



**Acoustics'08
Paris**
June 29-July 4, 2008

www.acoustics08-paris.org

euonoise

Acoustic cues to the voicing contrast in coda stops in the speech of 2-year-olds learning American English

Helen Hanson^a and Stefanie Shattuck-Hufnagel^b

^aUnion College, 807 Union St., Schenectady, NY 12308, USA

^bMassachusetts Institute of Technology, Research Laboratory of Electronics, Speech Communication Group, 77 Massachusetts Ave., Cambridge, MA 02139, USA
hansonh@union.edu

Applying acoustic analysis methods to recorded corpora of child speech has revealed important aspects of acoustic variation in the cues that children produce to distinctive feature contrasts. For example, two children from the Imbrie Corpus (2005) show a clear tendency to produce noise at the end of the nuclear vowel in words like *duck* and *cup*, which have [-voice] codas, but not in words like *tub* and *bug*, which have [+voice] codas. The noisy regions may occur at mid-vowel, at the end of the vowel or during the closure portion of the coda consonants. These observations raise important questions about models of the development process, including whether these developing speakers produce noise before voiceless codas because of their inability to time laryngeal and supralaryngeal gestures in adult-like ways, or because of their intention to provide perceptually salient cues to the [+voice] vs. [-voice] feature of the coda consonant. We describe a proposed labeling method to capture the salient facts about the location and type of the cues produced by child speakers to signal the voicing feature of a word-final stop consonant, and discuss some preliminary data.

1 Introduction

For decades (or even centuries) the primary source of data about phonological and phonetic development in children has been transcriptions made of child speech by adult listeners. With the advent of spectrograms in the 1940s and, more recently, widespread access to convenient computer-based tools for detailed acoustic analysis and to extensive and growing corpora of acoustic recordings of child speech, it has become possible to examine child productions in greater acoustic detail. The results of these more detailed analyses have been somewhat surprising in two respects. First, they have revealed a persistent pattern of behaviour in which the child produces systematic and reliable contrasts between classes of speech sounds which are not easily perceived by the adult listener. For example Kornfeld [3] observed that a child's rendition of *glass* and *grass* were both transcribed as [gwaes] by an adult listener, but 'spectrograms showed consistent differences ... in terms of F2 locus and duration of the "glide" segment' in the two word forms. This phenomenon, termed 'covert contrast', has been found for a wide variety of contrasts in a large number of languages (see [7] for a review). It has profound implications for our understanding of which phonological categories and contrasts are captured by the child's developing representational system, of the mechanisms that govern the child's choice of acoustic cues to signal those categories and contrasts, and of the concept of 'developmental delay' in the acquisition of phonology. These implications are still being worked out, along with the scope and nature of the covert contrasts themselves.

The second somewhat surprising insight gained from the combination of easily accessible acoustic analysis tools applied to acoustically recorded child speech corpora concerns the cues that a child uses to signal the distinctive feature contrasts that he has already mastered and that adult listeners can readily perceive. In this domain, initial analyses have begun to reveal that children sometimes employ cues which appear to be different from the adult pattern. For example, Shattuck-Hufnagel et al. [8] report that in a small sample of children from the Imbrie Corpus [2], i.e. two children between the ages of 2;6 and 3, the clear differences between voiced and voiceless stops often included a period of noise at the end of the vowel, presumably associated with the coda stop. Several typical examples of this noise are shown for utterances of the word *duck*, in Figures 1 and 2. In Fig. 1, the vowel becomes voiceless about halfway through. In Fig. 2, the vowel becomes noisy shortly before voicing ceases, and then the

noise continues into the closure. Although this vowel-final noise is not the only cue to the voiced-voiceless feature difference that these two children produced--they often produced, for example, long voice bars and epenthetic post-coda-release vowels for voiced stops, and heavy post-coda-release aspiration for voiceless stops---it is a particularly interesting cue for several reasons.

First, such noise has not been generally reported in this context for adult speakers of American English, although it has been reported as the phenomenon of 'pre-aspiration' in other languages and dialects. Gordeeva and Scobbie [1] summarize the situation as follows: "Preaspiration has been described as a co-ordinatory relationship between a vowel and a following voiceless segment (Laver, 1994)[5]. This involves an early offset of modal voicing in the vowel in anticipation of the wide opening of the vocal folds required by the voiceless segment (Ladefoged & Maddieson, 1996) [4]. The glottal opening is accompanied by variable-in-strength supraglottal turbulence that is often interchangeably termed as 'breathiness', 'aspiration' or 'whisper' (Laver, 1994, p.189-190)." Although the term 'pre-aspiration' implies that the noise is generated at the larynx, by airflow through the semi-approximated vocal folds, the term 'supraglottal' implies a different source. To date there has been little spectral analysis of such child tokens to determine whether the noise is generated at the larynx or by means of non-adult-like overlap between the region of voicing and turbulence noise generated by movement toward the supra-glottal constriction related to the upcoming coda consonant.

