



Durational Cues to Word Recognition in Spoken French

Ellenor Shoemaker

► To cite this version:

| Ellenor Shoemaker. Durational Cues to Word Recognition in Spoken French. 2011. <halshs-00683607>

HAL Id: halshs-00683607

<https://shs.hal.science/halshs-00683607v1>

Preprint submitted on 29 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Durational Cues to Word Recognition in Spoken French

Ellenor Shoemaker

Université Sorbonne Nouvelle – Paris 3/

Laboratoire de Phonétique et Phonologie (CNRS/UMR 7018)

Author Information

Institute du Monde Anglophone
5, rue de l'Ecole de médecine
75006 Paris FRANCE
+331.40.51.33.00

ellenor.shoemaker@univ-paris3.fr

Abstract

In spoken French, the phonological processes of liaison and resyllabification can render word and syllable boundaries ambiguous (e.g. *un air* ‘an air’ / *un nerf* ‘a nerve’, both [ẽ.nɛʁ]).

Production data have demonstrated that speakers of French vary the duration of consonants that surface in liaison environments relative to consonants produced word-initially (Wauquier-Gravelines 1996; Spinelli *et al.* 2003). Further research has suggested that listeners exploit these durational differences in the processing of running speech (Gaskell *et al.*, 2001; Spinelli *et al.*, 2003), though no study to date has directly tested this hypothesis. The current study examines the exploitation of duration in word recognition processes by manipulating this single acoustic factor while holding all other factors in the signal constant. The pivotal consonants in potentially ambiguous French sequences (e.g. /n/ in *un nerf*) were instrumentally shortened and lengthened and presented to listeners in two behavioral tasks. Results suggest that listeners are sensitive to segmental duration and use this information to modulate the lexical interpretation of spoken French.

The syllable has been shown to play a prominent role in the processing of continuous speech in French (Mehler, Dommergues, Frauenfelder, & Segui, 1981; Cutler, Mehler, Norris, & Segui, 1989). However, the frequent misalignment of syllable and word boundaries due to the phonological process of *enchaînement* (resyllabification) and *liaison* (linking) is problematic for syllable-based lexical access models which assume that the edges of words and syllables tend to coincide. *Enchaînement* occurs when a consonant-final word (W_1) is followed by a vowel-initial (V-initial) word (W_2). The coda of W_1 is resyllabified across the word boundary to become the onset of W_2 . The phrase *une amie* ‘a friend’ (feminine) is thus produced as [y.na.mi] where syllable and word boundaries are mismatched, instead of [yn.a.mi] where boundaries would be aligned. Liaison on the other hand concerns consonants in final position that are represented graphically¹, but are latent when the word is produced in isolation or before a consonant-initial (C-initial) word. The latent consonant is realized however before a V-initial word and then resyllabified through *enchaînement*. For example, the singular masculine indefinite article *un* is pronounced [ɛ̃] in isolation or before a consonant (e.g., *un stylo* [ɛ̃.sti.lo] ‘a pen’), however, when preceding a vowel onset in W_2 , as in *un ami* ‘a friend’ (masculine), the latent /n/ surfaces and is syllabified as the onset of *ami*. Accordingly, the sequence is syllabified [ɛ̃.na.mi] instead of [ɛ̃n.a.mi] where word boundaries would be respected. As a result, boundaries between syllables and words are often blurred (e.g., *un air* ‘an air’ and *un nerf* ‘a nerve’, both [ɛ̃.nɛʁ]).

Boundary misalignment has been shown to incur processing costs in French. Using a word-spotting task, Dumay, Frauenfelder and Content (2002) showed that reaction times were

¹ Except in cases of *epenthetic* liaison, as in *quatre* [z] *enfants* ‘four children’, in which a liaison consonant is introduced in production to resolve hiatus, but is not represented in the orthography of the W_1 .

significantly faster in identifying the word *lac* ‘lake’ embedded in the non-word *zun.lac*, where the illicit onset /n/ forces a syllable boundary between /n/ and /l/, than in *zu.glac*, where /gl/ is an allowed onset and therefore word and syllable boundaries are not necessarily aligned. It is the mismatch of word and syllable boundaries in the latter that these authors propose to be slowing reaction times. This work is consistent with research in numerous languages suggesting that listeners are sensitive to the fact that syllable boundaries generally coincide with word boundaries and exploit this fact in spoken word recognition (Content, Kearns, & Frauenfelder, 2001; Content, Meunier, Kearns, & Frauenfelder, 2001; Cutler & Norris, 1988; Norris, McQueen, Cutler, & Butterfield, 1997; Vroomen & de Gelder, 1997).

Given the prominent role of the syllable in French, the prevalence of liaison and resyllabification would presumably hinder speech processing and impede access to mental representations. Competition-based models of spoken word recognition (e.g. McClelland & Elman, 1986; Norris, 1994) propose that a set of candidate words consistent with acoustic (bottom-up) cues are simultaneously activated in a listener’s mental lexicon as the input is processed in real time. The competition process concludes when an optimal parse is achieved and the acoustic signal is segmented into non-overlapping words. In the case of liaison, however, multiple lexical interpretations can account for the same phonetic sequence (e.g. [ẽ.neʁ]), and therefore competition processes are assumed to be hindered.

A study by Gaskell, Spinelli, and Meunier (2002), however, suggested that resyllabification in French can in fact *facilitate* the recognition of V-initial words. A cross-modal priming study showed significant priming effects as compared to a control for V-initial words in three conditions: liaison (e.g., *un généreux italien* [ẽ.ʒe.ne.ʁø.zi.ta.ljẽ] ‘a generous Italian’),

enchaînement (e.g., *un virtuose italien* [ẽ.viʁ.tʁo.zi.ta.ljẽ] ‘an Italian virtuoso’) and a syllable-aligned condition (e.g., *un chapeau italien* [ẽ.ʃa.po.i.ta.ljẽ] ‘an Italian hat’). Reaction times were measured and participants recognized V-initial targets preceded by resyllabified consonants, in both liaison (e.g. *un généreux italien*) and *enchaînement* (e.g. *un virtuose italien*) conditions, significantly faster than targets that matched syllable boundaries (e.g. *un chapeau italien*). The authors’ suggestion was that resyllabification is somehow acoustically marked and that these acoustic cues facilitate the lexical competition process². This study extended previous findings by Wauquier-Gravelines (1996), who showed in a word-monitoring task that *éléphant* ‘elephant’ is recognized as readily in *un petit éléphant* [ẽ.pə.ti.te.le.fã] ‘a small elephant’, a liaison environment, as in *un joli éléphant* [ẽ.ʒo.li.e.le.fã], where no liaison is possible and boundaries are aligned.

Spinelli, McQueen, and Cutler (2003) also found that perceptual efficacy and processing are not hindered by liaison and resyllabification. These authors probed lexical access processes and revealed significant priming effects for both C-initial and V-initial words in globally ambiguous sentence pairs such as *c’est le dernier rognon*, ‘it’s the last kidney’, and *c’est le dernier oignon*, ‘it’s the last onion’, both [se.lə.dɛʁ.nje.ʁɔ.pɔ̃]. Spinelli *et al.* employed four priming conditions in a lexical-decision task: an ambiguous liaison condition (*c’est le dernier oignon*), an ambiguous non-liaison condition (*c’est le dernier rognon*), an unambiguous condition where liaison would not be possible (*c’est un demi rognon*, ‘It’s a half kidney’), and

² As pointed out by a reviewer of this article, one could posit an alternate explanation for the facilitation effects in that the third condition (e.g. *un chapeau italien*) represents a situation of hiatus, which could be more difficult to process than a consonantal onset.

finally an unambiguous baseline condition including an unrelated word where liaison would not be possible (*c'est un ancien nitrate*, 'It's an old nitrate').

Priming effects were observed for both V-initial (*oignon*) and C-initial (*rognon*) candidates in the ambiguous conditions. In other words, the ambiguity caused by liaison and subsequent resyllabification did not impair the lexical activation of the V-initial candidate. Furthermore, priming effects followed the intention of the speaker, i.e. priming effects were stronger for *oignon* than for *rognon* when the speaker intended *oignon*, and vice versa. Their results also suggested that words not intended by the speaker in ambiguous contexts (e.g., *oignon* when *dernier rognon* is produced) were activated, but not as strongly as in the intended production. Significantly, they did not find priming effects for *oignon* in the unambiguous condition where liaison is not possible (e.g., *demi rognon*), suggesting that solely the liaison environment allows for the activation of both consonant- and V-initial lexical candidates.

Spinelli *et al.* (2003) also invoke allophonic variation to account for the observed priming effects. They hypothesize that listeners exploit "subtle but reliable" (p. 248) variations in segmental duration to locate word boundaries and that access to mental representations is facilitated by these cues, concluding that durational differences are robust enough to "bias interpretation in the correct direction" (p. 250). In line with this hypothesis are French production data which have revealed significant differences in duration between liaison consonants (LC; e.g., /n/ in *un air*) and initial consonants (IC; e.g., /n/ in *un nerf*). These same authors found significant durational differences among five consonants that surface in liaison /n, t, ʁ, g, p/. Liaison consonants were on average 17% shorter than initial consonants. Measurements of the pivotal consonants revealed that ICs were on average 10 ms longer (difference range= 6 to 12 ms) than word-final, resyllabified consonants (however these authors did not report specific

differences between LCs and ICs for each of the five segments tested). Wauquier-Gravelines (1996) found similar results for /t/, which had an average closure duration of 50 ms in liaison position and 70 ms in initial position, though she did not find significant durational differences between liaison and word-initial /n/ (58 ms versus 61 ms, respectively). Gaskell *et al.* (2002) also found durational differences; the segments /t/, /ʁ/ and /z/ were significantly shorter when realized in liaison (mean 73 ms) than in word-initial position (mean 88 ms). Even more recently, Douchez and Lancia (2008) measured linguopalatal contact and found evidence of more sustained contact between the tongue and the palate in the articulation of ICs than LCs.

This line of research in French is supported by data showing that speakers in general tend to strengthen the articulation of segments at the edges of prosodic domains (Cho & Keating, 2001; Cho, McQueen, & Cox, 2007). Specifically, initial segments have been shown to be systematically longer than the same segment in medial or final position in English (Lehiste, 1961, 1972; Klatt, 1976; Gow & Gordon, 1995; Fougeron & Keating, 1997), French (Fougeron, 2001), Dutch (Shatzman & McQueen, 2006) and Italian (Tabossi, Collina, Mazzetti, & Zoppello, 2000).

