ACOUSTIC-PHONETIC CUES TO WORD BOUNDARY LOCATION: EVIDENCE FROM WORD SPOTTING

Nicolas Dumay[1], Alain Content[1,2] & Uli H. Frauenfelder[2]
(1) *Laboratoire de Psychologie Expérimentale – Université Libre de Bruxelles – Belgium*
(2) *Laboratoire de Psycholinguistique Expérimentale – Université de Genève – Switzerland*

ABSTRACT

This research examined acoustic-phonetic cues to word boundary location in French consonant clusters, and assessed their use in on-line lexical segmentation. Two word-spotting experiments manipulated the alignment between word targets and syllable boundaries. A perceptual cost of such misalignment was observed for obstruent-liquid clusters but not for /s/ + obstruent clusters. For the former clusters, the analysis of a corpus of utterances showed systematic variations in segment durations as a function of the lexical assignment of the pivotal consonant. We conclude that the availability of acoustic-phonetic cues to word boundary location in consonant clusters depends upon the cluster class. When available, these cues are exploited in on-line lexical segmentation of speech.

1. INTRODUCTION

When listening to speech, human listeners are confronted with the problem of locating words within the continuous speech stream. The absence of reliable word boundary markers in the speech signal –like systematic pauses– has led psycholinguists to search for other, more subtle word boundary cues that the listener could use to parse the incoming input. The syllabic structure of the speech signal appears to play a determining role in lexical segmentation. Several studies performed in English and Dutch provide evidence for this hypothesis. For English, Cutler and colleagues [1,2,3] have proposed and adduced experimental evidence that listeners rely on a Metrical Segmentation Strategy (MSS), in which strong syllable onsets constitute privileged alignment points for initiating a lexical search. In Dutch, Vroomen & de Gelder [4] assessed the activation of monosyllabic words embedded in longer carrier words. The authors measured the amount of cross-modal priming in a lexical decision task on visual words related or not to the embedded word. They clearly demonstrated the activation of embedded words aligned with a strong syllable onset [e.g. *boos* (angry) in *fram**boos*** (rasperry)] but found no evidence of activation of words that were misaligned with a syllable onset [e.g. *wijn* (wine) in *z**wijn*** (swine)]. In a word-spotting study also performed in Dutch, in which participants had to detect monosyllabic words embedded finally or initially in bisyllabic nonce strings, McQueen [5] showed that misalignment between word and syllable onsets [e.g. *rok* (skirt) in *fi.d**rok*** vs in *fim.**rok***] produced longer reaction times than misalignment between word and syllable offsets [e.g. *vel* (skin) in ***vel**m.brul* vs in ***vel**.brul*].

We have made a similar proposal for French [6,7]. However, unlike for English and Dutch in which segmentation is claimed to be restricted to strong syllables, we suggest that in French every syllable onset provides an alignment point for a lexical search. Evidence for this conclusion comes from a word-spotting experiment [8]. Participants had to detect monosyllabic CVC words (e.g. "*lac*") embedded initially or finally in bisyllabic nonce strings with aligned CVC.CVC (e.g. *zun.**lac*** or ***lac**.tuf*) or non-aligned CV.CCVC (e.g. *zu.g**lac*** or ***la**.**c**luf*) syllabic structure. According to the syllable onset hypothesis, only onset alignment is crucial. Hence, the processing cost should be greater for an onset misalignment (as in *zu.g**lac***) than for an offset misalignment (***la**.**c**luf*). This prediction was confirmed: a significant 74 ms effect was obtained for onset misalignment (*zun.**lac*** vs *zu.g**lac***), whereas a smaller and non-significant effect was found for offset misalignment.

One important shortcoming of a syllable-based segmentation strategy is its difficulty in handling potential resyllabification phenomena resulting from phonological processes applying across word boundaries. For example, in the case of French liaison, syllable boundaries are presumably not coterminous with word boundaries, and an incorrect lexical alignment would be made on the basis of syllable onsets. More generally, phonological theories assume that resyllabification occurs across word boundaries, based upon either the Maximal Onset Principle [9] or the Sonority Principle [10]. Accordingly, some psycholinguistic models of speech production [11] have also endorsed the notion of resyllabification. These models assume that speech production involves abstract syllabic gestural scores computed or accessed after resyllabification has occurred. Such a process would create frequent misalignments between syllable onsets and word onsets at the phonetic level, and hence a syllable onset alignment strategy would appear to be misguided and implausible.

