Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Journal of Phonetics 33 (2005) 27–46

Journal of Phonetics

www.elsevier.com/locate/phonetics

The role of juncture cues and phonological knowledge in English syllabification judgments

Melissa A. Redford*, Patrick Randall

The University of Texas at Austin, USA

Received 8 October 2002; received in revised form 28 April 2004; accepted 17 May 2004

Abstract

Listener syllabification judgments vary on words with VCCV sequences. This variability seems to present a challenge to phonological theory, which predicts invariant behavior following from rules or principles. Judgment variability might be better accounted for by positing that syllable production and perception are linked by the signal. According to such a hypothesis, variable listener judgments would result from variable speaker productions. To determine whether phonological or phonetic factors better account for listener syllabification judgments, different speakers produced nonsense words, listeners syllabified them, and then stimuli characteristics were used to predict syllabification. Results showed that, in spite of production variability, listener judgments were nearly invariant on words with medial sequences that formed illegal onset clusters to the second syllable and on words with first syllable stress, suggesting that these tokens were syllabified according to categorical phonological knowledge. Other judgments could not be similarly explained, but instead were best predicted by gradient phonetic patterns cuing juncture. The results disconfirm the hypothesis that syllable production and perception are directly linked via the signal. Instead, they suggest a two-step model of English syllabification in which listeners rely on juncture cues to determine syllable boundaries only after phonological knowledge fails to indicate a unique boundary location.

*Corresponding author. Department of Linguistics, University of Oregon, 1290 Eugene, OR 97403-1290, USA. Tel.: + 1 541 346 3789; fax: + 1 541 346 5961.
E-mail address: redford@darkwing.uoregon.edu (M.A. Redford).

0095-4470/$ - see front matter
doi:10.1016/j.wocn.2004.05.003

1. Introduction

Experimental studies show that listener syllabification judgments can vary moderately for many word-medial consonant sequences (Fallows, 1981; Treiman & Zukowski, 1990; Derwing, 1992), and extensively for some sequences, such as the /sC/ sequences of the English words *mister* [mIstər] and *dissuade* [dIsweid] (Davidsen-Nielsen, 1974; Treiman & Zukowski, 1990; Treiman, Gross, & Cwikiel-Glavin, 1992). This observation presents a challenge for phonological theory, which explains that syllabification proceeds according to a set of standard rules and principles. Such a theory requires further elaboration to account for listener variability. Accordingly, adherents to the theory draw a distinction between underlying (phonological) and surface (phonetic) syllable boundaries to explain variability in the syllabification of most sequences (e.g., Hyman, 1975; Selkirk, 1982; Ladefoged, 1993), and to conflicts between phonological principles or to the special status of particular phonemes to explain listener variability that characterizes the syllabification of problematic sequences, such as the English /sC/ sequences (e.g., Clements, 1990; Kaye, Lowenstamm, & Vergnaud, 1990).

Whereas differences in listener syllabification judgments are difficult to explain within a phonological theory, they may be easily explained within a phonetic theory that links the production and perception of suprasegmental units (e.g., Fowler, 1979; Lindblom, 1983; Tuller & Kelso, 1991). In such a theory, listener variability would be the direct result of measurable speaker variability. The type of inter-listener variability observed in experimental studies of syllabification could occur because listeners may repeat the tokens to themselves, and syllabify them according to their own unique productions. Intra-listener variability could occur if the naturally produced stimuli differ from token to token. In both cases, variability in listener judgments could be attributed to differences in the signal, whether these differences are produced internally (inter-listener variability) or externally (intra-listener variability). This would mean that the link between syllable production and perception is mediated by phonetic patterns, and so syllabification is phonetically motivated.

The present study was undertaken to test the idea that syllable production and perception are linked by the signal. Accordingly, the study examined whether phonetic factors provide better explanations of listener syllabifications than phonological factors.

1.1. Variability in syllabification judgments

Syllabification judgments may be most variable for certain types of word-medial consonant sequences, such as the English /sC/ sequence. Treiman et al. (1992) showed that although these sequences form legal English onsets (*spot*, *stop*, *sweet*, *sleep*), listeners often (65.5%[1]) split the consonants in word-medial position so that /s/ becomes an offset for one syllable and C the onset of another (i.e., VC.CV). Thus, the /sC/ sequences in Treiman et al.'s study often patterned with a control group of sequences that formed illegal English onsets—syllabified as VC.CV—and sometimes with a control group of sequences that formed legal onsets—usually syllabified as V.CCV.

[1] Averaged across 2 experiments and all /sC/ sequences.

Treiman et al. (1992) interpreted their results to support the idea that all /sC/ sequences may be illegal onsets (Kaye et al., 1990), despite their clear existence as word onsets in English and the fact that half of these sequences (e.g., /sw-/ and /sl-/) are also consistent with sonority sequencing, a cross-language segment sequencing preference that places obstruents before sonorants in onset clusters (Bell & Hooper, 1978). A simpler explanation might be that the nonsense words with /sC/ sequences were produced in a manner that more often patterned with the illegal than with the legal control sequences.

In the present study, we directly explored the link between syllable production and perception using the stimuli from Treiman et al. (1992). We chose these stimuli for the same reason that Treiman et al. developed them—because /sC/ sequences present a particular problem to phonological theories of syllabification. When /s/+stop sequences are treated as onset clusters, and /s/+sonorant sequences are split, the sonority sequencing principle is violated. We took this violation to imply that overall phonologically driven explanations of syllabification are inadequate, and that a phonetically driven explanation would be more successful. With respect to the problematic /sC/ sequences, the phonetic hypothesis was that these generate greater variability in boundary judgments than their controls because they are produced with more variability. More generally, the hypothesis was that words are syllabified based on their phonetic structure, not their phonological structure.

1.2. Phonetically motivated syllabification

Evidence for phonetically motivated syllabification comes from speech production and perception studies that explore phonetic cues to juncture. All aspects of the signal—amplitude, duration, and frequency modulation—have been implicated as juncture cues (e.g., Malmberg, 1955; Lehiste, 1960; Davidsen-Nielsen, 1974; Christie, 1977; Ainsworth, 1986; Boucher, 1988; Anderson & Port, 1994). Here, we focus on the most well-established cues: vowel and consonant duration patterns.

1.2.1. Vowel duration

A syllable is most commonly defined as a sound unit composed of one or more segments, organized around a sonorous peak (e.g., Ladefoged, 1993; Kenstowicz, 1994). In most instances, the sonorous peak is a vowel. Although vowels are at the center of syllables, their duration varies with factors that may bear on the perception of syllable boundaries. For instance, syllables with final consonants will very often have shorter vowels than those with no final consonant. This cross-language observation is called closed syllable vowel shortening, and may be general enough to provide a universal phonetic cue to syllabification (Maddieson, 1985). Munhall, Fowler, Hawkins, and Saltzman (1992) have suggested that this type of vowel shortening results from production factors, namely, coarticulation with the following consonant. The articulation of the vowel is truncated in order to accommodate the articulation of the consonant. Perceptually, the effect of syllable structure on vowel length could mean that a V1CCV2 sequence would be syllabified as V1C.CV2 when V1 is short, and as V1.CCV2 when V1 is long. Importantly, V1 duration is not defined here relative to V2 duration, but only with respect to different productions of V1. This means that syllabification may depend on V1 duration independently of V2 duration.

