Title
Compensation for Coarticulation of Fricative-stop Clusters as Mediated by Lexical and Temporal Effects

Journal
UC Berkeley PhonLab Annual Report, 9(9)

ISSN
2768-5047

Author
Chew, Vanessa

Publication Date
2013

DOI
10.5070/P72415w16p

Compensation for coarticulation of fricative-stop clusters as mediated by lexical and temporal effects

Vanessa Chew

Advisor

Keith Johnson

Reader

Sharon Inkelas

Honors Thesis
Department of Linguistics
University of California, Berkeley
April 22, 2013

Abstract

Previous work on compensation for coarticulation by Mann and Repp (1981) on alveolar-velar ambiguous stops has indicated that preceding [s] is linked to more velar percepts while preceding [ʃ] is linked to more alveolar percepts. In addition, syllabification has been found to have no effect on listeners' response, while temporal separation between the fricative and the stop produces significant change of categorical boundaries at small gap sizes.

However, methodological issues lend suspicion to these results, as do other experiments in the literature that strongly support the existence of perceptual grouping effects (Samuel & Pitt, 2003). Building on previous efforts employing listeners' natural bias for hearing real words (Ganong, 1980) to manipulate perception (Elman & McClelland, 1988), the present study uses this same Ganong effect to enforce novel syllable boundaries in order to investigate the interaction of lexicality and temporal separation on the magnitude and direction of compensation for coarticulation.

Three pairs of stimuli with ambiguous stops are studied: *ace dawn–ace gone* [eʲs.tɑn–eʲs.kɑn], *ash dawn–ash gone* [æʃ.tɑn–æʃ.kɑn], and *a stone–a scone* [ə.stoʷn–ə.skoʷn]. A fourth nonword pair, *astof–askof* [əstoʷf–əskoʷf] is also tested in an effort to probe the results of the initial experiments. It was found that syllabification and lexical status have great global influence over listeners' response. Lexical frequency and vowel context also produce changes in listeners' response that, while large, are limited in their scope. Surprisingly, the effect of temporal separation is found to be particularly sensitive to lexicality.

The mechanisms that could underlie all these phenomena are discussed, as are the implications of these results on current speech perception theory.

Acknowledgements

Great appreciation goes, of course, to my advisor and director of this thesis, Keith Johnson, Professor of Linguistics and Director of the Phonology Lab. Thank you for your constant support and willingness to puzzle through all the obstacles I encountered in the course of pursuing this one stray thought. Thank you for opening the lab to me so freely and unhesitatingly. Without you, my ideas would never have come into fruition.

I am immensely grateful for my second reader, Sharon Inkelas, Professor of Linguistics and Chair of the Department of Linguistics. Thank you for your ready acceptance to be my second reader and for your invaluable advice from the initial summary to the full draft. This work was very much improved by your input.

My heartfelt thanks go to Ronald Sprouse, Information Systems Analyst, for setting me up in the lab and teaching me how to use all the equipment there. Thank you also for providing your voice and expertise with the electropalatography apparatus on such short notice.

I am indebted to Grace Neveu; thank you for letting me use your voice in the synthesis phase of this research and for your general patience as my first-line guinea pig. I would be remiss in not thanking Melinda Fricke, who taught me the workings of an experiment and bestowed statistics and writing advice throughout the journey. A quick word of thanks goes to Anne Vilain, who first introduced me to the world of phonetics and speech perception and encouraged me to delve deeper.

To all my subjects, you have my eternal gratitude. Thank you all for coming into the lab week after week, not knowing how long and exhausting the experiment was going to be, but nevertheless doing me a huge favor by enduring them without complaint. Thank you for still being my friends and encouraging me even after I put you through all that. I could not be more appreciative:

From the chemistry department, fellow compatriots in the lab and in problem-set drudgery: Allison, Caroline and Phuong;

From the UC Men's and Women's Chorales, from jazz to Broadway and the Alto IIs, my fellow fake-men: Amanda, Anna, Charlotte, Christine, Cristy, Kelsey W., Pandu, Ryan, Sowmya;

Friends of friends of friends: Elizabeth, Hamilton;

And of course, last but not least, from the linguistics department, fellow undergrad majors and grad students, united in the belief that we do Science, and wait, what did you say, could you pronounce that again, please: Auburn, Emily, Erin, Grace, Hemisha, Jeff, Jevon, Kelsey N., Kevin, Linda, Matt, Melinda, Michelle, Nico, Roslyn, Sarah, Whitney.