However, a few studies have suggested the existence of acoustic-phonetic cues, mainly of durational nature, that depend upon the lexical assignment of the intervocalic pivotal consonants [12,13,14]. In addition, listeners have been shown to be sensitive to such cues. When Dutch listeners hear ambiguous two-word utterances such as *"die pin"* vs *"diep in"* [di(#)p(#)ɪn] in a forced-choice task, they exploit durational cues in the pivotal consonant and in the following vowel to make segmentation decisions [13].

The present research examined the nature of these word-boundary cues in consonant clusters in French, and assessed their use in on-line lexical segmentation and word recognition. Two word-spotting experiments were carried out in which monosyllabic CVC words had to be detected in bisyllabic carriers. We specifically compared two phonetic realizations of phonologically identical carriers in which the syllable assignment of the pivotal consonant was manipulated (VC.CV or V.CCV). Experiment I assessed the effect of initial and final misalignment between word and syllable boundaries in OBstruent + LIquid (OBLI) consonant clusters. Experiment II compared OBLI clusters to /s/ + OBstruent (SOB) clusters. Experiment III analysed durational properties of the carriers and evaluated the generality of the durational differences in a production study.

2. EXPERIMENT I

2.1. Method

2.1.1. Participants. Thirty-two students participated in the experiment. All were native speakers of French without hearing or speech disorders.

2.1.2. Materials. To examine the effect of onset misalignment, 32 VCCVC bisyllabic nonce strings with a medial OBLI cluster *(i.e. bl, gl, fl, pl, gʀ, dʀ, bʀ, tʀ, kʀ, fʀ* and *pʀ)* were constructed. Each contained a target CVC word at its end (e.g. *roche* [ʀɔʃ] "rock" in [ikʀɔʃ]), with the constraint that the final CCVC of the carrier was also a word (*croche,* [kʀɔʃ], "eighth note"). Two versions of each carrier were obtained. They were extracted from the end of longer two-word phrases that differed mainly in the word boundary location (i.e. ...VC#CVC, as in "*magique roche*" or ...V#CCVC, as in "*demi-croche*"). Thus, in the aligned condition, the onset of the target CVC word corresponded to an intended word boundary ([ik#**ʀɔʃ**]), whereas in the misaligned condition, target onset did not correspond to the intended word boundary ([i#k**ʀɔʃ**]). In order to prevent lexical biases from favoring one particular interpretation, non-target portions of the carriers also formed possible words. Finally, 32 similar VCCVC strings such that the CVC portion did not form a word served as filler trials.

Similarly, to examine the effect of offset misalignment, 32 CVCCV nonce carriers were devised, using again OBLI medial clusters (i.e. *pl*, *kl*, *bl*, *gl*, *tʀ*, *dʀ* and *vʀ*). The target words were the initial CVC (e.g. *tante* [tɑ̃t] "aunt" in [tɑ̃tʀu]), the constraint being in this case that the initial CV of the target was also a word (*temps* [tɑ̃] "time"). Again, two acoustic-phonetic versions of each carrier were extracted from longer paired two-word phrases (e.g. [**tɑ̃t**#ʀu], from "*tante roublarde*", or [**tɑ̃**#**t**ʀu], from "*temps troublant*"), and 32 CVCCV filler strings were built along the same principles.

All phrases were produced by a male native speaker of French (N.D.) who pronounced the phrases naturally, without separating the two words. The phrases were recorded in a sound-attenuated booth, and digitalized at a 44 kHz sampling rate. The carrier strings were extracted using Sound Designer software, and after amplitude normalization, they were stored at a 22 kHz sampling rate with 16-bit resolution onto a Macintosh PowerPC.

2.1.3. Procedure. Participants were required to press a button as quickly as possible whenever they had heard a CVC word in a previously specified position, and to repeat the word aloud immediately after. Carriers were presented through headphones directly from the computer disk, and reaction time collection was controlled by the computer. Carriers and fillers were mixed randomly in two experimental lists, alignment being counter-balanced across the two lists. As the position of the target was fixed within a given block of trials, the order of block presentation was counter-balanced across subjects.

