀ (Hz), total duration of the syllable (msec), and RMS amplitude (dB). Mean f₀ and f₀ range were also extracted from each syllable, but because they displayed similar trends to those seen in maximum f₀, only maximum f₀ will be reported. Results will be discussed separately for f₀, duration, and amplitude measures. For the initial analysis, the stress pattern of the adult target was assumed to be the stress pattern attempted by the child.

**F₀ Measures**

The means and standard deviations of pitch measures as a function of age are shown in Table 1. Stress1 refers to the monosyllable; Stress2 refers to the stressed syllable in the disyllable; and Unstress refers to the unstressed syllable in the disyllable. Comparison of Stress1 and the unstressed syllable constitutes the interword comparison and comparison of Stress2 and the unstressed syllable constitutes the intraword comparison.

In the interword comparison, the f₀ contour of the monosyllable is contrasted with its unstressed phonetically matched counterpart. The maximum f₀ should be greater in the stressed syllable than in its unstressed counterpart if children are marking a stress contrast. The summary data in Table 1 support this trend at all age ranges. The difference in maximum f₀ was greater in the stressed syllable than in the unstressed syllable. The magnitude of the difference was larger than in the interword comparison: 52 Hz at 18 months, 67 Hz at 24 months, and 66 Hz at 30 months.

The means and standard deviations of absolute f₀ values were considerably reduced in the adult data compared to the child data. However, similar relative trends were observed. The difference in maximum f₀ between stressed and unstressed syllables was 37 Hz in the interword comparison and 23 Hz in the intraword comparison. One interesting difference between children and adults was that children marked disyllables with greater f₀ accentuation than monosyllables; the reverse was true for the adults.

Univariate Analyses of Variance (ANOVAs) were conducted on the f₀ difference scores for interword and intraword comparisons. The difference score represents the difference in value between the stressed and unstressed syllable for each matched pair. Two tests were conducted, one for the main effect of stress and one for the main effect of age. The variable of subject was nested within age, because preliminary investigation revealed that some of the variability in difference scores was related to individual subject differences rather than age differences. Results indicated no significant age-related effects: Interword Comparison: F(3,24) = .34; p = .80; Intraword Comparison: F(3,24) = 1.10; p = .37. As there were no significant age effects, the main effect of stress was performed with all ages combined. The main effect of stress was significant for both comparisons: Interword Comparison: F(1,24) = 32.92; p < .001; Intraword Comparison: F(1,24) = 34.78; p < .001.

**Duration Measures**

The means and standard deviations of syllable duration as a function of age are shown in Table 2. In the interword comparison (Stress1 vs. Unstress), the mean duration of the monosyllable was greater than its unstressed phonetically matched counterpart at all age ranges. The magnitude of the difference was 99 msec at 18 months, 99 msec at 24 months, and 89 msec at 30 months. This resulted in a ratio between the unstressed and stressed syllable of approximately .75 across the three age ranges. The intraword comparison (Stress2 vs. Unstress) does not serve as an appropriate vehicle for comparing duration; phrase-final lengthening effects and possibly other uncontrolled seg-

<table>
<thead>
<tr>
<th>Stress1</th>
<th>Stress2</th>
<th>Unstress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>18 mo</td>
<td>388 (78)</td>
<td>44</td>
</tr>
<tr>
<td>24 mo</td>
<td>377 (118)</td>
<td>57</td>
</tr>
<tr>
<td>30 mo</td>
<td>360 (80)</td>
<td>63</td>
</tr>
<tr>
<td>Adults</td>
<td>223 (44)</td>
<td>36</td>
</tr>
</tbody>
</table>

Note. Stress1: monosyllable; Stress2: stressed syllable of the disyllable; Unstress: unstressed syllable of the disyllable.
mental factors result in the duration of the unstressed syllable being greater than the stressed syllable in the child data. This underscores the importance of controlling additional phonetic factors when examining stress correlates.

Again, the adult interword data show trends similar to the child data. The mean syllable duration was greater in the monosyllable than in its unstressed counterpart. The absolute difference (157 msec) was greater than in the child sample and the ratio of the unstressed to the stressed syllable was smaller: .51 compared to .75. The difference in ratio is principally due to the increased length of the unstressed syllable in the child data. By 30 months of age, the children’s duration values are 7% greater than the adult values for Stress1, 16% greater for Stress2, but 55% greater for unstressed syllables. The respective values for the 18 and 24 month olds were 27% and 27% greater for Stress1, 22% and 39% greater for Stress2, and 88% and 88% greater for Unstress.

