How Word Onsets Drive Lexical Access and Segmentation: Evidence from Acoustics, Phonology and Processing

David W. Gow Jr.[1], Janis Melvold [1], and Sharon Manuel [2]

[1]Massachusetts General Hospital/ Harvard Medical School
[2]Massachusetts Institute of Technology

ABSTRACT

We will argue that the beginnings of words are perceptual "islands of reliability" in connected speech, and that their perceptual and temporal properties allow them to drive critical aspects of spoken word recognition including lexical segmentation. This argument rests on three generalizations derived from research in speech science, phonology, and psycholinguistics. We suggest that word onsets differ from other parts of words in that: (1) they offer more robust and redundant acoustic evidence about phonetic features, (2) they are generally protected from phonological assimilation, neutralization and deletion and therefore show less lawful variation in surface realization, and (3) they may activate lexical representations which facilitate word perception and thus diminish listeners' dependence on veridical acoustic-phonetic processing of other portions of words. These properties of word onsets allow them to drive lexical segmentation by facilitating the recognition of items that begin with clear onsets. The implications of these findings for several models of lexical segmentation and spoken word recognition are discussed.

1. INTRODUCTION

Research has demonstrated that lexical segmentation is affected by both acoustic-phonetic [e.g. 1,2,3] and lexical [e.g. 4] factors. Theorists have responded to this state of affairs by proposing models that range from positing that segmentation is triggered by acoustic-phonetic factors [2], to models that posit segmentation as a purely lexical process with no direct role for these factors [5]. Recently, several new approaches have been suggested that attempt to integrate lexical and prelexical approaches into a single model [3,4]. The basic premise behind these approaches is that lexical segmentation is the result of recognizing a word, and that prelexical factors affect segmentation indirectly by facilitating recognition. In this paper we will consider the nature of the prelexical factors that appear to be relevant to segmentation. We will also consider how they might be suited to drive known spoken word recognition processes.

There is good reason to believe that the beginnings of words are the loci of acoustic-phonetic factors affecting segmentation. It has been noted that most hypothesized prelexical juncture cues, including phoneme lengthening, glottalization or laryngealization of vowels, aspiration of voiceless stops and the occurrence of strong syllables occur *after* word juncture, at the beginnings of words [3]. This distribution is somewhat surprising from the perspective of a purely prelexically-driven view of segmentation because it does not allow listeners to spot an onset until after they have heard it. Because recognition could only begin after word boundaries had been identified, and boundaries could not be identified until after the onset of a word, recognition would necessarily be delayed, and would involve some backtracking to recover the onset. This distribution is less surprising though if onsets facilitate lexical access, as they play a central role in most models of lexical access [3,5,6].

In the sections that follow we will draw on evidence from several fields to examine the claim of one segmentation model, the Good Start Model [3], that word onsets are perceptual islands of reliability in normal reduced connected speech, and that they are well-suited to drive lexical mechanisms that allow listeners to recognize words when other portions of the speech stream are less salient or reliable.

2. ACOUSTIC-PHONETIC CHARACTERISTICS OF ONSETS

Here we review several factors that have the effect of enhancing the acoustic properties of phonetic features at the beginnings of words. These factors include syllable-position, stress patterns, word-position as an independent factor, and juncture markers. Taken together, these factors suggest that word-initial phonemes ought to be more easily and readily identified than phonemes that occur later in words.

First, word-initial consonants benefit from also being syllable-initial. Generally, phonetic features appear to be more clear for syllable-initial consonants than they are for syllable-final consonants [7,8]. For example, place of articulation for stop consonants is conveyed both by formant transitions and by release bursts. Whereas both of these acoustic cues are produced for syllable-initial consonants, syllable-final consonants are often made without a release burst. With respect to voicing, it is well known that "voice onset time" (VOT) is a perceptually salient cue for the distinction between English voiced stops (short VOTs) and voiceless stops (long VOTs) for *syllable-initial* consonants. These VOT differences are particularly important, as both phonologically voiced and voiceless stops may have some vocal fold vibration during the oral closure phase. The long VOT of syllable-initial voiceless stops derives from the fact that for these stops, the vocal folds are apart (and thus unable to vibrate) at the release of the stop. However, for English *syllable-final* stops, devoicing is often achieved by constricting, rather than opening, the glottis. Thus, VOT is not a cue for voicing for syllable-final stops, which may not be released in any case. Glottal constriction

made near the time of oral implosion for a syllable-final stop can be a cue for voicelessness, because if it is timed early with respect to the oral implosion, it will effectively cut off the formant transitions. Unfortunately, this strategy has the effect of decreasing the amount of formant transition available for place identification. In any case, the fact that the phonological distinction between voiced and voiceless stops is lost in syllable-final position for many languages suggests that voicing cues are not as salient in syllable-final position.