However, though systematic durational differences have been found, they may not be robust enough in natural speech to allow listeners to disambiguate spoken French. Using the same recordings of globally ambiguous phrases used in the Spinelli *et al.* (2003) study, Shoemaker and Birdsong (2008) more directly tested the perception of liaison by employing a forced-choice identification task in which 15 native speakers of French and 15 late learners of L2 French were asked to differentiate ambiguous phonemic content, e.g. [il.na.o.kẽ.nɛʁ] produced as either *Il n'a aucun air* or *Il n'a aucun nerf*.

Participants in both groups performed at chance (native speaker mean accuracy, 53.2 %; non-native speaker mean accuracy, 52.7%) suggesting that, though durational differences may allow for the activation of V-initial candidates in the word recognition process, these differences are not sufficiently robust to systematically guide listeners in disambiguation.

More recently, Tremblay (2011) used eye-tracking to investigate the parsing of locally ambiguous real- and non-word sequences containing the pivotal consonant /z/ by both native and non-native speakers of French. Results from this study suggest that distributional cues (i.e. the frequency with which /z/ appears in liaison position relative to lexical-word initial position) influenced processing more than durational differences between /z/ produced in liaison (e.g. *fameux élan* ‘infamous swing’) and /z/ produced as a lexical onset (e.g. *fameux zélé* ‘infamous zealous one’). Put differently, acoustic cues were not sufficient to guide listeners’ eye fixations when the speech signal required temporary disambiguation.

The exploitation of durational variation in word recognition has been demonstrated in numerous languages. Davis, Marslen-Wilson, and Gaskell (2002) were among the first to demonstrate that lexical access processes in English are sensitive to subtle durational differences linked to prosodic boundaries, specifically between a mono-syllabic word and the same phonemic sequence embedded in a longer word (e.g. *cap* and *captain*). This study employed a cross-modal priming task in which participants heard a sequence such as /kæp/ produced either as mono-syllabic *cap* or as the first syllable of *captain*. The results showed differential activation for the two productions, with more activation of the shorter word when participants were presented with mono-syllabic productions, and conversely more activation of the longer word when participants were presented with a portion of the disyllabic sequence.

More recently, Shatzman and McQueen (2006) found that the online interpretation of ambiguous sequences in Dutch can be influenced by segment duration alone. This study investigated the recognition of sequences rendered ambiguous by the lexical assignment of /s/ (e.g. *eens pot* ‘once jar’ and *een spot* ‘a spotlight’). In a production sample, they found variation in several acoustic factors in the ambiguous pairs including segment duration, closure duration of the stop consonant, the duration of the entire word, root mean square energy of /s/, and root mean square energy of the stop consonant following /s/. However, the duration of the segment /s/ alone was significantly predictive of participants’ performance on an identification task that also tracked eye movements. In a second experiment in this study, the researchers instrumentally manipulated the duration of /s/ by both shortening and lengthening the segment. Even stronger correlations between responses and the duration of /s/ were found with these manipulated stimuli, confirming that duration was a sufficient and reliable cue to segmentation.

In the case of spoken French, durational differences between consonants that surface in liaison environments and the same consonant in initial position have been demonstrated in production, however no research to date has confirmed that these durational differences influence lexical interpretation. Thus Spinelli *et al.*’s (2003) suggestion that duration influences lexical access in French remains conjectural as their study did not demonstrate directly that duration was guiding participants’ responses. Moreover, Shoemaker and Birdsong (2008) did not find significant correlations between segment duration and response patterns using the same stimuli in an identification task.

The current study addresses this research gap by using French stimuli in which this single acoustic factor is manipulated in the same physical utterance while holding all other acoustic factors constant. In this way it is possible to directly test listeners’ exploitation of durational

variation in cases of potential lexical ambiguity. To this end, the current study includes both an AX discrimination task and a forced-choice identification task employing stimuli in which the pivotal consonants in ambiguous phrases (i.e., /n/ in [ẽ.nɛʁ], *un air* or *un nerf*) are instrumentally shortened and lengthened while the rest of the utterance remains unaltered. An AX discrimination task is used to tap lower-level acoustic processing and is motivated by the assumption that segmental duration represents an effective cue to segmentation and lexical access only to the extent that this cue is perceptually salient to listeners. A forced-choice identification task is used to investigate the exploitation of segmental duration in higher-level lexical decision processes by determining the extent to which segmental duration modulates lexical interpretation.

Specific questions asked about the processing of liaison in French are thus the following:

- (1) What are the thresholds of perceptual saliency with respect to durational differences between consonants that surface in liaison and word-initial consonants for speakers of French? (2) Are acoustic differences sufficient to override ambiguity in globally ambiguous liaison contexts? (3) To what extent does segmental duration modulate lexical access and the segmentation routines of speakers of French?

Method

Materials

Of the six segments that surface in liaison environments in French, /g, n, p, ʁ, t, z/, three, /n, t, z/, were chosen to be included in this study. Firstly, these three segments allow for the testing of a sonorant consonant (/n/) as well as two classes of obstruent, a stop (/t/) and a fricative

(/z/), which allows us to investigate whether the systematicity and/or robustness of durational differences that arise in environments of liaison vary as a function of consonant class.

Second, these three segments were chosen due to their frequency of occurrence in environments of liaison in contemporary spoken French. According to Durand and Lyche (2008), who analyzed a corpus of speech samples from over 600 native French speakers from the francophone world, these three consonants are the most commonly realized in liaison environments. Out of 9920 realizations of liaison in this corpus, /z/ accounts for 4544 realizations, /n/ for 3689, and /t/ for 1665. The liaison consonants /g, p, ʁ/ collectively account for the remaining 22 realizations. Previous studies have shown similar proportions. According to Léon (1992) roughly 50% of liaisons are realized with /z/, while /n/ and /t/ each account for approximately 25% of liaison occurrences. Realizations of liaison with /g, p, ʁ/ make up less than 1% combined.

Two lists of 12 globally ambiguous pairs of two- and three-word phrases incorporating these three segments were created: one list containing real words and one non-words containing final position. (See Appendices I and II.) For the real-word stimuli, four V-initial words for each of the three consonants were selected such that the realization of these words preceded by a word that triggers liaison gives rise to a sequence that is ostensibly homophonous. For example, the word *air* ‘air’ [ɛʁ] preceded by *un* [ɛ̃], the singular masculine indefinite article, yields a phonemic sequence consistent with both *un air* ‘an air’ and *un nerf* ‘a nerve’, [ɛ̃.nɛʁ].

A second list of non-words was also created. Four V-initial sequences with transparent orthography and conforming to French phonotactics were created for each of the three liaison consonants such that the realization of these sequences in liaison environments creates two

homophonous non-word sequences. For example, the nonce sequence *épeu* [e.pø] preceded by *un* [ɛ̃] yields phonemic content consistent with both *un épeu* and *un népeu* [ɛ̃.ne.pø]. All target words, both real and nonce, were included in two syntactic environments: determiner + noun (e.g. *un air*) and pre-posed adjective + noun (e.g. *un grand ami* ‘a great friend’).

Non-words were included in addition to real lexical items in order to investigate whether top-down information such as lexical frequency, syntactic and semantic plausibility play a role in the perception of duration in this environment. Lexical frequency was in particular thought to pose a problem for the pairs of real-word targets. Due to the limitations of finding real lexical items that meet the above criteria, many of the real-words pairs greatly differed in lexical frequency. For example, according to the *Lexique 3* database (www.lexique.org; New, Pallier, Ferrand & Matos, 2001) *air* showed 588 occurrences per million words, while its C-initial counterpart *nerf* showed 30. Overall, V-initial words had an average of 119.3 occurrences per million and C-initial words had an average of 14.3 occurrences per million. Thus in the creation of the current stimuli, we hypothesized that removing at least some lexical information through the use of non-words is one way to balance the stimuli with respect to lexical frequency and therefore more directly test the perceptual effects of durational differences between ICs and LCs.

A further consideration in the selection and formation of target pairs was the possible effect of lengthening due to stress placement. Stress accent is fixed (i.e., never lexical) in French and consistently falls on the final syllable of a word in isolation or the final syllable of a phrase. Furthermore, stress accent in French is signaled by duration in utterance-final position (e.g., Delattre, 1951, 1966; Jun & Fougeron, 2002). For this reason, there was some concern that the number of syllables in the target word in which the consonant appears could present an additional lengthening factor in the production of stimuli. In order to address lengthening, this

factor was incorporated into the design of the experiment as an additional independent variable. Half of the target words consisted of one-syllable words and the other half consisted of two-syllable words.

Six native speakers of French recorded the target phrases embedded in the carrier sentence, *je vais dire un/les _____ bleu(s)* ‘I am going to say a/the blue _____’, so as to maintain intonation as constant as possible. A post-nominal adjective was used in order to avoid any lengthening effects that might occur if the target word appeared phrase-finally. Each target phrase was recorded three times, giving a total of 72 real-word tokens and 72 non-word tokens (3 consonants x 8 tokens x 3 repetitions) for a total of 144 tokens read by each of the six speakers. This resulted in a total of 864 tokens from which production values were established.

Acoustic analyses

Four tokens (all /t/s; two real words and two non-words) were removed from acoustic analysis because the speaker inserted either a pause or a glottal stop before the consonant under investigation. In total measurements were thus taken from 860 tokens. All acoustic measurements were made from spectrogram and waveform displays in Praat sound-editing software (Boersma & Weenink, 2007). Measurements of the segment /n/ were taken from the offset of the preceding vowel to the onset of the following vowel where there was an obvious reduction in amplitude and waveform complexity. Measurements of /t/ were taken from the beginning of the closure to the beginning of the release³. Measurements of /z/ were taken where

³ Previous production studies on liaison consonants have included measurements for both the closure and voice onset time (VOT) for /t/. VOT has indeed been shown to vary in the same direction as the closure (e.g. longer for ICs than for LCs; Dejean de la Bâtie, 1993; Wauquier-Gravelines, 1996). However, in the current production sample we found no significant variation in VOT. Furthermore, Fougeron (2001) did not find consistent VOT variation in either /t/ or /k/ at four prosodic levels in French (syllable-initial, word-initial, accentual and intonational phrase-initial), but did find consistent durational differences for the closure. We concluded that Fougeron’s data on acoustic variation within a prosodic hierarchy, combined with a lack of differentiation in our own production sample, render it unlikely that VOT represents a systematic and reliable cue to liaison in spoken French.

there was a clear transition in amplitude complexity and periodicity between the preceding and following vowels. Measurements were made at zero crossings wherever possible. Table 1 gives mean durations and standard deviations of LCs and ICs for each of the three segments. The mean duration for all LCs was 84.44 ms and the mean duration for all ICs was 100.66 ms.