Thank you all for your kind words and your belief in me and my research.

Table of Contents

List of Figures

1 Introduction

1.1 Compensation for coarticulation

The phenomenon of compensation for coarticulation has been extensively studied over the past half century, being of particular interest to those seeking to develop models of speech perception. Compensation for coarticulation describes the differences in perception of ambiguous sounds due surrounding events. Typically, these are speech events: varying segments drive perception in different directions. To illustrate, a commonly-used example of compensation for coarticulation involves the phones [d] and [g]. When sounds with acoustic signals ambiguous between [d] and [g] are preceded by segments [l] and [ɹ] in the sequence [aLCa], where L is either of the two liquids and C is the ambiguous stop, listeners are far more likely to interpret the mystery stop as /d/ if the preceding liquid is [l] and /g/ if the preceded by [ɹ]. That is, listeners are biased towards hearing /alga/ and /arda/ sequences than the alternatives /alda/ and /arga/ (Lotto & Kluender, 1998).

There exist two competing theories to explain this. The first involves gesture perception (Fowler, 2006) and takes into account one's knowledge of one's articulators when producing such sequences in order to compensate for another person's speech. In the previous example, [d] has an alveolar place of articulation: the tongue tip is raised to the alveolar ridge to form the occlusion, while [g] has a velar place of articulation in which the tongue body is raised to the velum. Since American English /r/ usually involves some type of retroflexion or bunching of the tongue tip, the listener is able to extrapolate that the following [d] will be pronounced farther in the back of the mouth than at the usual alveolar ridge as this happens when he produces this sequence himself. Therefore, a sound that is ambiguous between [d] and [g] is more likely to be categorized as /d/ as the perceptual category space of /d/ is expanded posteriorly following [ɹ]. Similarly, [l], with its alveolar articulation, pulls the tongue body forward, leading to the category space of /g/ being expanded anteriorly.

The second theory relates to the listener's familiarity with auditory feedback via spectral contrast

(Lotto & Kluender, 1998). The segment [d] differs from [g] most notably in F3: [d] has a formant transition that has a higher starting position before descending to the vowel formant, while [g] is characterized by the 'velar pinch', with F3 starting below the vowel formant instead. American English /r/ is typically characterized by low F3. An ambiguous stop heard immediately after [ɹ] is likely to have an F3 higher than that of [ɹ]. Due to this contrast, listeners are likely to associate this sound with the category that has the higher F3: /d/. The liquid [l], with its relatively higher F3, will cause a contrast effect in the opposite direction. Since the following sound is likely to have a lower starting F3, listeners tend to perceive a velar pinch and categorize it as /g/. These contrasts are illustrated diagrammatically in Figure 1.1.

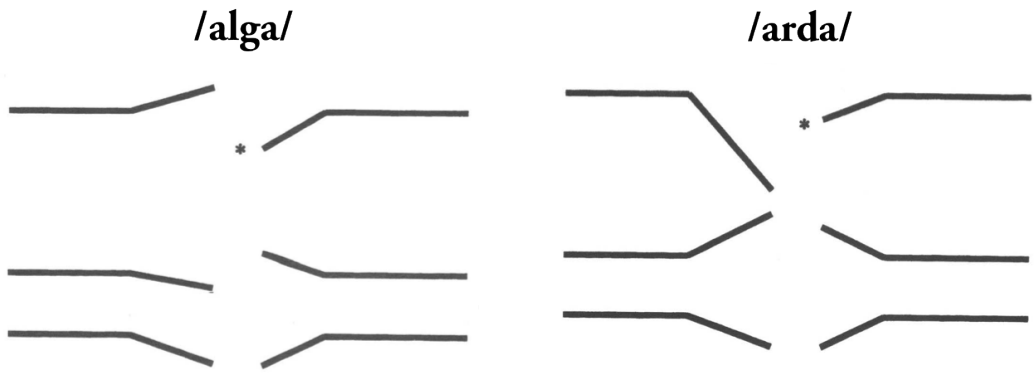

Figure 1.1: Formants in perceived /alga/ and /arda/ (Lotto & Kluender, 1998). The asterisks indicate the F3 transition of the same ambiguous stop.

