Published in final edited form as:
J Acoust Soc Am. 2007 October ; 122(4): 2328–2339. doi:10.1121/1.2769624.

Dynamic spectral structure specifies vowels for children and adults[a]

Susan Nittrouer[b]
Speech and Hearing Science, Ohio State University, Columbus, Ohio 43210

Abstract

When it comes to making decisions regarding vowel quality, adults seem to weight dynamic syllable structure more strongly than static structure, although disagreement exists over the nature of the most relevant kind of dynamic structure: spectral change intrinsic to the vowel or structure arising from movements between consonant and vowel constrictions. Results have been even less clear regarding the signal components children use in making vowel judgments. In this experiment, listeners of four different ages (adults, and 3-, 5-, and 7-year-old children) were asked to label stimuli that sounded either like steady-state vowels or like CVC syllables which sometimes had middle sections masked by coughs. Four vowel contrasts were used, crossed for type (front/back or closed/open) and consonant context (strongly or only slightly constraining of vowel tongue position). All listeners recognized vowel quality with high levels of accuracy in all conditions, but children were disproportionately hampered by strong coarticulatory effects when only steady-state formants were available. Results clarified past studies, showing that dynamic structure is critical to vowel perception for all aged listeners, but particularly for young children, and that it is the dynamic structure arising from vocal-tract movement between consonant and vowel constrictions that is most important.

I. INTRODUCTION

The question of how listeners assign vowel labels to acoustic signals has always been more problematic than the question of how consonant labels are assigned. Not long after experimenters began developing tools to independently manipulate the various acoustic properties of speech signals, the phenomenon of categorical perception was noted. This phenomenon was originally reported for experiments conducted in the 1950s involving consonant labeling where all acoustic properties were held constant in stimuli, except for one that was varied in discrete steps of equal size across a continuum (Liberman *et al.*, 1957). Results of such experiments are typically marked by flat labeling functions along the regions of the continua where stimuli are heard as belonging to a single phonemic category, and by sharp functions at category boundaries. Discrimination results show poor performance for stimuli in the flat, within-category regions, and contrastively good performance for stimuli straddling category boundaries. Experiments with vowels, however, failed to replicate the effects observed for consonants. Instead, labeling for vowel-like stimuli with formant frequencies varying along linear continua showed gradual slopes for the entire length of the continua, and discrimination among stimuli was good for all stimulus comparisons (Fry *et al.*, 1962). So it appeared in those early days of experimental speech perception that sensory processing for signal portions largely associated with vowels differed from processing for consonant-related signal portions.

[a] Portions of this work presented at the 149th Meeting of the Acoustical Society of America, Vancouver, May, 2005.

[b] Electronic mail: nittrouer.1@osu.edu.

Although there may have been no generally agreed upon perspective of *how* listeners process vocalic signals, the commonly accepted wisdom about *what* they process has always been that the frequencies of the first two or three steady-state formants are used by listeners in making vowel judgments (e.g., Ferrand, 2007). But this position faces challenges at even the most basic level: Continuous speech rarely has stretches that could be called steady state, and so the first challenge faced by the above-offered position has to do with the notion of "steady-state" syllable segments. Even if we modify the theoretical position to state that listeners make vowel judgments based on "target" formant frequencies, meaning the extremes in formant frequencies associated with specific vowels, other challenges arise. For example, target formant frequencies for the same vowel vary greatly across speakers due to differences in vocal-tract size and geometry: Not only are there length differences, but the ratio of oral cavity length to pharyngeal cavity length is very different for men, women, and children (e.g., Fant, 1973). Of course, problems of this sort have been handled with several proposals of how listeners might normalize across acoustic variation to derive stable phonetic representations. To deal with the problem of speaker variability, for example, various versions of speaker normalization have been offered (e.g., Gerstman, 1968; Syrdal and Gopal, 1986), and all share the theme that listeners must adjust their expectations of formant-to-vowel relations based on an individual speaker's acoustic characteristics, which become apparent by listening to that speaker. Although tests of speaker normalization have met with modest success (e.g., Ladefoged and Broadbent, 1957), these proposals alone cannot account for another problem, and that is the variability associated with phonetic context. This problem is illustrated in Fig. 1 showing spectrograms of "dad," "dud," and "bab," produced by an adult, male speaker. It can be seen that target F2 is more similar for "dud" and "bab" than for "dad" and "bab," even though "dad" and "bab" are heard as containing the same vowel. To handle the variability that arises due to phonetic context, the idea was proposed that listeners normalize their expectations of what formant frequencies should be for each vowel based on that context (e.g., Lindblöm and Studdert-Kennedy, 1967). So, if we combine notions of speaker and context normalization we must propose that listeners apply both kinds of normalization in the course of vowel perception. In this model, perception becomes complicated and might even be considered "mentalistic."

In contrast to these approaches, Strange and her colleagues explored the possibility that vowels are recognized not based on static formant frequencies recovered from specific regions of the signal, but rather based on patterns of change into and out of those specified signal regions. In a series of experiments these investigators replaced middle portions of natural consonant-vowel-consonant (CVC) syllables with equal amounts of silence (Jenkins *et al.*, 1983; Strange *et al.*, 1983, 1976). The syllables used in these experiments were obtained from several speakers, and typically nine different vowels were represented across the samples. Listeners heard both the syllables with silent centers (i.e., the "vowel-less" syllables) as well as those extracted syllable centers presented in isolation. Results showed that listeners were more accurate in their vowel recognition for the vowelless stimuli than for the steady-state sformants. These findings led these investigators to conclude that "…dynamic acoustic information distributed over the temporal course of the syllable is utilized regularly by the listener to identify vowels" (Strange *et al.*, 1976, p. 213). Specifically, Strange and her colleagues contended that it is the spectral change arising from the vocal tract moving from articulatory consonant configurations to vowel configurations, and back again, that specifies vowel identity. This conclusion has been supported by the work of others (e.g., Assmannn and Katz, 2000; Fox, 1989), although some investigators have found that vowels can be labeled equally well with either static formants or dynamic formant transitions (e.g., Diehl *et al.*, 1981).

Nearey and Assmannn (1986) offered an alternative description of the notion that vowels are dynamically specified. They examined a variety of vowel productions, and observed that the syllabic regions specifically associated with vowel production have inherent spectral change that is uniquely associated with the vowel being produced. They further observed that these patterns of inherent spectral change are preserved in consonant contexts. From those observations, Nearey and Assmann concluded that it is this inherent dynamic information that accounts for accurate vowel recognition, regardless of whether isolated vowels or vowels in consonant contexts are heard. Part of the support for their conclusion came from the finding that high error rates were observed when listeners heard only 30-ms portions of the signal region that could be termed the vowel target. Their tests showed that listeners needed stretches of vowel samples longer than 30 ms, presumably so they could hear the inherent spectral change. Nearey and Assmann did not present listeners with stimuli in which the middle portions were replaced by silence—the classic vowel-less stimuli. Nonetheless, the juxtaposition of their findings with those of Strange and colleagues offers two alternative views of dynamic vowel information: One view suggesting that it is precisely the movement between consonant syllable margins and vocalic targets that provides the relevant dynamic information, and the other suggesting that it is the production of the isolated vowel that provides that information. These alternatives influenced the design of stimuli used in this experiment.

But the above-described work was all done with adults as listeners. Murphy *et al.* (1989) were the first investigators to examine the perception of vowel-less syllables by children, using synthetic versions of/bæd/and/bΛd/. One relevant finding from that study was that many children had difficulty labeling syllables when the deleted center portions were left silent. When those deleted portions were replaced with white noise, none of the children had difficulty labeling the vowels, prompting the conclusion that many children are unable to integrate two signal sections across a silent interval of even a few tens of milliseconds. Furthermore, the finding that all children performed accurately when listening to the syllables with filled centers led to the conclusion that children can use dynamic signal portions arising from the movement of the vocal tract between consonant- and vowel-related configurations for making vowel judgments, just as adults do.