[TABLE 1 HERE]

The durations of pivotal consonants were examined across all six speakers in a repeated-measures analysis of variance (ANOVA) treating Word Type (real and nonce), Onset Type (LC and IC), Word Length (one syllable and two syllables) and Consonant (/n/, /t/, and /z/) as within-subject variables. This analysis revealed a no effect of Word Type: $F(1,5) = 1.934$, n.s.; a significant effect of Onset Type: $F(1,5) = 163.78$, $p < .0001$; a significant effect of Word Length: $F(1,5) = 373.68$, $p < .0001$; as well as a significant effect of Consonant: $F(1,5) = 25.17$, $p < .0001$. There was also a significant interaction between Onset Type and Consonant: $F(2,10) = 24.28$, $p = .0001$, suggesting that the durational differences between LCs and ICs were not equivalent among the three segments. In addition, there was a significant interaction between the factors Word Length and Consonant: $F(2,10) = 10.473$, $p = .0035$, suggesting that the durations of each segment varied according to whether they appeared in one-syllable words or two-syllable words. The effect of Word Length did not however interact with Onset Type $F(1,5) = .003$, n.s., indicating that, though the durations of the individual segments varied according to word length, the durational differences between LCs and ICs remained equivalent across one- and two-syllable words. Furthermore, post-hoc pairwise comparisons showed that the difference between mean durations of LCs and ICs was significant for each individual segment: /n/ ($F(1,5) = 12.51$, $p = .0166$), /t/ ($F(1,5) = 20.79$, $p = .0061$), and /z/ ($F(1,5) = 73.69$, $p = .0004$).

Not only do the current production data establish the basis for the stimuli to be used in the behavioral tasks, but they add to a body of research demonstrating systematic durational differences between consonants that surface in liaison environments and their lexical-word-initial counterparts (e.g. Douchez & Lancia, 2008; Spinelli *et al.*, 2003; Wauquier-Gravelines, 1996). These data also add to research showing acoustic variation at boundaries at multiple levels of prosodic organization in spoken French (Fougeron, 2001, 2007; Fougeron *et al.*, 2003; Nguyen *et al.*, 2007; Spinelli, Welby & Scheagis, 2007).

Stimuli

From this production sample, a set of experimental stimuli was created by enhancing the durational differences between LCs and ICs through instrumental manipulation. In order to determine which value the duration of the manipulated consonants should take, the distribution of durations from the production sample was examined. Following methodology laid out in Shatzman and McQueen (2006), the factor by which the segments were manipulated was the standard deviation (SD) in each respective production condition. The use of the SD ensures that the durations of manipulated stimuli – although exaggerated – represent points that fall within a reasonable durational distribution and therefore represent viable instances of allophonic variation in spoken French. In addition, SDs are calculated for each particular consonant in each condition. The SD factor is therefore context-specific in that it takes into account any possible variation due to inherent durational differences owing to consonant class.

A three-step durational continuum of manipulated stimuli was created which included (1) a shortened consonant representing a LC, (2) a baseline consonant representing durations intermediate to those of LCs and ICs, and (3) a lengthened consonant representing an IC. For each of the three segments /n, t, z/, six separate measurements were calculated: The shortened

(liaison) version of each token represented the mean duration for all instances of that consonant in the *liaison* environment *minus* one SD from that particular mean. The value for the midpoint of the continuum (baseline version) represented simply the mean duration across all instances (LCs and ICs) of each consonant. Finally, the value for the lengthened (word-initial) version of the consonant represented the mean duration for that consonant in *word-initial* position *plus* one SD from that particular mean. Again, since significant differences were found between the segmental durations in one-syllable and two-syllable words, a different continuum of durations was calculated for each of these two conditions. The resultant durations used as target values in the manipulated stimuli are presented in Table 2. Values in parentheses represent the percentage difference from the production mean in that particular condition.

[TABLE 2 HERE]

Using the target durations calculated in Table 2, tokens were subsequently edited using Praat speech-editing software. The recordings from one of the six speakers who participated in the production portion of the study were chosen to be manipulated for use in the behavioral tasks. Although recordings were made and measurements were taken from both the V-initial and C-initial member of each lexically ambiguous minimal pair, only the C-initial member of each pair was manipulated for the sake of limiting the number of stimuli to be employed in the perceptual portion of the experiment. In other words, though multiple tokens of both *un air* and *un nerf* were recorded and included in the acoustic analyses reported above, only one token of *un nerf* was chosen to be instrumentally altered⁴.

⁴ It is worth noting that the manipulation of only the C-initial token could affect the results of the perceptual tasks in that there may be other acoustic cues in a C-initial utterance that would not be consistent with a V-initial utterance (vowel duration, F0, intensity of the burst in the /t/ stimuli, etc.), and could therefore bias participants' perception. Indeed, by lengthening the pivotal consonant in the C-initial token we are in fact providing enhanced support for a C-initial interpretation, while by shortening the same consonant in the C-initial token we are providing one

Durations of /t/ were manipulated by either deleting a portion of the closure as needed to shorten the consonant or by inserting a segment of silence into the closure to lengthen the consonant. Durations for /n/ and /z/ however were manipulated by identity splicing. Again, following methodology laid out in Shatzman and McQueen (2006), middle portions of /n/ and /z/ were deleted leaving approximately 20 ms of the initial and final portions of the segment. A portion of a version of the same segment from another version of the same word from the same speaker was then spliced into the recording in order to attain the desired duration. All splices were made at zero crossings in an effort to avoid any acoustic artifacts such as clicks, buzzes or other audible distortions that could occur in the splicing process.

The manipulation of these phrases resulted in 36 real-word sequences (12 phrases x 3 manipulated versions) and 36 non-word sequences that are therefore phonemically identical in their content but differ as to the precise acoustic phonetic realization of the pivotal consonant.

Experiment 1

An AX discrimination task incorporating pairs of instrumentally manipulated stimuli taken from the three-step continuum of duration was used to investigate the saliency of durational differences by establishing thresholds of noticeability between LCs and ICs at the acoustic level. Experiment 1A employs real-word stimuli, while Experiment 1B employs non-word stimuli.

Participants

Thirty-six native speakers of French recruited from the Parisian metropolitan area took

particular acoustic cue that may conflict with other cues already present in the signal. However, if participants do succeed in using this one single cue (segmental duration) to interpret shortened tokens as instances of liaison in the identification task, this gives even more evidence of the robustness of this cue in that it can actually override incongruent acoustic information.

part in Experiment 1 (18 in Experiment 1A and 18 in Experiment 1B). Participants were paid for their participation and reported no hearing or vision impairments.

Materials

Stimuli consisted of pairs of phrases drawn from the three-step continuum of manipulated sequences described above. Each token on the three-step durational continuum was paired with the other two manipulated versions of that same token on the continuum as well as with a duplicate version of the same sound file. This resulted in nine pairings for each of the 12 manipulated phrases, where 1 represents a shortened token, 2 represents a baseline token, and 3 represents a lengthened token. Of the nine stimulus pairings three were identical (1_1, 2_2, 3_3) and six were different (1_2, 1_3, 2_1, 2_3, 3_1, 3_2)⁵. Of the six different pairs, two pairs were separated by two steps on the durational continuum (1_3, 3_1) and four were separated by one step (1_2, 2_1, 2_3, 3_2).

Procedure

Participants were tested individually in a quiet room. The experimental protocol was created using E-Prime experimental software (Schneider, Eschman & Zuccolotto, 2002) and presented on a laptop computer. Stimuli were presented binaurally through headphones. Participants were instructed that they would hear pairs of phrases in French and to indicate whether the two phrases were identical or different by pressing on the keyboard either 1 or 2 respectively. No direction was offered as to what parameters responses should be based on. Participants were instructed to respond quickly, but not so quickly as to sacrifice accuracy. Each

⁵ There was some concern that the uneven distribution of same and different pairs in the stimuli would result in a response bias in the discrimination task, however, given that the responses in the current task are analyzed using d' , response bias is explicitly factored out of the analysis of responses.

experimental trial consisted of one pair of manipulated stimuli separated by an inter-stimulus interval of 250 ms. Individual trials were separated by a 2000 ms pause. Before beginning the experiment, participants completed a training portion consisting of 14 trials. Items included in the training portion were not included in the experimental portion. Each of the 9 pairs of 12 stimuli was presented in random order 6 times, resulting in a total of 648 trials presented in one block. There were no visual stimuli to accompany the auditory stimuli. No feedback as to the accuracy of responses was given in either the training or the experimental portion. Testing lasted approximately 50 minutes.

Results: Experiment 1A

Only responses for the six different pairs (1_2, 1_3, 2_1, 2_3, 3_1, 3_2) are used in the statistical analyses that follow. Mean percentage correct for all same pairs (1_1, 2_2, 3_3) was 86.1%. Analyses of Variance with both subjects (F1) and items (F2) were performed.

D' ('d-prime') scores are given for each of the different pairs in Table 3. D' , a sensitivity index used in signal detection theory, is calculated by transforming participant hit rates (H) and false alarm (F) rates to z-scores and then calculating the difference: $d' = z(H) - z(F)$. In this way, d' reflects not only accuracy (as measured by correct responses and correct rejections), but also factors out any possible response bias (as measured by incorrect responses and false alarms). Since d' is considered to be a more sensitive measure of discrimination than accuracy, only d' was used in the statistical analyses which follow. (See Appendix III for mean percentage correct on different pairs.)

[TABLE 3 HERE]

Mean d' scores for real-word stimuli were submitted to a repeated-measures ANOVA with Pair (1_2, 1_3, 2_1, 2_3, 3_1, 3_2) and Consonant (/n/, /t/, /z/) as within-subject variables.