The occurrence of such noise events at the ends of vowels in children learning American English has the potential to be informative about the mechanisms of phonological development. For example, analysis of this vowel-final noise cue in child-caretaker pairs of speakers may shed light on the extent to which children imitate the precise cue patterns of the adults who speak to them. In contrast, if the adults do not produce these cues, even in speech to children, it may provide evidence for the hypothesis that speakers (including developing speakers) have the option of enhancing phonological feature contrasts by providing additional acoustic cues, particularly in certain contexts [9]. It may be that adult speakers produce a small amount of noise in this context, and children who are still learning the patterns of contextual phonetic variation from the adult language around them exaggerate the noise cue that they hear. Finally, it is possible that the noise we observe in this context arises not from an active decision on the part of the child to produce a noise cue to the feature [-voice], but rather because of an immature ability to coordinate the various laryngeal and supralaryngeal movements and configurations that are required to make a series of speech

sounds. That is, it may be that, particularly in coda position, the supralaryngeal articulators close too slowly or too early, resulting in overlap between the voiced nuclear vowel and the occurrence of turbulence noise at the still-partial constriction. An alternative account in terms of immature articulatory control might be that the laryngeal gestures associated with the offset of voicing for [-voice] stop consonants in English begins too early or occurs too slowly, again resulting in temporal overlap between periodic vibration and noise generation at the glottis.

Disentangling these potential accounts will require careful acoustic analysis of the temporal and spectral relationships among the various cues to the voicing distinction in an appropriate set of child utterances. In this study, we present a method for labelling and analysing the cues to voicing in child productions of coda stops, which will allow us to subsequently address some of the questions noted above. We focus on coda consonants because the developmental patterns of acoustic cues to features in this structural location has not been much studied, and because their apparent lagging development may provide evidence that bears on these questions.

2 Methods

The labeling effort reported here has as its eventual goal the use of acoustic analysis techniques to explore the cues that child speakers provide to signal the contrast between voiced and voiceless coda stops in a corpus of CVC words in the Imbrie Corpus. Methods will be described in terms of 1) the corpus of recorded child utterances and 2) the acoustic labeling methodology.

2.1 Corpus

The child speech tokens analysed here are taken from the Imbrie Corpus [2]. The corpus consists of recordings from 10 children learning American English, made at approximately 1-month intervals during a 6-month period beginning at age 2;6 to 3;x. (Complete details about the subjects and the recordings can be found in [2], which is available online.) A set of 20 target words were elicited multiple times in each session, via interactive play using pictures and objects that were removed from and replaced in a travel bag as they were named. Subsequently each token of a target word saved in a separate computer file, categorized by whether the token was elicited in a naming task or occurred in a longer statement during general conversation between the experimenter and the child. There was a wide range of total numbers of tokens of each target word elicited during each recording session, from zero to more than 20, and the number of approximately-monthly sessions per child ranged from 5 to 7.

2.2 Acoustic Labeling

The Praat software was used to display and label the data. By a process of trial and error we determined that the critical facts about these CVC monosyllables could be

captured with five labeling tiers: one for the beginnings and ends of noisy regions, one for the beginnings and ends of voiced regions, one for the releases of the stops, one for irregular phonation, and one for comments; see Figure 3.

On the noise tier we label the beginning and end of noise regions, including a) aspiration and later noise associated with the release of the onset consonant, b) noise associated with the end of the vowel, and c) aspiration and later noise associated with the release of the coda consonant. For this label we do not attempt to determine the source of the noise, but only its beginning and end points. On the voice tier we label the beginning and end of voiced regions, including the nuclear vowel, the voice bar and epenthetic voicing produced after the release of the coda. On the release tier we label the onset of the release burst for each consonant, with multiple release bursts coded as e.g. onset2, onset3 etc. On the IPP tier we insert a single ‘ipp’ label (for ‘irregular pitch periods’), and ‘glot’ when the speech is quite apparently glottalized. We do not attempt to determine the beginning and end of these regions because this has proven difficult to do reliably. Finally, in the comments tier we label any other events that seem potentially significant, such as changes in vowel amplitude or voice quality, overlap with another speaker, or other extraneous acoustic signals not produced by the child’s vocal tract.