Vowel duration also varies with lexical stress. Stressed vowels are usually longer than their unstressed counterparts (Klatt, 1976), and there is psycholinguistic evidence to show that syllables with longer vowels are associated with more consonants than those with shorter vowels (Treiman & Danis, 1988; Treiman & Zukowski, 1990; Derwing & Nearey, 1991). This type of evidence suggests that speakers and listeners syllabify a V1CCV2 sequence V1C.CV2 when V1 is long and V2 is short, and as V1.CCV2 when V1 is short and V2 is long. Interestingly, vowel duration variation due to stress is expected to have the opposite effect on syllabification as variation in V1 duration due to syllable structure. With stress, V1 attracts more consonants as duration increases. With syllable structure, V1 is more likely to be associated with consonants as duration decreases.

In the present study, the structure of the nonsense word stimuli parallel the medial V1CCV2 sequences just described. For this reason, we are able to distinguish between syllable structure and stress as factors in boundary production and perception by measuring V1 and V2 duration and their relationship to different syllabifications of alike sequences.

A potential problem may arise in the interpretation of the results, though, since vowel duration is not uniquely correlated with either syllable structure or stress, and is correlated with additional variables, such as vowel quality. Conversely, phonological variables such as lexical stress and vowel quality are phonetically encoded not only by vowel duration, but also by amplitude and frequency changes. It is therefore possible that these phonological variables may explain syllabification independently of vowel duration. To assess this possibility and test the effect of additional phonological factors on syllabification, lexical stress and vowel quality were transcribed and included as variables in the analyses.

1.2.2. Consonant duration

Boucher (1988) and Tuller and Kelso (1991) have argued that consonant duration is actually a more important cue to syllable boundary perception than vowel duration. Boucher asked speakers to produce VstopV sequences with stress either on the first or second syllable. Syllable boundary location varied with word boundary location (e.g., *free Danny* vs. *freed Annie*). He found that vowel duration was a less reliable cue to syllabic division than the closure duration of the stop consonant. This finding is consistent with the Tuller and Kelso's (1991) perception experiment, which showed that stop closure duration was the most accurate predictor of boundary judgments for VstopV sequences (e.g., *apa*).

The Boucher (1988) and Tuller and Kelso (1991) experiments suggest a difference in the relative importance of vowel and consonant durations for the syllabification of medial consonants. Such a difference has not been established for the syllabification of intervocalic consonant sequences, even though these sequences are characterized by distinct duration patterns depending on whether they are split or syllabified as an onset cluster. For instance, Christie (1977) varied boundary location for a VC1C2V sequence (*help us nail* vs. *help a snail*), and analyzed consonant durations at the boundary. He found that when the sequence was divided VC1.C2V, C1 and C2 were nearly equal in duration, but when the sequence was divided V.C1C2V, C1 was longer than C2. These findings correspond with those of Haggard (1973), who reported a negative correlation between the duration of successive consonants in a cluster, especially in onset position.

To summarize, the primary purpose of the study was to test the hypothesis that differences in production account for variable syllabification judgments. We expected that different vowel and consonant duration patterns would predict different syllabifications of the same type of stimuli better than phonological variables.

2. Methods

To test the hypothesis that syllable production and perception are linked via the signal we asked different speakers to produce disyllabic nonsense words with medial consonant sequences that were expected to generate variability in listener syllabifications. Next, phonetically trained listeners transcribed the tokens to determine stress and vowel quality, and acoustic measurements were made to capture the phonetic patterns that have been described as juncture cues. Finally, groups of listeners were asked to syllabify the nonsense words. The analyses were then conducted in three parts. First, phonetic and phonological patterns were analyzed to assess speaker differences in production. Second, listener boundary judgments were analyzed to assess judgment variability with respect to phonological categories. Lastly, the phonetic patterns were used to predict variable boundary judgments. The relationship between the phonetic patterns and listener judgments were analyzed according to the type of medial consonant sequence syllabified to assess the generalizability of the results. The methods are described in more detail below.

2.1. Stimuli

The stimuli were disyllabic nonsense words with medial two-consonant sequences that were meant to test the effects of legality and sonority sequencing on syllabification. The words had 4 different types of consonant sequences, half of which began with /s/ and half of which did not. Type 1 sequences—/s/ + stop sequences—were legal English onsets that were inconsistent with sonority sequencing. Type 2 sequences—/s/ + sonorant sequences—were legal onsets consistent with sonority sequencing. The other two nonsense word types had sequences that functioned as controls: Type 3 sequences formed legal English onsets and Type 4 sequences did not. There were 17 words in each of the four categories, yielding a total of 68 words. The words are listed by Type in Table 1.

Two male and two female native English speakers produced the stimuli. Both males and one female were American-English speakers. The other female was a Canadian-English speaker. All speakers were between 20 and 30 years old. The speakers read the written form of the nonsense words from a randomized list and produced each of them in the carrier phrase, "Say ____ eight times." Speakers were not instructed on word pronunciation. The goal was to elicit further variability and individual differences in production. We expected that this variability in production would encourage further variety in boundary judgments.

The sentences were recorded using a Shure SM48 microphone and a PC waveform editor developed in the Speech Perception Laboratory at the University of Texas in Austin. The root-mean squared (RMS) amplitude of each stimulus sentence was normalized across speakers using the in-house editor. RMS amplitude was calculated for the latter part of the frame sentence and then scaled so that the average amplitude of the entire utterance was consistent across all speakers and all utterances. This ensured that listeners heard the stimuli produced by different speakers at a similar level.

Table 1
The written form of the nonsense word stimuli are shown, broken down by medial consonant sequence type

Type 1	Type 2	Type 3	Type 4
baspinge	baslinge	bapringe	chigfoon
chuskeem	chusleem	chupleem	dagmaste
daskeft	dasmeft	datreft	emlafe
destibe	deswibe	degribe	envorse
fiskanch	fismanch	fiklanch	gaktibe
geskint	geswelve	gefrelve	ganteeled
gespelve	geswint	gekrint	gethloove
huspoit	husloit	hudroit	hapkeet
jastoped	jasnoped	jadwoped	iptheen
kestibe	kesnibe	kethribe	lemgeeve
nuspeem	nuslange	nufleem	minveesh
nustange	nusweem	nuglange	monlave
teskang	tesmang	teblang	munleebs
vuspobe	vusluct	vuthwobe	obgorm
vuspuct	vusnobe	vutwuct	objenk
zeskib	zesmib	zegwib	vumroove
zusteeg	zusneeg	zushreeg	yadlorn

Type 1 were /s/ + stop sequences; Type 2 were /s/ + sonorant sequences; Type 3 were legal control sequences; and Type 4 were illegal control sequences.