Both theories have a wide body of evidence supporting them, but above all, they are both products of a loop between a person's auditory and perceptual systems, as perception is not simultaneous with audition: these events are separated and one occurs after the other. The presence of such a distance suggests that this process of compensation for coarticulation might also be subjected to other psychological speech-processing effects, not just immediate and neighboring phonemes, that may intervene at any point during the loop.

1.2 Context effects: A previous investigation

In 1981, Mann and Repp, having accidentally discovered the compensation for coarticulation phenomenon in fricative-stop sequences in a previous experiment (Mann & Repp, 1980), conducted

precisely such an investigation. They confirmed the presence of what was known as the "context effect" and explored the effect of an increase in temporal separation and the insertion of a perceptual syllable boundary between fricative and stop. The fricatives studied were [s] and [ʃ]; the stops were alveolar /d/ and velar /g/. Their previous work had indicated that preceding [s] caused more [k] percepts while preceding [ʃ] caused more [t] percepts.

The study of temporal separation yielded unexpected results. In the preceding [s] condition, there was no significant change in the pooled percent /d/ responses from 75 to 375 ms. In the preceding [ʃ] condition, however, there was a decrease in /d/ responses from a gap size of 75 ms to 150 ms, but no significant difference for longer gap sizes of 150–375 ms, though the percent /d/ responses were still higher at 375 ms than for the corresponding preceding-[s] responses, indicating that *some* compensation effect was still at work even over gap sizes far longer than any natural utterance would have. The reason underlying the decrease in the difference in responses for the two fricatives is evident: the effect of a preceding sound only lasts for so long and a threshold boundary is expected. Less clear is the persistence of a context effect even at long gap sizes, which Mann and Repp eventually attribute to a general response bias.

With regards to perceptual syllabification, initial experiments showed that the insertion of a perceptual syllable boundary by adding an initial vowel produced significant change, i.e. FCV versus VFCV, where V is a vowel, F is either of the two fricatives, and C is a token from the continuum of stops. However, it was unclear *where* exactly the syllable boundary was and whether this was a crucial factor in perception. A subsequent experiment was thus conducted using the same set of VFCV stimuli, in which Mann and Repp instructed listeners to interpret the stimuli as VF-CV, with a syllable boundary between the fricative and the stop, or as V-FCV, with the syllable boundary before the fricative. In the first condition, listeners were told to classify the following stop as /d/ or /g/, while in the second condition, to classify the stop as /t/ or /k/, thereby driving the desired syllabification perception by relying on English phonotactics, which do not allow fricative-voiced stop sequences. This led to the conclusion that the "cognitive variable of subjective syllable division" had no effect, and that it was solely the presence of a "stimulus component preceding the fricative noise", i.e. the vowel, which caused the shift in the alveolar-velar perceptual boundary.

But this conclusion is hardly satisfying. Mann and Repp initially ascribed the FCV and VFCV difference to a syllabification effect, conducting the VF-CV versus V-FCV experiment in the expectation that it would confirm this hypothesis. The results indicated otherwise, but no explanation was offered as to why the presence of a vowel, if not provoking syllabification, would cause a shift

in the perceptual boundary. The two current competing theories of gesture perception and spectral contrast, hinted at in their earliest forms in the 1981 paper, do not account for this. A possible theory invoked by the authors is that cues for following stop place articulation are contained within the fricative itself, specifically in its second half. Yet studies of this phenomenon have provided contradictory results: when asked to identify stops after steady-state [s], listeners responded with more alveolar /t/ percepts in Schwartz (1967), but with more /k/ percepts in Malécot and Chermark (1966). In any case, Mann and Repp's 1981 finding that the fricative effect persists through gap sizes of 375 ms, as mentioned above, is also longer than the persistence of this "noise-offset-cue" effect [200 ms, as seen in Repp, Liberman, Eccardt, and Pesetsky (1978)], discounting the applicability of this theory.

