HOMONYMS AND CLUSTER REDUCTION IN THE NORMAL DEVELOPMENT OF CHILDREN'S SPEECH

S. McLeod, J. van Doorn, and V. Reed

School of Communication Disorders
The University of Sydney

ABSTRACT - As children are learning to speak, they sometimes reduce consonant clusters to a single element, and as a result produce a homonym (e.g. they say "top" for "stop"). There is some evidence in the literature which suggests that even though these words may sound the same, they have differences which can be detected with acoustic analysis, indicating that the children are making covert distinctions between the two contexts. The purpose of this study was to make acoustic comparisons of homonym pairs produced as a result of cluster reduction by a group of 16 young children (2;0 to 2;11 years). Duration and relative energy of aspiration for stops /k/ and /t/, duration and spectral distribution for fricative /s/, and voice onset time (VOT) for /k/ were measured in several word-initial contexts. Results showed that for word-initial /s/ plus stop clusters which had been reduced to a stop, the aspiration duration for the stop in the cluster target word was significantly less than that for the singleton target word. No other time or spectral measures reached statistical significance. The results have been interpreted in terms of phonological and speech motor development in children.

INTRODUCTION

The speech of very young children is quite frequently characterised by homonyms which occur as a result of the child producing only one member of an initial consonant cluster (e.g. "tick" meaning "stick"). It has been suggested that when children produce homonyms in this way, they make subtle distinctions that are not perceptible to a listener, but are detectable acoustically (Scobbie, 1995; Weismer, 1984), or physiologically (Gibbon, 1990). Information about covert differences between the production of words containing consonant clusters and those containing their singleton counterparts has the potential to provide useful insight into children's developing phonological systems.

In order to interpret any findings in the developmental data obtained from acoustic features of children's speech it is important to have information about the equivalent features in adult speech. Acoustic studies on consonant clusters for English speaking adults concentrate on temporal features. They consistently report that, in general, consonants have significantly shorter duration in cluster environments than in singleton contexts (Klatt, 1976; O'Shaughnessy, 1974; Umeda, 1977). Particularly relevant to the present study are findings that VOT is reduced for stops in /s/ plus stop environment (Davidsen-Neilsen, 1969; Klatt, 1975; O'Shaughnessy, 1974), and that duration of /s/ is less when /s/ is the first element in two element consonant clusters (Haggard, 1972; Weismer & Elbert, 1982). A reported exception to the generally shorter duration of consonants in cluster context is VOT for stops in stop plus /l,r/ clusters, where VOT increases (Menyuk & Klatt, 1975; O'Shaughnessy, 1974). It has been acknowledged that the durational changes of consonants in cluster environments could have several origins: phonological conditions of the language; automatic consequences of coarticulatory speech motor mechanisms; and perceptual constraints (Klatt, 1975; 1976; O'Shaughnessy, 1974).

Temporal data for consonant clusters in children's speech generally reflects the same trends as adult data (see Weismer, 1984 for a review) - shorter VOT for stops in /s/ plus stop context (Scobbie, 1995), shorter duration for /s/ as the first element in two element clusters (Gilbert & Purves, 1977; Hawkins, 1979; Weismer, 1984; Weismer & Elbert, 1982), and longer VOT for voiceless stops in stop plus /l,r/ clusters (Menyuk & Klatt, 1975). Whenever children's speech is compared with adult speech, there are universal findings of longer segment durations with increased variability (Kent, 1976; Weismer, 1984). There has been some debate about the nature of the variability that is so characteristic of children's speech: whether it is a mathematical consequence of slower speech with longer segment durations, or whether it is an independent consequence of a less mature speech motor system (Smith, 1994). On balance there is evidence to suggest that the degree of variability cannot be entirely accounted for by the longer segment durations found in children's speech, but is a direct consequence of a more variable speech motor production system (Kent, 1992; Smith, 1994).