Sussman (2001), on the other hand, used a different approach to investigate vowel perception by adults and children (4 to 5 years of age),[1] and obtained a somewhat different result. Stimuli were synthetic syllables designed to sound like/bib/and/bæb/, with 40-ms transitions on either side of 280-ms regions of steady-state formants. Stimuli were manipulated in several ways. First, stimuli were presented as vowel-less stimuli. Children performed slightly (but not significantly) more poorly than adults in this condition, which would have been predicted based on the finding of Murphy *et al.* (1989) because there was nothing filling the silent centers. But it was another stimulus manipulation that was most important to Sussman's conclusions. In that manipulation, Sussman crossed the steady-state stimulus regions with the formant transitions at the syllable edges. In one condition the transitions and steady-state portions all supported the labeling of the same vowel (the *congruent* condition). In the other condition, 220-ms sections of the steady-state vocalic portions were removed and reinserted between the margins of the other stimuli (the *conflicting* condition). Results showed that all listeners performed with close to perfect accuracy in the *congruent* condition. When asked to label the stimuli in the *conflicting* condition, all listeners responded with the label associated with the 220-ms steady-state section, but children did so slightly (but again not significantly) more often than adults. From this result, Sussman concluded that listeners of all ages, but especially children, weight steady-state formants most strongly in vowel recognition. Unfortunately, there are several reasons to worry that these stimuli may not inform us about how listeners recover vowel quality when hearing natural speech. In particular, common sense tells us that if a stimulus has a steady-state region as long as 220 ms then briefer stimulus portions, whether static or dynamic in nature, will have little chance of influencing decisions. And as Nearey and Assmannn's work (1986) emphasized, perfectly steady-state regions of this length are not

[1] Sussman (2001) also included 5- and 6-year-old children with specific language impairments, but results for those listeners are not discussed here.

found in natural speech, not even in natural vowels produced in isolation. Nonetheless, results of Sussman's experiment helped to renew interest in the question of what information child and adult listeners use in making vowel decisions.

The current study extended the work of Murphy *et al.* (1989) and Sussman (2001) in order to examine the patterns of vowel perception by children. In particular, this study was designed to cross the kind of vowel contrast being examined (i.e., vowel height or vowel frontedness) with the expected magnitude of influence on vowel formant frequencies, as well as on transitions, introduced by the consonant context (i.e., expected to be highly influential or not very influential). A strong motivation behind the design of stimuli in this experiment was the lack of articulatory complexity in the stimuli used by Murphy *et al.* and Sussman, and so a lack of opportunity for strong coarticulatory effects. Murphy *et al.* had examined a contrast between two vowels that differed primarily in tongue frontedness (/æ/vs/Λ/), whereas Sussman examined a contrast between two vowels that primarily involved jaw height (/i/vs/æ/). Both used a bilabial consonant (/b/), although Murphy *et al.* also used an alveolar at the final syllable margin. Because the lips can move largely independently of the jaw and tongue, little coarticulatory effect would be predicted for the /b/ consonantal frame. For the present study it was hypothesized that canonical vowel formant frequencies might be preserved more or less robustly depending on the type of contrast (height versus frontedness), and that different consonant contexts should affect those frequencies, as well as formant transitions, to different extents. Specifically, the hypothesis was that the perception of vowel contrasts differing in whether the tongue body is more fronted or backed (such as/æ/vs/Λ/) should be more influenced by the type of consonant context (whether it involves the lips or tongue tip) than the perception of vowel contrasts that differ in jaw height (such as/i/vs/æ/) because the tongue body is tightly linked to tongue tip movement whereas the jaw is not. In other words, consonant contexts that constrain tongue movement (such as/dVd/) should influence decisions about vowels differing in frontedness more than consonant contexts that do not involve the tongue (such as/bVb/). These potential relations of contrast type and consonant context were considered in stimulus design for this experiment.

Stimuli were created to be as natural as possible in this experiment. To achieve this goal, natural productions from adult speakers were used. Because Murphy *et al.* (1989) had observed that some children had difficulty integrating syllable portions across a silent gap, deleted syllable centers were replaced with coughs in the vowel-less stimuli. Coughs were considered more ecologically valid than white noise, and using them replicated the practice of Warren (1970; Warren and Obusek, 1971), who used coughs to replace syllable portions in the study of phoneme restoration. Perceptual impressions of such stimuli typically are that someone is saying a word, and someone else coughs at the same time, partially masking the word production. That is, the cough streams off from the string of words being heard. This impression is well-documented in the work of Warren, and was obtained in the current study as well. Results of vowel decisions for these stimuli were compared to decisions for steady-state vowel formants. Children, as well as adults, served as listeners because we were explicitly interested in how children label vowels, and if they differ from adults in how they do so.

A methodological problem that needs to be anticipated in experiments involving vowel-less stimuli is the possibility that vowel labeling may be highly accurate overall. For example, Strange *et al.* (1976) reported that listeners correctly recognized vowels 90.5% of the time when tokens from a single talker (producing nine vowels) were presented. Murphy *et al.* (1989) reported correct recognition of better than 90% for both adult and child listeners when vowel decisions were binary, and Sussman (2001) obtained recognition scores of close to 90%. Overall rates of responding can be diminished by using tokens from many different speakers (more than the three used in this experiment) or by forcing listeners to select from more than two vowel choices. On the other hand, if great uncertainty is incorporated into stimulus design and results show age-related differences it is hard to determine if the source of those differences is perception itself, or the fact that children have difficulty listening under conditions of high uncertainty (e.g., Wightman and Kistler, 2005). Because of that concern about eventual interpretation, stimulus uncertainty was restricted in this experiment, even though stimuli were designed to maximize coarticulatory effects. There was no way to know ahead of time if overall performance would be high, as in earlier experiments, or if it would be diminished substantially in this experiment precisely because stimulus design was meant to maximize coarticulatory effects.

II. METHOD

A. Listeners

Forty eight adults, 55 seven year olds, 56 five year olds, and 62 three year olds came to the laboratory to participate. All adults were between 20 and 40 years of age. All 5 and 7 year olds were between −1 and +5 months of their birthdays. Three year olds were between 3 years, 5 months, and 3 years, 11 months.[2] No more than a 40%-to-60% split between male and female participants in each age group was permitted. All participants were native English speakers with no histories of speech, language, or hearing problems. Children were required to have had fewer than six episodes of otitis media before their third birthdays. All participants were required to pass hearing screenings of the frequencies 0.5, 1, 2, 4, and 6 kHz presented at 25 dB HL to each ear separately. Children needed to perform at or better than the 30th percentile on the Goldman-Fristoe Test of Articulation 2, Sounds-in-Words subtest (Goldman and Fristoe, 2000), and adults needed to read at or better than an 11th grade reading level on the Wide Range Achievement Test—Revised (Jastak and Wilkinson, 1984). Although superficial, this brief screening provided some evidence that our adults had normal language. We took the fact that none of the children participating in the experiment were being seen for language or reading intervention as evidence that they had typical language for their ages. Our screening procedures resulted in some attrition for 3 year olds: 8 three year olds simply refused to cooperate with even the screening procedures, seven failed the hearing screening, and five failed the Goldman-Fristoe Test of Articulation. Consequently, only 42 three year olds were left to participate.

Different listeners were assigned to participate with each of the four contrasts: (1)/bIb/vs/bæb/; (2)/bæb/vs/bΛb/; (3)/dId/vs/dæd/; and (4)/dæd/vs/dΛd/. Table I shows how many participants in each age group participated in testing with each contrast. Numbers for 3 year olds do not sum to 62 because some 3 year olds did not pass the pretest (reported in Sec. III).

2B. Stimuli

Natural stimuli were used in this experiment, and presented both with and without consonantal context. Three vowels were used:/ I/,/æ/, and/Λ/. This is fewer than have been used in vowel labeling experiments with adults, but more than were used with children in the experiments of Murphy *et al.* (1989) or Sussman (2001). The choice of these three vowels meant that they could be paired so that listeners were being asked to make a decision about vowel height (/I/ vs/æ/) or a decision about frontedness (/æ/vs/Λ/). For this experiment, the consonant contexts of /dVd/and/bVb/were used. These two contexts were selected to vary the extent of contextual influence on vowel formant frequencies. This can be seen in Fig. 1, where/Λ/F2 frequency in/ dΛd/and/æ/F2 frequency in/bæb/are similar at the middle of the syllables, but the direction of transitions into and out of those vowel targets differ for the two syllables.

[2] In spite of frequent attempts to do so, we have never been able to train 3 year olds younger than 3 years, 5 months to perform a labeling task reliably, and so for 3 year olds we use children toward the higher end of their age bracket.