Second, many word-initial consonants benefit from the fact that they are stressed. In a large corpus of spontaneous English conversation, 74% of content word tokens (37% of all word tokens) had some stress on the first syllable [9]. Stress enhances the acoustic manifestation of phonetic properties such as voicing by increasing VOT for syllable-initial voiceless stops. In addition, the duration of the hold phase of consonants is increased for stressed syllables [10]. This duration effect is also seen for syllable-initial versus syllable-final consonants [11]. Increased duration of consonant constriction phases should particularly enhance the saliency of those features which are typically characterized by relatively steady states, e.g. [continuant], and for fricatives, [place features]: longer closures mean the feature is present for more time. Furthermore, all other things being equal, increasing the constriction phases should decrease the amount by which a consonant is "coproduced" with adjacent segments, thus decreasing acoustic coarticulatory effects. Finally, longer durations have the effect of providing more time before the next phoneme dominates the signal for the listener to process the signal.

Of course, not all syllable-initial consonants are word-initial, not all stressed syllables are word-initial, and not all word-initial syllables are stressed.[1] If the factors of syllable-position and stress are held constant, then is there any special additional advantage to being in word-initial position? Several studies suggest there is. For example, word-initial, prestressed consonants are longer than prestressed, syllable-initial consonants in other positions in a word [10], and voiceless, syllable-initial consonants in stressless syllables have longer VOT than they do in word-medial position [12]. It has also been shown that both stressed and nonstressed word-initial syllables show less of a within-category trend towards vowel reduction than matched non-initial syllables [13].

Finally, it may be that glottalization of word-initial vowels, which is often considered to be simply a juncture cue, also has the function of enhancing the identification of the vowel quality. One effect of glottalization is to decrease the bandwidths of the formants. It may be that listeners are better able to quickly determine the formant frequencies, and thereby identify the vowel quality, with such narrowed formants.

In summary, while much work needs to be done to tease apart the various factors that can affect the acoustic saliency of the broad range of phonetic features, there is evidence to suggest that word beginnings are especially favorable sites for identification of phonemes.

[1] In fact, in many languages (e.g. French) stress is fixed on post-initial syllables in a word.

3. PHONOLOGICAL PROCESSES

In this section we will argue that word onsets are phonologically more stable than are segments in other positions, and thus provide particularly invariant sources of information about underlying phonological form. Phonological processes applying to underlying forms result in lawful phonological variation between underlying and surface forms of a word. Processes of assimilation, neutralization, and deletion are commonly triggered by the phonological conditions produced by affixation of morphemes to a stem or by construction of syntactic constituents.

A survey of phonological processes across languages suggests some interesting asymmetries between the behavior of word-initial and word-final segments. For instance, in a number of languages, including Russian and German, word-final stop consonants are regularly devoiced (d->t; b->p; g->k). This results in alternations within an inflectional paradigm, as illustrated in the following example from German:

underlying form of stem		**nom. sg.**	**nom. pl.**
rad	'wheel'	rat	räder
rat	'advice'	rat	räte

This type of process neutralizes a phonemic distinction (voicing) within the language, in a particular context. The counterpart to this kind of process, namely neutralization at the word-initial boundary, appears to be relatively rare in languages of the world.

It is well known that phonological processes apply not only within, but across word boundaries as well. In certain types of noun phrases in English, for example, a final stop consonant of the first word may assimilate to the following word-initial consonant (or possibly delete): **bat cave --> /bæ k/ cave**. What we don't find are word-initial consonants assimilating to a preceding word-final consonant: **back tap --> back /kæ p/**.

Our generalization is strengthened by evidence of asymmetries in phonological variation among dialects and from studies of diachronic phonological change. For instance, it appears that the word-final boundary is the locus of most of the phonological variation between Standard American English (SE) and Black English (BE) dialects. SE words ending in a consonant cluster, for example, often have the final member of a consonant cluster absent in BE (**child-->/tʃɑjl/; guest --> /gɛs/**; a word-initial consonant belonging to a cluster does not delete. In a number of Nonstandard English dialects, a word-final coronal stop consonant (**/d,t/**) may delete even when they do not belong to a cluster (**rabid --> /ræ bI/; rabbit --> /ræ bI/**)[14].