This set of labels is designed to enable us to analyse the presence or absence of various release bursts as well as noisy, voiced and irregularly phonated regions, their temporal duration, and their degree of overlap. Some preliminary results are described in Section 3 below.

3 Results

Our initial labeling effort has revealed several observations that tell us something about the noise observed at the end of the vowel. For example, in this limited corpus the noisy region occurs primarily in association with voiceless codas; it does not always occur at precisely the same location with respect to the end of the nuclear vowel; and it sometimes makes up the entire vocalic nucleus.

3.1 End-of-vowel noise occurs largely for [-voice] codas

As reported in [8] these two children often produced noise at the end of the vowel in association with voiceless codas, but almost never with voiced codas. The distribution of this cue for these two coda voicing conditions is shown in Table 1, showing data from that study for three of the 6 sessions for each child. Note that the occurrence of end-of-vowel noise in predominantly voiceless coda contexts is found for both velar and labial codas. This makes it less likely that the noise is generated simply by early or prolonged supralaryngeal gestures. This might be a plausible account of noise before a velar stop, because the tongue dorsum is known to move relatively slowly. However, it is a less-plausible account for the labial stops, because the lips can move much more quickly.

	Voiceless coda		Voiced coda	
	/p/	/k/	/b/	/g/
C01	21/34	38/40	3/35	1/29
C09	25/39	30/30	6/20	2/24

Table 1: Tokens with noise at the end of the vowel

3.2 Location of the noise within the voiced portion is variable

In some tokens the noise overlapped substantially with the voiced portions of the vowel, while in others it occurred only after voicing ceased. The latter observation would be consistent with incomplete closure at the supralaryngeal constriction. An example is shown in Figure 2. In other cases the noisy region occurred near the midpoint of the vowel, as shown in Figure 1. This variation in the location of the noisy region might support the hypothesis that the appearance of this cue is largely due to incomplete control of the articulatory apparatus. On the other hand it might be interpreted as evidence that the child intends to produce some noise as a cue to the [-voice] feature of the coda stop, and is not concerned to control the precise location of its occurrence.

3.3 Full de-voicing of the vowel nucleus occurs only between a [-voice] onset and a [-voice] coda

The four target words examined for this study included *bug*, *duck*, *cup* and *tub*. Among them, these four words include all possible combinations of [+voice] and [-voice] in both onset and coda position. We noted that one of the words, *cup*, was often produced with an entirely voiceless vowel, i.e. it was whispered, as illustrated in Figure 4. Again, this might result from one of two possible mechanisms: the child's immature ability to turn the laryngeal vibration on and off quickly in such a short vowel with a preceding long aspirated VOT and a following noise cue at the end of the vowel, or the child's intention to provide strong cues to the voicelessness of the stops in onset and coda positions of the word. The fact that very few tokens of the other three target words were produced with no voiced region in the vowel, i.e. with 'whispered' voice quality, would be consistent with either of these hypotheses.

4 Discussion

The proposed labelling method for noise, voicing and glottalized regions as well as release events has revealed several interesting observations about the cues to the feature [voice] in stop consonants produced by two children in the third year of development. Once we know the extent and acoustic shape of the cues used by child speakers to signal voicing contrasts (and other contrasts) in various structural positions, it will be possible to pose a number of important questions about this distribution that have significant implications both for theories of how cues to distinctive feature contrasts can vary, for models of the

speech production planning process, and for models of the development of this process in early childhood. For example:

-To what extent are the child's distinctive feature cue patterns different from those of the adult---particularly the child's adult caretaker?

-How do the cues vary with the context in which a sound is produced? E.g. onset vs coda consonants, pre-stressed-vowel vs. pre-reduced-vowel consonants, word-onset vs other consonants, morphemic vs non-morphemic consonants, singleton consonants vs. consonants in clusters.

-How do the cue patterns used by the child change during development?

-How are the acoustic cues used by caretakers speaking to children different from or similar to acoustic cues in adult-to-adult 'clear speech'?

-Is the end-of-vowel noise considered in this study produced at the larynx, or does it represent turbulence noise generated at a narrowing supralaryngeal constriction? It is often assumed that adults abduct (spread) their vocal folds to inhibit voicing in onset consonants, but may adduct (i.e. strongly approximate) them to inhibit voicing in codas, but it is possible that children are using spreading instead of adduction to inhibit voicing in codas.

Such questions, which can be addressed in part by close acoustic analysis of data labelled in the proposed manner, have the potential to help distinguish among models of the development of the production planning process in children.