Stimuli differed from those of Murphy *et al.* and Sussman in several ways. The stimuli in both of those earlier experiments were synthetic; these stimuli were natural. Murphy *et al.* used only a/æ/vs/Λ/ contrast and Sussman used only a/æ/vs/i/contrast. In this experiment, both vowel contrasts were examined, with /I/substituted for the /i/ of Sussman's experiment to make the difference in formant frequencies between vowels in each contrast more similar. The stimuli in those earlier experiments did not include a condition with a /dVd/ context, which provides tongue constraints on either side of the vowel target: Murphy *et al.* used a /bVd/ context and Sussman used a/bVb/ context. Because /b/ involves lip and jaw gestures, it would not be expected to affect vocalic F2 very much, if at all. In the current experiment, the /dæd/vs/dΛd/ contrast was the one expected to have vocalic formant frequencies affected most strongly by consonantal context because the /æ/vs/Λ/ contrast largely involves a distinction in the fronting of the tongue body, and the production of/d/involves the tongue. In fact, this prediction is readily observed in the F2 frequencies reported in Tables II and III: For the /æ/and/Λ/ vowels produced either in isolation or in the /bVb/ context, the differences in F2 are roughly 400 Hz. However, when these vowels were produced in the /dVd/ context that difference is reduced to roughly 240 Hz, mainly because F2 in /dΛd/ is raised considerably, reflecting tongue fronting. Consequently, we would expect listeners to have a particularly rough time labeling vowels based on target formant frequencies, in the absence of consonantal context, for this particular contrast.

Vowel and syllable samples were recorded by three adult, male speakers who had fundamental frequencies (f0s) of roughly 130 Hz, with no extreme breathiness or fry. Each speaker was recorded producing five samples each of /bIb/,/bæb/,/bΛb/,/dId/,/dæd/,/dΛd/,/I/,/æ/, and/Λ/. First, samples of the six syllables were obtained in five randomized sets, and then samples of the three isolated vowels were obtained in five randomized sets. Speakers were instructed to produce all syllables as similar in duration as possible, and to produce all isolated vowels as similar in duration as possible. These instructions resulted in syllables with roughly 300 ms of vocalic portion and isolated vowels of roughly 200 ms. Speakers were also instructed to produce samples with similar inflectional patterns and similar f0. Samples were digitized as they were produced at a 22.05-kHz sampling rate with 16-bit digitization. If a sample was produced with vocal fry, with f0 extremely different from 130 Hz, or with duration extremely different from the targets of 300 ms (for vocalic portions of syllables) and 200 ms (for isolated vowels), the speaker was asked to repeat the production. All productions were analyzed to derive whole syllable or vowel duration, duration of vocalic portion only (for syllables), f0, and frequencies of the first three formants at vocalic onset, vocalic middle, and vocalic offset (i.e., just before closure). The two productions of each type from each speaker that were most representative of that production type from that speaker were used as stimuli in the listening experiment. Table II displays mean F1, F2, and F3 frequencies at the onset of the vocalic portion, vocalic middle, and the offset of the vocalic portion for the syllables, and Table III shows mean formant frequencies for vowels spoken in isolation, at the middle of the sample. Although the temporal middle of a speech sample does not always line up perfectly with the “target” (i.e., most extreme) formant frequencies, it is the best single time frame that provides a close estimate of those frequencies for all formants. A separate speaker, a woman, was recorded producing several coughs, and sections of these coughs were used in stimulus creation as well. A cough from a female speaker helped to ensure the perceptual impression that whole syllables were being heard, with someone else coughing during their production. Six kinds of stimuli were created, and are described in the following section. rms amplitude of all stimuli was equalized.

Flat stimuli—Three kinds of 150-ms stimuli were created to sound like isolated vowel productions. (1) 150-ms stretches of the vowels produced in isolation were obtained by finding the temporal middle of the productions, and taking 75 ms on either side of that (the “isolated vowel” stimuli); (2) the two pitch periods closest to the temporal middle of each isolated vowel were reiterated ten times to produce 150-ms stretches of steady-state formants (the "reiterated vowel" stimuli); and (3) the two pitch periods closest to the temporal middle of the vocalic portions of the syllables were reiterated ten times to produce 150-ms stretches of steady-state formants (the "reiterated syllable" stimuli). These last two types of stimuli were created to address Nearey and Assmannn's (1986) hypothesis that vowel-inherent dynamic spectral change supports vowel labeling more strongly than steady-state information: The isolated vowel productions exhibited inherent spectral change; the reiterated pitch periods did not. The inclusion of reiterated pitch periods from the isolated vowel and from syllable centers meant that some stimuli had formant frequencies consistent with isolated vowel production, and some had formant frequencies consistent with those found in coarticulated syllables. All waveform editing was done with cuts made at zero crossings. Figure 2 shows examples of the three kinds of flat stimuli: a section of a vowel produced in isolation, reiterated pitch periods from that vowel, and reiterated pitch periods from a syllable with that vowel as the nucleus.

Dynamic stimuli—(1) Whole, unaltered syllables were used (the "whole syllable" stimuli); (2) the middle 50% of the syllable was replaced with a cough section (the "50% cough" stimuli); and (3) all of the syllable except for the first and last three pitch periods was replaced with a cough section (the "pitch period" stimuli). For the 50%-cough stimuli, 150-ms stretches consisted of cough, on average. There were generally 10 pitch periods preceding that stretch of cough (75 ms) and 10 pitch periods following that stretch of cough (again, roughly 75 ms). For the pitch-period stimuli, there was generally 255 ms of cough separating the three pitch periods on either side. Thus, about 85% of the syllable was replaced with cough. Figure 3 shows examples of the three kinds of dynamic stimuli: a whole, unaltered syllable, the same syllable with the center 50% replaced by the cough, and the same syllable with all but the first and last three pitch periods replaced by the cough.

C. Equipment and materials

All speech samples were recorded in a sound-proof booth, directly onto the computer hard drive, via an AKG C535 EB microphone, a Shure M268 amplifier, and a Creative Labs Soundblaster 16-bit analog-to-digital converter. A waveform editor 3WAVED; Neely and Peters, 1992b3 was used for recording and editing. An acoustic analysis program based in MATLAB was used for spectral analysis, and SPECTO 3Neely and Peters, 1992a3 was used to make spectrograms.

Perceptual testing took place in a sound-proof booth, with the computer that controlled the experiment in an adjacent room. The hearing screening was done with a Welch Allen TM262 audiometer and TDH-39 earphones. Stimuli were stored on a computer and presented through a Creative Labs Soundblaster card, a Samson headphone amplifier, and AKG-K141 headphones. The experimenter recorded responses with a keyboard connected to the computer. Two hand-drawn pictures (8 in.×8 in.) were used to represent each response label in each experiment. These included a baby's bib for *bib*, bubbles for *bub*, a baby babbling for *bab*, a child showing his mother a picture that he made (and supposedly saying "Look what I did.") for *did*, a man for *dad*, and a firecracker with the fuse burnt for *dud*. Game boards with ten steps were also used with children: They moved a marker to the next number on the board after each block of test stimuli. Cartoon pictures were used as reinforcement that the child had completed another block of stimuli, and these pictures were presented on a color monitor after completion of each block of stimuli. A bell sounded while the pictures were being shown and served as additional reinforcement.

D. Procedures

Listeners came to the laboratory for one session. Every participant was first given the hearing screening, followed by either the speech screening (for children) or the reading screening (for adults). Next, the flat and dynamic stimuli were presented, in randomized orders across the participants. Presentation with the flat stimuli included the 150-ms segments of isolated vowel, the reiterated pitch periods from the isolated vowel, and the reiterated pitch periods from the syllable center. Presentation with the dynamic stimuli included the whole syllables, the 50% cough syllables, and the pitch period syllables. Presentation with both kinds of stimuli consisted of ten blocks of all 36 stimuli (3 stimulus types ×2 vowels×3 speakers×2 tokens).

