Acoustic correlates of phrase-internal lexical boundaries in Dutch

Taehong Cho & Elizabeth K. Johnson

Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands
`Taehong.Cho@mpi.nl` & `Elizabeth.Jonhson@mpi.nl`

Abstract

The aim of this study was to determine if Dutch speakers reliably signal phrase-internal lexical boundaries, and if so, how. Six speakers recorded 4 pairs of phonemically identical strong-weak-strong (SWS) strings with matching syllable boundaries but mismatching intended word boundaries (e.g. *reis # pastei* versus *reispas # tij*, or more broadly $C_1V_2(C)\#C_2V_2(C)C_3V_3(C)$ vs. $C_1V_2(C)C_2V_2(C)\#C_3V_3(C)$). An Analysis of Variance revealed 3 acoustic parameters that were significantly greater in S#WS items (C_2 DURATION, RIME$_1$ DURATION, C_3 BURST AMPLITUDE) and 5 parameters that were significantly greater in the SW#S items (C_2 VOT, C_3 DURATION, RIME$_2$ DURATION, RIME$_3$ DURATION, and V_2 AMPLITUDE). Additionally, center of gravity measurements suggested that the [s] to [t] coarticulation was greater in *reis # pa[st]ei* versus *reispa[s] # [t]ij*. Finally, a Logistic Regression Analysis revealed that the 3 parameters (RIME$_1$ DURATION, RIME$_2$ DURATION, and C_3 DURATION) contributed most reliably to a S#WS versus SW#S classification.

1. Introduction

The goal of speech from the speaker's perspective is perhaps best described by the Jakobson, Fant & Halle's classic statement "We speak to be heard in order to be understood" [1]. For a successful delivery of the speaker's message, however, the listener must be able to parse the incoming flow of the transmitted speech into meaningful chunks—i.e., words, as intended by the speaker, a phenomenon known as *lexical segmentation*. Listeners use various cues in lexical segmentation that come from their knowledge of the language (e.g., lexical stress, phonotactics, segmental transitional probability, lexical frequency), contextual information (e.g., discourse, syntactic and semantic structures), as well as bottom-up information present in the acoustic signal.

Bottom-up cues are likely to play an important role in lexical segmentation, especially when listeners are faced with otherwise ambiguous utterances such as *paper # forms* versus *pay # performs*. By producing a phrase boundary between intended words, speakers can help listeners parse an otherwise ambiguous utterance. This strategy works because word boundaries align with phrase boundaries, which in turn are characterized by acoustic-phonetic cues such as preboundary lengthening, boundary tones and domain-initial strengthening. Thus, if listeners were to capitalize on these cues, lexical segmentation could be facilitated. Indeed, word monitoring experiments have shown that listeners use precisely this sort of information to disambiguate strings such as *pay-per-forms* (e.g., in [*The butler*] [*with the highest pay*] [*performs the most*], where brackets mark phrase boundaries) [2].

Although the existence of bottom-up acoustic-phonetic cues to phrase boundaries is well established, attempts to demonstrate that phrase-internal lexical boundaries are similarly marked have met with mixed success. By definition, phrase-internal lexical boundaries cannot be marked by boundary tones or pauses. Consequently, it is not surprising that the current literature contains conflicting evidence regarding the existence of acoustic cues to phrase-internal lexical boundaries. On one hand, it seems that acoustic markers of lexical boundaries become attenuated, or perhaps disappear completely, phrase-internally. Prosodic lengthening becomes unreliable [3] or speaker-specific [4], and word-level domain-initial strengthening becomes inconsistent [cf. 5]. On the other hand, phrase-internal lexical boundaries are sometimes present in the signal. For example, Quené [6] found Dutch sequences such as *diep # in* versus *die # pin* are reliably differentiated by stop closure duration. Indeed, psycholinguistic studies have suggested that English listeners can use word-level acoustic-phonetic cues to distinguish between phonemically identical strings (*two # lips* vs. *tulips*, [7]) or temporally ambiguous sequences (*ham vs. hamster*, [8]). Moreover, English-learning twelve-month-olds infants appear to use phrase-internal word boundary cues to parse ambiguous strings (*dog # maligns* vs. *dogma # lines*, [9]). The authors of these studies have suggested, either tacitly or directly, that the prosodically-driven durational pattern plays a pivotal role in marking a phrase-internal lexical boundary (e.g., [8]). However, in spite of the fact that the importance of the phonetic cues to a phrase-internal lexical boundary for perception has been increasingly recognized, larger-scale systematic phonetic studies that illustrate how speakers signal phrase-internal lexical boundaries holistically are not readily available in the field.