2.2. Results and discussion

Analyses were based upon reaction times (RTs) for correct responses, measured from the end of the target. Words embedded in final position were on average detected 98 ms faster in the aligned than in the misaligned condition. Similarly, words embedded in initial position were on average detected 140 ms faster when their offset was aligned than when it was not. Both effects were highly reliable in separate analyses of variance (ANOVAs; final embedding: $F1(1,31) = 8.94$, $p < .01$; $F2(1,31) = 24.49$, $p < .001$; initial embedding: $F1(1,31) = 18.15$, $p < .001$; $F2(1,31) = 19.07$, $p < .001$). Moreover, although the alignment effect tended to be larger for initial word detection, the alignment effects did not differ statistically from one another, as shown by the absence of interaction between Alignment and Position ($F1(1,31) = 1.01$; $F2(1,62) = 1.26$, $p < .3$). Error rates were very similar across conditions. ANOVAs performed on error rates revealed no reliable effect or interaction.

Table 1. Mean reaction time (RT) from the end of the target and error rate (Er) in Experiment I (** indicates that the effect is statistically significant by participants and items).

Target onset	Carrier type	RTs (ms)	Ers (%)
Aligned	ik#**ʀɔʃ**	869	13
Misaligned	i#k**ʀɔʃ**	967	14
Difference		98**	1
Target offset			
Aligned	**tɑ̃t**#ʀu	916	14
Misaligned	**tɑ̃**#**t**ʀu	1056	16
Difference		140**	2

The finding of perceptual costs at target onset and offset indicates that, in OBLI clusters, acoustic-phonetic cues to the assignment of the pivotal consonant are present and are exploited on-line by the listener. Interestingly, contrary to our previous experiments [8], similar processing costs were observed for onset and offset misalignment.

3. EXPERIMENT II

Experiment II aimed at assessing the generality of these findings by comparing OBLI clusters with SOB clusters. While OBLI clusters are unanimously assigned to syllable onset position in phonological analyses, and comply with both the Maximal Onset Principle and the Sonority Principle, SOB clusters admit divergent syllabifications. Because it was impossible to find a sufficient number of SOB target words satisfying the required constraints, the comparison was restricted to the final alignment condition.

3.1. Method

3.1.1. Participants. Forty-eight students participated in the experiment. All were native speakers of French without hearing or speech disorders.

3.1.2. Materials and procedure. Thirty-two bisyllabic CVCCV or CCVCCV nonsense strings with a medial SOB cluster (i.e. *st, sp, sk* and *sf*) were constructed. Each carried a target CVC or CCVC word at the beginning (e.g. *race*, [ʀas] "race" in [ʀasty]), with the constraint that the initial C(C)V of the target was also a word (*rat*, [ʀa] "rat"). As in Experiment I, two versions of each carrier were extracted from the beginning of paired two-word phrases that differed in the word boundary location (i.e., VC#CV as in "*race tuméfiée*" or V#CCV as in "*rat stupéfait*"). Note that the lexical status of the final CCV and CV portions of the carriers was impossible to control for.

For the OBLI condition, 32 target words and carrier strings were devised as in Experiment I but in the present experiment both CVC (e.g. *tante* [tɑ̃t] "aunt") and CCVC (e.g. *grappe* [gʀap] "cluster") words served as targets. Other characteristics of the materials and procedure were identical to Experiment I, except that the speaker was different.

3.2. Results and discussion

In the reaction time analysis, a significant interaction between alignment and cluster type was obtained ($F1(1,47) = 3.83$, $p < .057$; $F2(1,62) = 9.62$, $p < .01$). As shown in Table 2, a reliable effect of Alignment was found for OBLI clusters (98 ms; $F1(1,47) = 5.65$, $p < .025$; $F2(1,31) = 21.80$, $p < .0001$). In contrast, no Alignment effect was obtained for SOB clusters (6 ms; $Fs < 1$). Error rates for the OBLI clusters were slightly lower in the aligned than in the misaligned condition, and the reverse trend occurred for SOB clusters. However, none of these trends was reliable. Thus, in summary, the results of Experiment II replicate the previous findings for OBLI clusters with a different naïve speaker, but fail to reveal similar alignment effects in the case of SOB clusters.

Table 2. Mean reaction time (RT) from the end of the target, and error rate (Er) in Experiment II (* indicates that the effect is statistically significant by participants only).