Before testing with each stimulus type, participants were presented with either the isolated vowels or the whole syllables, depending on stimulus type to be presented, and asked to match them to the pictures used to represent them. All listeners, regardless of age, were given the same instructions. For practice with the isolated vowels, listeners were told that they would not be hearing all of the word, but they should still point to the correct picture. The pictures were introduced one at a time by the experimenter, and the appropriate label told to the listener. Generally there was a brief discussion about how the picture related to the word (e.g., *bub* is a little bit of *bubble*), and listeners were given practice responding to live voice. Listeners were asked to point to the appropriate picture, and say the word they heard, as the form of responding. Having both kinds of response helped to ensure that the listener was matching what was heard to the corresponding picture. Next, listeners heard all 12 natural, unedited stimuli, either isolated vowels or whole syllables (2 vowels×3 speakers×2 tokens), and had to respond to 11 of the 12 stimuli correctly to proceed to testing. During this practice activity, the experimenter listened to the stimuli under headphones. If a child had some initial difficulty matching pictures to stimuli, feedback was provided for up to two blocks of the 12 stimuli. By the third block of stimuli the child had to be able to respond with no feedback. If the listener could not do so, that listener was not tested in that condition (either flat or dynamic stimuli). This pretest served as a check that all listeners were able to match words (or word portions) to pictures. When testing started, listeners were told to continue doing the same thing that they had been. No more feedback was provided at this point, and the experimenter removed her headphones so that she did not know which stimulus was being presented on any given trial.[3]

It is well-documented that children have difficulty recognizing single phonemes (e.g., Routh and Fox, 1984; Liberman *et al.*, 1974; Walley *et al.*, 1986), and so it may be that they have difficulty following instructions to label them. So, in this experiment, listeners were not asked to label individual vowels, but rather were asked to decide between words containing those vowels. Moreover, uncertainty escalates as choices increase, so only two words can be used in each task with children. These considerations contributed to decisions regarding how to modify the experimental task from that used with adults.

The dependent measure in this experiment was the percentage of stimuli in each condition recognized as containing the originally produced vowel. Statistical analyses were performed on percent correct scores to examine patterns across contrasts, listener age, and stimulus types. Arcsine transforms were used because percentages were generally close to 100% correct, and so were highly skewed and kurtotic. A screening of the transformed data showed that this transformation mitigated the effects of skewness and kurtosis sufficiently to allow analyses of variance (ANOVAs) to be performed.

[3] It has been common practice in this laboratory to require data from any one listener to meet specific criteria in order for those data to be included in the final analyses; typically that criterion is that the best exemplars of the stimuli be labeled with 80% accuracy. That sort of criterion is established to ensure that all listeners, particularly children, maintain general attention during testing, and so any observed age-related differences cannot be attributed to a generalized failure on the part of children to stay on task. An explicit criterion was not established in this experiment, other than the pretest criterion, because one goal was to examine labeling accuracy across stimulus types and conditions. However, the fact that all listeners responded with well above 80% accuracy to whole-syllable stimuli, which might be considered the best exemplars in this experiment, is an indication that all listeners maintained general attention to the task during testing.

III. RESULTS

Some of the 3 year olds had difficulty reaching the criterion for participation in the pretest. For the /bIb/vs/bæb/contrast, 3 three year olds could not reach this criterion for isolated vowels, but all reached criterion for whole syllables. For the/bæb/vs/bΛb/contrast, all 3 year olds reached criterion for isolated vowels, but two did not do so for whole syllables. For the /dId/ vs/dæd/ contrast, 2 three year olds could not reach the oriterion for isolated vowels, and a different 2 three year olds could not reach criterion for whole syllables. For the /dæd/vs/dΛd/ contrast, only 1 three year old could not reach criterion for isolated vowels, but all reached criterion for whole syllables. None of the older children or adults had difficulty reaching the criterion to participate in the actual testing for any of the contrasts.

Tables IV–VI show percent correct responses (and standard deviations) for each contrast, age group, and stimulus type; the Appendix shows means for each group, in each condition. Three general trends are apparent: (1)/dæd/vs/dΛd/showed less accurate recognition than other contrasts; (2) the pitch period stimuli showed less accurate recognition than other types of stimuli; and (3) younger listeners were less accurate than older listeners. A three-way ANOVA confirmed that these three factors (contrast, stimulus type, and age) did indeed show significant effects for percent correct vowel recognition: contrast, $F(3,173) = 62.34$, $p<0.001$; stimulus type $F(5,865) = 266.84$, $p<0.001$; and age, $F(3,173) = 42.86$, $p<0.001$. However, there were also significant interactions: Contrast×Stimulus type, $F(15,865) = 26.13$, $p30.001$; Contrast×Age, $F(9,173) = 2.49$, $p = 0.011$; Stimulus type×Age, $F(15,865) = 8.18$, $p<0.001$; and the three-way interaction of Contrast×stimulus type ×Age, $F(45,865) = 1.57$, $p=0.011$. Consequently, these results do nothing to identify how children’s performance varied across stimulus type and contrast, or how it differed from that of adults. To look at these questions, results for each stimulus type were examined separately to ascertain the effects of contrast and listener age on performance, and to see where interactions occurred. These analyses were critical to determining if the extent of coarticulation affected vowel labeling.

Table VII shows results of two-way ANOVAs performed on data for each stimulus type separately, with contrast and age as the main effects. For each stimulus type there were significant effects of contrast and age, indicating that all groups generally performed more poorly for the/dæd/vs/dΛd/contrast than for other contrasts and that younger listeners generally performed more poorly than older listeners. Of greater relevance for this study, there were significant Contrast×Age interactions for flat stimuli, but not for dynamic stimuli. In other words, children’s perception (compared to adults’ perception) was disproportionately affected by the contrast only when stimuli were flat. When stimuli were dynamic, all listeners were similarly affected by the extent of coarticulatory effects in those stimuli. Children performed disproportionately more poorly than adults for flat stimuli involving the/dæd/vs/dΛd/contrast than for the other contrasts. This is precisely the contrast in which are found the greatest coarticulatory constraints of the consonantal context on target formant frequencies for the vowel. It is tempting to hypothesize that this result indicates that young children’s perception is disrupted in the face of heavy coarticulatory effects. That may indeed be true, as we find here that young children’s labeling of flat stimuli was generally diminished for the/dæd/vs/dΛd/contrast compared to conditions in which coarticulatory effects were not as strong. One piece of evidence for this assertion is the finding that 3-year-olds’ labeling of isolated vowel productions of/æ/and/Λ/was poorer when those same stimuli were presented along with reiterated pitch periods from the syllables /dæd/and/dΛd/than when they were presented along with reiterated pitch periods from/bæb/and/bΛb/: the Appendix shows that mean recognition of these isolated vowels was 95.67 percent correct in the/bæb/vs/bΛb/condition, but only 83.44 percent correct in the/dæd/vs/dΛd/condition. So, young children encountered difficulty labeling any stimuli heard as isolated vowels when some of those stimuli were reiterated pitch periods from heavily coarticulated syllables.

A metric used to examine the apparent finding that young children have disproportionate difficulty labeling vowels from steady-state formants when coarticulatory effects are strong, as in the/dæd/vs/dΛd/contrast, was Cohen's *d*. This statistic is the difference between any two means, normalized by the pooled standard deviation associated with those means (Cohen, 1988). While straightforward in computation, Cohen's *d* serves as a robust index of effect size. In this case, it can index the magnitude of coarticulatory effects on vowel labeling. To this end, means of percent correct vowel recognition were compared between a contrast with little expectation of coarticulatory effects (/bIb/vs/bæb/) and the contrast with the greatest expectation of coarticulatory effects (/dæd/vs/dΛd/). Two stimulus types were examined: a set of flat stimuli, the reiterated syllable stimuli, and a set of dynamic stimuli, the 50% cough stimuli. These stimulus types were selected because they best represent what listeners are likely to hear in the real world: We normally hear whole words rather than isolated vowels, and so any acoustic consequences of coarticulation are available, even if some syllable portions are masked. Table VIII shows Cohen's *d* for each age group, for each of these syllable types, computed as the difference between mean percent correct scores for the/bIb/vs/bæb/and/dæd/vs/dΛd/contrasts, divided by the pooled deviations of those means. What we find is that all three groups of children showed greater coarticulatory effects on their vowel labeling when only steady-state formant frequencies were presented (i.e., reiterated syllable stimuli) than when dynamic syllable components were presented (i.e., 50% cough stimuli). This clearly was not the case for adults, whose performance held up well with the reiterated syllable stimuli derived from /dæd/and/dΛd/.