In addition, the measure of significance for the FCV versus VFCV condition was calculated based on data pooled across both fricatives, [s] and [ʃ]. This raises some immediate questions. Given that /ʃC/ sequences are illegal in English words (discounting very recent vocabulary additions such as "shtick", "schpiel" and "schmuck" that are still identified by most English speakers as borrowings), it would not be surprising to find that the presence or absence of a syllabification effect varies with the fricative's place of articulation. Indeed, qualitative examination of the graphs of this experiment seem to indicate that the change in boundary for /s/ is larger than the corresponding change in /ʃ/. Precisely why this would be so is unclear. It is possible that listeners also separate /ʃC/ clusters the same way as they do /ʃ.C/ sequences, whereas /sC/ clusters are grouped in contrast to the separation of /s.C/. Whatever the reason may be, further data broken down in detail is needed to determine if listeners truly associate /ʃ/ with stops in a cluster to the same degree as they do phonotactically legal /s/, as Mann and Repp seem to indicate.

More controversial, perhaps, is the method in which the study investigated the effects of the two methods of syllabification of the same VFCV stimuli. Although some subjects were surprised that the same set of stimuli had been presented twice after having being told that they would be hearing separate VF-CV and V-FCV tokens, this fact, while seemingly endorsing the effectiveness of the syllabification instructions, is at best tenuous as evidence for two markedly different listening conditions. In particular, memory effects could have played a large role in the syllabification perception of the stimuli as they were heard back to back, thereby confounding the results. As such, the null result obtained by the experiment lends itself to some skepticism.

1.3 Direction of the current study

The study described here proposes another look at the experiments mentioned above. This re-examination is motivated by the following questions: first, is there a way to enforce syllabification boundaries before or within fricative-stop clusters more rigorously than to simply instruct listeners to place them there? That is, what perceptual mechanism or bias can be taken advantage of in order to assure the perception of differing syllabification conditions? Second, will this more rigorous syllabification method produce significant change in listeners' perceptual boundaries, and if so, how much and in which direction? Third and lastly, how do strong segment association biases modulate the effects of temporal separation between fricative and stop? Which will dominate: gap size or segment clustering?

Much work has been done on the various conditions that affect a listener's degree of compensation for coarticulation. This has ranged from the strictly phonetic, i.e. changing the segments before and after the ambiguous sound under study, to higher-level word and sentence structure, and even to external non-speech events. Only a few have, however, been employed in such a manner as to regulate syllabification and to investigate the resulting outcomes. In the next section, some of these conditions and their corresponding theories are reviewed and shown to be amenable to the task of answering the questions posed above. The innovations and expectations of the current study are also discussed.

2 Background

2.1 Lexical knowledge in perception

A landmark study in the landscape of speech perception was conducted in 1980, when Ganong constructed two continua varying in voice onset time (Ganong, 1980): *tash–dash* [tʰæʃ]–[dæʃ] and *task–dask* [tʰæsk]–[dæsk]. Notably, for each set of continua, one endpoint was a real word (*dash, task*) while the other was a nonword (**dask, *tash*). The experiment hypothesized that listeners would be biased towards hearing real words instead of nonwords, and that their categorical boundaries would shift correspondingly. Indeed, this effect was found: though the immediate context of the two continua were identical, that is, the vowel [æ] followed the ambiguous stop, listeners heard more voiced stops in *tash–dash* than in *task–dask*. In addition, this lexical effect was stronger at the phoneme boundary than at the continua ends, indicating that speech perception takes into account both lexical and auditory streams of information.