This analysis revealed a main effect of Pair: $F(5, 85) = 7.292, p < .0001$, $F(5, 55) = 9.655, p < .0001$. Post-hoc pairwise comparisons revealed that pairs separated by two degrees on the durational continuum (1_3 and 3_1) were discriminated better than pairs separated by one degree (1_2, 2_1, 2_3, 3_2). A main effect of Consonant was also observed: $F(2, 34) = 45.727, p < .0001$ (OK), $F(2, 22) = 20.328, p < .0001$. A significant interaction between the two factors was also observed: $F(10, 170) = 3.434, p = .0004$, $F(10, 110) = 2.723, p = .005$, indicating that the effect of continuum Pair was not equivalent among all three segments. Post-hoc pairwise comparisons among the three segments showed that participants discriminated /n/ significantly better than both /t/ ($p = .0002$) and /z/ ($p < .0001$) and /t/ better than /z/ ($p = .0122$).

Results: Experiment 1B

Two participants were removed from analysis in Experiment 1B due to the fact that they responded *same* to all 648 trials of the experiment. The following analyses therefore include 16 participants. Mean d' scores for non-word stimuli (see Table 3) were also submitted to a repeated-measures ANOVA with Pair (1_2, 1_3, 2_1, 2_3, 3_1, 3_2) and Consonant (/n/, /t/, /z/) as within-subject variables. This analysis revealed a main effect of Pair: $F(5, 75) = 7.459, p < .0001$, $F(5, 55) = 19.260, p < .0001$. As with the real-word data, post-hoc pairwise comparisons revealed that non-word pairs separated by two degrees on the durational continuum (1_3 and 3_1) were discriminated better than non-word pairs separated by one degree (1_2, 2_1, 2_3, 3_2). A main effect of Consonant was observed in the by-item analysis only: $F(2, 30) = .944$, n.s., $F(2, 22) = 3.999, p = .033$. There was no significant interaction between the two factors: $F(10, 150) = 0.830$, n.s., $F(10, 110) = 1.657$, n.s., suggesting that the effect of Pair was equivalent among the three segments. As was observed in Experiment 1A, post-hoc pairwise comparisons in Experiment 1B showed that pairs separated by two degrees on the durational

continuum were discriminated better than most pairs separated by one degree. Contrary to responses observed for real-word stimuli, however, all three consonants were discriminated equally well in non-word stimuli. Pairwise comparisons further showed that the effect of Consonant observed in the by-item analysis is due to the fact that /n/ was discriminated significantly better than /z/ ($p = .0067$), however, this was the only significant difference among the segments.

D' scores for Experiment 1A were then compared to scores from Experiment 1B in order to investigate a possible effect of lexicality on sensitivity to durational difference in this context. . Experiment (1A and 1B) was treated as a between-subjects factor in the F1 analysis and as a between-items factor in the F2 analysis. This analysis revealed no significant effect of Experiment: $F1(1,33) = 1.221$, n.s., $F2(1,22) = 2.439$, n.s., and no significant interactions with any other factors. This result suggests that sensitivity to segmental duration in this phonological environment is not enhanced when lexical information is removed.

Discussion: Experiments 1A and 1B

The objective of the AX discrimination tasks in Experiments 1A and 1B was to investigate the off-line perceptual saliency of durational variation in consonants at word boundaries in possible environments of liaison. Stimuli were employed in which the duration of pivotal consonants was instrumentally manipulated according to a three-step durational continuum.

The results of Experiments 1A and 1B showed that listeners are sensitive to durational differences at the acoustic level between ICs and LCs, but only when these differences are greatly exaggerated with respect to what would be a normal distribution of allophonic variation. Pairs separated by two degrees on the durational continuum showed significantly higher

discrimination than pairs separated by one-degree, which showed in effect a lack of consistent discrimination, as indicated by d' scores inferior to 1. In addition, there was a great deal of variation across participants, as evidenced by substantial standard deviations.

It is worth noting that the physical differences in duration exhibited in the current stimulus sample are substantial in relation to what are considered to be thresholds of noticeability (see for example Huggins, 1972), however sensitivity seems to be mitigated by other factors. One factor which could possibly render the discrimination of these durational differences more difficult than established thresholds of noticeability would suggest is the fact that the durational differences in the current stimulus sample represent within-category variation, which has been shown to be less perceptually salient than variation which crosses categorical boundaries (Peperkamp, Pettinato, & Dupoux, 2003). This fact may have rendered discrimination more challenging. One further factor is the position of the segment in the utterance. Klatt and Cooper (1975) showed that just-noticeable differences (JNDs), i.e. the amount by which a stimulus must be altered in order to produce a noticeable variation in the sensory experience of the perceiver, depend on a segment's position within a sentence; durational changes in initial segments, for example, tend to be more perceptually salient and therefore exhibit smaller JNDs than utterance-medial and utterance-final segments. The manipulated segments in the current stimuli appear in utterance-medial position and therefore may be more difficult to perceive.

Low discrimination scores, coupled with the extent of individual variation observed among participants, suggest that acoustic differences in this phonological environment may not represent an extremely robust processing cue in natural speech. This conclusion is supported by results from the cross-modal priming task used in Spinelli *et al.* (2003). These authors concluded that, though subphonemic detail may be robust enough to *activate* V-initial candidates and to

bias the perceptual system toward the word(s) intended by the speaker, it may not be robust enough to allow the system to *de-activate* unintended words (p. 250). When there is not sufficient acoustic information in the signal, activated candidates not intended by the speaker may ultimately be ruled out by higher-level information such as context. This is further supported by the results of Shoemaker and Birdsong (2008) who found that listeners failed to discriminate the same ambiguous phrases used in Spinelli *et al.* (2003) in a forced-choice identification task. Spinelli *et al.* employed an online priming task to demonstrate the activation of V-initial candidates in environments of liaison, Shoemaker and Birdsong failed to show that listeners have enough information in the signal to discard C-initial competitors.

However, we must also entertain the possibility that the relatively low discrimination scores do not in fact reflect a purely acoustic phenomenon, but rather are linked to higher level processing. It is feasible that in the processing of one-degree pairs, listeners are classifying what they hear according to ‘type’ and not to acoustic information. Put differently, stimuli which differ by one degree necessarily represent a comparison between an ambiguous baseline form (2 on the continuum) and a form that is greatly enhanced and therefore a robust exemplar of either a LC or IC (1 or 3 on the continuum, respectively). Participants may in fact hear the difference, but ultimately conclude that both stimuli represent either tokens of a liaison environment or tokens of C-initial words and therefore respond ‘same’. Conversely, the stimuli separated by two degrees on the continuum represent robust exemplars of each type of consonant and therefore were completely unambiguous as to their categorization, or ‘type’ and elicited more *different* responses. However, if participants were classifying stimuli solely on the basis of ‘type’ and not on the basis of acoustic information, we would expect to see no significant difference in response patterns between the one-degree pairs and the pairs of stimuli which were actually the same (1_1,

2_2, 3_3), which was not the case. Participants responded ‘same’ to the one-degree stimulus pairs 68.6% of the time, while they responded ‘same’ to the identical pairs 86.1% of the time, a difference which was statistically significant. This difference suggests that they were using acoustic information, and not categorization according type, to qualify the manipulated stimuli. (We thank an anonymous reviewer for pointing out this possibility.)

The perceptual differences among the three segments also warrant discussion. Given that perceptual saliency in this task follows the same pattern as the robustness of durational variation between LCs and ICs observed in the production sample and therefore in the manipulated stimuli (i.e. /n/ > /t/ > /z/; see Tables 1 and 2), performance differences in the discrimination task likely represent an artifact of the particular production sample collected for this study. While we maintain that the current production sample is generalizable to a larger population in terms of the presence of durational differences between LCs and ICs, further research would be needed to establish that the differences in robustness observed among the three consonants are indicative of a general phenomenon.

We now discuss briefly an unpredicted result that emerged from the data in both Experiments 1A and 1B, namely an effect of stimulus order. We observed that pairs in which the order of presentation was a shorter token followed by a longer token (e.g. 1_2, 1_3, 2_3) were discriminated better than pairs in which a longer token was followed by a shorter token (e.g. 2_1, 3_1, 3_2). This result is difficult to account for if we consider the fact that the same two manipulated stimuli in both presentations (e.g. 1_2 and 2_1) are differentiated by the same absolute duration along the three-step durational continuum. However, two possible explanations present themselves. A first potential explanation for why sequences such as 1_3 were discriminated better than sequences such as 3_1 could be due to the fact that the former imposed

lower demands on working memory than the latter. The duration of a shortened stimulus (1) would be shorter than the total duration of a lengthened stimulus (3). If we assume that participants held the first stimulus of the pair in working memory while they heard the second stimulus and subsequently made a decision, it could feasibly have been more difficult to hold the longer stimulus in working memory than the shorter stimulus, thus resulting in lower discrimination rates for pairs with decreasing length than for pairs with increasing length.

A second potential explanation involves an analysis of discrimination performance as a function of the proportional change between the two stimuli in each pairing. Perceptible variation in segmental duration can be characterized in relational terms in addition to absolute values. *Weber's law* is a general law of psychophysics which states that JNDs can be determined in terms of a proportion of the original stimulus value. This law predicts, for example, that a physical change in duration would be more perceptible in a shorter segment than in a longer segment due to the fact that the absolute value of the change represents a larger proportion of the former than the latter.

Taking a simplified example based on the current stimulus set, let us say that a shortened consonant (1 on the continuum) is 50 ms, while a lengthened consonant (3 on the continuum) is 100 ms. Therefore, the percentage change in the pairing 1_3, i.e. the proportion of the absolute durational difference between 1 and 3 relative to the absolute duration of 1, is an increase of 100% (50 ms + 50 ms); however in the pairing 3_1, the percentage change, i.e. the absolute durational difference between 3 and 1 relative to the absolute duration of 3, is a decrease of 50% (100 ms – 50 ms). Therefore, if duration is assessed relativistically, it is plausible that the change in 1_3 would be easier to perceive than the change in 3_1.

In order to test an effect of stimulus order based on the relative change between the two tokens in each experimental trial, we performed post-hoc simple regression analyses comparing the percentage change between stimuli with both d' scores and percentage correct for each different pair, all of which revealed either significant main effects or effects that just missed significance (Table 4). This indicates that across all six different pairs, the percentage of durational difference from the first member of the pair to the second member of the pair predicted both mean accuracy rates and d' scores, suggesting that the perception of duration in this phonological environment can be framed in terms of differential thresholds as well as absolute values.