Another useful way of viewing these data is to examine trends across syllable types for each contrast, for each age group separately. These analyses should provide information about how listeners of each age group dealt with the degraded stimuli compared to the unprocessed stimuli: Were different age groups more or less affected either by having truly steady-state formants (as in the reiterated conditions) rather than isolated vowels (with inherent spectral change) or by having coughs replace vocalic centers? To help answer these questions, a series of matched *t*-tests were conducted for each listener age, for each contrast. These *t*-tests were computed for the following comparisons: (1) isolated vowel versus reiterated vowel; (2) isolated vowel versus reiterated syllable; (3) whole syllable versus 50% cough; (4) whole syllable versus pitch period; (5) isolated vowel versus whole syllable; and (6) reiterated syllable versus 50% cough. The last comparison was considered critical because it involved the acoustic portions that children are likely to hear in their daily lives. Again, arcsine transformations were used, and an alpha level of 0.01 was applied, rather than 0.05, due to the high probability of obtaining significant differences by chance when so many comparisons are made. A Bonferroni correction may also be applied, and with six contrasts per condition this would mean that significance at the 0.01 level would be reached only when obtained *p* values are less than 0.002. Results reaching this significance level are noted by an asterisk.

/bIb/vs/bæb/

All three children's groups performed more poorly for the pitch-period stimuli than for the whole syllables: 3 year olds, $t(8) = 7.08$, $p < 0.001$*; 5 year olds, $t(13) = 7.42$, $p < 0.001$*; and 7 year olds, $t(12) = 3.75$, $p = 0.003$. No other differences were observed.

/bæb/vs/b^b/

All four age groups showed poorer performance for the pitch-period stimuli than for the whole syllables: 3 year olds, $t(9) = 7.43$, $p < 0.001$*; 5 year olds, $t(12) = 10.02$, $p < 0.001$*; 7 year olds, $t(11) = 4.77$, $p < 0.001$*; and adults, $t(11) = 3.25$, $p = 0.008$. In addition, 5 year olds performed more poorly for isolated vowels than for whole syllables, $t(12) = 3.78$, $p = 0.003$. The Appendix shows that 3 year olds performed similarly to 5 year olds with both isolated vowels and whole syllables: The mean difference between the two conditions was 3 percentage points for both groups. However, this comparison for 3 year olds did not quite reach criterion for reporting statistical significance because of the fewer degrees of freedom.

/dɪd/vs/dæd/

Again, all four age groups showed poorer performance for the pitch-period stimuli than for the whole syllables: 3 year olds, *t*(8) =6.12, *p*<0.001*; 5 year olds, *t*(13) =9.02, *p*<0.001*; 7 year olds, *t*(11) =6.64, *p*<0.001*; and adults, *t*(11) =5.60, *p*<0.001*. In addition, 5 year olds performed more poorly for the 50% cough stimuli than for whole syllables, *t*(13) =3.25, *p*=0.006. In this case, the failure to find a similar result for 3 year olds may be attributable both to fewer degrees of freedom and to the fact that there was not as great of a difference between the two conditions for 3 year olds: mean differences between the two conditions were 3 and 8 percentage points for 3 and 5 year olds, respectively.

/dæd/vs/d^d/

Once again, all four age groups showed poorer performance for the pitch-period stimuli than for the whole syllables: 3 year olds, *t*(9) =15.14, *p*<0.001*; 5 year olds, *t*(14) =25.81, *p*<0.001*; 7 year olds, *t*(17) =22.46, *p* <0.001*; and adults, *t*(11) =17.58, *p*<0.001*. For this vowel contrast, all three children's groups also showed poorer performance for the 50% cough stimuli than for whole syllables: 3 year olds, *t*(9) =4.03, *p*=0.003; 5 year olds, *t*(14) =6.31, *p*<0.001*; and 7 year olds, *t*(17) =5.67, *p*<0.001*. Five and 7 year olds also performed more poorly for the reiterated syllable stimuli than for isolated vowels, 5 year olds, *t*(14) =3.36, *p*=0.005; and 7 year olds, *t*3173 =4.78, *p*<0.001*. In this case, the failure to find a similarly significant effect for 3 year olds is likely due to fewer degrees of freedom: For all three children's groups there was a 4 percentage point difference between conditions.

IV. DISCUSSION

This experiment was conducted primarily to examine whether children rely more on static or dynamic spectral structure in vowel perception, and whether their perceptual strategies for vowels differ from those of adults. Stimuli were designed to differ in which aspect of vowel quality was manipulated within a contrast ×height or frontedness× and in the extent of consonant-vowel coarticulation. In some cases, only flat formants heard as isolated vowels were presented to listeners and in other cases only formant transitions heard as syllables, either with or without overlaid coughs, were presented to listeners. Overall results showed that listeners were able to recognize vowels with either static or dynamic spectral information, supporting the conclusion of Diehl *et al.* (1981), but not without possible caveats. In particular, earlier experiments showing a preference in adult listeners for dynamic signal components included more vowels in each contrast (e.g., Jenkins *et al.*, 1983; Strange *et al.*, 1983, 1976), and so performance was generally lower than what was found here. That makes it difficult to compare adults' results from this experiment to those of earlier experiments. With adults' performance consistently close to 100% correct, our ability to observe potential differences across syllable types was constrained. But the focus of this experiment was not on how adults would do, but rather on how children would perform, compared to adults. The finding that listeners of all ages performed as well as they did with only dynamic syllable structure suggests that this structure informs listeners substantially about vowel quality. Certainly this result argues against the notion that children rely more than adults on "…the formant steady states for vowel identification in difficult conditions," as suggested by Sussman (2001, p. 1179). If that were the case we would have expected children to do far worse than adults explicitly with the dynamic stimuli that had syllable centers replaced with a cough. Instead, children were somewhat less accurate than adults, but this age-related discrepancy in performance was not specific to stimuli preserving only dynamic structure. All listeners used the entire spectral pattern for vowel labeling, as long as some minimal amount was available: Listeners had difficulty when only three pitch periods on either side of the cough were heard, which is reasonable given that the very notion of a “dynamic” property suggests that a sufficiently long stretch of signal must be input in order for a perceiver to recognize the change. In fact it was precisely the “formant steady states” in the difficult condition that children had particular difficulty with (i.e., the reiterated pitch periods from the/dæd/vs/dΛd/contrast).

The finding that children were more affected than adults by the vowel contrast being heard when stimuli were flat, rather than dynamic, indicates how strongly children rely on dynamic spectral structure to recognize vowels. With the dynamic signals, children showed similar trends across contrasts to those of adults. With the flat stimuli, children performed much more poorly with the contrast involving tongue frontedness in the/dVd/context than with the other contrasts. This trend was found even for the natural vowels produced in isolation: Apparently, children were so hindered by hearing only the flat formants that their abilities to recognize vowels in the absence of consonantal contexts were generally disturbed. This result demonstrates how important dynamic signal components are for young children listening to speech.

In general, these findings supported the conclusions of others asserting that vowel identity is specified in the dynamic spectral structure of the entire syllable (e.g., Jenkins *et al.*, 1983; Strange *et al.*, 1983; 1976). At the same time, this study significantly extends that work with adults by demonstrating that children similarly use the dynamic spectral structure of whole syllables to make vowel decisions. Although Murphy *et al.* (1989) had demonstrated this result, the present study uncovered it for a greater number of contrasts, and served to mediate between the contradictory find of Murphy *et al.* and Sussman (2001). With a contrast involving greater coarticulation between consonant and vowel than those used in either of those studies, and stimuli constructed to be as natural as possible, it became clear that children can and do use information arising from coarticulated segments to make vowel decisions.

This study was unable to provide specific support for Nearey and Assmannn’s (1986) claim that vowel-inherent spectral change is critical to vowel recognition, but instead found further support for the claim that dynamic spectral structure arising from articulatory movements between consonant and vowel constrictions is most important. Of course, there has accumulated over the years substantial evidence showing how important dynamic structure is to speech perception, for both consonant and vowel recognition. A few of the many examples regarding consonant perception include a study by Harris (1958) on fricative perception, one by Kewley-Port *et al.* (1983) on syllable-initial stop perception, and one by Nittrouer (2004) on syllable-final stop perception. In all these studies, dynamic structure spanning some portion of the syllable was found to be more important to decisions of consonant identity than either static spectral or temporal structure. If one cobbles together the conclusions of these many studies a theoretical framework emerges based on the idea that dynamic structure in the speech signal plays a fundamental role in phonetic recognition. It is not difficult to fit Nearey and Assmannn’s finding into that framework: Given a large stimulus set of many spectrally similar vowels, as Nearey and Assmannn used, dynamic structure inherent in the production of isolated vowels would be expected to provide information to help disambiguate between those vowel choices.