2.2. Production characteristics

2.2.1. Phonological variables

Speakers introduced the random factors of stress and vowel quality in their productions of the nonsense word stimuli. Since these factors may also contribute to syllabification behavior (e.g., Pulgram, 1970; Selkirk, 1982), they were included as variables in the analyses. Four phonetically trained listeners transcribed lexical stress and the two vowels for each the nonsense word stimuli. Transcribers controlled the presentation of the stimuli, and could listen to the same token several times if they liked. However, since only broad transcriptions of the tokens were required, the transcribers usually only listened once or twice to the same stimulus. To foster inter-transcriber consistency, all transcribers were given a photocopy of Ladefoged's (1993) guidelines for transcribing American English vowels (p. 31). The guidelines present the IPA and an example English word meant to illustrate the pronunciation of the symbol. In spite of the guidelines, one transcriber used [a] in preference to [ɑ], two transcribers used [e] in preference to [eɪ], and several used [o] in preference to [oʊ]. After transcription, vowels were labeled as either tense or lax according to Ladefoged (p. 87). Specifically, vowels transcribed as [i, e, a, ɑ, o, u] or diphthongs were considered to be "tense", and those transcribed as [ɪ, ɛ, æ, ʌ, ʊ] were considered to be "lax". The reduced mid-central vowel [ə] was also considered to be "lax". Inter-transcriber reliability was extremely high: listeners agreed on stress placement for 97% of the tokens, and on the quality of vowel 1 and vowel 2 for 91.5% and 88% of the tokens, respectively. Disagreements were resolved in favor of the majority judgment.

2.2.2. Phonetic variables

Vowel and consonant duration measures were recorded based on visual inspection of the acoustic waveform with auditory confirmation. The visual cues used to establish segment boundaries were as follows: (1) Fricative consonant boundaries were indicated by the onset and offset of aperiodic, fricative energy. (2) Stop consonant boundaries were indicated by the onset of closure and the offset of a release burst and/or aspiration. In the control tokens with illegal onset consonant sequences, there were sometimes sequences of two stop consonants. When this occurred, and no release burst indicated the end of one stop and the beginning of another, a boundary was established at the midpoint of closure. (3) Sonorant consonant (nasals and liquids) boundaries were signaled by a change in the waveform as well as a sudden drop (or rise) in amplitude. (4) Vowel boundaries were delimited by the previously described consonantal boundaries and by a sudden rise or drop in amplitude. Segment durations were obtained by subtracting the time of offset from the time of onset.

In addition to overall segment duration measures, the duration of stop aspiration was measured separately from closure duration. This additional phonetic variable was controlled in the analysis of the syllabification of /s+stop/ sequences, as it is known to affect syllabification of these sequence types (Christie, 1974; Davidsen-Nielsen, 1974).

All acoustic measurements were made by the first author, and so the criteria were always the same. Many of the tokens were measured repeatedly and at different times to ensure measurement reliability. In addition, all measurements were made blind to the syllabification results, and so any measurement errors that may have occurred would have occurred randomly across the stimuli. Any resulting noise in the data would only have served to weaken the analyses.

2.3. Syllable boundary judgments

We used a modified version of the written slash-insertion technique (Treiman & Danis, 1988) to record listener syllable boundary judgments. Forty-eight, native English-speaking undergraduate students were instructed to write down and split apart the nonsense words that they heard embedded in the carrier phrase. They were given the example, "Say *super* eight times," and asked how they would split apart the word *super*. When it was clear that they had understood the task, a maximum of 4 students were placed in a sound-attenuated booth to listen, via headphones, to the stimuli produced by one speaker. Different groups of 12 students listened to different speakers. Such a nested design maximized the likelihood that listeners would syllabify the words based on the (external) acoustic characteristics of the stimuli. We reasoned that if listeners heard multiple tokens of a single word produced by different speakers (a more traditional within-subjects design), they may form a prototype for the word, and make syllable boundary decisions based on their internal productions of that prototype.

2.4. Analyses

Three separate sets of analyses were conducted, each of which was primarily associated with its own method. In the first set of analyses, speaker-induced differences in the production of the stimuli were considered. A multivariate analysis of variance (MANOVA) was performed to determine which of the speaker-controlled variables differed by speaker and as a function of stimulus type. This omnibus analysis was followed up with individual univariate analyses of variances (ANOVAs) in order to explore the patterns more thoroughly. The resulting p values were corrected using the Bonferroni method to avoid Type 1 error inflation.

Boundary judgments were analyzed next to examine whether or not listener syllabifications were as variable as we expected based on our review of the psycholinguistic literature. Listener boundary judgments were scored according to whether C1 was placed with the first (0) or second (1) syllable. Judgments where C1 was placed both in the first and second syllables were rare, representing only 2.3% of all judgments, and so were excluded from the analyses. The second consonant of the sequence was not scored as all listeners almost always placed this consonant with the second syllable. The relative frequency of different judgments (VC.CV and V.CCV) was considered as a function of several categorical phonological variables. Nearly invariant judgments that appeared to be conditioned only by phonological factors were considered to be explained by these variables, and so excluded from the further analyses.

The final set of analyses considered whether or not truly variable boundary judgments were explained according to the gradient acoustic characteristics previously discussed as juncture cues. Multi-level logistic regression analyses were used, in which the dependent variable—boundary judgment—was nested within listeners, which were nested within speakers. The nested structure of the model reflected the experimental design, in which groups of listeners heard stimuli produced by different speakers. The primary independent predictor variables were segment duration patterns. The specific patterns of interest were (1) vowel duration patterns, and (2) consonant duration patterns. These patterns were captured by two variables (e.g., V1 and V2 duration). The first of these variables were transformed from continuous to categorical by dividing the continuous range into terciles. Such a transformation provided a more generalized test of the interaction between the first and second segment durations than would be possible by centering and multiplying continuous variables. In addition to the segment duration patterns, stop aspiration was used as a control predictor variable in Type 1 stimuli, which had an intervocalic /s/ + stop sequence.

3. Results

The hypothesis that syllable production and perception are linked via phonetic patterns was proposed to explain why listener syllable boundary judgments vary from token to token and from sequence type to sequence type. The analyses examined the inherent assumptions of the hypothesis, namely, that speakers' productions and listeners' syllabifications are highly variable under all conditions, and so not explained by categorical phonological variables. In addition, the analyses directly explored the link between production differences and variability in syllabification by trying to predict listener behavior according to gradient phonetic patterns. The results show that although inter- and intra-speaker productions were variable, listener syllabifications were sometimes best described as nearly invariant and due to categorical phonological factors. However, listener judgments on half of the tokens could not be accounted for by categorical factors. For these tokens, judgment variability was largely explained by differences in segment duration patterns.