Another important investigation in the literature concerns perceptions of words themselves as affected by noise (Warren, 1970). From the sentence of natural speech "The state governors met with their respective legislatures convening in the capital city" was excised the first [s] of *legislatures*, as well as portions of the adjacent sounds that might provide acoustic transitional cues to the deleted sound. The empty space was then replaced by a cough. Subjects were told that they would hear a sentence with a cough occurring at some point in the middle; later, they would be given a typewritten copy of the sentence and asked to circle the the exact position of the cough. They were also asked if the cough completely replaced the circled letters. All listeners reported that the cough did not replace the circled letters; furthermore, none of the subjects correctly circled the position of the cough, with the errors being evenly distributed before and after its actual position. This demonstrated that listeners must have perceptually restored the missing phoneme on top of the noise of the cough, and that they did so very early in the process of perception. Interestingly, this effect occurs for replacements

with pure tones as well as white noise, but if the gap from the excised sound is left silent, phoneme restoration does not occur, indicating that silence is also a salient cue in speech perception.

Unifying the two above-mentioned phenomena are studies conducted by Samuel (Samuel, 1996). He created stimulus sets in which some had complete replacement of phonemes by noise, while some merely had noise added to the signal. 90 words and pseudowords were used: the pseudowords were matched to a real word by syllable length, stress pattern, and "critical phoneme identity and position"; in general, only one phoneme was changed per syllable. It was found that phoneme perception and restoration were better in stimuli involving real words compared to pseudowords (nonwords), even though this lexical effect was somewhat fragile. However, this fragility might have been due to the pooled results of shorter and longer stimuli. In 2006, Pitt and Samuel tested word-nonword continua of varying lengths: among the shorter stimuli pairs were *miss–*mish* and **wiss–wish*, while the longer stimuli pairs included *arthritis–*arthritish* (Pitt & Samuel, 2006). They found that the Ganong effect was greater for longer words, but was extremely context-dependent for shorter words.

It follows that perceptual restoration might also follow this trend. Yet whether it would or would not depends on the effect of lexical knowledge on speech processing in listeners. Alongside these discoveries have arisen many theories of speech perception attempting to address how a listener's perceptual system might integrate both pre-existing lexical knowledge and immediate, on-line auditory information. Current theories are generally split by the question of *when* lexical knowledge becomes inserted into the auditory process: some theories posit that lexemes affect prelexical perception, while others assert that the two streams are separate. These models are presented later in Chapter 6. Whatever the theories may be, there is one line of experimental inquiry for which lexical biases cannot be ignored, that any speech perception model must take into account.

2.2 Lexeme-driven compensation for coarticulation

A question that arises from Ganong effect and perceptual restoration studies is: Can lexical biases really be so strong as to be comparable to stimuli with real segments? To answer this, researchers investigated the effects of lexical bias on another perceptual phenomenon. As implied by its name, lexeme-driven compensation for coarticulation is highly relevant to the present study. In these experiments, listeners hear a stimulus that should activate a lexical response; however, instead of making a judgement on that stimulus itself, another perceptual ambiguity is presented and judgements made on the latter instead. If a positive result is obtained, this would therefore suggest that the lexical

response is strong enough to condition subsequent perceptual processing.

The investigative paradigm of this line of inquiry represents a combination of previous well-studied effects. Pioneering work was conducted by Elman and McClelland (1988) who designed an experiment in which two words would be presented to the listener. The first would be a stimulus ending with an ambiguous fricative between [s] and [ʃ] that was meant to invoke the Ganong effect (Ganong, 1980). Two stimuli were used in this first slot: the words *foolish* and *Christmas*, which had corresponding nonword variants if the final fricative was replaced with its counterpart at the other end of the continuum: **foolis*, **Christmash*. The second slot would be filled by a continuum of words varying in place of articulation that was meant to invoke compensation for coarticulation as described by Mann and Repp (1981): *tapes–capes* [$t^h e^j$ps]–[$k^h e^j$ps]. Subjects were asked to identify the initial sound of the second word: either [t] or [k].

If the Ganong effect was extremely strong, the categorical boundaries for the ambiguous stops following the fricatives would be shifted from each other. On the contrary, if lexical biases were weak, the expectation would be that the boundaries would be exactly the same. Elman and McClelland found the latter result: listeners tended to hear *tapes* after /ʃ/ from *foolish* and *capes* after /s/ from *Christmas*. The shift of categorical boundaries was of a comparable magnitude to Mann and Repp's 1981 study that had compared the same set of fricatives and stops. It is important to emphasize that the 'active' properties of the relevant fricatives were completely illusory in this study: phonetically, the exact same ambiguous fricative was presented, which should have resulted in the same categorical boundaries for the following continuum of stops. However, listeners perceived all cues associated by experience with /s/ or /ʃ/, thereby *overriding* actual auditory input; the perceived cues then drove subsequent compensation for coarticulation.