5 Conclusion

A proposed method to standardize the labeling of acoustic cues to the voicing contrast in coda stops for child CVC utterances has begun to reveal new facts about the cues that children produce. The application of this labeling convention to samples from both the growing number of acoustically recorded corpora of child speech, and the similarly expanding number of corpora of adult speech, (including adult speech to children) may resolve a number of outstanding theoretical issues. By applying a consistent scheme of labeling acoustic events to these various speech types, and using these labels to guide quantitative acoustic analysis, it may be possible to determine precisely what is happening in the child's vocal tract, and what those events imply about a model of the child's developing ability to produce adult-like cue patterns to distinctive feature contrasts.

Acknowledgments

This work was supported in part by NIH/NIDCD Grant DC00075 to Kenneth Stevens. We gratefully acknowledge the valuable input provided by Dr. Stevens as well as by Dr. Katherine Demuth. Any errors are our own.

References

- [1] O. Gordeeva and J. Scobbie, "Non-normative preaspirated voiceless fricatives in Scottish English: phonetic and phonological characteristics", *Queen Margaret College Speech Science Research Center Working Paper WP-12*, Edinburgh (2007)
- [2] A. Imbrie (2005). "Acoustical study of the development of stop consonants in children," Ph.D. thesis, Cambridge, MA: MIT. <http://hdl.handle.net/1721.1/27198>
- [3] J. Kornfeld, "What initial clusters can tell us about a child's speech code", *MIT RLE Quarterly Progress Report 101* (1971)
- [4] P. Ladefoged and I. Maddieson, *The Sounds of the World's Languages*. Oxford, UK: Blackwell Publishers (1996)
- [5] J. Laver, *Principles of Phonetics*. Cambridge: Cambridge University Press (1994)
- [6] M. Macken and D. Barton Macken, "The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants", *Journal of Child Language* 7: 41-74 (1980)
- [7] J. Scobbie, F. Gibbon, W.J. Hardcastle, P. Fletcher, "Covert contrast as a stage in the acquisition of phonetics and phonology", *Queen Margaret College Working Papers in Speech and Language Sciences* 1, 43-62 (1996)
- [8] S. Shattuck-Hufnagel, K. Demuth, H. Hanson and K. Stevens, "Acoustic cues to voicing contrasts in coda stop consonants in 2-3-year old speakers of American English" (submitted)
- [9] K. Stevens and S.J. Keyser, "Quantal theory, enhancement and overlap." Plenary talk at Colloquium on Quantal Theory, presented in Paris, France, July 2006. To appear in Special Issue of *J. Phonetics* (In press).

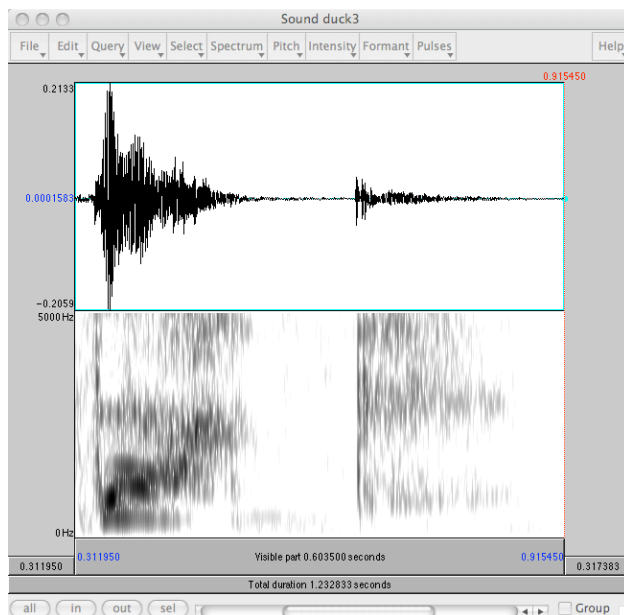


Figure 1: Subject C01, "duck3"

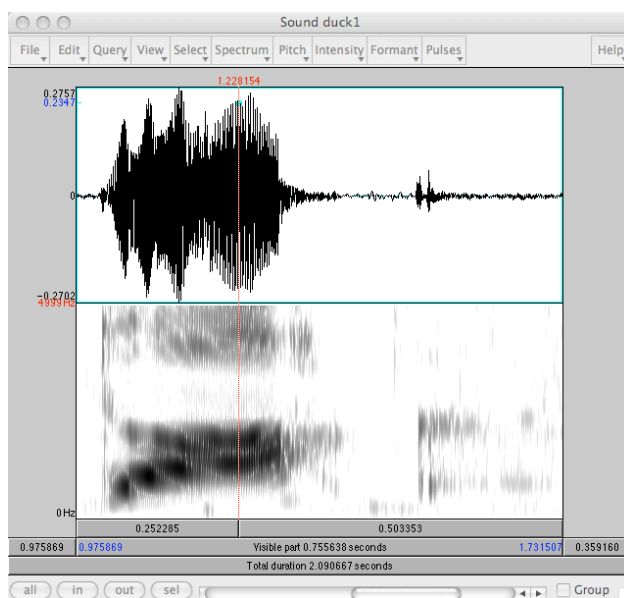


Figure 2: C09, "duck1"