An ANOVA conducted on the duration difference scores for the interword comparison showed no significant age effect \( [F(3,24) = .94; p = .44] \). Thus, the increased duration difference score seen for the adult subjects was not shown to be statistically significant. A significant main effect of stress was obtained with all age groups combined \( [F(1,24) = 59.55; p < .001] \). Duration values for the intraword comparison were not subject to statistical analysis due to the confounding effect of phrase position.

### Amplitude Measures

The means and standard deviations of relative amplitude measures are indicated in Table 3. These values represent the difference in RMS amplitude (dB) between the stressed and unstressed syllables in the intraword comparison. The results show that the amplitude of the stressed syllable is greater than the unstressed syllable at all age ranges with a tendency for greater contrast with increasing age.

An ANOVA conducted on the amplitude difference scores showed no significant age effect \( [F(3,24) = 2.91; p = .06] \).

### Individual Differences

The small sets of tokens obtained per child, along with the inherent variability of young children’s speech, contribute to the large variance seen in the children’s data (see SDs in Table 1–3). However, individual differences in the use of stress can be examined for systematic factors contributing to this variability. Individual subjects’ patterns of phonetic parameter usage were plotted to explore for graphic age-related trends or phonetic trading relationships. Figure 2 presents the maximum \( f_0 \) and duration difference scores for the interword comparison by individual subject. Intensity measures were not obtained in the intraword comparison. Figure 3 presents the maximum \( f_0 \) and intensity difference scores for the intraword comparison by individual subject. Duration scores are not shown due to the confounding effect of phrase position.

The results indicate that there are considerable differences among individual subjects in how they utilize phonetic parameters. There is some suggestion of age-related trends for the child subjects, although these effects are not consistent across all phonetic parameters. For example, in the interword comparison, two children at 18 months and two children at 24 months display negative or minimal \( f_0 \) difference scores, in contrast to no children at 30 months. In the intraword comparison, one child at 18 months and one child at 24 months display minimal \( f_0 \) difference scores, in contrast to no children at 30 months. However, these effects are not observed for the duration difference score in the interword comparison where one to two children at each age range show minimal difference differences.

There also is some indication that some subjects recruit certain phonetic parameters over others. For example, in the interword comparison, Subjects 2 and 3 differentiate the monosyllable and its unstressed counterpart by duration but not by \( f_0 \) accentuation, whereas Subject 19 uses \( f_0 \) accentuation but not duration. Interestingly, the same type of “trading” effects also are apparent with the adult subjects—Subject 24 does not appear to mark stress by \( f_0 \) in either the inter- or intraword comparisons but rather marks stress by duration and intensity. Though intriguing, these data must be interpreted with caution because they are based on a relatively limited data set (6–8 tokens) per subject. Overall, the individual subject results support the group findings; that is, there is considerable variability in how children mark
stress but they do not appear to be qualitatively different from adults.

**Listening Experiment**

Table 4 displays group results from the listening study. Seventy-one percent of tokens spoken by the 18-month-old group, 69% of tokens spoken by the 24-month-old group and 77% of tokens spoken by the 30-month-old group were perceived as having correct stress placement. Of the remaining tokens, the majority were coded as unreliable; that is, listeners did not agree on stress placement. Only 11% of tokens produced by 18-month-olds, 5% of tokens produced by 24-month-olds, and 3% of tokens produced by 30-month-olds were consistently perceived as incorrectly stressed; that is, produced with level or misplaced stress. Some children produced a high proportion of unreliable or incorrectly stressed tokens: Two 30-month-olds, two 24-month-olds, and two 18-month-olds produced at least four of their eight tokens with unreliable or incorrect stress placement, accounting for over 50% of the incorrect tokens.