But the explanation behind this effect was contested by Pitt and McQueen (1998). They argued that Elman and McClelland had failed to take into account a confounding factor that could explain the positive compensation for coarticulation results: the transitional probabilities of the phonemic sequences. On a language-wide scale, /s/ is more likely to occur after /ə/ while /ʃ/ is more likely to occur after /ɪ/, which would account for the percept of /s/ in *Christmas* and /ʃ/ in *foolish*. Listeners can learn, through experience, the difference between [t] after [s] and [t] after [ʃ]; the same occurs for [k]. They can then take these transitional probabilities into account to create a compensation for coarticulation effect for the [t]–[k] ambiguous stop after the [s]–[ʃ] ambiguous fricative. Pitt and McQueen tested this hypothesis by creating two sets of stimuli: the first contained nonwords with transitional probability biases; the second had real words which were matched in terms of

transitional probabilities. They found that compensation for coarticulation occurred with the first set of stimuli, but not for the second set.

Following this, Magnuson, McMurray, Tanenhaus, and Aslin (2003b) presented evidence disputing Pitt and McQueen's conclusions and instead supporting Elman and McClelland's original hypotheses. Their experimental paradigm used the words *bliss* /blɪs/ and *brush* /brʌʃ/ with the *tapes–capes* continuum. Their words opposed the transitional probability bias towards /s/ in [əs]~[ʌs] and towards /ʃ/ in [ɪʃ] against the lexical bias of hearing words *bliss* and that *brush* compared to nonwords **blish*, **bruss*. This was an effort to decouple transitional probabilities from lexical biases in the original *Christmas* and *foolish* stimuli. Which bias listeners weighted more heavily would produce opposite results in the compensation for coarticulation task. They found that more *tapes* responses were obtained after *brush* and that, correspondingly, more *capes* responses were obtained after *bliss*, indicating that the lexical bias was much greater than any bias arising from transitional probabilities.

Around the same time, Samuel and Pitt (2003) also conducted a series of experiments that took into account phonotactic probabilities and arrived at the same conclusion: that there is lexical influence on sublexical perception and processing. Together, these results suggest that the lexical status of stimuli is a variable that can be manipulated. Considering that listeners prefer to hear words instead of nonwords when a segment is ambiguous, and that this effect is as strong as if the actual, unambiguous segment had been presented instead, it follows that these phenomena can be exploited to manipulate perception of real words in stimuli. This provides a tool to re-analyze Mann and Repp's study (1981), which enforced syllabification of nonword stimuli with instructions. Instead of relying on mere instruction, syllable boundaries could be enforced with the Ganong effect; that is, there would be only one way of syllabifying the stimuli such that two real words would be heard. These perceived lexemes, which exert a bias on subsequent prelexical processing as shown by Magnuson and colleagues (2003a), would then affect perception of the following stop, leading to compensation for coarticulation effects.

2.3 Perceptual grouping

The discussion now turns to the existing literature on syllabification as it affects perception. To begin, how great a role does syllabification play in the perception of an auditory stream? Just as the previous section investigated listeners' preferences for real words, how large a bias is there towards well-formed, or ''real'' syllables in a language, and what causes it?

Massaro and Cohen (1981) had, very early on, established that phonotactically illegal clusters were disfavored by listeners. Building on this study, an experiment was later designed to test if this bias was due to phoneme sequence frequency in the language or due to listeners' innate knowledge of their language's phonotactics (Pitt, 1998). Pitt created a number of stimuli that had test consonant+liquid onsets, with the test consonants being [s], [d], [t] and [g], and the liquids being a continuum between [l] and [ɹ]. The nucleus of the syllable was the vowel [æ], preventing the occurrence of any real English words in the stimuli. One of the hypotheses formed was that if phoneme sequence frequency dominated, the response for the continuum [gl]–[gɹ] would be rather asymmetric, as word-initial /gr/ clusters are far more frequent in the English lexicon than /gl/ clusters. On the other hand, if listeners were only concerned with phonotactic legality, the response would be symmetric. The symmetric result supported the latter theory. Moreover, findings for the other clusters did not scale based on frequency; rather, they scaled in a somewhat binary fashion: if the cluster was legal, there was a bias; if it was illegal, there was no bias. Pitt also tested these clusters in word-medial position and found the same result. Therefore, one can conclude that phonotactic legality is a large factor in the segmentation and recognition of the speech stream: a type of "top-down" influence.