[TABLE 4 HERE]

In the following section we examine whether durational differences between LCs and ICs are exploited in higher level processing, specifically lexical access and the segmentation of continuous speech. While the AX discrimination task is employed to probe thresholds of acoustic saliency, an identification task is employed to investigate the use of segmental duration in higher-level lexical access processes.

Experiment 2

Experiment 2 consists of a forced-choice identification task also employing manipulated tokens taken from the three-step durational continuum of stimuli described above. Experiment 2A employs real-word stimuli, and Experiment 2B employs non-word stimuli. Specifically, this task investigates whether the lexical interpretation of ambiguous sequences such as [ẽ.nɛʁ] is modulated by the duration of the pivotal consonant when all other acoustic information in the signal is held constant. If segmental duration is indeed robust enough to guide listeners in disambiguation then stimuli containing a shortened segment will elicit a significantly higher

proportion of V-initial (liaison) responses, while stimuli containing a lengthened segment will elicit a higher proportion of C-initial responses. Furthermore, the baseline stimuli, which represent segmental durations intermediate to LCs and ICs will not offer participants enough acoustic information to guide responses and will therefore elicit a guessing strategy resulting in performance levels at chance.

Participants

Participants in Experiment 2A were the same as in Experiment 1A. Participants in Experiment 2B were the same as in Experiment 1B.

Procedure

Experiment 2 immediately followed Experiment 1, however participants were given the opportunity to take a short break between the two experiments. Each experimental trial had the following structure. Participants heard one of the three manipulated phrases from the durational continuum presented binaurally through headphones. Stimuli were presented without a carrier frame, thus eliminating any potential priming effects from context. At the offset of the auditory stimulus, two words appeared on the computer screen. The two visual targets represented the V-initial and C-initial interpretations of the final word in each ambiguous sequence. For example, when the auditory stimulus is a manipulated version of the sequence [ẽ.nɛʁ], (1) *air* and (2) *nerf* are visual targets. Participants were instructed to indicate which of the two words presented on the screen was present in the phrase they had heard by pressing on the computer keyboard either (1), corresponding to the word on the left of the screen, or (2), corresponding to the word on the right of the screen. Visual targets were counter-balanced across participants in order to offset any possible bias toward the left-hand visual target that might occur from reading effects. Half of the participants were presented with the V-initial (liaison) target on the left of the screen and the

other half were presented with the C-initial target on the left of the screen. There was no delay between the offset of the auditory stimulus and the presentation of the visual targets. Each of the 36 stimuli (i.e., three manipulated versions of each of 12 tokens) was presented randomly six times resulting in a total of 216 trials presented in one block. Participants completed a training portion consisting of 14 trials before beginning the experiment. Items included in the training portion were not included in the experimental portion. Individual trials were separated by 2000 ms. Experiment 2 lasted approximately 20 minutes.

Results: Experiments 2A

Table 5 shows the mean proportion of V-initial (i.e., ‘liaison’) responses calculated for manipulated stimuli in each of the three continuum conditions: the shortened (LC) version, the baseline version, and lengthened (IC) version. Mean response proportions for real-word stimuli were submitted to a repeated-measures ANOVA with Condition (LC, Baseline, and IC) and Consonant (/n/, /t/, /z/) as within-subject variables. This analysis revealed a main effect of Condition: $F(2, 34) = 25.01, p < .0001, F(2, 22) = 10.713, p = .0006$, as well as a main effect of Consonant: $F(2, 34) = 26.89, p < .0001, F(2, 22) = 9.62, p = .001$. A significant interaction between the two factors was also observed: $F(4, 68) = 4.35, p = .0034, F(4, 44) = 3.332, p = .0343$, suggesting that the effect of continuum Condition was not equivalent among all three segments.

Results: Experiment 2B

Two participants were removed from analysis due to the fact that they chose either the V-initial or the C-initial response for all 216 trials of the experiment. The following analyses therefore include data from 16 participants.

Mean responses for non-words were also submitted to a repeated-measures ANOVA with

Condition (LC, Baseline, and IC) and Consonant (/n/, /t/, /z/) as within-subject variables. This analysis revealed a main effect of Condition: $F1(2, 32) = 4.19, p = .0243, F2(2, 22) = 21.36, p < .0001$, and no effect of Consonant: $F1(2, 32) = 1.61, n.s., F2(2, 22) = 2.10, n.s.$ There was no interaction between the two factors: $F1(4, 64) = 1.88, n.s., F2(4, 44) = 1.99, n.s.$, indicating that, unlike the real-word stimuli, the effect of Condition did not vary among the three segments in the non-word stimuli.

[TABLE 5 HERE]

A repeated measures ANOVA was also conducted on response proportions in Experiment 2A and Experiment 2B in order to compare the processing of real-word stimuli with that of non-word stimuli. Experiment (2A and 2B) was treated as a between-subjects factor in the F1 analysis and as a between-items factor in the F2 analysis. No significant effect of Experiment was revealed: $F1(1,33) = 3.341, n.s., F2(1,22) = 2.35, n.s.$ A significant interaction was however observed between the factors Experiment and continuum Condition in the by-subject analysis only: $F1(2,66) = 4.074, p = .0215, F2(2,44) = 2.82., n.s.$, suggesting that the effect of continuum condition was not equivalent between the two experiments. This result is consistent with discrimination results from discrimination tasks in Experiments 1A and 1B. The removal of lexical information did not have a significant effect on the lexical assignment of pivotal consonants, indicating that listeners generalize the perception of durational differences between ICs and LCs to never before seen items.

Discussion: Experiments 2A and 2B

The results of Experiment 2A and 2B suggest that the duration of the pivotal consonant alone can indeed modulate the lexical interpretation of sequences rendered ambiguous by liaison and resyllabification in spoken French. Given that the sequences used in the identification task

were phonemically and acoustically identical, differing only in the duration of the pivotal consonants under investigation, the data suggest that participants used this single acoustic cue in the localization of word boundaries. Shortened consonants elicited significantly more V-initial (liaison) responses relative to both baseline and lengthened consonants, while lengthened consonants elicited significantly fewer V-initial responses (and therefore more C-initial responses) relative to both baseline and shortened consonants. In addition, baseline consonants elicited roughly the same proportion of V-initial and C-initial responses, indicating a guessing strategy on the part of participants presumably due to a lack of sufficient acoustic information in the signal. As was observed in the AX discrimination task, there was also a great deal of individual variation as evidenced by large standard deviations. This again brings into question the consistency with which this single acoustic cue is exploited in natural (unmanipulated) speech. However, the fact that participants responded in the predicted direction in the identification task demonstrates that segmental duration does have cue value in the processing of liaison environments in spoken French. Though generalizations to the processing of natural speech given the current results must be made with caution, these results offer strong evidence that durational differences between LCs and ICs are exploited in the decoding of continuous speech.

The results of Experiment 2B, which employed non-word stimuli, are in line with those of Experiment 2A, which employed real-word stimuli. Listeners are more likely to interpret a shortened consonant as a resyllabified LC and a lengthened consonant as a (lexical-) word-initial segment, even when lexical information is removed. One further issue that merits discussion is an apparent bias toward C-initial responses in the identification tasks. Post-hoc analyses revealed that this response bias was significant only for the segment /t/ in both real words ($p = .0047$) and

non-words ($p = .0243$). As noted above, the production tokens that were manipulated were all C-initial (e.g. *un nerf*), and therefore it is possible that there are other acoustic cues present in the utterance that would be consistent with the C-initial token (e.g. F0, the duration of surrounding vowels, etc.). Therefore, the direction of the bias is not surprising. However, we can only speculate as to why this bias would be significant for /t/ and not for /n/ and /z/. One possibility is the presence of some sort of acoustic information that is more salient in /t/ than in /n/ and /z/, the obvious candidate being voice onset time (VOT), the time between consonant release and the onset of voicing following voiceless plosives. Given that the manipulation of /t/ stimuli included only the shortening and the lengthening of the closure, it is feasible that VOT in these stimuli signaled to the listener that the segment was produced as lexical word-initial. Some previous research has shown that VOT can be longer in ICs than in LCs (Dejean de la Bâtie, 1993; Wauquier-Gravelines, 1996). Therefore, we would predict that if VOT had an effect on the response pattern, stimuli with longer VOTs would elicit a higher proportion of C-initial responses. However, post-hoc correlations between VOT values in the current stimulus sample and the proportion of t-initial responses were not significant ($r = .193$, n.s.).

One non-phonetic based explanation for the C-initial bias also presents itself. Of the three consonants used in this study, /t/ is less frequent in environments of liaison relative to both /n/ and /z/ according to recent corpus data (Durand & Lyche, 2008). Following a usage-based model of liaison (e.g. Bybee, 2001), participants could be sensitive to this distribution and therefore prefer t-initial words over their V-initial counterparts in the perceptual tasks. Furthermore, Durand and Lyche found that liaison with /t/ was not realized in several instances where the syntactic environment would predict its obligatory realization (proposed adjective + noun; e.g. *grand honneur* ‘great honor’), while /n/ and /z/ were realized 100% of the time in their respective

obligatory condition (determiner + noun), suggesting that the realization of /t/ in liaison in contemporary French may be more variable relative to the other two segments in this study⁶, possibly also biasing listeners toward a C-initial interpretation. In fact, based on this corpus, Durand and Lyche propose that the syntactic environment preposed adjective + noun can no longer be considered to trigger ‘obligatory’ liaison. However, following this same logic, we would expect to see a V-initial response bias for stimuli containing /z/, given that /z/ is by far the most common segment that surfaces in liaison, and that there are very few z-initial words in French. Unfortunately, reasons as to why participants preferred /t/ in lexical word-initial position over liaison position are merely speculative at this point. There may be some other as yet unidentified acoustic cue in the /t/ stimuli that led to the C-initial response bias, or there may be top-down information available to listeners that we are unable to pinpoint. Further research would be required to pinpoint the phenomena underlying this effect.