In the present study, we re-visit the issue of phrase-internal lexical boundaries in Dutch. Although past studies have reported that Dutch reliably marks lexical boundaries, these studies confounded word-level effects with position-in-syllable effects [6]. For example, the /p/ in *diep # in* occupies a word-final position whereas the /p/ in *die # pin* occupies a word-initial position (cf. [10]). Therefore, in the present study, we studied the production of phonemically identical two-word strings with matching syllabic structure (e.g. *reis # pastei* vs. *reispas # tij*). However, our goal was not to examine a particular phonetic event limited to a domain-initial or a domain-final position, but to explore the holistic acoustic phonetic hallmarks that may differentiate the phonemically identical three-syllable strings. The question was whether or not (and if so, how) speakers differentiate potentially ambiguous strings in the absence of linguistic (e.g., phonotactic, contextual) or phrase boundary cues.

2. Method

Four phonemically identical strings carrying a SWS (strong-weak-strong) stress pattern were chosen for use in this study.

The intended parse of each string was ambiguous in the sense that the word boundary could be placed after either the first or second syllable: (1) *dis # kopie* ('table' # 'copy') versus *disco # pi* ('disco' # 'pi(π)'), (2) *reis # pastei* ('trip' # 'pastry') versus *reispas # tij* ('passport' # 'tide'), (3) *vee # totaal* ('cattle' # 'total') versus *veto # taal* ('veto' # 'language') and (4) *zee # coupon* ('sea' # 'coupon') versus *zeekoe # pon* ('sea cow' # 'nightie'). Note that the onsets of the second and third syllables (C_2, C_3) are stops. The four target strings were embedded in six sentences each. The sentential positioning of the two-word SWS targets was varied in order to inhibit speakers from developing a repetitive strategy that might ensue from reading the target sequences in a fixed position across sentences (see Table 1).

Six native Dutch speakers were recorded (4 females, 2 males). Sentences were read in pairs (alternating S#WS and SW#S) to enable speakers to maintain a constant speech rate and inflection across sentence pairs. (In a pilot study with no such pairing, speakers produced the sentences in a highly variable manner. This resulted in no measurable boundary marking, presumably due to the noise added by inconsistent patterns of production). The sentences were read in a moderate child-directed voice in order to facilitate their use in subsequent perceptual studies. (The perceptual study is not reported here). Finally, to obtain consistent phrasal accent patterns across sentences, the speakers were instructed to always accentuate both words in the target sequence.

Table 1: An example set of sentences.
a. Is **reis-pastei** (reispas-tij) bij jou bekend?
b. Zo'n **reis-pastei** (reispas-tij) is mooi om te zien.
c. Zij kon allen maar **reis-pastei** (reispas-tij) praten .
d. Hij had nog nooit van **reis-pastei** (reispas-tij) gehoord.
e. We wachten op reispas-tij **reis-pastei** (reispas-tij).
f. Het boek gaat over **reis-pastei** (reispas-tij).

Before the actual recording session, speakers were introduced to discourse situations in which the two-word noun phrases could have occurred. This step was necessary because although all target sequences consisted of real lexical items, the meanings of the noun phrases were somewhat nonsensical. In addition, speakers practiced producing the intended renditions as naturally as possible (with proper accent and no discernible phrasal boundary). Two tokens of each sentence pair was read from each of two scripts. The two scripts had different orders with respect to carrier sentences and lexical boundaries. In total, 1153 tokens were collected (4 lexically different target strings x 2 parsings x 6 sentence types x 4 repetitions x 6 speakers). After the recording session, both authors listened to each token to ensure that all target words were accented and produced connectedly with no discernible phrase juncture separating the words. Tokens were excluded from further analysis if either author was unsure about its accent pattern and/or lack of connectedness. In total, 88 tokens were excluded, yielding a total of 1065 tokens for analysis.