3.1. Production characteristics

In the present study, we sought to maximize differences in production within and between speakers by allowing individual interpretations of the written form of the nonsense words and including different types of medial consonant sequences. To establish whether these manipulations had the desired effect, we used the ratio of vowel 1 duration to vowel 2 duration (V1:V2) to capture the vowel duration pattern, and the ratio of consonant 1 duration to consonant 2 duration (C1:C2) to capture the consonant duration pattern. The ratios were combined with the stress and vowel quality judgments and used as dependent variables in a multivariate analysis of variance (MANOVA) with Speaker and Sequence Type as fixed factors.

The 5 variables (V1:V2, C1:C2, stress, V1 quality, V2 quality) jointly varied systematically with Speaker [$F(15,714) = 14.54$, $p < .001$], indicating individual differences in the production of the same stimuli (see Table 2). Univariate analyses adjusted for multiple comparisons showed variation between Speakers for V1:V2 [$F(3,240) = 9.12$, $p < .001$] and stress [$F(3,240) = 156.03$, $p < .001$], but not for C1:C2 [$F(3,240) = 3.61$, $p = .07$], V1 quality [$F(3, 240) = 3.25$, $p = .113$], or V2 quality [$F(3,240) = .23$, $p = 4.377$]. The different Sequence Types were also distinct according to the phonetic and phonological measures used [$F(15,714) = 9.56$, $p < .001$], but not significantly different across Speakers [Speaker × Type, $F(45,1200) = 1.04$, $p = .394$]. Such a result indicated a relatively uniform effect of Sequence Type on production (see Table 3). Univariate analyses adjusted for multiple comparisons showed variation between Sequence Types for C1:C2 [$F(3,240) = 44.74$, $p < .001$], V1 quality [$F(3,240) = 5.63$, $p < .01$], and V2 quality [$F(3, 240) = 6.05$, $p < .01$], but not for V1:V2 [$F(3,240) = 1.29$, $p = 1.392$] or stress [$F(3,240) = .43$, $p = 3.658$].

Table 2
The means, standard deviation (S.D.), and variances are shown for the phonetic and phonological measures that could influence syllabification judgments

Speaker	Statistics	V1:V2	C1:C2	Stress	V1 qual	V2 qual
1	Mean	0.86	1.48	0.11	0.23	0.53
	S.D.	0.41	1.16	0.32	0.43	0.50
	Variance	0.17	1.34	0.10	0.18	0.25
2	Mean	0.57	1.92	1.00	0.22	0.58
	S.D.	0.30	1.10	0.00	0.42	0.50
	Variance	0.09	1.20	0.00	0.17	0.25
3	Mean	0.62	2.05	0.88	0.14	0.56
	S.D.	0.45	1.50	0.33	0.35	0.50
	Variance	0.20	2.25	0.11	0.12	0.25
4	Mean	0.53	1.91	0.92	0.06	0.52
	S.D.	0.34	1.29	0.27	0.24	0.50
	Variance	0.11	1.67	0.07	0.06	0.25

The measures are presented by Speaker and collapsed across all nonsense word tokens (Table 1). V1:V2 and C1:C2 are the ratios of V1 to V2 duration and C1 to C2 duration, respectively. Stress was coded as either on the first (0) or second (1) syllable. Vowel quality was coded as either lax (0) or tense (1).

Table 3
The means, standard deviation, and variances are shown for the phonetic and phonological measures that could influence syllabification judgments

Type	Statistics	V1:V2	C1:C2	Stress	V1 qual	V2 qual
1	Mean	0.68	1.11	0.70	0.25	0.50
	S.D.	0.42	0.41	0.46	0.44	0.50
	Variance	0.18	0.17	0.21	0.19	0.25
2	Mean	0.70	2.28	0.71	0.25	0.49
	S.D.	0.39	0.96	0.46	0.43	0.50
	Variance	0.15	0.92	0.21	0.19	0.25
3	Mean	0.64	2.80	0.73	0.13	0.42
	S.D.	0.45	1.77	0.45	0.33	0.50
	Variance	0.20	3.13	0.20	0.11	0.25
4	Mean	0.57	1.16	0.77	0.03	0.77
	S.D.	0.33	0.52	0.42	0.17	0.42
	Variance	0.11	0.27	0.18	0.03	0.18

The measures are presented by Sequence Type and collapsed across all Speakers. Type 1 nonsense words had medial /s/ + stop sequences; Type 2 had medial /s/ + sonorant sequences; Type 3 had medial consonant sequences that could form legal onsets to the second syllable (control legal); and Type 4 had sequences that could not (control illegal).

Table 4
The percent of V.CCV syllabifications are shown averaged across listeners and the different stimuli types

Speaker	Type 1 (%)	Type 2 (%)	Type 3 (%)	Type 4 (%)
1	5.9	6.9	23.0	1.0
2	70.8	73.3	90.7	4.7
3	49.7	38.0	70.1	2.7
4	50.5	58.3	83.1	4.4

Type 1 stimuli had medial /s/ + stop sequences; Type 2 had medial /s/ + sonorant sequences; Type 3 had medial consonant sequences that could form legal onsets to the second syllable (control legal); and Type 4 had sequences that could not (control illegal).

These results establish production differences between speakers and sequence types. Next, we consider whether these differences were paralleled by variability in listener syllabifications.

3.2. Boundary judgments

As expected, there was a great deal of variability in listener syllabification judgments for most types of medial consonant sequences and for different speakers. Table 4 points to this variability by showing the percent of V.CCV syllabifications for different Sequence Types averaged across listeners within a Speaker. The Table shows that syllabification of Types 1 and 2 sequences (the /sC/ sequences) was more variable than syllabification of Type 3 (control legal) for tokens produced by 3 of the 4 speakers. Such a result is consistent with our expectations and with

Treiman et al. (1992). It is also consistent with the argument that the Sonority Sequencing Principle (SSP) does not adequately explain listener syllabifications.

On the other hand, Table 4 also shows that listeners sometimes make (nearly) invariant syllabification judgments. The control illegal sequences (Type 4) were almost always split across all speakers. This result might suggest that Type 4 stimuli were produced in a more uniform manner than the other stimuli Types, but the data shown in Table 3 argue against this possibility. As shown in the Table, Type 4 tokens were produced with as much variability as the tokens of other Types that elicited variable syllabification judgments.

In light of the variable productions of Type 4 stimuli, the invariant listener syllabifications suggest that listeners determined boundary location for these tokens according to categorical knowledge rather than according to the acoustic properties of the stimuli themselves. Listeners probably surmised that because the medial consonant sequences could not form legal onset clusters to the second syllable, they must be split. Such use of phonotactic knowledge suggests that boundary judgments may sometimes be explained by categorical phonological variables rather than the continuous phonetic ones. To examine this possibility, the variable judgments associated with Types 1, 2, and 3 were considered as a function of the other two phonological variables, stress and vowel quality.