General Discussion

The current series of experiments examined the role of segmental duration in resolving lexical ambiguities that can arise in environments of liaison in spoken French. The phonological processes of liaison and resyllabification often render word and syllable boundaries ambiguous, in that vowel-initial words (e.g. *air*) are in effect masked by consonants that surface in liaison and are resyllabified across the word boundary to become the onset of the following syllable. Misalignment between word and syllable boundaries could be presumed to hinder the spoken word recognition system in that multiple lexical candidates may be consistent with the input (e.g. [ẽ.nɛʁ] is consistent with both *un air* and *un nerf*). In line with a body of work showing that

⁶ The reader will recall as well that in the current production sample, the four tokens that were removed from acoustic analysis were all tokens of /t/. These tokens were removed because speakers either inserted a pause or a glottal stop at the word boundary instead of realizing /t/ in liaison.

speakers tend to strengthen the left edge of prosodic boundaries (Klatt, 1976; Cho *et al.*, 2007; Cho & Keating, 2001), speakers of French vary the duration of LCs relative to ICs—LCs are systematically shorter than the same segment in lexical word-initial position. We examined here whether listeners are sensitive to this cue in interpreting continuous speech.

Experiments 1A and 1B examined the perceptual saliency of durational variation between ICs and LCs at the acoustic level through the use of an AX discrimination task which employed real- and non-word stimuli in which the durations of pivotal consonants had been instrumentally manipulated according to a three-step continuum of segmental duration. Though pairs of these manipulated stimuli represent differences in absolute duration well above what would normally be considered just-noticeable differences, perceptual sensitivity to differences was only robust when segments were separated by two-degrees on the continuum, i.e. by durations two standard deviations above or below mean durational values from a production sample of six native speakers. This suggests that allophonic differentiation in this context may only be perceptually salient when these differences represent extreme values within a normal distribution of speech.

The results of Experiments 1A and 1B revealed that sensitivity to these durational differences also depends on the order in which the stimuli are presented. Stimulus pairs in which a shorter token is followed by a longer token (e.g. 1_3) were discriminated better than the reverse order (e.g. 3_1), suggesting that the perceptual system is sensitive to the proportion of change from the first stimulus to the second stimulus. This finding coupled with research showing that the perception of duration also varies as a function of the duration of surrounding segments (Diehl *et al.*, 1980; Miller, 1987; Summerfield, 1981), speech rate (Summerfield, 1975) as well as the length of the utterance in which the segment appears (Kawai & Carrell, 2005), points to a relativistic assessment of durational cues and implies that these segments are qualified in relation

to both previous and following acoustic information. We again reiterate that working memory could also have played a role in the observed stimulus-order effects, though further research would be needed to substantiate this hypothesis.

Experiments 2A and 2B explored the exploitation of segmental duration in the lexical interpretation of sequences rendered ambiguous by the prosodic position of pivotal consonants. Manipulated stimuli from the three-step durational continuum were employed in a forced-choice identification task. The results of the identification tasks demonstrate that utterance interpretation in spoken French can be influenced by segment duration in that responses were guided by the duration of the pivotal consonant alone. Participants interpreted a shortened consonant significantly more often as an instance of liaison and a lengthened consonant significantly more often as lexical word-initial. Furthermore, responses were mostly at chance for baseline stimuli where pivotal consonants represent durations intermediate to those of LCs and ICs.

The current results make three important empirical contributions. First, data from our production sample add to a body of research demonstrating acoustic variation at boundaries at multiple levels of prosodic organization in spoken French (Douchez & Lancia, 2008; Fougeron, 2001, 2007; Fougeron *et al.*, 2003; Spinelli, Welby & Scheagis, 2007; Welby, 2003). The phonological processes of liaison and resyllabification are acoustically marked by speakers.

Second, the results provide the first direct evidence that durational cues are exploited in the comprehension of spoken French, extending previous research on the perception and processing of liaison. Priming data from unmanipulated speech (Spinelli *et al.*, 2003; Gaskell *et al.*, 2002) showed that ambiguity in liaison environments does not impede lexical access to V-initial candidates, even though the misalignment of syllable and word boundaries caused by liaison and resyllabification would be predicted to complicate this process. These authors

suggested that listeners differentiate liaison environments from non-liaison environments based in part on the duration of pivotal consonants, however in neither study did they test this hypothesis explicitly. The current data directly support the claim that segmental duration affects the lexical interpretation of liaison environments by employing manipulated input in which this single acoustic cue is altered while holding all other information in the signal constant. While we have shown that listeners exploit this single acoustic cue in the interpretation of ambiguous sequences, we must point out that the current tasks reflect offline behavior, and therefore any generalizations made about the online processing of speech must be made cautiously. Our results demonstrate that listeners *can* use segmental duration in interpreting the speech signal, but they do not explicitly demonstrate that they *do* use it the processing of natural speech. However the results of Spinelli *et al.* (2003), who used an online task (cross-modal priming), suggest that listeners do make use of subphonemic information in accessing lexical representations in cases of ambiguity.

It is further worth noting that segmental duration is only one of several acoustic factors that may vary in this phonological environment. Other cues can interact and co-vary with the duration of the pivotal consonant; the duration of the pre-boundary vowel has been shown to be shorter before LCs than before ICs (Spinelli *et al.*, 2003), and some research has suggested that the duration of the closure and of the following burst are both shorter for liaison /t/ compared with word-initial /t/ (Dejean de la Bâtie, 1993). Therefore, though durational cues may be systematically present in the signal, listeners do not need to rely solely on durational information when recognizing words in natural speech.

Third, the results support more broadly a view in which subphonemic, non-categorical detail directly modulates segmentation and recognition processes. Speakers give listeners

phonetic cues to the location of prosodic boundaries and surface acoustic detail is integrated into the input to the word recognition system. The exploitation of segmental duration at word boundaries in French, as well as in Dutch (Quené, 1992; Salverda *et al.*, 2003; Shatzman & McQueen 2006), English (Davis, Marslen-Wilson, & Gaskell, 2002) and Italian (Tabossi, *et al.*, 2000), can be added to a body of research demonstrating that lexical access is modulated by other phonetic and phonological cues in the signal such as non-contrastive aspiration and pre-vocalic glottal stops (Nakatani & Dukes, 1977), voice onset time (Andruski, Blumstein & Buton, 1994), F0 contours (Welby, 2003), distribution of lexical stress (Cutler & Butterfield, 1992) and phonotactics (McQueen, 1998).

The current data directly demonstrate the exploitation of fine-grained acoustic detail in spoken French, however they do not speak to the fundamental question of *how* this information is integrated into the processing of speech. Researchers are in agreement that acoustic information in the speech signal serves to activate representations stored in a mental lexicon and that word recognition processes entail competition among these activated candidates. There is further agreement that multiple sources of information play concurrent roles in both the initial activation and ultimate selection of words (and by extension parses of sequences of words). Cues come from the sources mentioned above, as well as from higher-level information such as lexical frequency (Grosjean, 1980), context (Tyler & Wessels, 1983), lexicality (Mattys, White & Melhorn, 2005) and neighborhood density (Luce & Pisoni, 1998). Extending this line of inquiry, more recent research has begun to explore the weighting of multiple segmentation strategies in an attempt to formulate a hierarchy of these cues based upon the saliency of each individual cue in speech processing (Mattys 2003, 2004, Mattys *et al.*, 2005).

One way to model the use of segmental duration is to propose that acoustic detail

operates at a pre-lexical level, intermediary to the speech signal and abstract lexical representations. Connectionist models such as TRACE (McClelland & Elman, 1986) for example propose that acoustic input activates distinctive features of phonemes, which then excite phonemes, which in turn excite all words that contain these phonemes. A strictly phoneme-based interpretation of this model cannot immediately account for the incorporation of non-contrastive detail in the activation of lexical candidates in that, for example, a shorter liaison /n/ and a longer word-initial /n/ are both still categorized as /n/. However, the pre-lexical processing level could, in principle, also be sensitive to acoustic detail that is not contrastive. A mechanism that is sensitive to position-specific realizations of phonemes could be posited at this level which would bias the speech recognition system toward either syllable-initial or lexical word-initial phonemes. In the case of [ẽ.nɛʁ], variation in duration would either increase the activation of *air* for a shorter /n/ or increase the activation of *nerf* for a longer /n/.

One alternative to phoneme-based recognition is to posit that the spoken word recognition system stores and makes use of multiple episodic traces of lexical representations. Exemplar-based models of spoken word recognition (Goldinger, 1996; Johnson, 1997; Pisoni, 1997) propose that lexical units are in effect stored as detailed acoustic traces in long-term memory. The lexicon is therefore made up of stores of multiple exemplars of words with varying acoustic detail; allophonic variation and fine-phonetic detail are stored at the lexical level with the mental representation of each word. Acoustic input is then matched against these exemplars with no intervening pre-lexical processing, allowing phonetic detail to be directly mapped onto lexical units. Exemplar models are also supported by research showing that not only cues to speaker identity, but also speech rate, are retained in long-term memory and have an effect on

word-recognition tasks (Bradlow, Nygaard & Pisoni, 1999; Pisoni, 1997; see Lachs, McMichael & Pisoni, 2003 for a review).

Specifically in the case of liaison, an exemplar approach could account for the current data, if we consider the real-word data only. Longer consonants would be stored as examples of consonant-initial words while shorter consonants would be stored as (resyllabified) coda consonants. Thus a relatively longer /n/, for example, would better match stocked representations of *nerf* than representations of *un* in the sequence [ẽ.nɛʁ]. However, data in this study showing that participants used segmental duration in the localization of word boundaries in never before seen non-word sequences as well are not in line with a strictly exemplar-based model.

One further possibility is a model in which prosodic structure is evaluated concurrently with segmental information, thus involving parallel processing at multiple levels of linguistic organization. Consequently, instead of a word recognition system based primarily on low-level information, segmental and suprasegmental analyses are in effect undertaken simultaneously in the processing of the speech signal (Cho *et al.*, 2007; Norris *et al.*, 1997; Shatzman & McQueen, 2006). Specifically, acoustic cues to word and syllable boundaries serve to favor lexical candidates whose boundaries align with the prosodic structure predicted by these cues.

On this view, the goodness of fit of a particular realization of a phoneme is based not only on the distribution of acoustic properties that traditionally qualify the segment, but also on the perceived goodness of the segment relative to a particular prosodic position. This type of model therefore functions in a probabilistic fashion, in that acoustic information reflects the likelihood that the duration of a particular segment corresponds to a predicted prosodic boundary based on previously calculated distributions of acoustic factors. In the case of liaison, for example, a longer consonant would suggest a preceding word boundary, while a shorter

consonant would suggest a preceding syllable boundary, and by extension resyllabification of a (latent) coda consonant.