Acoustic measurements included (1) the duration of the rime in all three syllables ($RIME_1$, $RIME_2$, $RIME_3$), (2) the closure duration of onsets C_2 and C_3 in the second and the third syllable (C_2DUR, C_3DUR), (3) VOT of C_2 and C_3, (4) the peak amplitude during vowels (V_1AMP, V_2AMP, V_3AMP), and (5) the peak burst amplitude during the first 15 ms of the burst for C_2 and C_3 (C_2BURST, C_3BURST). In addition, the spectral center of gravity (COG) was measured for the coda [s]'s for the string *rei[s]pa[s]tij*, in order to examine the spectral characteristics of [s] that may vary with a lexical boundary. Since COG for the fricative [s] is likely to indicate place of articulation and degree of constriction, these measures can be used to infer how coda [s]'s place of articulation varies as a function of a lexical boundary, and, more specifically, the degree to which [s] is coarticulated with the [t] in *pas (#) tij*.

3. Results and discussion

The results will be presented from two different perspectives. First, the results of repeated measures Analyses of Variance (ANOVAs) with Boundary Location (S#WS versus SW#S) as an independent factor will be discussed in terms of how each acoustic parameter under investigation is conditioned by the presence or absence of a lexical boundary. Due to space limits, only overall main effects will be reported with data pooled across all four pairs of lexical strings. Then, the results will be re-considered in terms of the extent to which each acoustic parameter contributes to the classification of strings as S#WS versus SW#S. For this analysis, a binary logistic regression analysis was performed with Boundary Location as a *dependent* variable. In the final section of the results, we will report the effect of Boundary Location on the acoustic realization of [s] in the string *rei(#)pas(#)tij* and discuss it in terms of consonantal coarticulation between [s] and [t].

3.1. ANOVAs

The results are summarized in Table 2. First, consider the duration parameters in the vicinity of the lexical boundary in S#WS. $RIME_1$ and C_2DUR are both significantly longer in S#WS (versus SW#S). The lengthening of $RIME_1$ in S#<u>WS</u> is ascribable to either the boundary effect (i.e., word-level preboundary lengthening) or the effect of the number of syllables (i.e., monosyllabic words are longer than polysyllabic words). In contrast, the lengthening of C_2DUR in S#<u>WS</u> is likely to be due to postboundary domain-initial strengthening rather than the number of syllables in the word because C_2 occurs in disyllabic words in both conditions. In contrast with lengthening of C_2DUR, however, C_2VOT is significantly shorter in S#<u>WS</u> (versus S<u>W</u>#S). This finding is in line with the domain-initial VOT shortening in Dutch reported by Cho [11], who proposed that while English stops are produced with longer VOTs domain-initially, domain-initial Dutch stops are produced with shorter VOT. Cho argued that this effect may reflect the enhancement of language-specific features ([+spread glottis] versus [-spread glottis]).

Turning to the durational correlates in the vicinity of the boundary between the second and the third syllables, both $RIME_2$ and $RIME_3$ as well as C_3DUR are all shorter in the absence (S#<u>WS</u>) versus presence (S<u>W</u>#S) of a lexical boundary. The shorter $RIME_2$ may be due to lack of preboundary lengthening in S#<u>WS</u> (it precedes no boundary whereas its counterpart S<u>W</u>#S does). But it may also be due in part to directionality of accentual lengthening: when a stressed syllable in a trochaic word is accented (S<u>W</u>#S) accentual lengthening generally spreads to the second syllable (left to right) whereas the leftward effect of accentual lengthening on an iambic word (S#<u>WS</u>) is, if anything, weaker [12]. This could explain the shorter $RIME_2$ in S#<u>WS</u>. The shorter $RIME_3$ is most likely due to the polysyllabic shortening in S#<u>WS</u> (versus SW#<u>S</u>).