3.2.1. Stress

Speakers produced the disyllabic nonsense words sometimes with first syllable stress and sometimes with second syllable stress. Speaker 1 produced the majority of her tokens with first syllable stress, whereas Speakers 2, 3, and 4 produced a majority with second syllable stress (see Table 2). Given this information, and the large differences observed on boundary judgments for different speakers (see Table 4), it would appear that stress alone may explain a large portion of the variance in syllabification. This possibility was confirmed by considering the percentage of V.CCV syllabification responses for Types 1, 2, and 3 as a function of first or second syllable stress. When stress occurred on the first syllable, V.CCV syllabifications were rarely observed (Type 1, 2.7%; Type 2, 6.7%; Type 3, 9.1%). Syllabification judgments remained variable, however, on tokens that received second syllable stress. For these tokens VC.CV syllabifications were nearly as frequent as V.CCV (Type 1, 44.7% vs. 55.3%; Type 2, 46.9% vs. 53.2%; Type 3, 16.9% vs. 83.1%). As before, the tokens with the form that elicited more nearly invariant responses (i.e., those with first syllable stress) were produced with as much variability as those that elicited more variable responses (i.e., those with second syllable stress). The variances for V1:V2 and C1:C2 for tokens produced with first syllable stress were $\sigma^2 = 0.23$ and 1.249, respectively. The variances for tokens produced with second syllable stress were $\sigma^2 = 0.10$ and 1.616, respectively.

Overall, it would appear that a portion of the variable listener boundary judgments associated with Types 1–3 is explained by stress, a categorical phonological variable. A strong–weak stress pattern elicited nearly invariant VC.CV syllabifications. However, a weak–strong pattern elicited variable syllabifications. Next, we examine whether this remaining variability would be explained by vowel quality.

3.2.2. Vowel quality

Excluding syllabification judgments that were explained by the categorical phonological variables of legality and stress, 1656 variable judgments remained to be explained. These

Table 5
The percent of V.CCV syllabifications are shown as a function of V1 and V2 quality

VOWEL	Quality	Type 1 (%)	Type 2 (%)	Type 3 (%)
1	Lax	49.0	49.2	81.8
	Tense	77.2	71.3	91.7
2	Lax	54.0	50.0	82.0
	Tense	56.6	56.5	84.4

The data are averaged across listeners and the three stimuli types that elicited variable syllabifications. Type 1 stimuli had medial /s/ + stop sequences; Type 2 had medial /s/ + sonorant sequences; and Type 3 had medial consonant sequences that could form legal onset to the second syllable (control legal).

judgments were made on 138 or 53.49% of the original tokens produced for the study. Of these, 82.61% had a reduced or lax first vowel and 52.17% had a reduced or lax second vowel. Table 5 shows the percent of V.CCV for the tokens as a function of V1 and V2 quality and medial sequence type. What is evident from the Table is that neither the quality of V1 or V2 explained boundary judgments in the way that legality or stress did. That is, different vowel qualities were not uniquely associated with a certain type of boundary judgment.

A multi-level regression analysis in which V1 and V2 Quality were used to predict boundary judgments on stimuli that were nested within speaker (see Section 2.4) showed that the interaction between V1 and V2 Quality, not represented in Table 5, did not explain syllabification behavior either ($B = 0.067$, $t(1577) = 0.918$, $p = 0.36$). As might be expected from the descriptive data shown in Table 5, neither main effect (V1 quality or V2 quality) was significant.

In summary, roughly half of the listener boundary judgments (1608 out of 3264) could be almost completely explained in terms of two categorical phonological variables, legality and stress. Sequences that could not form legal onset clusters to the second syllable were split as were tokens that received first syllable stress. The other phonological variables that are traditionally thought to affect syllabification—sonority sequencing and vowel quality—appeared not to. This may mean that listeners based the other half of their boundary decisions (those on tokens with second syllable stress and medial sequences that could form legal onset clusters to the second syllable) on the gradient acoustic characteristics of the stimuli. This possibility is explored in the next section. Multi-level regression analyses were performed to determine whether or not segment duration patterns, a much discussed juncture cue, could explain the variable syllabifications that were unexplained by the SSP and vowel quality.

3.3. *Linking production differences and variable syllabifications*

Multi-level regression analyses were performed in which segment duration variables were used to predict listener boundary judgments on stimuli that were nested within Sequence Type and within Speaker. The individual variables for the vowel duration pattern were the durations of the first and second vowels (V1 and V2) of the disyllabic nonsense word. The variables for the consonant duration pattern were the durations of the first and second consonants (C1 and C2) in

Table 6
Range, mean, and standard deviation in milliseconds are shown for the categorical variables in the analyses

Tercile	V1 duration			C1 duration		
	Range	Mean	S.D.	Range	Mean	S.D.
1st	30–79	60.21	11.3	20–120	96.9	18.8
2nd	80–100	88.2	7.6	130–150	139.9	8.0
3rd	110–270	151.3	65.4	160–250	182.3	22.3

The terciles group the continuous variables into short (1st tercile), medium (2nd tercile), and long (3rd tercile) categories.

the medial, two-consonant sequence. In both cases, the first of the two variables was transformed into a categorical variable that represented the continuous range in terciles (Table 6). This transformation allowed for a better test of the interaction between the variables (see Section 2.4). The results for the vowel and consonant duration patterns are presented individually.

3.3.1. Vowel duration patterns

The 3-way interaction between V1, V2, and Sequence Type was significant [$X^2(1) = 5.00$, $p < .05$], but this significant effect was due to minor changes in the slope of the predictor variables as shown in Fig. 1.

Of greater interest than the differences between the Types is the fact the slopes of the regression lines were fairly shallow across all types, indicating that vowel durations were not particularly good predictors of boundary judgments. This result may not be surprising given the close relationship in English between vowel quality and duration and the finding that vowel qualities were not significant predictors of boundary judgments. However, in these data the relationship between vowel duration and vowel quality was not simple: A 4-way interaction between V1 duration, V2 duration, V1 quality, and V2 quality [$X^2(1) = 4.503$, $p < .05$] resulted when Vowel Qualities were added to the model. This interaction showed that vowel durations were good predictors of boundary judgments in the subset of data where both V1 and V2 were lax vowels (Fig. 2). In particular, listeners were more likely to judge that the medial sequence was an onset to the second syllable when a lax V2 was short (relative to a lax V1) than when it was long. When either or both of the vowels were tense, vowel duration appeared to have little or no effect on syllabification.

Overall, the results suggest that a subset of the variable listener syllabifications were explained by the vowel duration pattern. In parallel with the idea that categorical responses result from categorical factors, one might expect that a gradient acoustic pattern induces probabilistic responses. This expectation was confirmed here in that vowel durations explained the variable data in a probabilistic manner rather than in a categorical one. Next, we evaluate the accuracy with which consonant durations predicted the variable listener behavior.