Support for a this model of word recognition comes from online behavioral tasks probing the continuous uptake of acoustic detail in speech processing. Returning to the Shatzman & McQueen (2006) study mentioned above, comparisons of eye movements between the two experiments in this study showed that fixations started significantly later when participants were listening to unaltered stimuli in Experiment 1 than when listening to the exaggerated versions in Experiment 2. This finding suggests that when there is less acoustic information available, listeners wait for more information to accumulate before fixating the lexical target. The exaggerated durations of the pivotal segment /s/ in Experiment 2 offered enough acoustic information to allow the listener to choose the target before letting further input accumulate. The difference in time course between the two experiments suggests that segmental duration is indeed used in a probabilistic fashion, i.e. that it is evaluated in relation to both acoustic information in the signal itself as well as in relation to a listener's prior knowledge of the distributions of acoustic information in that language.

In sum, while we are unable to distinguish among the above accounts of the exploitation of subphonemic detail in spoken word recognition outlined above, the current pattern of data supports the hypothesis that listeners are sensitive to segmental duration in French and use this information to modulate the lexical interpretation of ambiguous liaison sequences. Speakers give listeners subtle acoustic cues as to the prosodic position of pivotal consonants, which listeners can exploit to locate word and syllable boundaries. As Shatzman and McQueen (2006) point out, "finding a difference in the acoustic properties of speech stimuli is not sufficient to conclude that participants use that particular difference in lexical disambiguation" (p. 14). Although the

particular cue investigated here may not be the most robust cue in naturally produced speech, we have demonstrated that listeners can use this cue when it is salient in the signal.

APPENDIX I: REAL-WORD STIMULI

Consonant	Vowel-initial (liaison) target	Consonant-initial target	Pronunciation
/n/	<i>un hectare</i> ‘a hectare’	<i>un nectar</i> ‘a nectar’	[ẽ.nɛk.taʁ]
	<i>un aval</i> ‘a support’	<i>un naval</i> ‘a naval officer’	[ẽ.na.val]
	<i>un air</i> ‘an appearance’	<i>un nerf</i> ‘a nerve’	[ẽ.nɛʁ]
	<i>un œuf</i> ‘an egg’	<i>un neuf</i> ‘a nine’	[ẽ.nœf]
	<i>un grand assaut</i> ‘a big assault’	<i>un grand tasseau</i> ‘a big bracket’	[ẽ.gʁɑ̃.ta.so]
	<i>un grand ami</i> ‘a great friend’	<i>un grand tamis</i> ‘a big sieve’	[ẽ.gʁɑ̃.ta.mi]
/t/	<i>le grand Est</i> ‘the big East’	<i>le grand test</i> ‘the big test’	[lə.gʁɑ̃.tɛst]
	<i>un grand acte</i> ‘a great act’	<i>un grand tact</i> ‘a great (sense of) tact’	[ẽ.gʁɑ̃.takt]
	<i>les aunages</i> ‘the measurements by aune’ ⁷	<i>les zonages</i> ‘the zonings’	[le.zo.naʒ]
	<i>les ailés</i> ‘the winged ones’	<i>les zélés</i> ‘the zealous ones’	[le.ze.le]
/z/	<i>les ailes</i> ‘the wings’	<i>les zèles</i> ‘the zeals’	[le.zɛl]
	<i>les aines</i> ‘the groins’	<i>les Zens</i> ‘the Zen ones’	[le.zɛn]

⁷ An *aune* is an antiquated measure of length roughly equivalent to 119 cm.

APPENDIX II: NON-WORD STIMULI

Consonant	Vowel-initial (liaison) target	Consonant-initial target	Pronunciation
/n/	<i>un auvis</i>	<i>un nauvis</i>	[ẽ.no.vi]
	<i>un épeu</i>	<i>un népeu</i>	[ẽ.ne.pø]
	<i>un upe</i>	<i>un nupe</i>	[ẽ.nyp]
	<i>un ade</i>	<i>un nade</i>	[ẽ.nad]
/t/	<i>un grand auvis</i>	<i>un grand tauvis</i>	[ẽ.gʁã.to.vi]
	<i>un grand épeu</i>	<i>un grand tépeu</i>	[ẽ.gʁã.te.pø]
	<i>un grand upe</i>	<i>un grand tupe</i>	[ẽ.gʁã.typ]
	<i>un grand ade</i>	<i>un grand tade</i>	[ẽ.gʁã.tad]
/z/	<i>les auvis</i>	<i>les zauvis</i>	[le.zo.vi]
	<i>les épeus</i>	<i>les zépeus</i>	[le.ze.pø]
	<i>les upes</i>	<i>les zupes</i>	[le.zyp]
	<i>les ades</i>	<i>les zades</i>	[le.zad]

APPENDIX III: PERCENTAGE CORRECT ON AX DISCRIMINATION TASKS
(Experiments 1A and 1B)

		Pair					
		1 2	1 3	2 1	2 3	3 1	3 2
Experiment 1A (Real words)							
/n/	<i>M</i>	36.9	89.2	52.7	52.8	81.4	44.6
	<i>SD</i>	25.8	15.1	25.9	25.5	16.6	23.9
/t/	<i>M</i>	42.2	86.3	42.5	56.0	74.0	34.0
	<i>SD</i>	21.3	14.1	17.3	21.9	18.9	23.1
/z/	<i>M</i>	20.1	60.2	14.6	25.4	36.2	12.0
	<i>SD</i>	14.7	23.8	8.5	13.7	18.2	8.7
		<i>M</i>	33.0	78.5	36.4	44.8	63.9
		<i>SD</i>	16.2	14.4	13.4	15.8	15.1
Experiment 1B (Non-words)							
/n/	<i>M</i>	19.9	64.7	29.0	46.9	60.3	39.1
	<i>SD</i>	26.4	29.8	24.5	29.9	36.2	32.9
/t/	<i>M</i>	22.5	65.6	25.3	41.7	55.0	15.1
	<i>SD</i>	25.5	34.0	25.6	28.4	33.3	18.9
/z/	<i>M</i>	11.5	50.3	17.4	33.8	31.8	21.9
	<i>SD</i>	23.7	32.3	22.1	30.7	27.6	25.0
		<i>M</i>	18.1	60.1	23.9	40.8	49.1
		<i>SD</i>	23.3	29.4	21.8	26.8	28.9

Acknowledgements

The current project was funded by Department of French and Italian at the University of Texas at Austin, as well as by the University of Texas at Austin Graduate School. I gratefully acknowledge David Birdsong, Carl Blyth, Randy Diehl, Jean-Pierre Montreuil, and Marta Ortega-Llebaria for their extensive commentary and discussion throughout the course of the project. I am further indebted to Elsa Spinelli at the Laboratoire de Psychologie et Neurocognition at the Université Pierre Mendes, Grenoble, France for her help and generosity during the initial stages of the project. I also wish to thank three reviewers for their insightful comments and invaluable suggestions.

References

- Allopenna, D., Magnuson, S. & Tanenhaus, M. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419-439.
- Andruski, J., Blumstein, S. & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52(3), 163-187.
- Boersma, P. & Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.6.36) [Computer program].
- Bradlow, A., Nygaard, L. & Pisoni, D. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, 61 (2), 206–219.
- Bybee, J. (2001). Phonology and language use. Cambridge: Cambridge University Press.
- Cho, T. & Keating, P. (2001). Articulatory and acoustic studies on domain- initial strengthening in Korean. *Journal of Phonetics*, 29 (2), 155-190.
- Cho, T., McQueen, J. & Cox, E. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35 (2), 210-243.
- Content, A., Kearns, R. & Frauenfelder, U. (2001). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language*, 45, 177-199.
- Content, A., Meunier, C. Kearns, R. & Frauenfelder, U. (2001). Sequence detection in pseudo-words in French: Where is the syllable effect? *Language and Cognitive Processes*, 16(5/6), 609–636.
- Cutler, A., Mehler, J., Norris, D. & Segui, J. (1989). Limits of bilingualism. *Nature*, 340, 229-230.

Cutler, A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access.

Journal of Experimental Psychology: Human Perception and Performance, 14 (1), 113-121.

Cutler, A. & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31, 218-236.

Davis, M., Marslen-Wilson, W. & Gaskell, M. (2002). Leading up the lexical garden-path: segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 218-244.

Dejean de la Bâtie, B. (1993). Word boundary ambiguity in spoken French. Unpublished doctoral dissertation. Monash University, Victoria, Australia.

Delattre, P. (1951). Principes de phonétique française: à l'usage des étudiants anglo-américains. Middlebury, Vermont: Ecole française d'été, 2 ed.

Delattre, P. (1966). *Studies in French and comparative phonetics*. The Hague: Mouton & Co.

Diehl, R., Souther, A. & Convis, C. (1980). Conditions on rate normalization in speech perception. *Perception & Psychophysics*, 27, 435-443.

Douchez, C. & Lancia, L. (2008). Caractéristiques articulatoires des consonnes de liaison: Etude pilote. *Journées d'Etudes sur la Parole*, Avignon, France.

Dumay, N., Frauenfelder, U., & Content, A. (2002). The role of the syllable in lexical segmentation in French: Word-spotting data. *Brain and Language*, 81, 144-161.

Durand, J & Lyche, C. (2008). French liaison in the light of corpus data. *French Language Studies*, 18, 33-66.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, 29 (2), 109-135.

- Fougeron, C. & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 106 (6), 3728-3740.
- Fougeron, C., Bagou, O., Stefanuto, M., Frauenfelder, U. (2003). Looking for acoustic cues of resyllabification in French. *15th International Congress of Phonetic Sciences*, Barcelona.
- Gaskell, M. & Marslen-Wilson, W. (2001). Lexical ambiguity and spoken word recognition: bridging the gap. *Journal of Memory and Language*, 44, 325-349.
- Gaskell, G., Spinelli, E. & Meunier, F. (2002). Perception of Resyllabification in French. *Memory and Cognition*. 30 (5), 798-810.
- Goldinger, S. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166–1183.
- Gow, D. & Gordon, P. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, 21 (2), 344-359.
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, 28, 267-283.
- Huggins, A. (1972). Just noticeable differences for segment durations in natural speech. *Journal of the Acoustical Society of America*, 51, 1270-1278.
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145-165). San Diego: Academic Press.
- Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, 14. 147.172.