In addition to the acoustic studies on correctly produced consonant clusters in children's speech, there are also reports on acoustic characteristics of reduced clusters, where children have substituted stops for /s/ plus stop word-initial clusters (Bond & Wilson, 1980; Catts & Khami, 1984). In a cross sectional study of a group of 5 children aged 1;10 to 3;0 years, Bond & Wilson (1980) found that two children used long lag VOT for the substituted stop, one child used short lag VOT, and the other two used both. In contrast, Catts & Khami (1984) performed a longitudinal study with measurements for 6 children, initially aged 1,9 to 2;10 years, over a period of 5-17 months (until correct cluster production was achieved), and found that the children consistently produced short lag stops as substitutes for clusters. Neither of these studies made a direct comparison of VOT for the singleton and substituted cluster contexts to establish whether the VOT for long lag stops used as a cluster substitutes was shorter (but still in the long lag range) than that of stops in singleton context.

The discovery of any acoustic subtleties in children's speech will provide valuable insight into the nature of children's developing phonological systems. One debate in theories of child phonology centres around the nature of children's underlying representations (URs), particularly whether or not children's URs are identical to adult surface forms. Maxwell (1984) provides an excellent summary of the two viewpoints with an overview of phonological theories that had been proposed prior to 1984. Since then, Spencer (1988) has suggested a more elaborate model that allows for the child's UR to be different from the adult surface representation at two levels: first via a perceptual filter to establish what Spencer calls the child's "input representation", and secondly at a level where the child encodes input representations as "output underlying representations" through a "realisation process". If acoustic analysis is able to uncover covert contrasts of perceptually similar phonemes in different contexts, then it will provide a way to test some aspects of existing theories of developmental phonology.

This paper is an account of an acoustic analysis of homonyms produced as a result of cluster reduction in young children's speech. The aim of the investigation was to establish whether there is evidence of a developmental stage in speech production where children acoustically mark the substituted consonant in reduced consonant clusters, and if so to establish whether the marking is consistent with phonological features of the ambient language.

METHOD

Subjects

The 16 children who participated were aged from 2;0 to 2;11 years at the commencement of the study. On the basis of a series of standardised tests all subjects were judged to be within normal limits for their hearing, speech, language and cognition. All children had acquired voicing contrast for word-initial stop consonants prior to the start of the study.

Speech corpus

The speech corpus from which homonyms were identified consisted of 6 separate recordings (one recording per month) of 12 sets of word triplets for 16 children. Each word triplet contained a word with an initial consonant cluster and two of its potential homonyms. This study used only those word triplets that contained the clusters /kl/, /st/, /sk/, /sn/, /sw/. Thus the 8 word triplets were {clap, cap, lap}; {click, kick, lick}; {snail, sail, nail}; {snow, sew, no}; {ski, key, sea}; {stick, tick, sick}; {sweet, wheat, seat}; {swing, wing, sing}.

Recording procedure

Longitudinal data was recorded at monthly intervals over a period of 6 months for each child. At each monthly recording session the children performed a speech production task in which they said 12 sets of three words. Computer graphics of the 36 words were presented in a sequence which kept separate any potential homonyms. The words were elicited spontaneously as far as possible, where single words were produced in response to a computer graphic. If necessary an additional spoken prompt from an adult speaker was provided. The recordings were made on metal audio cassette tapes with a Marantz CP430 recorder using a Shure TP ETP Radio Microphone under quiet conditions in the children's homes.

Speech tokens

In order to identify any homonyms produced by the children, all single word productions in the speech corpus (36 words x 16 children x 6 visits) were transcribed by one of the experimenters using broad and narrow transcription. Those words which had the same broad transcriptions were considered to be homonyms. Only those homonyms with target clusters /kl/, /st/, /sk/, /sn/, /sw/ were used for acoustic analysis in this study. The actual homonyms which were measured were "click" realised as "kick" (16); "clap" as "cap" (8); "stick" as "tick" (5) and as "sick" (7); "ski" as "key" (7) and as "sea" (3); "snow" as "sew" (8), "snail" as "sail" (6); "sweet" as seat (1), and "swing" as "sing" (2).

Acoustic analysis procedure

The selected homonyms were digitised at 20kHz and stored as individual computer files using a Kay CSL 4300B. The standard waveform, spectrogram, energy contour and LPC spectra facilities available on the CSL were used to make the duration and spectral measures. The actual acoustic measures were (a) duration and energy of aspiration for /t/ and /k/ where the target cluster /s/ plus stop was realised as the stop consonant, (b) duration and frequency spectra of /s/ where the target cluster /s/ plus /n,m,t,k,w/ was reduced to the first element /s/, and (c) VOT of /k/ when the target cluster /kl/ was reduced to /k/.