3.3.2. Consonant duration patterns

The 3-way interaction between C1, C2, and Sequence Type was not significant, but the interaction between C1 and C2 was [$X^2(1) = 12.21$, $p < .001$], as shown in Fig. 3. What is also

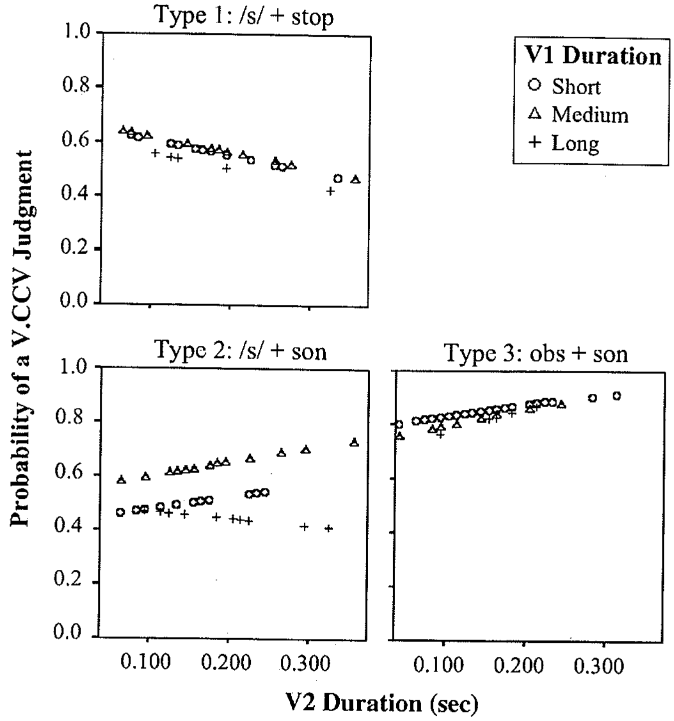

Fig. 1. The probability that V1 and/or V2 durations predicted V.CCV syllabification is shown for stimuli with second syllable stress and medial consonant sequences that could form legal syllable onsets.

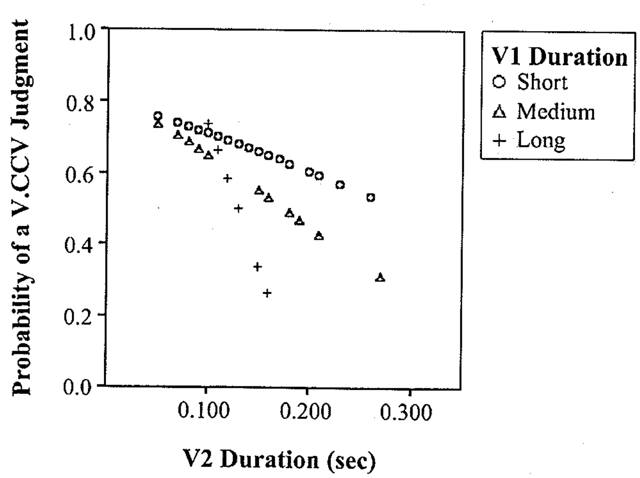

Fig. 2. The probability that V1 and/or V2 durations predicted V.CCV syllabification is shown for stimuli with two lax vowels, second syllable stress, and medial sequences that could form legal syllable onsets.

evident from the Figure is that consonant durations were good predictors of boundary judgments. The slopes of the regression lines were steep, indicating that different patterns of C1C2 duration were distinctive with respect to the judgments. The most salient pattern evident in the Figure is

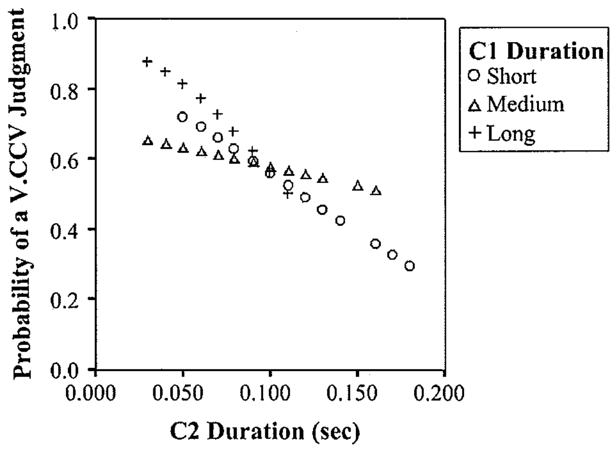

Fig. 3. The probability that C1 and/or C2 durations predicted V.CCV syllabifications is shown for stimuli with second syllable stress and medial consonant sequences that could form legal syllable onsets.

that when C1 was long and C2 was short listeners were especially likely to judge that the intervocalic sequence formed an onset cluster to the second syllable. Conversely, when C1 was short and C2 was long listeners were especially likely to split the intervocalic sequence.

Although the judgments on different Sequence Types were explained by the same patterns (viz. the lack of an interaction with Type), the fact that stop duration included aspiration could have obscured a more important cue to /s/ + stop sequences syllabification, namely, stop aspiration. Accordingly, the analysis was repeated on Type 1 tokens with stop aspiration duration included as an additional predictor variable in the model. In this analysis, the effect of consonant duration disappeared, but Stop Aspiration was found to be a highly significant predictor of boundary judgments, albeit in the presence of the consonant durations and their interactions [$B = -7.19$, $t(505) = -6.842$, $p < .001$]. Consistent with the C1C2 duration pattern, greater stop aspiration duration increased the likelihood that listeners would split the medial consonant sequence.

In summary, the results indicate that the variable syllabification judgments for iambically stressed Types 1, 2, and 3 stimuli were largely explained by phonetic factors and, in particular, by consonant duration patterns. The intervocalic consonant sequences were syllabified as onsets to the following syllable when the first consonant in the sequence was longer than the second. The sequences were split when C1 and C2 were more equal in duration, or when the consonant sequence did not form a legal onset to the second syllable. However, in the /s/ + stop sequences the duration pattern reflected the presence versus absence of stop aspiration, which was the better predictor of judgments for this stimulus type.

4. Discussion

The primary hypothesis tested in this study was that variable syllable boundary judgments can be attributed to differences in production, which are manifest in the signal. The assumptions that led to this hypothesis were that (1) syllable production and perception are directly linked by the signal, and (2) syllabification is phonetically motivated, not phonologically motivated. Whereas boundary judgments on half of the stimuli were best explained by gradient phonetic patterns, and especially by consonant duration patterns, judgments on the other half of the variably produced tokens were nearly invariant and best explained by categorical phonological factors. These results counter assumptions (1) and (2) and suggest instead that listeners use both phonological and phonetic information to make decisions about syllable boundary location.

In the remainder of this section, we focus on the specific phonological and phonetic patterns that explained the syllable boundary judgments, and how they did so. From this discussion, we develop a basic model of English syllabification.