- Kawai, N. & Carrell, T. (2005). Discrimination of phoneme length differences in word and sentence contexts. *Journal of the Acoustical Society of America*, 118 (3), 2033.
- Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59 (5), 1208-1221.
- Klatt, D. & Cooper, W. (1975). Perception of Segment Duration in Sentence Contexts. In A. Cohen & S. G. (Eds.) *Structure and Process in Speech Perception*. Berlin: Nootboom Springer-Verlag.
- Lachs, L., McMichael, K., Pisoni, D. (2003). Speech perception and implicit memory: Evidence for detailed episodic encoding of phonetic events. In J. Bowers & C. Marsolek (Eds.), *Rethinking implicit memory* (pp. 215-235). Oxford: Oxford University Press.
- Lehiste, I. (1961). Acoustic studies of boundary signals. *Proceedings of 4th ICPhS* (pp. 178-187).
- Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 6 (2), 2018-2024.
- Léon, P. (1992). *Phonétisme et prononciations du français avec des travaux d'application et leurs corrigés*. Paris: Nathan.
- Local, J. (2003). Variable domains and variable relevance: interpreting phonetic exponents. *Journal of Phonetics*, 31, 321–339.
- Luce, P. & Pisoni, D. (1998). Recognizing spoken words: the neighborhood activation model. *Ear & Hearing*, 19 (1), 1-36.
- Mattys, S. (2003). Stress-based speech segmentation revisited. In *Proceedings of Eurospeech: The 8th Annual Conference on Speech Communication and Technology*. (pp. 121–124). Geneva.

- Mattys, S. (2004). Stress versus coarticulation: Towards an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 397-408.
- Mattys, S., White, L. & Melhorn, J. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134, 477-500.
- McClelland, J. & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18 (1), 1-86
- McQueen, J. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, 39, 21-46.
- McQueen, J., Norris, D. & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, 25 (5), 1363-1389.
- Mehler, J., Dommergues, J.Y., Frauenfelder, U. & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20(3), 298-305.
- Miller, J. (1987). Rate-dependent processing in speech perception. In A. Ellis (Ed.), *Progress in the Psychology of Language*, (pp. 119- 57). Hillsdale, NJ: Erlbaum.
- Nakatani L. & Dukes, K. (1977). Locus of Segmental Cues to Word Juncture. *Journal of the Acoustical Society of America*, 62, 714-719.
- New B., Pallier C., Ferrand L. & Matos R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L'Année Psychologique*, 101, 447-462.
- Nguyen, N., Wauquier, S., Lancia, L. & Tuller, B. (2007). Detection of liaison consonants in speech processing in French: Experimental data and theoretical implications. In P.

- Prieto, J. Mascaró & M.-J. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 3–23). Amsterdam: John Benjamins.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52 (3), 189-234.
- Norris, D., McQueen, J., Cutler, A. & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, 34 (3), 191-243.
- Peperkamp, S., Pettinato, M. & Dupoux, E. (2003). Allophonic variation and the acquisition of phoneme categories. In B. Beachley, A. Brown & F. Conlin (Eds.) *Proceedings of the 27th Annual Boston University Conference on Language Development: Volume 2* (pp. 650-661). Somerville, MA : Cascadilla Press,
- Pisoni, D. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. Mullennix (Eds.), *Talker variability in speech processing*. (pp. 9-32). San Diego: Academic Press.
- Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*, 20 (3), 331-350.
- Salverda, A., Dahan, D. & McQueen, J. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90, 51-89.
- Schneider, W., Eschman, A. & Zuccolotto, A. (2002). E-Prime User's Guide. Pittsburgh: Psychology Software Tools Inc.
- Shatzman, K. & McQueen, J. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception and Psychophysics*, 68 (1), 1-16.
- Shoemaker, E. & Birdsong, D. (2008.) La résolution de la liaison par des locuteurs natifs et non-natifs. *Acquisition et Interaction en Langue Etrangère*, 27, 43-62.

Spinelli, E., McQueen, J. & Cutler, A. (2003). Processing Resyllabified Words in French.

Journal of Memory and Language, 48 (2), 233-254.

Spinelli, E., Welby, P. & Schaegis, A. (2007). Fine-grained access to targets and competitors in phonemically identical spoken sequences: the case of French elision. *Language and Cognitive Processes*, 22 (6), 828 – 859.

Summerfield, A. (1975). How a full account of segmental perception depends on prosody and vice versa. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception*. New York: Springer-Verlag,

Summerfield, A. (1981). On articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 1074–1095.

Tabossi, P., Collina, S., Mazzetti, M., & Zoppello, M. (2000). Syllables in the processing of spoken Italian. *Journal of Experimental Psychology: Human Perception and Performance*, 26 (2), 758–775.

Tremblay, A. (2011). Learning to parse liaison-initial words: An eye-tracking study. *Bilingualism: Language and Cognition*, 14, 257-279.

Tyler, L. & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception and Psychophysics*, 34, 409-420.

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.

Wauquier-Gravelines, S. (1996). *Organisation phonologique et traitement de la parole continue: Contraintes prosodiques et phonologiques de l'accès lexical*. Unpublished doctoral dissertation. Université Paris 7.

- Welby, P. (2003). French intonational rises and their role in speech segmentation. In *Proceedings of Eurospeech: The 8th Annual Conf. on Speech Communication and Technology*, (pp. 2125–2128). Geneva.
- Zwitserslood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32, 25-64.

Table 1

Mean Segmental Durations in Milliseconds of Consonants in Liaison Position and Initial Position

		/n/	/t/	/z/
Real Words	Liaison Position	88.14	68.94	92.82
	<i>SD</i>	21.01	23.47	16.50
	Initial Position	110.51	86.54	105.10
	<i>SD</i>	23.79	25.80	15.14
Nonce Words	Liaison Position	91.21	70.02	90.27
	<i>SD</i>	20.58	27.32	18.78
	Initial Position	113.78	85.83	103.48
	<i>SD</i>	27.90	32.08	21.87

Table 2

Segmental Durations in Milliseconds Used in the Manipulation of Experimental Stimuli

		Continuum condition		
		/n/	/t/	/z/
Two-syllable tokens	Shortened (LC)	64.83	44.88	72.82
		(-21.46%)	(-29%)	(-18.64%)
	Baseline	94.04	71.08	94.97
		(0)	(0)	(0)
	Lengthened (IC)	129.20	105.14	117.17
		(+22.43%)	(+33.12%)	(+14.26%)
One-syllable tokens	Shortened (LC)	75.53	46.09	75.11
		(-21.94%)	(-39.24%)	(-22.21%)
	Baseline	107.78	84.71	100.86
		(0)	(0)	(0)
	Lengthened (IC)	145.23	123.32	128.11
		(+22.29%)	(+32.15%)	(+18.47%)

Note. Percentage difference from the production mean in each condition is given in parentheses.

Table 3
Mean D-prime (d') Scores on AX discrimination Tasks (Experiments 1A and 1B)

		Pair					
		1 2	1 3	2 1	2 3	3 1	3 2
Experiment 1A (Real words)							
/n/	<i>M</i>	1.05	3.11	1.15	1.17	2.20	0.94
	<i>SD</i>	0.93	1.31	0.97	0.74	1.01	0.60
/t/	<i>M</i>	0.62	2.26	0.84	1.28	1.56	0.22
	<i>SD</i>	0.80	1.18	0.89	0.88	0.88	0.40
/z/	<i>M</i>	0.69	2.06	0.25	0.64	0.87	-0.03
	<i>SD</i>	0.62	1.00	0.63	0.49	0.65	0.42
		<i>M</i>	0.76	2.18	0.75	0.99	1.39
		<i>SD</i>	0.61	1.06	0.75	0.55	0.61
Experiment 1B (Non-words)							
/n/	<i>M</i>	0.59	2.19	0.59	1.17	1.72	0.81
	<i>SD</i>	0.77	1.07	0.74	0.76	1.58	1.30
/t/	<i>M</i>	0.53	2.29	0.53	1.05	1.35	-0.11
	<i>SD</i>	0.87	1.51	0.62	0.97	1.03	0.83
/z/	<i>M</i>	0.81	1.78	0.43	1.08	0.90	0.53
	<i>SD</i>	1.30	0.97	0.74	0.75	1.06	0.87
		<i>M</i>	0.48	1.97	0.53	1.06	1.15
		<i>SD</i>	0.64	1.02	0.52	0.61	1.00

Table 4

Correlations between mean accuracy rates and d' scores on AX discrimination tasks (Experiments 1A and 1B) as a function of percentage change within each different pair.

Exp 1A		
Accuracy * percent change	$r = .852$	$p = .0313$
d' scores * percent change	$r = .930$	$p = .0071$
Exp1B		
Accuracy * percent change	$r = .765$	$p = .0766$
d' scores * percent change	$r = .880$	$p = .0208$

Table 5

Mean Proportion of 'Liaison' (Vowel-initial) Responses in Forced-choice Identification Tasks (Experiments 2A and 2B)

		Continuum condition		
		Shortened (LC)	Baseline	Lengthened (IC)
Experiment 2A (Real words)				
/n/	<i>M</i>	71.06*	53.01	17.11**
	<i>SD</i>	25.62	10.48	18.17
/t/	<i>M</i>	68.29*	44.21	17.31**
	<i>SD</i>	17.28	11.62	10.59
/z/	<i>M</i>	59.95*	44.44	32.18**
	<i>SD</i>	19.08	12.86	18.51
		<i>M</i>	47.22	22.54**
		<i>SD</i>	6.25	13.56
Experiment 2B (Nonce words)				
/n/	<i>M</i>	73.02*	56.07	25.91**
	<i>SD</i>	23.45	17.12	22.47
/t/	<i>M</i>	54.43	39.73	19.53**
	<i>SD</i>	23.35	22.98	19.53
/z/	<i>M</i>	62.21***	56.33	31.77**
	<i>SD</i>	23.34	26.34	29.26
		<i>M</i>	53.01	28.55**
		<i>SD</i>	19.84	22.60

*Above chance performance (50%) at $p < .05$.

**Below chance performance (50%) at $p < .05$.

*** Just missed significance at $p = .054$