Measurements

Aspiration duration of stops /k/ and /t/ was defined as being equivalent to voice onset time, with
Onset: the time of the release of the plosive as indicated by a sharp spike in the waveform, and
Offset: the time of the first appearance of periodicity in the waveform.
Voice onset time for /k/ was defined using the same criteria.

Aspiration energy (measured over 20ms in the middle of aspiration following a stop burst) was measured relative to the energy of the following vowel (measured mid-vowel). Measurements were taken from cursor readings on energy contours generated by the CSL 4300B.

Duration of fricative /s/ was measured as the time between onset and offset, with
Onset: The simultaneous appearance of high frequency energy above baseline in the waveform and wideband spectrogram.
Offset: The return to baseline of the high frequency signal, or the appearance of periodicity in the waveform for the onset of voicing of the following vowel.

The spectral properties of /s/ were sampled at two points in the middle of the waveform, and LPC spectra (up to 10kHz) were generated for each sample. The spectra were compared visually to look for evidence of differences of spectral distributions for the two members of each homonym pair.

Measurement error and reliability

Reliability for phonetic transcription was established by having ten percent of the entire speech corpus transcribed by another judge. The phoneme-by-phoneme agreement for broad transcription was 0.92. For acoustic analysis, the measurement error was dominated by the ability of the experimenter to locate segment boundaries according to the operational definitions given above. Generally the accuracy was within 5 ms, or one pitch period for children's speech. Speech tokens where this accuracy could not be achieved because onset and offset boundaries were too difficult to identify were omitted from measurement.

Statistical Analysis

The aim of this experiment was to make paired comparisons of homonyms produced by the same child within the same session. Thus comparisons of consonants in singleton and cluster-substitute contexts were made using paired two-sample t-tests for all measurements of duration and energy.

RESULTS

Aspiration following /k/ and /t/

For 12 measured token pairs, the mean aspiration duration of stop consonants in those target words which contained consonant clusters (26.3±18.5 ms) was found to be significantly shorter (P<.0001) than that for stop consonants in their counterpart singleton target words (80.2±40.4 ms), using a one-tailed paired two-sample t-test. The mean ratio of vowel energy to aspiration energy was larger for cluster target words (23.2) than for singleton target words (10.3) (indicating a lower aspiration energy in the cluster targets), but the large degree of variability in the results did not bring this result to significance. Figure 1 shows a typical comparison of stop consonant /t/ in the two different linguistic contexts (cluster target vs singleton target).

Acoustic features of /s/

The mean duration for /s/ in the reduced clusters (201.8±76.7ms) was not significantly different from the singleton context (200.4±70.1ms) for 25 measured token pairs across several target cluster environments. Additionally, visual inspection of overlaid spectra of /s/ from the paired contexts did not reveal any consistent differences for specific cluster targets, nor for specific subjects. Figure 2 shows a typical comparison of the /s/ spectra in cluster target vs singleton target context. The spectra do not display any distinctive differences in frequency distribution for the two different articulation contexts.

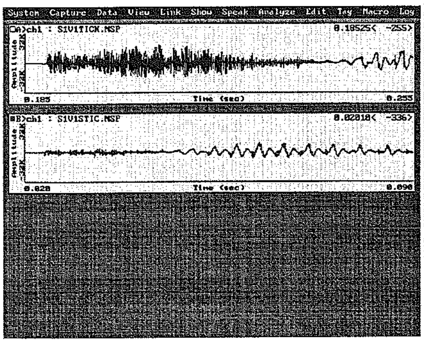

Figure 1. Aspiration following stop burst for /t/ in contexts of word initial singleton (top trace) and reduced cluster (bottom trace).

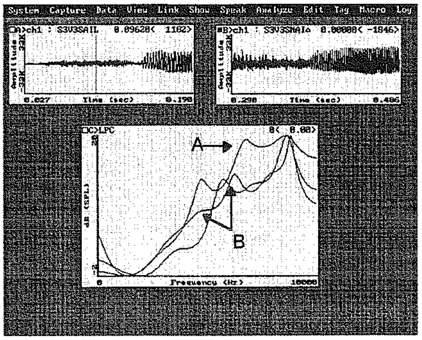

Figure 2. Sample LPC spectra of /s/ in contexts of word initial singleton (A) and reduced cluster (B).