	Carrier type	RTs (ms)	Ers (%)
OBLI			
Aligned	**tɑ̃t**#ʀu	826	19
Misaligned	**tɑ̃**#**t**ʀu	924	23
Difference		98**	4
SOB			
Aligned	**ʀas**#ty	762	18
Misaligned	**ʀa**#**s**ty	768	14
Difference		6	4*

4. EXPERIMENT III

In order to examine the nature of the acoustic-phonetic differences between the two versions of each carrier, segment durations for all carriers from both experiments were estimated from waveforms. When the pivotal consonant was assigned to the offset of the first word (...VC#CV...), OBLI strings showed a substantial and reliable lengthening of the pre-boundary vowel (Speaker 1: 29%; Speaker 2: 20%) and of the liquid consonant (S1: 41%; S2: 47%) for both speakers, as well as a substantial and reliable lengthening of the obstruent for S1 only (30%). In contrast, no such durational differences were observed in SOB carriers.

Experiment III assessed the generality of these findings by collecting production data from eight naïve speakers of French. They were recorded pronouncing a subset of the two-word phrases used in the previous experiments, both in isolation and in sentence context.

4.1. Method

4.1.1. Materials and procedure. The materials were composed of 48 pairs of two-word phrases (CVC#CV… vs CV#CCV…); 32 pairs with an OBLI cluster (16 OBstruent + /l/ and 16 OBstruent + /ʀ/) used in Experiment I, and 16 with a SOB cluster from Experiment II.

The 96 phrases were mixed randomly in two balanced blocks. Production in isolation consisted in reading the phrases naturally without separating the words. For production in sentential context, the relevant phrases appeared just after the verb as one of its complement. Sentences containing paired phrases were matched for length in syllables and for the position of the relevant phrase. In half the sentences the final constituent was attached to the head noun of the phrase as in "*Aurélie est une **tante rou**blarde du Brabant*", matched with "*Il naquît en ces **temps trou**blants d'avant-guerre*", while in the other half the final constituent was attached to the verb as in "*J'ai souffert d'une **mince ca**taracte comme un martyr*", matched with "*Il nettoie la **main sca**rifiée à l'aide d'éther*". Speakers were required to read the sentence silently before saying it aloud.

4.2. Results and discussion

Table 3. Mean segmental durations (ms) as a function of the consonant cluster, the word boundary location (WBL) and the production context. Each cell represents 128 data points.

Production	C_2C_3	WBL	C_1	V_1	C_2	C_3	V_2
in isolation	OB + /l/	C#C	106	143**	89	61**	101
		#CC	105	115	83	47	99
	OB + /ʀ/	C#C	88	135**	63	99**	85
		#CC	91	111	59	82	79
	/s/ + OB	C#C	95	119	95	88	73
		#CC	96	116	89	83	83
in sentence	OB + /l/	C#C	68	122**	68	49**	97
		#CC	66	102	67	41	94
	OB + /ʀ/	C#C	59	108**	48	87**	81
		#CC	56	89	50	74	73
	/s/ + OB	C#C	61	101	73	81	67
		#CC	63	96	77	77	74

For the phrases with an OBLI cluster, the lengthening of the pre-boundary vowel and of the liquid was clearly and reliably confirmed. Both in isolation (23% for V_1; 25% for C_3; see Table 3) and in sentential context (20% for V_1; 19% for C_3), the pre-boundary vowel and the liquid were found to be longer when the word boundary was located within the cluster. As for the stimuli of Experiment II, no reliable difference emerged for SOB clusters.

5. GENERAL DISCUSSION

This research examined the acoustic-phonetic cues to word-boundary location in two classes of consonant clusters in French.

It clearly appears that the presence of such cues depends upon the nature of the cluster. In SOB clusters, durational analyses revealed no reliable difference due to the word boundary location. In contrast, analyses of OBLI clusters revealed a substantial and systematic lengthening of the pre-boundary vowel and of the liquid when the obstruent was word-final. Thus, the OBLI results indicate that phonetic encoding in speech production does not lead to complete resyllabification. In contrast to the predictions by Levelt and Wheeldon [11], it appears that, at least for OBLI clusters, the production of a VC#CV sequence requires syllabic gestural scores that differ from those activated during production of the paired V#CCV sequence, in which the obstruent is assigned to the second syllable at the lexical level.