In addition to the biased resolution of ambiguous phonemes, listeners may also transform unambiguous acoustic signals to fit phonotactic expectations, an effect that is perhaps even more astonishingly powerful. When French listeners were asked to freely transcribe stimuli which began with initial consonant clusters [dl] or [tl], both of which are illegal in French (Segui, Frauenfelder, & Hallé, 2001), researchers found that the clusters were more frequently transcribed with phonotactically legal /kl/ or /gl/ clusters instead. A forced-choice task confirmed this result. Frauenfelder and Segui (1989) subsequently instructed listeners to hear the [dl] and [tl] stimuli as their actual phonemes, but the rate of correct responses with the illegal clusters was poor. On the contrary, when listeners were given instructions to erroneously hear the [dl] and [tl] stimuli as legal clusters /gl/ and /kl/ respectively, they obtained an enormously positive response.

To verify that the perception of the illegal consonant clusters was due only to phonotactic knowledge and not due to any other acoustic or phonetic reason, Hallé, Segui, Frauenfelder, and Meunier (1998) conducted a phonemic gating experiment. This type of experiment paradigm involves letting subjects hear progressive longer portions of the stimulus (in this case, the [dl]-initial nonwords), but at every stage, asking them to make a judgement. They found that a sharp change from mainly /d/-responses to mainly /g/-responses occurred around 60 ms. That is, the amount of [l] that was heard was not enough to trigger the illegal cluster transformation response when only the first 60

ms of the word was heard. This is a rather short window of acceptability, implying that for any sequence involving illegal clusters, a legal representation will definitely be preferred to the extent that the probability of hearing the illegal sequence is negligible. On the other hand, if the illegal consonant cluster is preceded by a vowel, the bias against hearing /l/ after /d/ disappears (Moreton, 2002); this is easily explained by the fact that the additional vowel allows listeners to assign /d/ and /l/ to separate syllables, which then makes the cluster acceptable to the phonotactics of French.

But this suggests that without phonotactic cues to word boundaries, listeners have much difficulty segmenting the speech signal. In response, Kirk (2001) demonstrated that structure is assigned to the auditory input using rankings of universal constraints to determine where syllables should begin or end based on prior experience with well-formed syllables in the language. Once a well-formed syllable onset has been identified, the possible words that begin with that onset are activated. This method of lexeme assignment by syllabification is much more parsimonious than other models using individual phonemes as the number of possibilities is dramatically restricted, which reduces cognitive load. Viewed from another perspective, this finding asserts that syllabification occurs *before* word-recognition. This patterns with previous cross-linguistic studies by Dupoux, Pallier, Kakehi, and Mehler (2001) and Dehaene-Lambert, Dupoux, and Gout (2000), who showed that Japanese listeners tend to hear illusory vowels in consonant clusters while French listeners do not, owing to the differing legalities of consonant clusters in the two languages. Together, this body of literature suggests that syllabification is a sublexical process which takes place at the level of other sublexical processes, notably including the key phenomenon of the current study, compensation for coarticulation.