Finally, it is worth emphasizing that what is generally termed the “steady-state” syllable region is actually part of the whole dynamic structure of that syllable. When the beginning and ending dynamic components of a syllable are presented to listeners in normal temporal proximity, it is likely the case that listeners “hear through” the silence (or white noise or cough, whatever the case may be). These stimulus portions provide information about both vowel and consonant. But upon hearing only brief, steady-state portions of the vowel target, we would not expect listeners to recognize syllable margins in that same way because there is not enough information available about what consonants were at those margins. With this in mind, the current study serves as a reminder that human speech perception does not unfold as a process of extraction of static spectral slices from the acoustic speech signal, matched against internally stored templates of individual phonemes or features. Rather, speech perception very much depends upon the dynamic structure of the signal, spanning temporal slices generally affiliated with several phonetic segments.

Acknowledgements

This work was supported by Research Grant No. R01 DC000633 from the National Institute on Deafness and Other Communication Disorders. The author wishes to thank Sandy Estee for help with data collection, and Amy Bartlett, Chris Chapman, and Joanna Lowenstein for help with manuscript preparation.

References

Cohen, J. Erlbaum. 2. Hillsdale, New Jersey: 1988. Statistical power analysis for the behavioral sciences.

Assmannn PF, Katz WF. Time-varying spectral change in the vowels of children and adults. J Acoust Soc Am 2000;108:1856–1866. [PubMed: 11051512]

Diehl RL, McCusker SB, Chapman LS. Perceiving vowels in isolation and in consonantal context. J Acoust Soc Am 1981;69:239–248. [PubMed: 7217522]

Fant, G. Speech Sounds and Features. MIT; Cambridge, MA: 1973.

Ferrand, CT. Speech Science: An Integrated Approach to Theory and Clinical Practice. Allyn and Bacon; Boston: 2007.

Fox RA. Dynamic information in the identification and discrimination of vowels. Phonetica 1989;46:97–116. [PubMed: 2608728]

Fry DB, Abramson AS, Eimas P, Liberman AM. The identification and discrimination of synthetic vowels. Lang Speech 1962;5:171–189.

Gerstman LJ. Classification of self-normalized vowels. IEEE Trans Audio Electroacoust 1968;AU-16:78–80.

Goldman, R.; Fristoe, M. Goldman Fristoe 2: Test of Articulation. American Guidance Service; Circle Pines, MN: 2000.

Harris KS. Cues for the discrimination of American English fricatives in spoken syllables. Lang Speech 1958;1:1–7.

Jastak, S.; Wilkinson, GS. The Wide Range Achievement Test-Revised. Jastak Associates; Wilmington, DE: 1984.

Jenkins JJ, Strange W, Edman TR. Identification of vowels in ‘vowelless’ syllables. Percept Psychophys 1983;34:441–450. [PubMed: 6657448]

Kewley-Port D, Pisoni DB, Studdert-Kennedy M. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. J Acoust Soc Am 1983;73:1779–1793. [PubMed: 6223060]

Ladefoged P, Broadbent DE. Information conveyed by vowels. J Acoust Soc Am 1957;29:98–104.

Liberman AM, Harris KS, Hoffman HS, Griffith BC. The discrimination of speech sounds within and across phoneme boundaries. J Exp Psychol 1957;54:358–368. [PubMed: 13481283]

Liberman IY, Shankweiler D, Fischer FW, Carter B. Explicit syllable and phoneme segmentation in the young child. J Exp Child Psychol 1974;18:201–212.

Lindblöm BE, Studdert-Kennedy M. On the role of formant transitions in vowel recognition. J Acoust Soc Am 1967;42:830–843. [PubMed: 6075568]

Murphy WD, Shea SL, Aslin RN. Identification of vowels in ‘vowel-less’ syllables by 3-year-olds. Percept Psychophys 1989;46:375–383. [PubMed: 2798031]

Nearey TM, Assmannn P. Modeling the role of inherent spectral change in vowel identification. J Acoust Soc Am 1986;80:1297–1308.

Neely, ST.; Peters, JE. SPECTO User’s Guide. 3Boys Town National Research Hospital; Omaha, NE3: 1992a.

Neely, ST.; Peters, JE. WavEd User's Guide. 3Boys Town National Research Hospital; Omaha, NE3: 1992b.

Nittrouer S. The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. J Acoust Soc Am 2004;115:1777–1790. [PubMed: 15101656]

Routh DK, Fox B. 'Mm…is a little bit of May:' Phonemes, reading, and spelling. Advances in Learning and Behavioral Disabilities 1984;3:95–124.

Strange W, Jenkins JJ, Johnson TL. Dynamic specification of coarticulated vowels. J Acoust Soc Am 1983;74:695–705. [PubMed: 6630725]

Strange W, Verbrugge RR, Shankweiler DP, Edman TR. Consonant environment specifies vowel identity. J Acoust Soc Am 1976;60:213–224. [PubMed: 956528]

Sussman JE. Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions? J Acoust Soc Am 2001;109:1173–1180. [PubMed: 11303931]

Syrdal AK, Gopal HS. A perceptual model of vowel recognition based on the auditory representation of American English vowels. J Acoust Soc Am 1986;79:1086–1100. [PubMed: 3700864]

Walley AC, Smith LB, Jusczyk PW. The role of phonemes and syllables in the perceived similarity of speech sounds for children. Mem Cognit 1986;14:220–229.

Warren RM. Perceptual restoration of missing speech sounds. Science 1970;167:392–393. [PubMed: 5409744]

Warren RM, Obusek CJ. Speech perception and phonemic restorations. Percept Psychophys 1971;9:358–362.

Wightman F, Kistler D. Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. J Acoust Soc Am 2005;118:3164–3176. [PubMed: 16334898]