Results were examined in terms of the proportion of correct tokens produced per individual subject out of the total number of tokens (typically eight tokens). Individual proportions were arcsin transformed and subjected to a one-way ANOVA with age as the between-subject factor. There were no significant differences in stress accuracy across age [F(2,19) = .66; p = .53]. Separate analyses were conducted on the acoustic parameters of those tokens perceived as correctly stressed by at least four out of five judges and those tokens perceived as incorrectly stressed (level or misplaced) or unreliable. Incorrectly stressed and unreliable tokens were grouped together since there were insufficient numbers in each of these groups to analyze them separately. Table 5 presents the mean maximum f0 and duration difference scores for the interword comparison and the mean maximum f0, duration, and amplitude difference scores for the intraword comparisons according to listener judgments. Duration scores are reported for the intraword comparison because they offer some information on what parameters may have been influencing listener judgments. Some general tendencies can be observed, although caution must be exercised in interpreting these results due to the fewer numbers of tokens coded as incorrect or unreliable. For all acoustic parameters, a strong separation can be seen between values for tokens perceived as correct and tokens perceived as incorrect or unreliable. Tokens coded as incorrect were produced with very little f0 and duration differences between stressed and unstressed syllables in the interword condition. The only exception was in the duration comparison where 18-month-old children were still seen to show duration differences.

---

6 The grouping together of unreliable and incorrect tokens was considered justified, since for tokens to be coded as unreliable, two or three judges had to perceive them as incorrectly stressed. In acoustic terms, the unreliable tokens probably represent less extreme variants of the misplaced or level stress tokens.
differences between productions of the monosyllable and the unstressed syllable of the disyllable. In the intraword comparison, there was very little $f_0$ accentuation and intensity difference between the stressed and unstressed syllables on tokens coded as incorrect. The unstressed syllable was much longer in duration than the stressed syllable indicated by the negative duration difference scores.

To determine whether significant age effects were present when listener responses were taken into consideration, additional statistical tests were employed. Acoustic values for only those tokens perceived as correctly stressed were submitted to a second set of ANOVAs. All adult tokens had been judged to be correctly stressed. No significant age effects were present for maximum $f_0$ in the interword comparison: $F(3,24) = .83; p = .49$; maximum $f_0$ in the intraword comparison: $F(3,24) = 2.18; p = .12$; or duration in the interword comparison: $F(3,24) = .49; p = .70$. However, there was a significant age effect present for intensity: $F(3,24) = 3.00; p = .05$. This also was the parameter that

TABLE 4. Listener's coding of stress pattern in two-syllable words: Percentages of tokens coded as correct, level, misplaced, or unreliable.

<table>
<thead>
<tr>
<th></th>
<th>Correct</th>
<th>Level</th>
<th>Misplaced</th>
<th>Unreliable</th>
<th># of words</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 mo</td>
<td>71</td>
<td>2</td>
<td>9</td>
<td>18</td>
<td>45</td>
</tr>
<tr>
<td>24 mo</td>
<td>69</td>
<td>2</td>
<td>3</td>
<td>27</td>
<td>64</td>
</tr>
<tr>
<td>30 mo</td>
<td>77</td>
<td>2</td>
<td>3</td>
<td>20</td>
<td>64</td>
</tr>
</tbody>
</table>

TABLE 5. Means and standard deviations of maximum $f_0$, duration, and amplitude difference scores on the basis of listener judgments.

<table>
<thead>
<tr>
<th>Listener judgments</th>
<th>Correct stress</th>
<th>Incorrect stress</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Interword comparison</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_0$ measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 mo</td>
<td>46</td>
<td>(108)</td>
</tr>
<tr>
<td>24 mo</td>
<td>68</td>
<td>(92)</td>
</tr>
<tr>
<td>30 mo</td>
<td>65</td>
<td>(66)</td>
</tr>
<tr>
<td>Duration measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 mo</td>
<td>108</td>
<td>(195)</td>
</tr>
<tr>
<td>24 mo</td>
<td>126</td>
<td>(224)</td>
</tr>
<tr>
<td>30 mo</td>
<td>110</td>
<td>(129)</td>
</tr>
<tr>
<td>Intraord comparison</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_0$ measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 mo</td>
<td>83</td>
<td>(149)</td>
</tr>
<tr>
<td>24 mo</td>
<td>89</td>
<td>(109)</td>
</tr>
<tr>
<td>30 mo</td>
<td>82</td>
<td>(76)</td>
</tr>
<tr>
<td>Duration measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 mo</td>
<td>-34</td>
<td>(152)</td>
</tr>
<tr>
<td>24 mo</td>
<td>-7</td>
<td>(157)</td>
</tr>
<tr>
<td>30 mo</td>
<td>-3</td>
<td>(129)</td>
</tr>
<tr>
<td>Amplitude measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 mo</td>
<td>4.2</td>
<td>(7.7)</td>
</tr>
<tr>
<td>24 mo</td>
<td>5.9</td>
<td>(4.4)</td>
</tr>
<tr>
<td>30 mo</td>
<td>7.1</td>
<td>(5.0)</td>
</tr>
</tbody>
</table>
approached significance in the first analysis. Post hoc comparisons (Tukey HSD) revealed that 18-month-old children used amplitude to mark stress significantly less than 30-month-old children and adults.