Finally, we turn to the amplitude measures during the vowel and the first 15 ms. of the burst. The V_2AMP is smaller in S#WS (versus SW#S), and C_3BURST is smaller in S#WS (versus SW#S). It appears that the V_2 is *less* loud word-initially (S#WS) than word-finally (SW#S), which again may be accounted for by the directionality of accentuation: the accent-induced increase in loudness may spread less to the left in the iambic S#WS than to the right in the trochaic SW#S. The weaker C_3BURST word-medially (S#WS) than word-initially (SW#S) can be taken to be word-medial lenition.

Table 2. A summary table of repeated measures ANOVAs. For the sake of consistency, the *Descriptions* column describes the patterns for S#WS as compared to those for SW#S. The last column notes which one of the four different lexical pairs show the same direction in the means. (1) = dis-kopie, (2) = reis-pastei, (3) = vee-total, (4) = zee-coupon.

Measure	Descrip-tion	F-ratio [1,23]	S#WS means (s.e.)	SW#S means (s.e.)	Notes
C_2DUR (ms.)	longer ***	30.87	80.01 (3.28)	71.55 (3.15)	all
C_2VOT (ms.)	shorter *	4.67	25.22 (1.92)	26.09 (1.94)	all but (4)
C_3DUR (ms.)	shorter ***	24.94	106.32 (10.53)	114.59 (10.11)	all
C_3VOT (ms.)	*n.s.*	0.76	22.33 (11.65)	21.91 (1.72)	*n.s.*
$RIME_1$ (ms.)	longer ***	56.46	222.96 (11.22)	206.99 (10.66)	all
$RIME_2$ (ms.)	shorter ***	22.99	116.29 (10.63)	130.99 (11.15)	all
$RIME_3$ (ms.)	shorter *	7.97	247.14 (11.87)	251.52 (12.03)	all
V_1AMP (dB)	*n.s.*	2.20	77.03 (0.32)	76.91 (0.43)	*n.s.*
V_2AMP (dB)	smaller *	6.61	70.89 (0.477)	71.40 (0.37)	all
V_3AMP (dB)	*n.s.*	1.24	78.05 (0.61)	78.20 (0.59)	*n.s.*
C_2BURST (dB)	*n.s.*	0.01	66.29 (0.82)	66.31 (0.74)	*n.s.*
C_3BURST (dB)	larger *	7.08	63.07 (0.72)	62.32 (0.89)	all but (2)

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

To sum up, the results of the ANOVAs indicate that the potentially ambiguous S(#)W(#)S strings used in this study were disambiguated by speakers, as reflected in eight acoustic parameters including C_2VOT, C_3DUR, $RIME_1$, $RIME_2$, $RIME_3$, V_2AMP and C_3BURST. It may not be possible to pinpoint the sources of variation in these measures as some of the effects are related in a complex way to multiple phonetic events (e.g., domain-initial strengthening, domain-final lengthening, directionality of accentual lengthening, and polysyllabic shortening). What emerges, however, is that speakers signal a lexical boundary by marshalling various acoustic parameters in a complex way, even in the absence of both higher-order linguistic cues and larger boundary cues. In the next section, we further explore the extent to which each of the 8

abovementioned acoustic parameters contributes to the classification of the strings as SW#S versus S#WS.

3.2. Logistic Regression

A binary logistic regression analysis was performed with Boundary Location (S#WS versus SW#S) as a dependent variable. Only the eight acoustic parameters that were significantly influenced by Lexical Boundary in our ANOVAs (see Table 2) were entered as independent variables.