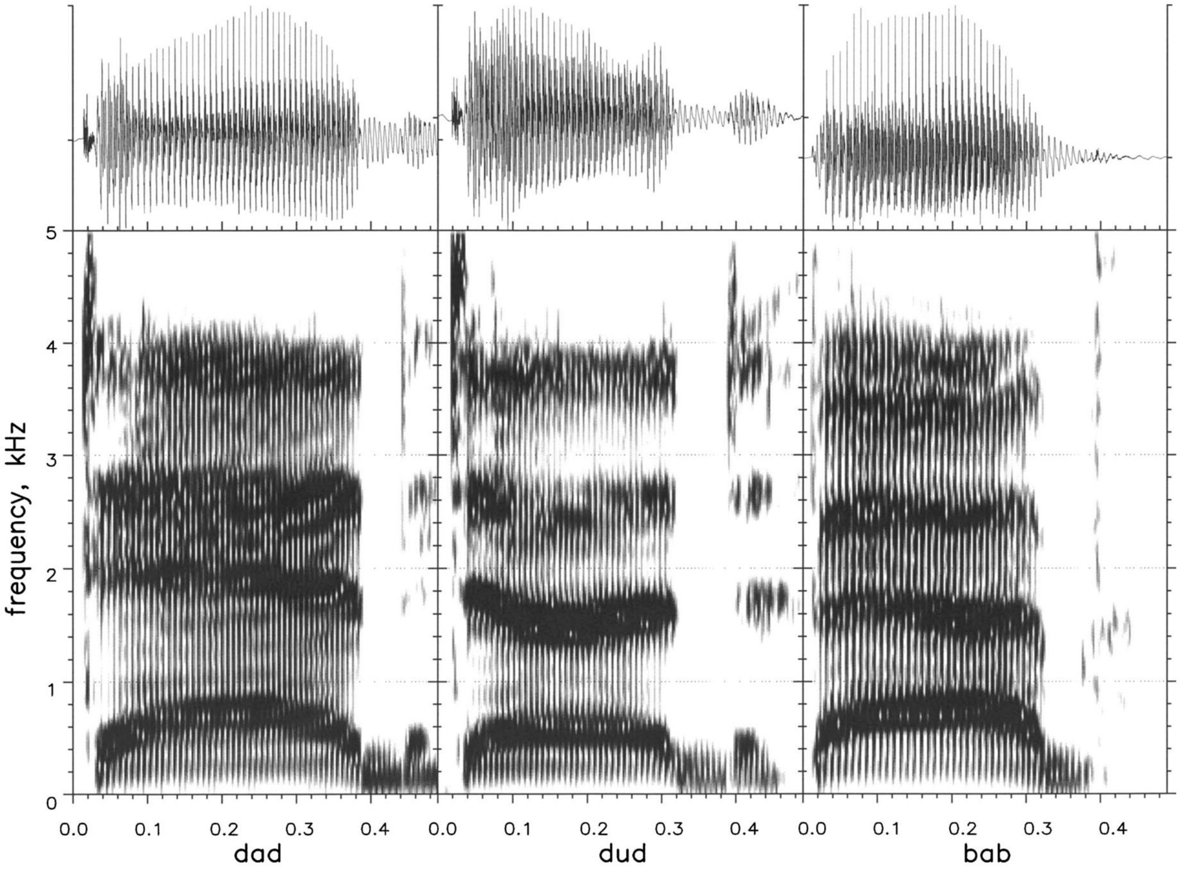

FIG. 1.
Spectrograms of adult, male speaker saying “dad,” “dud,” and “bab.”

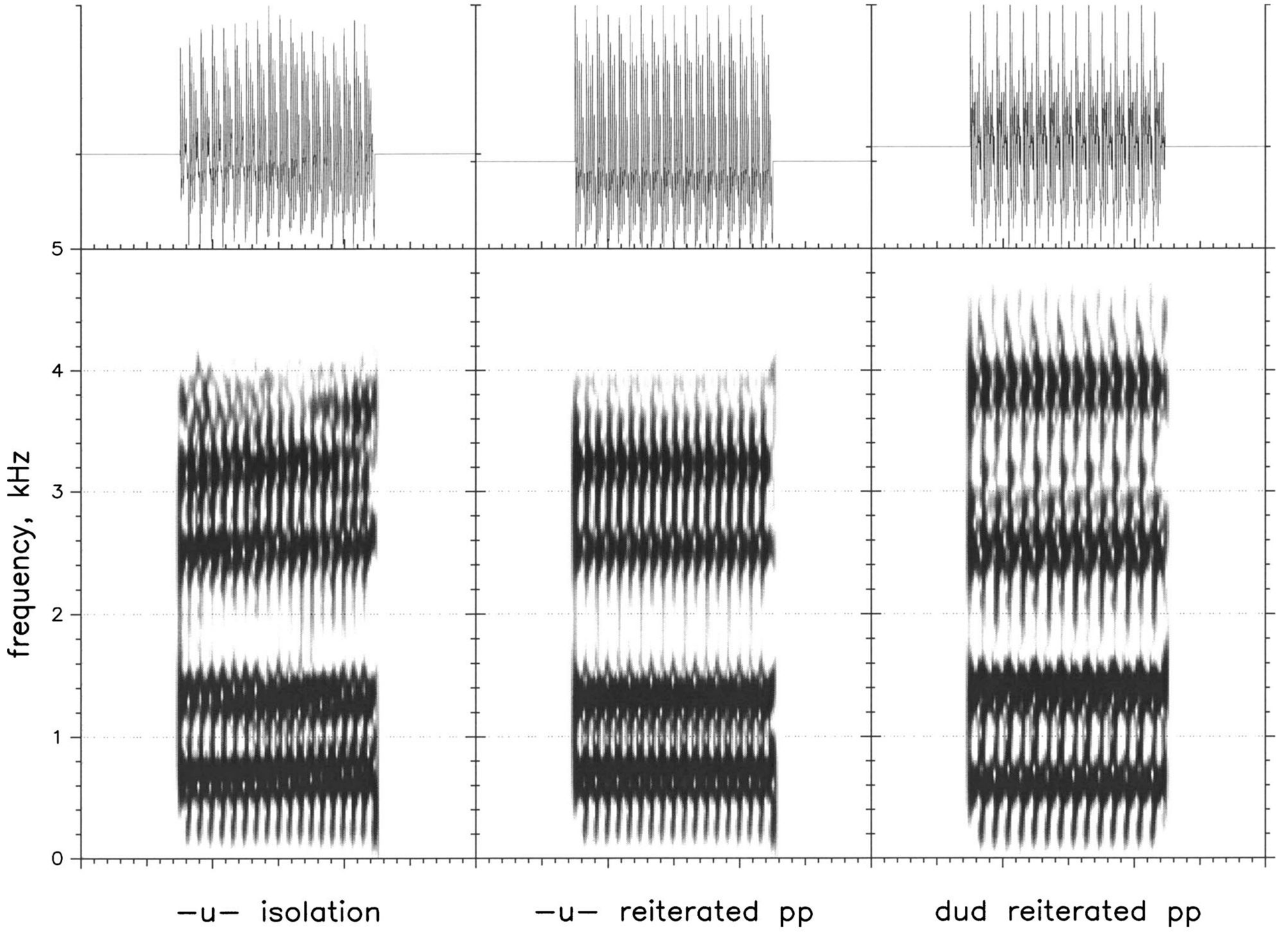

FIG. 2.
Spectrograms of the three examples of the flat stimuli: 150 ms of the vowel/Λ/produced in isolation (labeled here as -u-); reiterated pitch periods from the vowel/Λ/produced in isolation; and reiterated pitch periods from/Λ/produced in the syllable “dud.”

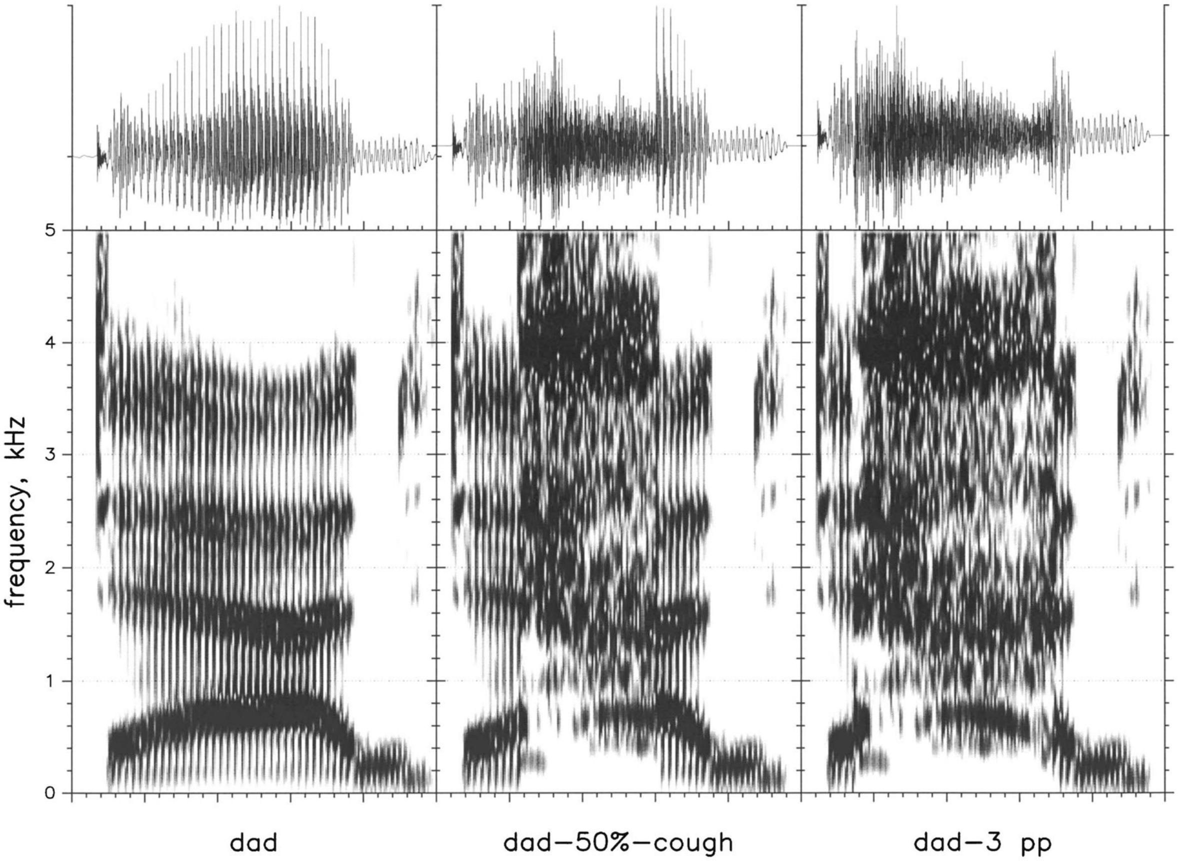

FIG. 3.
Spectrograms of three examples of the dynamic stimuli: the whole syllable “dad;” the same syllable with the center 50% replaced by a cough; and the same syllable with all except the first and last three pitch periods replaced by a cough.