Voice Onset Time for /k/

For 24 token pairs the mean VOT was greater for the target cluster context (90.8±54.6ms) than for the singleton target context (72.3±44.2ms), but the difference did not reach significance due to the large degree of variability. Inspection of individual speech tokens reveals that the direction of the change in VOT from one context to the other (positive or negative) was inconsistent, thus illustrating little conclusive evidence of a consistent trend towards a larger VOT in the cluster target context.

DISCUSSION

The search for acoustic markers in consonants which are used in the context of a reduced cluster compared with their singleton context revealed one statistically significant result: reduced VOT for stop consonants /t/ and /k/ in /s/ plus stop cluster target context compared with singleton context. The high level of significance is supported by the individual data, where all tokens except one show at least 40% reduction in VOT. It is interesting to note that the range of VOTs for the intended clusters is 18-58ms, with about a third of the tokens having VOTs still in the long lag region (>25ms) even though the VOT was significantly less than that of the singleton counterpart. If the data from the present study had been analysed in a manner similar to that used by Bond & Wilson (1980) or Catts & Khami (1984) where the VOTs were simply categorised as short lag or long lag, it would have supported the Bond & Wilson results, where some of the reduced cluster VOTs were in the long lag region, and some were short lag. However, paired comparisons of singleton and intended cluster productions found that even the long lag VOT tokens had VOTs which were at least 40% shorter than the singleton production for the same child at the same recording session.

The children's ability to consistently reduce VOT during their attempts to produce an /s/ plus stop cluster supports the notion that the children are demonstrating a phonological awareness of the features of the stop consonant as the second member of an /s/ plus stop cluster (i.e. a deaspirated stop with short lag VOT). In Spencer's terms (Spencer, 1988) it is likely that their internal representations are equivalent to the ambient language surface forms. However, even though the VOT was shorter for stop consonants used as intended clusters, sometimes it was still in the long lag range. There are two possible interpretations of this finding: either that the long lag VOTs are in fact within the range which can be found in the correct production of /s/ plus stop clusters, and thus the long lag VOT is simply a reflection of the expected range of VOTs in /s/ plus stop clusters for young children; or, using the Spencer model (Spencer, 1988) it may be that those children who are producing long lag VOT have a correct internal representation, but they have not yet achieved a correct output underlying representation, and are realising the stop with a shorter VOT, but which is not yet in the short lag VOT range.

The negative and highly variable findings for /s/ duration did not reflect any trends towards the "adult-like" shorter duration of /s/ as the initial element of two element clusters. It is interesting to note that the significance of shorter /s/ duration in adult speech is not well understood (Klatt, 1976). Although /s/ duration has been reported as being more than 30ms shorter in cluster context than in singleton context, and thus could well be perceptible, it is not formalised as a phonological rule of the language. The shorter duration could indeed be partly or wholly due to physiological efficiencies in the adult speech motor system. The highly variable result for /s/ duration in this study has two likely origins: one experimental and one physiological. Experimentally, the use of isolated words as speech tokens resulted in perceptible differences in rate of production of the same word by the same child, and this will have resulted in fluctuations in segment duration. Physiologically, speech segment duration is more variable in children's speech as a result of slower speech, and a less mature speech production mechanism (Kent, 1976; Smith, 1994). The highly variable speech segment data for /s/ may well be masking any evidence that children possess a phonological awareness of the cluster that they are trying to produce. In fact, the reported high degree of variability in duration of /s/ in correctly produced /s/ clusters by children (Weismer, 1984) would support this notion.

The result for the VOT of /k/ as the first element in /kl/ clusters could be interpreted in exactly the same way as /s/ duration, where acoustic noise due to experimental conditions of speech production, and due to the immaturity of children's speech production systems, may be masking any acoustic evidence of their phonological awareness of the cluster. Once again the role of increased VOT of /k/ as the first member of /k/ plus /l,r/ clusters in adult speech is not fully understood. It is difficult to know whether the observed increase found in adult speech is even perceptually detectable. Answering questions about the developmental features of children's speech in this instance may not be possible until more is known about the significance of acoustic durational measures in adult speech.