Discussion

The major finding of this study is that children between 18 and 30 months of age mark differences in stress by $f_0$, duration, and intensity parameters. When data were analyzed on the basis of the target stress contour, there were no age related effects for any of the acoustic variables examined, suggesting that both adults and children as young as 18 months function similarly in their use of these phonetic parameters. There were, however, significant main effects of stress, verifying that stressed and unstressed syllables are marked differently in terms of acoustic parameters.

Graphs of the findings (see Figures 2 & 3) do suggest some age-related effects that were not supported by statistical tests. Nesting of subjects within age combined with the variability of the children's data may have contributed to the reduced power in the F Tests and a lack of sensitivity to smaller effects. However, our conclusion that children use all three phonetic parameters to mark stress is borne out by the controlled nature of this study. Random factors that could otherwise statistically veil small age effects have been controlled in this study by strict sample selection criteria. Further calibration of the data, such as norming acoustic contrasts in terms of subject means, may reveal more subtle age-related findings.

This study did not support Pollock et al.'s (1993) finding that 2-year-old children employ duration only in marking stress. This study showed that children as young as 18 months differentiate stressed and unstressed syllables by all acoustic parameters, although there was some indication that the use of intensity to mark stress increased with age. The discrepancy in findings may reflect differences in methodology. This study employed children's productions of familiar words in contrast to Pollock et al.'s (1993) study, which involved imitated productions of nonsense words. Children's control over phonetic parameters may emerge sooner in natural speech than in imitated forms in an experimental task. This is consistent with Hochberg's (1988b) finding that children demonstrate mastery of the stress system earlier in spontaneous speech than in an imitation task. In addition, Pollock et al. (1993) examined stress marking in both trochaic and iambic disyllabic words, whereas this study focused only on stress marking in trochaic disyllabic words. Because the metrical pattern most frequently heard and produced by young children acquiring English is trochaic, it may be hypothesized that phonetic parameters are more adequately controlled in these types of words.

This study used two types of comparisons to determine stress marking: an interword and intraword comparison. Children tended to display greater $f_0$ differences between stressed and unstressed syllables in the intraword comparison than in the interword comparison. That is, they marked the stressed syllable of the disyllable with higher $f_0$ than the monosyllable. This may reflect a pattern of tone declination that makes earlier elements in the phrase of higher pitch than later elements. The same trend was not seen with the adult subjects, however, suggesting that children may particularly exploit $f_0$ to mark accentuation in two-syllable words.

In the interword comparison, all subjects produced the phonetically matched monosyllable with greater duration than the unstressed syllable of the disyllable. In the intraword comparison, children displayed longer mean durations for the unstressed syllable than the stressed syllable. It is interesting to hypothesize that the increased $f_0$ accentuation associated with the stressed syllable in children's two-syllable productions may be due to the fact that duration plays such an ambiguous role in these words. In order to mark prominence, children may need to exploit other parameters of stress, such as $f_0$, because they cannot control duration sufficiently to create stress contrast. In contrast, the adults in this study made less use of $f_0$ but greater use of intensity and duration to create stress contrast in their two-syllable productions.

The use of intensity to mark stress was only measured in the intraword comparison. All subjects marked the stressed syllable with higher overall intensity than the unstressed syllable. There was a tendency for the intensity difference to increase with age, a trend that was borne out in statistical analyses based on listener judgments. Intensity has generally been underplayed as a cue in stress perception, but the finding of age-related differences in intensity suggests that this parameter may deserve closer attention in stress measurement studies.