TABLE I

The number of participants of each age, in each contrast.

	/bIb/vs/bæb/	/bæb/vs/bΛb/	/dId/vs/dæd/	/dæd/vs/dΛd/
3 year olds	9	12	11	10
5 year olds	14	13	14	15
7 year olds	13	12	12	18
Adults	12	12	12	12

Appendix_ Percentages of correct vowel recognition, in each condition for each contrast.

	Age	Isolated vowel	Reiterated vowel	Reiterated syllable	Whole syllable	50% cough	Three pitch periods
/b(b/vs/bæb/	3	98.83 (2.86)	98.17 (2.99)	97.83 (2.04)	98.67 (1.73)	94.44 (9.99)	86.22 (13.11)
	5	99.07 (1.94)	98.93 (2.09)	98.21 (2.67)	99.29 (1.49)	96.86 (4.20)	86.14 (11.14)
	7	99.85 (0.55)	99.38 (0.96)	98.85 (2.15)	99.69 (0.75)	98.00 (4.18)	89.38 (13.38)
	Adult	100.00 (0.00)	99.33 (1.30)	98.67 (2.10)	99.83 (0.58)	100.00 (0.00)	99.42 (1.51)
/bæb/vs/b#b/	3	95.67 (3.65)	95.42 (4.54)	96.58 (4.38)	99.00 (1.41)	97.90 (2.13)	85.40 (9.24)
	5	95.77 (4.48)	96.00 (4.22)	95.46 (6.10)	99.54 (0.88)	98.54 (2.47)	91.46 (4.20)
	7	99.67 (0.78)	99.3) (1.30)	99.83 (0.58)	100.00 (0.00)	99.67 (0.78)	96.17 (3.95)
	Adult	99.67 (0.78)	99.50 (0.90)	99.33 (1.56)	100.00 (0.00)	99.83 (0.58)	98.83 (1.34)
	Age	Isolated vowel	Reiterated vowel	Reiterated syllable	Whole syllable	50% cough	Three pitch periods
/d(d/vs/dæd/	3	95.56 (4.03)	95.00 (5.94)	94.11 (3.95)	96.78 (4.74)	91.78 (6.08)	78.67 (11.70)
	5	99.50 (1.24)	99.33 (1.30)	97.83 (2.59)	99.07 (1.69)	91.57 (10.21)	73.93 (16.00)
	7	100.00 (0.00)	99.83 (0.58)	99.33 (1.56)	99.50 (0.90)	92.83 (10.78)	70.42 (18.11)
	Adult	99.83 (0.58)	99.25 (1.76)	97.92 (4.06)	100.00 (0.00)	99.67 (0.78)	93.75 (5.59)
/dæd/vs/d#d/	3	83.44 (19.04)	85.11 (16.31)	79.22 (18.79)	96.40 (3.86)	84.40 (13.62)	52.00 (5.52)
	5	92.80 (11.31)	96.40 (6.29)	88.80 (10.35)	98.67 (2.13)	93.00 (4.14)	56.07 (5.16)
	7	98.72 (2.11)	99.78 (0.65)	94.61 (4.60)	99.22 (1.26)	95.11 (5.05)	56.83 (8.81)
	Adult	99.25 (1.29)	99.42 (1.24)	96.83 (3.69)	99.92 (0.29)	98.08 (2.71)	75.58 38.363

TABLE II

Formant frequencies for syllables produced by three speakers, two tokens per speaker.

	F1			*F2*			F3		
	Onset	**Middle**	**Offset**	**Onset**	**Middle**	**Offset**	**Onset**	**Middle**	**Offset**
/bæb/	449 (58)	721 (26)	567 (100)	1575 (68)	1694 (132)	1339 (80)	2360 (76)	2498 (143)	2437 (149)
/bIb/	370 (9)	438 (11)	424(57)	1762 (46)	1830 (71)	1518 (52)	2512 (106)	2596 (140)	2476 (118)
/bΛb/	463 (50)	624 (14)	510 (50)	1217 (47)	1278 (45)	1199 (35)	2505 (121)	2633 (104)	2559 (91)
/dæd/	388 (41)	704 (49)	488 (47)	1852 (65)	1708 (150)	1612 (69)	2649 (174)	2494 (141)	2616 (217)
/dId/	327 (22)	417 (42)	350 (77)	1913 (141)	1967 (134)	1784 (98)	2681 (146)	2624 (115)	2641 (128)
/dΛd/	395 (37)	599 (25)	427 (42)	1712 (109)	1461 (108)	1633 (54)	2584 (177)	2588 (159)	2667 (112)

TABLE III

Formant frequencies for isolated vowels produced by three speakers, two tokens per speaker.

	***F*1**	***F*2**	***F*3**
/æ/	786 (26)	1694 (98)	2524 (168)
/I/	452 (14)	1963 (113)	2627 (77)
/Λ/	671 (16)	1314 (64)	2487 (106)

TABLE IV

Mean percent correct vowel recognition for each contrast, across all listeners and syllable types.

/bIb/vs/bæb/	/bæb/vs/bΛb/	/dId/vs/dæd/	/dæd/vs/dΛd/
97 (6)	97 (6)	94 (3)	89 (4)

TABLE V

Mean percent correct vowel recognition for each age group, across all contrasts and syllable types.

Adults	7 year olds	5 year olds	3 year olds
98 (2)	95 (4)	93 (6)	90 (9)

TABLE VI

Mean percent correct vowel recognition for each syllable type, across all contrasts and listeners.

Flat			Dynamic		
Isolated vowel	**Reiterated vowel**	**Reiterated syllable**	**Whole syllable**	**50% cough**	**3 pitch periods**
98 (7)	98 (5)	96 (7)	99 (2)	96 (7)	80 (18)

TABLE VII

Results of a two-way ANOVA for all four contrasts, for each stimulus type separately. Denominator degrees of freedom are shown in the first line of each section, and reflect the fact that both contrast and age are between subjects factors. Numerator degrees of freedom are shown below. Exact *p* values are shown if less than 0.05. NS indicates nonsignificant effects.

Stimulus type	**df**	**F**	**p**
	Flat stimuli		
Isolated vowel	177		
Contrast	3	15.04	<0.001
Age	3	19.74	<0.001
Contrast×Age	9	2.44	0.012
Reiterated vowel	177		
Contrast	3	3.86	0.011
Age	3	18.20	<0.001
Contrast×Age	9	2.32	0.017
Reiterated syllable	177		
Contrast	3	26.45	<0.001
Age	3	15.40	<0.001
Contrast×Age	9	2.07	0.035
	Dynamic stimuli		
Whole syllable	181		
Contrast	3	3.50	0.017
Age	3	12.48	<0.001
Contrast×Age	9	0.47	NS
50% cough	181		
Contrast	3	16.80	<0.001
Age	3	19.36	<0.001
Contrast×Age	9	0.86	NS
Pitch periods	181		
Contrast	3	108.87	<0.001
Age	3	38.64	<0.001
Contrast×Age	9	1.75	NS

TABLE VIII

Cohen's *d* for mean percent correct vowel recognition in the/bIb/vs/bæb/and the/dæd/vs/dΛd/contrasts for the reiterated syllable and 50% cough stimuli. This metric provides an index of the magnitude of coarticulatory effect on listeners' responses.

	Reiterated syllable	**50% cough**
3 year olds	1.39	0.84
5 year olds	1.25	0.93
7 year olds	1.18	0.62
Adults	0.61	1.00