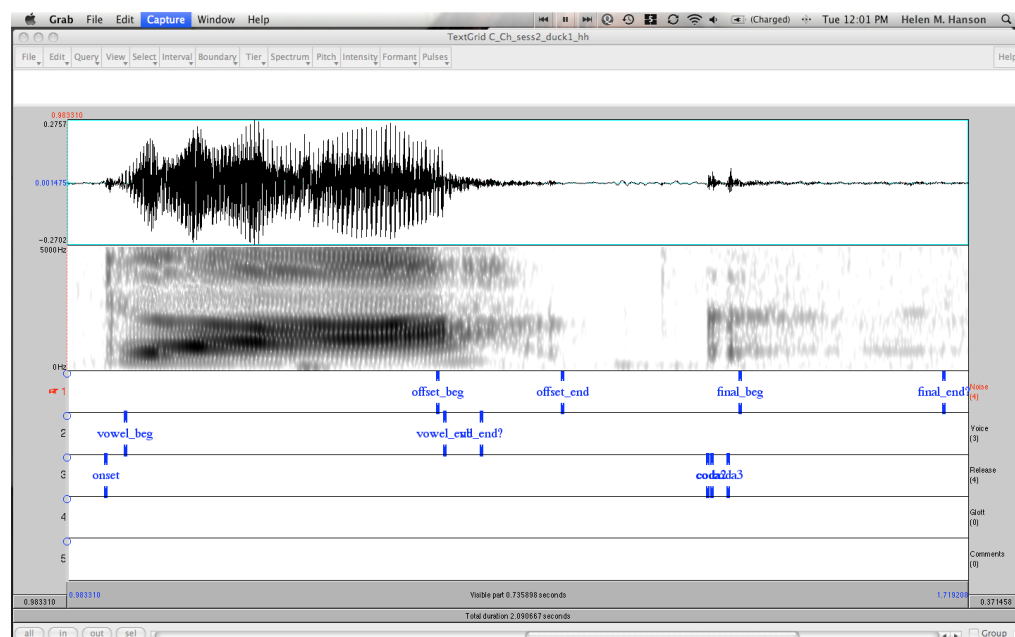


Figure 3: C09, "duck3"

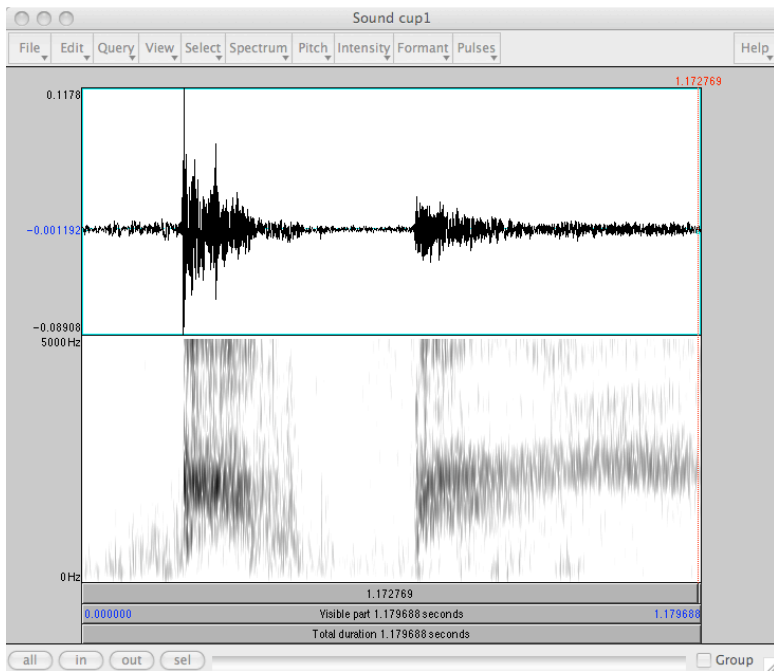


Figure 4: C01, "cup1"