Interestingly, the cue most implicated in previous stress perception studies in English (Fry, 1955, 1958; Morton & Jassem, 1985), $f_0$, was the parameter that children appeared to control to the greatest extent. Children did utilize duration differences, but the unstressed syllable was still relatively long, supporting the assertion that children have difficulty reducing the length of the unstressed syllable (Allen & Hawkins, 1980; Pollock et al., 1993). Children also displayed amplitude differences but, as mentioned above, out of the three phonetic parameters, these were most susceptible to age effects.

The stimulus selection and matching criteria used in this study were designed to control for the additional acoustic variation that may result from phrase position and other phonetic factors. However, even given this control, variability was present in the acoustic measures. Some of the variability may be explained by individual subject differences. Across all age ranges, there was great diversity in how individual subjects employed different phonetic parameters. Some subjects employed certain acoustic parameters over others, that is, employed duration only rather than $f_0$, or employed $f_0$ only rather than duration. Such trading effects also were observed with the adult subjects, supporting the overall group findings that children perform similarly to adults in marking stress.

The results of the listening study indicated that approximately 70% of the children's productions were perceived as correctly stressed. The remaining productions were either not consistently coded in terms of stress placement (unre-
liable) or were perceived as incorrectly stressed (misplaced or level). As mentioned earlier, inaccurate stress placement may reflect: (a) insufficient control over phonetic parameters in marking stress, or (b) incorrect use of the lexical stress pattern. It seems unlikely that children in this study were experiencing difficulty with the lexical stress pattern since all disyllabic words had trochaic stress, the metrical form that is considered to be unmarked in the acquisition of English (Allen & Hawkins, 1980). It seems likely that listeners’ perceptions of misplaced and level stress reflect children’s inadequate control over phonetic parameters, an interpretation supported by the acoustic findings of this study. The tokens perceived as correct had different phonetic qualities than those perceived as incorrect. The former were characterized by a coming together of all three parameters: $f_0$, intensity, and duration, whereas the latter were characterized by increased duration, but minimal $f_0$ and intensity accentuation on the syllable perceived as stressed. One would predict that if misplaced stress reflected inaccurate rule usage, the same phonetic parameters would be employed for both correct and incorrect tokens but they would be applied to different syllables. At this point these conclusions are tentative; further research is needed to separate phonetic from phonological factors related to inaccurate stress placement in children.

The results of the listening study also indicated that some children displayed more incorrect forms than other children, a finding similar to that of Wijnen, Krikhaar, & den Os (1994) who note in their study of Dutch stress that stress shift appears to be an idiosyncratic quality of children’s speech. Presently, there is a need for research to address why certain subgroups of children display different types of prosodic behavior, such as misplaced stress, and whether this behavior is related to other areas of language development.

In sum, the results of this study suggest that children as young as 18 months are able to control the phonetic features of stress well enough to yield perceptually identifiable stress patterns in most of their productions. Nevertheless, children still display considerable variability in their productions and a proportion of their productions are perceived as incorrectly stressed. This study supports other speech production studies that suggest quite early phonological sophistication in children’s use of phonetic parameters (Smith, 1978). Despite this sophistication, however, development can be characterized as involving increased stability in production.

Acknowledgments

This study was supported by grants from NIDCD: DC00033 and DC00525. The authors wish to thank David Snow for his significant contribution toward an earlier version of this study; Marilyn Vihman, Bruce Smith, and two other anonymous reviewers for valuable comments on the manuscript.

References


Appendix

Text of sentences read by adult subjects

1. I was upset that they had chosen her when they could have chosen me.
2. I laughed and laughed, the funniest things were the monkeys.
3. The names I prefer like 'Dominique,' 'Decklan,' and 'David,' all start with the letter 'D.'
4. My sister had twins in December, they are beautiful babies.
5. The little girl started to cry, she clung to her mommy.
6. Close to the finishing line Mary fell, she bruised her knee.
7. Her husband is a bee-catcher, he is good at catching bees.
8. I had a very bad cold but I could not find a tissue.
9. I was very poor, I couldn't afford to give him any money.
10. Washed up on the beach, like a piece of driftwood, was an old shoe.
11. My father is forgetful, he is always losing his keys.
12. The best letter-writer in the world is my daddy.

Note. Some of the sentences the adults read were more than single sentences in the strictest sense. These variations did not appear to influence the adult’s speaking patterns.