The present results also provide strong evidence that the acoustic-phonetic word boundary cues in OBLI clusters are exploited in on-line lexical segmentation. In two word-spotting experiments, onset and offset misalignment between the target word and the syllabic structure produced significant perceptual costs. Such findings confirm results obtained in English and Dutch with off-line discrimination tasks [12,13,14] and extend them to French clusters with an on-line word recognition task.

Our hypothesis of lexical segmentation at syllable onsets predicts an onset misalignment effect. Thus, in the case of finally-embedded targets [e.g. i.k**ʀɔʃ**], recognition of the target word [**ʀɔʃ**] is assumed to be delayed because an incorrect segmentation decision is initially taken and subsequently needs to be revised. In contrast, for the recognition of the initially-embedded target [e.g. **tɑ̃.t**ʀu], another interpretation is proposed. We assume that recognition is delayed because the target word [**tɑ̃t**] must compete with the word aligned with the onset of the second syllable [tʀu] for the pivotal consonant [t]. Since the competitor is activated, the decision to assign the pivotal consonant to the target word receives less support and takes more time.

According to the syllable-onset hypothesis, final misalignment effects should disappear if there is no lexical competitor for the pivotal consonant. Such an interpretation is supported by the lexical segmentation data obtained in English and Dutch. In both languages lexical activation was shown to be modulated by the number of activated words beginning with the next strong syllable [3,15]. Moreover, the interpretation may also explain the absence of final misalignment effect [e.g. **la.k**lyf] in Dumay et al. [8] since in this experiment the second syllable [klyf] of the carrier never corresponded to a word. Current experiments are directly examining the modulation of final misalignment effects by the number of competitors, and should provide a further test of our hypothesis.

In conclusion, the present study indicates that the existence of phonetic cues to word boundary location in French consonant clusters depends upon the nature of the consonants. When available, as in obstruent-liquid clusters, such subtle but reliable cues are effective in on-line lexical segmentation and word recognition. The findings argue against the notion of resyllabification which would assume that a word-final obstruent is systematically re-assigned to the next syllable onset when followed by a liquid. They reinforce our belief that syllable onsets constitute relevant alignment points for lexical search.

ACKNOWLEDGMENTS

This research was supported by grants from the Direction Générale de la Recherche Scientifique – Communauté française de Belgique (A.R.C. 96/01-203) and from the Swiss FNRS (Project 1113-049698.96). N. Dumay (ndumay@ulb.ac.be) is research assistant at the Belgian National Fund for Scientific Research.

REFERENCES

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

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

[3] Norris, D., McQueen, J.M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 1209-1228.

[4] Vroomen, J., & de Gelder, B. (1997). Activation of embedded words in spoken word recognition. Journal of Experimental Psychology: Human Perception and Performance, 23, 710-720.

[5] McQueen, J.M. (1998). Segmentation of continuous speech using phonotactics. Journal of Memory and Language, 39, 21-46.

[6] Content, A., Kearns, R.K., & Frauenfelder, U.H. (1998). Boundaries versus onsets in syllabic segmentation. Journal of Memory and Language. Manuscript submitted for publication.

[7] Frauenfelder, U.H., & Content, A. (1999). The role of the syllable in spoken word recognition: Access or segmentation? Actes des 2èmes Journées d'Etudes Linguistiques, Nantes, in press.

[8] Dumay, N., Banel, M.H., Frauenfelder, U.H., & Content, A. (1998). Le rôle de la syllabe: segmentation lexicale ou classification? Actes des XXIIèmes Journées d'Etude sur la Parole, (pp. 33-36), Martigny, Suisse.

[9] Pulgram, E. (1970). Syllable, word, nexus, cursus. The Hague: Mouton.

[10] Clements, G.N. (1990). The role of sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), Between the grammar and the physics of speech. New York: Cambridge University Press.

[11] Levelt, W.J., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? Cognition, 50, 239-269.

[12] Barry, W.J. (1984). Perception of juncture in English. In M.P.R. van den Broecke & A. Cohen (Eds.), Proceedings of the Tenth International Congress of Phonetic Sciences, Utrecht 1983, pp. 529-536. Dordrecht, Cinnaminson: Foris.

[13] Quené, H. (1992). Durational cues for word segmentation in Dutch. Journal of Phonetics, 20, 331-350.

[14] Quené, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. Journal of the Acoutical Society of America, 94, 2027-2035.

[15] Vroomen, J. & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. Journal of Experimental Psychology: Human Perception and Performance, 21, 98-108.