In summary, these results raise questions about the role of acoustic measurement as a key to unlocking various theoretical standpoints on children's phonological development. It is interesting to note that the one highly significant result in this study occurred for the well defined phonological feature of stop clusters in English: that they are deaspirated in the context of second element in /s/ plus stop clusters. Other measurements in the present study were made on features where the phonological role of changes to the segment in cluster environment is not as well understood. High levels of variability in those measurements led to inconclusive interpretations of experimental findings. In these instances, it will continue to be difficult to interpret children's data until more is known about the nature of relevant acoustic differences in consonants as cluster elements in adult speech. Are changes in segment duration found for cluster context in adult speech phonological or coarticulatory in nature, and can they be perceptually distinguished? Insights into these questions will make interpretation of acoustic developmental speech data for children's speech more meaningful.

REFERENCES

Bond, Z.S. & Wilson, H.F. (1980) */s/ plus stop clusters in children's speech,* Phonetica, 37, 149-158.

Catts, H.W & Khami, A.G. (1984) *Simplification of /s/ + stop clusters: A developmental perspective,* Journal of Speech and Hearing Research, 27, 556-561.

Davidsen-Neilsen, N. (1969) *English stops after initial /s/,* English Studies 50, 321-340.

Gibbon, F. (1990) *Lingual activity in two speech disordered children's attempts to produce velar and alveolar stop consonants: Evidence from EPG data,* British Journal of Disorders of Communication, 25, 329-340.

Gilbert, J.H.V. & Purves, B.A. (1977) *Temporal constraints on consonant clusters in child speech production.,* Journal of Child Language, 4, 417-432.

Haggard, M. (1972) *Abbreviation of consonants in English pre- and post-vocalic* clusters, Journal of Phonetics, 1, 9-24.

Hawkins, S. (1979) *Temporal coordination of consonants in the speech of children: further data.,* Journal of Phonetics, 7, 235-267.

Kent, R.D. (1976) *Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies,* Journal of Speech and Hearing Research, 19, 421-447.

Kent, R.D. (1992) *The biology of phonological development,* in C.A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds) *Phonological development; models, research, implications,* (York Press: Timonium MD) pp. 65-90.

Klatt, D.H. (1975) *Voice onset time, frication and aspiration in word initial consonant clusters,* Journal of Speech and Hearing Research, 18, 687-703.

Klatt, D.H. (1976) *Linguistic uses of segmental duration in English: Acoustic and perceptual evidence,* Journal of the Acoustical Society of America, 59, 1208-1221.

Maxwell, E. (1984) *On determining underlying phonological representations of children: A critique of the current theories,* ASHA Monographs, 22, 18-29.

Menyuk, P. & Klatt, M. (1975) *Voice onset time in consonant cluster production by children and adults,* Journal of Child Language, 2, 223-231.

O'Shaughnessy, D. (1974) *Consonant durations in clusters,* IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-22, 282-295.

Scobbie, J.M., Hardcastle, W.J., Fletcher, P., & Gibbon, F. (1995) *Consonant clusters in disordered L1 acquisition. A longitudinal acoustic study,* Proceedings of ICPhS, Stockholm, Vol 2, 706-709.

Smith, B.L. (1994) *Effects of experimental manipulations and intrinsic contrasts on relationships between duration and temporal variability in children's and adult's speech,* Journal of Phonetics, 22, 155-175.

Spencer, A. (1988) *A phonological theory of phonological development,* in M.J. Ball (Ed) *Theoretical linguistics and disordered language,* (Croom-Helm: London) pp. 115-151.

Umeda, N. (1977) *Consonant duration in American English,* Journal of the Acoustical Society of America, 61, 846-858.

Weismer, G. (1984) *Acoustic analysis strategies for the refinement of phonological analysis,* ASHA Monographs, 22, 30-52

Weismer, G. & Elbert, M. (1982) *Temporal characteristics of "functionally" misarticulated /s/ in 4- to 6-year old children,* Journal of Speech and Hearing Research, 25, 275-287.