A number of observers [e.g. 15] have pointed out that these kinds of phonological variations, which result in lexical ambiguity, may complicate the process of speech segmentation and lexical access. Listeners must balance the need to maintain strict criteria for the evaluation of mappings between acoustic-phonetic information necessary to distinguish between minimally different words, with the need to maintain sufficient

flexibility to recognize words that have undergone some rule conditioned change.

These asymmetries could be an accidental or arbitrary property of phonology. Or, a more intriguing possibility, and the hypothesis that we will pursue, is that they represent constraints which the speech perception system or phonetic principles and processes (or a complex interaction between those two) impose on phonology.

4. TEMPORAL DYNAMICS OF LEXICAL EFFECTS

In this section we will argue that word onsets, as perceptual islands of reliability at the beginnings of words, are uniquely suited to guide the perception of words marked by reduction or lawful phonological variation in the context of connected speech. When listeners listen to high quality tokens of citation-form speech, word recognition may be a simple matter of one to one bottom-up mapping between the acoustic-phonetic and lexical representations. When listeners listen to fluent connected speech, this kind of veridical bottom-up mapping may be precluded by degradation of the signal by reduction and phonological variation. When this is the case, listeners must rely on word-level representations to facilitate perception. Indeed, lexical effects are most robust when bottom-up mapping is underdetermined [16]. There is ample evidence that lexical hypotheses play a role in word perception. Lexical effects have been found in the perception of phonetically ambiguous, incomplete or distorted words in tasks including phoneme categorization [17], phoneme monitoring [18], phoneme restoration [16], shadowing [19], mispronunciation monitoring [19], and the identification of words in noise [20].

Lexical effects depend on the presence of sufficient veridical bottom-up acoustic-phonetic analysis to generate a usable cohort of lexical candidates. We have already argued that word onsets are the richest and least variant parts of the word, and thus are best suited to support this kind of analysis in simple informational terms. We would also like to argue that they are well-suited to play this role because lexical effects are well-suited to affect the processing of the ends of words in connected speech, but not the beginnings. In this sense, they might be thought of as acting from left to right. Across paradigms, lexical effects tend to be more reliable and robust at the ends of words than they are at the beginnings of words[16].

While left-to-right lexical effects appear to be the norm in the perception of fluent connected speech, right-to-left effects do occur under some conditions. We believe that the pattern of those conditions may reveal the source of the left-to-right bias observed under the most naturalistic conditions. Several studies have demonstrated that listeners may access words successfully when their onsets have been altered. A classic demonstration of the lexical identification shift involved stimuli with word-initial phonetic ambiguities [17]. Identity priming was found in a task involving the offline identification of citation-form word tokens presented in noise when prime and probe items had different initial phonemes (e.g. HAND primes SAND) [20]. Similarly, derived citation form primes facilitate lexical decision for semantically related visual probe stimuli [21], and the same result has been found using an intramodal auditory priming paradigm with derived primes, and primes with phonetically ambiguous onset phonemes [22]. It should be noted that another study [23] failed to show priming by derived primes with word-initial segment features changed. These studies share several features. Two of the four studies showing right-to-left lexical effects required offline responses, and all four relied on the exclusive use of citation form tokens of prime words presented in isolation. The offline nature of some of the tasks, and the fact that all of the studies examined the perception of single words, loosen critical temporal constraints that are found in connected speech perception. In normal speech perception there is a premium on rapid word recognition because listeners must keep pace with speakers. Phoneme monitoring studies have shown that factors that might slow the recognition of one word actually affect how quickly listeners are able to recognize the word that follows it [24]. There is thus a risk of backing up the word recognition system and falling behind the speaker if too much revision is necessary. The issue of keeping pace is particularly acute in the case of right-to-left effects, where possible top-down lexical effects on acoustic-phonetic processing would require maintaining a detailed echoic memory representation of word onsets, and possibly reallocating attention that would otherwise be used to attend to new information as it is being received. A second potential source of these effects is the fact that citation form tokens do not show the degree of reduction shown by normal connected speech, and so normal lexical effects may not play a major role in speech perception. If this is the case, then priming may reflect strategic or task dependent processes that differ from those found under normal listening conditions.