The results reveal that when all eight parameters are taken into account, about 70% of the data classification (as S#WS or SW#S) was accounted for by the holistic acoustic signatures of lexical boundaries ($R^2 = 0.694$, $p<0.0001$, with 87.5% accuracy). As summarized in Table 3, however, $RIME_1$, $RIME_2$ and C_3DUR contributed most reliably to the classification (OR=1.116, 0.864 and 0.956, respectively, $p<0.05$). This implies that $RIME_1$, $RIME_2$ and C_3DUR are the best predictors of the lexical boundary: the presence of the boundary between the first and the second syllables (S#WS) is best marked by a *longer* $RIME_1$, but a *shorter* $RIME_2$; and the absence of the boundary between the second and the third syllables (S#WS) is most reliably marked by a shorter C_3DUR (as compared to SW#S).

Table 3. A summary table of a binary logistic regression analysis with Boundary Location as a dependent variable.

Measure	B (coef.)	Wald (1)	p-values	Odds Ratio (Exp(B))
C_2DUR (ms.)	.113	2.516	.113	1.120
C_2VOT (ms.)	-.133	1.690	.194	.875
C_3DUR (ms.)	-.045	5.776	.016	.956
$RIME_1$ (ms.)	.109	7.015	.008	1.116
$RIME_2$ (ms.)	-.146	8.112	.004	.864
$RIME_3$ (ms.)	-.001	.003	.955	.999
V_2AMP (dB)	.118	.128	.720	1.125
C_3BURST(dB)	-.288	3.127	.077	.750
constant: S#WS	6.813	.047	.828	909.256

3.3. Center of Gravity of [s] and C-to-C coarticulation

As shown in Table 4, COG of [s] in *rei[s]* did not vary with the presence or absence of the following lexical boundary. In contrast, COG of [s] in the second syllable (*reis pa[s]*) was significantly higher word-internally (followed by no lexical boundary as in *reis # pa[s]tei*) than word-finally (followed by a lexical boundary in *reispa[s] # tij*). The higher COG for [s] may be because [s] is produced with a relatively smaller front cavity. It is also conceivable that the channel between the palate and the tongue necessary for the frication noise of [s] may be smaller and thus produce a higher velocity jet of air, which, on striking the upper incisors, would produce a higher frequency. Given these possibilities, the higher COG for word-internal [s] can be interpreted as a consequence of greater coarticulatory influence from the following [t].

The phonemic status of [ʃ] is only marginal with little lexical support for [s]-[ʃ] contrasts in Dutch. Thus, there may be a substantial degree of freedom in the place of articulation of [s] often produced in the postalveolar region (which has been impressionistically observed by the authors, especially in

comparison with English [s]). Such greater degree of freedom, however, may be constrained by the type of following segment. When [s] is followed by a segment such as [t] that shares articulators with [s], its articulation likely resembles that of [t] (C-to-C coarticulation). The resemblance of [s] to [t] may mean that the wide range in place of articulation for [s] (possibly straddling both the alveolar and postalveolar regions) is more narrowed down to the alveolar region, which is the typical place of articulation for [t]. It is also plausible that the constriction for [s] becomes narrower in preparation for the complete closure for the following [t]. Both of these possibilities are likely to yield higher COG as the coarticulatory influence of [t] becomes greater. We suggest that the degree of such coarticulation is greater with no lexical boundary (S#<u>WS</u>) than across a boundary (S<u>W#S</u>) (cf. [13]).

In contrast to the [st] sequence, however, when [s] is followed by [p] as in *reis(#)pas*, the coarticulatory influence of the following [p] is likely to decline as [s] and [p] are produced by independent active articulators, the tongue tip/blade and the lips, in different places of articulations. This is presumably why [s] in *rei[s](#)pas* did not show differential COGs as a function of a lexical boundary.

Table 4. A summary table of repeated measures ANOVAs for the string *reis-pas-tij* with center of gravity during /s/.

Measure COG (HZ)	Descrip -tion	F-ratio [1,23]	S#WS means (s.e.)	SW#S means (s.e.)
rei[s]	n.s.	0.324	5157 (175)	5216 (254)
pa[s]	higher*	7.412	5730 (282)	5649 (266)

4. Summary and conclusion

In this paper, we have explored how the lexical boundaries of potentially ambiguous two-word strings (*reispas # tij* versus *reis # pastei*) are manifested in the acoustic signal. While such a potential ambiguity may be resolved with resort to higher-order linguistic information and/or placing a larger prosodic boundary between the intended lexical boundaries, the present study limited its scope of study to instances in which linguistic and phrasal boundary cues are absent. The results were considered from two inter-related perspectives: how each acoustic parameter under investigation is influenced by a lexical boundary (ANOVAs) and how each acoustic parameter contributes to the classification of two distinct parsings (logistic regression analysis).