4.1. Phonological and phonetic patterns explaining syllabification

The phonological factors that resulted in nearly invariant syllable boundary judgments were phonotactic patterns and lexical stress. Those sequences that formed illegal English onsets (Type 4) were almost always syllabified so that C1 became the offset of the first syllable, and C2 the onset of the second. Such syllabifications were also made when the consonant sequences formed legal English onsets (Types 1, 2, 3) and had first syllable stress. The very consistent pattern of syllabification for these tokens was not matched by a consistent pattern in production. Type 4 stimuli and stimuli with first syllable stress were as variable in their patterns of vowel and consonant durations as the stimuli that elicited variable syllabification judgments. So, with respect to legality and stress, the present results agree with those from psycholinguistics studies investigating the effect of phonological knowledge on syllabification and the role of stress in syllabification (Treiman & Danis, 1988; Treiman & Zukowski, 1990; Derwing & Nearey, 1991; Treiman et al., 1992; Pitt, 1998). Language-specific phonotactic knowledge dictated whether or not a consonant sequence could function as an onset cluster independent of the acoustics. Sequence-initial stressed vowels dictated that the consonant sequence be split, also independent of the other acoustic variables that could cue boundary judgment.

Whereas two of the categorical phonological factors explained nearly invariant syllable boundary judgments, two others did not: sonority sequencing and vowel quality. The Sonority Sequencing Principle (SSP), which specifies segment organization within a syllable, requires that the /s/ + stop (Type 1) sequences be split, and the obstruent + sonorant (Types 2 & 3) sequences be syllabified as onset clusters to the second syllable. However, English listeners sometimes syllabified Type 1 sequences as onset clusters, and Types 2 and 3 sequences as offset/onset. These variable syllabifications were best explained by specific patterns of vowel durations, consonant durations, and—in the case of /s/ + stop sequences—stop aspiration. Furthermore, the patterns that explained different syllabifications of /sC/ sequences were the same as those that explained different syllabifications of the legal control sequences. This result is important because it suggests that listeners do not treat /sC/ sequences differently from other consonant sequences, even though these sequences are sometimes given special status in syllabification theory (e.g., Kaye et al., 1990; Clements, 1990).

Syllabifications were also variable no matter what the quality of the first or second vowel. According to Pulgram (1970), English allows open syllables with tense vowels, but not with lax vowels. If the syllabic nucleus is a lax vowel, then the syllable is obligatorily closed with a consonant (e.g., *sleepy*→/sli.pi/ vs. *slippery*→/slIp.pri/). However, vowel quality in terms of tense/lax distinctions had no effect on syllabification. Further, when vowel duration explained listener syllabification behavior, longer V1 durations were associated with offset/onset judgments, and shorter ones with onset clusters, which is the opposite of what might be expected if vowel quality were to determine syllabification. It is also opposite to what might be expected if closed syllable vowel shortening provides a universal phonetic cue to syllabification, as was suggested by Maddieson (1985).

As noted several times already, variable syllable boundary judgments were explained by phonetic factors. However, not all of these patterns were equal in explanatory power. The vowel duration pattern, that is the relative duration of V1 and V2, was a fairly poor predictor of syllabification judgments. In fact, vowel duration patterns only predicted judgments on a subset of the variable data (Fig. 2), which was already a subset of the total data set. It is interesting to note, though, that when vowel durations explained boundary judgments, the pattern paralleled that which was identified with stress. That is, a trochaic-like, long-short vowel duration pattern was highly associated with VC.CV judgments, just as trochaic stress always resulted in such a judgment.

In contrast to the vowel duration pattern, the consonant duration pattern was a strong predictor of boundary judgments. This result is interesting because it generalizes findings on VCV sequences (Boucher, 1988; Tuller & Kelso, 1991) to more complicated sequences and syllable structures. As expected, the exact patterns that cue V.CCV and VC.CV syllabifications—long-short and long-long or short-short respectively—agree with work on phonetic cues to juncture at word and syllable boundaries (Christie, 1977; McCasland, 1977).

The particular effect of consonant durations on syllabification held for Types 1, 2, and 3 stimuli, but in Type 1 (/s/ + stop) stimuli the result was an artifact of the measurement procedure, which included stop aspiration in the overall duration of stop consonants. For these sequences, the independent predictor of syllable boundary judgments was stop aspiration duration. In agreement with Christie (1974) and Davidsen-Nielsen (1974), we found that listeners split /s/ + stop sequences with longer stop aspiration.

Overall the results point to how syllable boundary judgments might be made. Below, we continue in the direction indicated to develop a basic psycholinguistic model of syllabification.

4.2. *Towards a model of English syllabification*

A principle result to emerge from this study is that English-speaking listeners make both invariant and variable syllable boundary judgments. The invariant judgments can be explained by phonological factors and the variable judgments by phonetic patterns. One possible implication of this finding is that English syllabification is a two-step process. Listeners may first consider the phonotactics and stress of a particular word in assigning syllable boundary location, and consider the phonetic patterns only when boundary location remains ambiguous. Even though the study provides no direct evidence that phonological factors are considered prior to phonetic patterns in making syllabification judgments, and not simultaneously with the patterns, a serial model is hypothesized for two reasons. The first is economy. It is generally agreed in the literature on lexical access that language processing moves quickly from the input representation to more and more abstract representations (see Jusczyk & Luce, 2002, for a review). Further, it seems likely that syllabification—a process that divides words into smaller units between two phonemes—is only possible once phonemes have been decoded, and so it is a process that necessarily works on a more abstract level of speech representation than that which is originally available to the listener. Finally, if phonological information often specifies a unique syllable boundary, then nothing more is gained by backtracking to process the detailed phonetic input representation in each instance.

Secondly, a serial model may help explain inter-listener variability, which was otherwise not treated in the present study. If phonetic patterns are used secondarily in making boundary decisions, listeners may be forced to silently repeat a word in order to reconstruct a (by then) degraded phonetic representation. Such silent repetitions would introduce production differences into the representation. The syllable boundaries would then be made on the transformed, and now unique, representations, which would result in inter-listener variability.

The results of this study also suggest that phonological factors underlying syllabification might be language-specific rather than language-universal. Clearly, legality makes reference to English-language knowledge, as it is based on what occurs or does not occur at the edges of English words. The effect of stress on syllabification may also be language-specific. The trochaic pattern, which elicited nearly invariant boundary judgments in this study, has a special status in English for speech segmentation (e.g., Cutler, 1994), and this status may help explain its influence on syllable boundary judgments, especially if, as we argue below, there is an important parallel between syllable and word boundaries.