To summarize, due to the reduced nature of connected speech, listeners must rely on lexically mediated mechanisms to make sense of some parts of the speech stream. Research on such mechanisms shows that these processes work primarily from left to right in the perception of connected speech. The fact that right-to-left lexical effects play a role in the perception of citation-form tokens of individual words suggests that this left-to-right bias is the result of temporal constraints imposed by connected speech. It follows then, that onsets, as islands of perceptual reliability, are optimally located to facilitate the perception of words in connected speech through lexical effects. Within the context of a lexically-mediated model of segmentation such as the Good Start model [3], onsets may facilitate the recognition and thus segmentation of intended words in the speech stream. As a corollary, it is possible that the relative reduction and variability of non-onsets may limit the access of spurious embedded words (e.g. TAX in SYNTAX) by limiting early lexical activation and thus diminishing the possible role of lexical effects in their recognition.

5. SUMMARY

We have drawn attention to three tantalizing asymmetries between the beginnings of words and other parts of words that are consistent with the Good Start model [3]. The fact that patterns of acoustic-phonetic structure, phonological representation and lexical processing dovetail together so neatly hints at a possible interaction of constraints in the evolution of language.

6. REFERENCES

1. Nakatani, L.H., and Dukes, K,D.(1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America, 62,* 714-719.

2. Cutler, A., and Norris, D. (1988). The role of strong syllables in segmentation for lexical access. Journal of Experimental Psychology:Human Percep[tion and Performance, 14, 113-121.

3. Gow, D.W., and Gordon, P.C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 344-359.

4. McQueen , J.M., Norris, D., and Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 621-638.

5. McClelland, J.L., and Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

6. Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word recognition. *Cognition, 25,* 71-102.

7. Manuel, S.Y. (1991). Some phonetic bases for the relative malleability of syllable-final versus syllable-initial consonants. *Proceedings of the 12th International Congress of Phonetics Sciences, V,* 118-121.

8. Ohala, J. J. & Kawasaki, H. (1984). Prosodic phonology and phonetics, *Phonology Yearbook 1,* 113-128.

9. Cutler, A., and Carter, D.M. (1987).The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2,* 133-142.

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

11. Byrd, D. (1966). Influences on articulatory timing in consonant sequences. *Journal of Phonetics, 24,* 209-244.

12. Cooper, A. M. (1991). Laryngeal and oral gestures in English /P ,T ,K/. *Proceedings of the 12th International Congress of Phonetics Sciences, V,* 50-53.

13. Gow, D.W., and Gordon, P.C. (April, 1993). The distinctiveness of word onsets. Paper presented at the meeting of the Acoustical Society of America, Ottawa, Ontario.

14. Wolfram, W., and Fasold, R.W. (1974). *The study of social dialects in American English.* Englewood Cliffs. NJ: Prentice-Hall.

15. Frauenfelder, U., and Lahiri , A. (1989).Understanding words and word recognition: Can phonology help? In W.D. Marslen-Wilson (Ed.), *Lexical representation and process.* Cambridge: MIT Press, 319-341.

16. Samuel, A. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General, 12*5, 28-51.

17. Ganong, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110-125.

18. Segui, J. and Frauenfelder , U. (1986). The effect of lexical constraints on speech perception. In F. Klix and H. Hagendorf (Eds.), *Human memory and cognitive capabilities: Mechanisms and performance.* Amsterdam:North-Holland, 795-808.

19. Marslen-Wilson , W.D., and Welsh , A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 1*0, 29-63.

20. Slowiaczek, L.M., Nusbaum , H.C., and Pisoni, D.B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13,* 64-75.

21. Connine, C.M., Blasko, D.G., and Titone, D. (1993). Do the beginnings of words have a special status in auditory word recognition? *Journal of Memory and Language, 32,* 193-210.

22. Marslen-Wilson, W.D. (1993). Issues of process and representation in lexical access. In G. Altmann, and R. Shillcock (Eds.), *Cognitive models of language processes: The second Sperlonga meeting. Hillsdale: Lawrence Erlbaum,* 187-210.

23. Marslen-Wilson, W.D., and Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 576-585.

24. Foss , D.J., and Blank , M.A. (1980). Identifying the speech codes. *Cognitive Psychology, 12,* 1-31.