The most significant finding of this study was that, although the perceptibility of lexical boundaries may be highly attenuated in the absence of higher-order linguistic cues and phrasal boundaries, several acoustic parameters (C_2VOT, C_3DUR $RIME_1$, $RIME_2$, $RIME_3$, V_2AMP and C_3BURST) were significantly influenced by the presence of a lexical boundary. All of these acoustic correlates of lexical boundaries can be accounted for by prosodically-conditioned phonetic events such as domain-initial strengthening, domain-final lengthening, polysyllabic shortening, and predominant left-to-right accentual lengthening. In addition, a microscopic analysis for [s] suggested that the degree of C-to-C coarticulation in [st] sequence is constrained by a lexical boundary such that a lesser coarticulation was evident across a lexical boundary.

While speakers may produce all of these cues as correlates of lexical boundaries, a binary logistic regression

analysis indicated that a subset of these correlates (C_3DUR $RIME_1$, $RIME_2$, $RIME_3$) played a particularly important role in disambiguating the sequences. It remains to be seen whether listeners are sensitive to all eight lexical boundary correlates identified in our initial analysis, or whether a subset of these cues are weighed more heavily, as the logistic regression analysis suggested. Either way, the implication of the present study is that phrase-internal lexical boundaries are marked by non-contrastive fine-grained phonetic details. Further studies will be needed to determine how salient these cues are. If listeners can detect and interpret these cues accurately enough, subphonemic cues to lexical boundaries could play an important role in achieving the communicative goals of the speaker. This, in turn, would have important implications for models of speech production and perception.

6. References

[1] Jacobson, R, C., Fant, G. M., and Halle, M. (1952). *Preliminaries to speech analysis: The distinctive features and their correlates.* MIT Press, Cambridge, 1952.

[2] Christophe, A., Gout, A., Peperkamp, S., Morgan, J. Discovering words in the continuous speech stream: the role of prosody. *J. of Phonetics,* **31**: 585-598, 2003.

[3] Beckman, M.E. & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3: 255-309.

[4] Turk, A. & Shattuck-Hufnagel, S. Word-boundary-related duration patterns in English. *J. Phon.*, 28: 397-440, 2000.

[5] Keating, P. A., Cho, T., Fougeron, C. & Hsu, C. (2003). Domain-initial strengthening in four languages. In Local, Ogden, & Temple (eds.), *Papers in Laboratory Phonology 6*, 145-163, C.U.P., Cambridge, 2003.

[6] Quené, H. Durational cues for word segmentation in Dutch. *J. Phon.*, 20: 331-350.

[7] Gow, D.J. Jr. & Gordon, P.C. (1995). Lexical and prelexical influences on word segmentation. *JEP:HPP*, 21, 344-359.

[8] Salverda, A. P., Dahan, D. & McQueen, J. M. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90: 51-89, 2003.

[9] Johnson, E.K. Speaker intent influences infants' segmentation of potentially ambiguous utterances. *Proc. of the 15th ICPhS*, 1995-1998. Barcelona, Spain. 2003.

[10] Keating, P.A., Wright, R., & Zhang, J. (1999). Word-level asymmetries in consonant articulation. *UCLA Working Papers in Phonetics*, 97: 157-173.

[11] Cho, T. Lexical stress, phrasal accent and prosodic boundaries in the realization of domain-initial stops in Dutch. *Proc. of the 15th ICPhS*. Barcelona, Spain 2003

[12] Cambier-Langeveld, T. & Turk, A.E. A cross-linguistic study of accentual lengthening: Dutch *vs*. English. *J. Phon.*, 27: 255-280, 1999.

[13] Cho, T. Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *J. Phon.*, 32 2004.