To understand why language-specific knowledge, rather than universal principles, explains English syllabification, it is useful to contemplate how syllable structure might be learned. One significant possibility is that it is learned from word structure. This seems especially likely given that learning some of the most frequent words in English (e.g., *my cat is black and blue*) is tantamount to learning English syllable structure (CV.CVC.VC.CCVC.VCC.CCV). If syllable structure is learned from word structure, then syllable boundaries may be identified in the same way as word boundaries.[2] This suggestion would explain why legality, which is defined with respect to word boundaries, plays such a strong role in determining syllable boundaries. It would also explain why the phonetic cues that predicted syllabification in this study are also known to cue the location of word boundaries (e.g., Christie, 1977; Boucher, 1988; Pitt, 1998; Turk & Shattuck-Hufnagel, 2000). In particular, the consonant duration patterns, which explained the syllabification of legal sequences as either VC.CV or V.CCV, also create meaningful differences at word boundaries (e.g., *help us nail* vs. *help a snail*); and, the lack of stop aspiration in /s/ + stop sequences, which explained the syllabification of these sequences as onset clusters, parallels the way in which these sequences are produced word-initially.

Beyond the similarities between the determinants of syllable boundary location in this study and word boundary location in other studies, the results do not speak to the type of representations that listeners access when breaking words into smaller parts—the task assigned to the listeners in the present study. The suggestion here, that syllable boundary decisions are made by analogy to word boundaries, is merely the simplest explanation for the behavior observed. Future work will aim to explore the hypothesis more directly by examining how manipulations at word boundaries affect word-medial syllable boundary judgments.

[2]In fact, this is the approach advocated by Pulgram (1970) for defining English syllables.

5. Conclusion

In conclusion, the present study shows that the speech acoustics of English do not clearly reflect a syllable unit: the relationship between variable productions and different syllabifications is not simple. This could suggest that the unit of production is not necessarily the same as the unit of perception, and the relationship between what is produced and what is perceived is mediated by language knowledge. Categorical phonological factors, which may be language-specific, result in invariant syllable boundary judgments. When these factors do not suffice to determine boundary location, phonetic juncture cues appear to be used. From these results, a two-step model of English syllabification was proposed in which listeners are thought to base their initial syllable boundary decisions on an already abstract representation of the input, and then backtrack to a partially degraded acoustic representation when they must decide between alternate syllabifications of a single stimulus. Finally, whether using phonological or phonetic information, syllable boundary decisions may reference words and not syllables.

Acknowledgments

The authors are grateful to Kenneth de Jong, Alice Turk, and an anonymous reviewer for their insights and thought-provoking critiques on an earlier draft of the paper. The authors are also especially grateful to Randy Diehl for making this research possible. The data for this study were collected for part of the first author's doctoral dissertation. The bulk of the reanalysis and a portion of the writing was supported by an NIH National Research Service Award #F32-DC00459-02 granted to the first author.

References

Ainsworth, W. A. (1986). Pitch change as a cue to syllabification. *Journal of Phonetics*, *14*, 257–264.

Anderson, S., & Port, R. (1994). Evidence for syllable structure, stress and juncture from segmental durations. *Journal of Phonetics*, *22*, 283–315.

Bell, A., & Hooper, J. B. (1978). Issues and evidence in syllabic phonology. In A. Bell, & J. B. Hooper (Eds.), *Syllables and segments* (pp. 3–22). Amsterdam: North-Holland Publishing.

Boucher, V. J. (1988). A parameter of syllabification for V*stop*V and relative-timing invariance. *Journal of Phonetics*, *16*, 299–326.

Christie, W. M. (1974). Some cues for syllable juncture perception in English. *Journal of Acoustical Society of America*, *55*, 819–821.

Christie, W. M. (1977). Some multiple cues for juncture in English. *General Linguistics*, *17*, 212–222.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston, & M. E. Beckman (Eds.), *Between the grammar and the physics of speech* (pp. 283–333). New York: Cambridge University Press.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua*, *92*, 81–104.

Davidsen-Nielsen, N. (1974). Syllabification in English words with medial sp, st, sk. *Journal of Phonetics*, *2*, 15–45.

Derwing, B. (1992). A 'pause-break' task for eliciting syllable boundary judgments from literate and illiterate speakers: preliminary results for five diverse languages. *Language and Speech*, *35*, 219–235.

Derwing, B. L., & Nearey, T. M. (1991). The "vowel-stickiness" phenomenon: three experimental sources of evidence. From *Proceedings of the XIIth international congress of phonetic sciences*. Aix-en-Provence: U de Provence.

Fallows, D. (1981). Experimental evidence for English syllabification and syllable structure. *Journal of Linguistics*, *17*, 309–317.

Fowler, C. (1979). "Perceptual centers" in speech production and perception. *Perception and Psychophysics*, *25*, 375–388.

Haggard, M. (1973). Correlations between successive segment durations: values in clusters. *Journal of Phonetics*, *1*, 111–116.

Hyman, L. (1975). *Phonology: theory and analysis*. New York: Holt, Rinehart & Winston.

Jusczyk, P., & Luce, P. (2002). Speech perception. In H. Pashler, & S. Yantis (Eds.), *Steven's handbook of experimental psychology, Vol. 1: sensation and perception* (pp. 493–536). New York: Wiley.

Kaye, J., Lowenstamm, J., & Vergnaud, J.-R. (1990). Constituent structure and government in phonology. *Phonology*, *7*, 193–231.

Kenstowicz, M. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell.

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

Ladefoged, P. (1993). *A course in phonetics*. Fort Worth: Harcourt Brace and Jovanovich.

Lehiste, I. (1960). An acoustic–phonetic study of internal open juncture. *Phonetica*, *5*, 1–54.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 217–245). New York: Springer.

Maddieson, I. (1985). Phonetic cues to syllabification. In V. Fromkin (Ed.), *Phonetic linguistics. essays in honor of Peter Ladefoged* (pp. 203–221). New York: Academic Press.

Malmberg, B. (1955). The phonetic basis of syllable division. *Studia Linguistica*, *9*, 80–87.

McCasland, G. P. (1977). English stops after /s/ at medial-word boundary. *Phonetica*, *34*, 218–228.

Munhall, K., Fowler, C., Hawkins, S., & Saltzman, E. (1992). "Compensatory shortening" in monosyllables of spoken English. *Journal of Phonetics*, *20*, 225–239.

Pitt, M. A. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics*, *60*, 941–951.

Pulgram, E. (1970). *Syllable, word, nexus, cursus*. The Hague: Mouton.

Selkirk, E. (1982). The syllable. In H. Van der Hulst, & N. Smith (Eds.), *The structure of phonological representation* (pp. 337–383). Dordrecht: Foris.

Treiman, R., & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language*, *27*, 87–104.

Treiman, R., Gross, J., & Cwikiel-Glavin, A. (1992). The syllabification of /s/ clusters in English. *Journal of Phonetics*, *20*, 383–402.

Treiman, R., & Zukowski, A. (1990). Toward an understanding of English syllabification. *Journal of Memory and Language*, *29*, 66–85.

Tuller, B., & Kelso, J. A. S. (1991). The production and perception of syllable structure. *Journal of Speech and Hearing Research*, *34*, 501–508.

Turk, A., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, *28*, 397–440.