In fact, a preliminary investigation into the interaction between syllabification and compensation for coarticulation is encompassed within Samuel and Pitt's study in 2003. They studied the effects of real words and pseudowords (real words that had had one segment altered) on the perception of the following stop. Their results indicated that the real word, *abolish*, yielded a much weaker compensation effect compared to the pseudoword **aboliss*. Furthermore, stimuli of pure VC syllables yielded a still stronger effect compared to the pseudoword condition. From these results can be inferred that the more word-like a stimulus is, the more strongly it binds its fricative, increasing the pertinence of the syllable break in the fricative-stop gap and thereby reducing compensation for the following stop; alternatively, that compensation is highly dependent on the perceptual grouping (or lack of) for the fricative-stop sequence. Later, Kingston, Kawahara, Mash, and Chambless (2007) conducted a follow-up inquiry in which Japanese and English speakers' sensitivities to perceptual

grouping were compared, and found that there was no significant difference in their responses for phonotactically illegal consonant clusters whether or not a preceding vowel was appended to allow syllabification. However, their study used only syllables and not words: the much greater magnitude of compensation for coarticulation in the VC condition, as shown by Samuel and Pitt (2003), could have overwhelmed any vowel-addition effects. This confounding factor could have been removed, or at least its influence reduced, by using real words instead, allowing the perhaps more nuanced differences between the two languages to percolate to the surface.

2.4 Purpose and innovations of the current study

Instead of enforcing syllabification via instruction such as in Mann and Repp (1981), a possible solution is to exploit the Ganong effect, which has been shown to produce comparable effects to non-Ganong stimuli, in order to affect the perception of the following alveolar-velar ambiguous stops. In addition, the sequences tested in previous studies of lexeme-driven compensation for coarticulation have placed their syllable boundaries only between the fricative and the stop (e.g. *Christmas capes*). At present, as mentioned at the end of the previous section, there have been no studies investigating the influence of *lexeme*-driven syllabification on subsequent stop percepts across the fricative-stop boundary. Given that listeners tend to syllabify segments at early stages of processing, it follows that this prelexical-level syllabification will have a major effect on perceptual grouping, a process that is also prelexical. Furthermore, since perceptual grouping is a function of the association between fricative and stop, it stands to reason that the temporal separation between the segments will change perceptual grouping and thus affect compensation.

This study thus investigates compensation for coarticulation of fricative-stop sequences by modifying two variables: the lexemes used in the stimuli as well as the temporal separation between fricative and stop. Real words are used to enforce syllabification by having the alternative syllabification possibility be composed of nonwords and thus disfavored by listeners. A novel component of the study can be found in its treatment of perceptual grouping: instead of merely presenting stimuli in which the fricative and stop are separated, the experiment considers a word-enforced condition grouping both fricative and stop. Moreover, temporal separation *within* a group also provides an opportunity to compare the magnitude of lexical versus perceptual grouping biases. The exact stimuli and paradigms of the current investigation are described in the following experimental chapters.

3 Experiment 1: Syllabification

3.1 Aims

Experiment 1's objective is to replicate Mann and Repp's 1981 compensation data in which [s] and [ʃ] have varying effects on their following stops. In addition, in an effort to enforce syllable boundaries more effectively, the Ganong effect is used to change the perceptual grouping of fricatives and stops; the corresponding categorical boundary will be studied. The real words used in the study are the following pairs: (1) *ace dawn, ace gone*: [eʲs.tɑn]–[eʲs.kɑn]; (2) *ash dawn, ash gone*: [æʃ.tɑn]–[æʃ.kɑn]; and (3) *a stone, a scone*: [ə.stoʷn]–[ə.skoʷn]. The voicing differences of the acoustics of target perceived phonemes /t/ and /k/ versus /d/ and /g/ are irrelevant as phonetically, English bare /d/ or /g/ are realized with voice onset times (VOT) similar to the stops in clusters /st/ or /sk/.

There are two hypotheses that can formed about the results of this experiment, based on the independent variable being manipulated. Comparing pairs (1) and (2), it is the final fricative of the first word that is changed: they have different places of articulation. However, the perceptual grouping does not vary as both word pairs have a syllable boundary between fricative and stop, with any other alternative yielding nonwords. We expect that there will be a significant difference between the effect of [s] and that of [ʃ], as has been demonstrated multiple times in the literature (Samuel, 2011). In this case, there will be more [k] (velar /k/ or /g/) responses for perceived category /s/ (phonetically corresponding to [s]) and more [t] (alveolar /t/ or /d/) responses for /ʃ/. This result holds true whether we posit gesture perception or spectral contrast as the attributing factor behind compensation for coarticulation.

For pairs (1) and (3), syllabification is the independent variable that is changed. In *ace dawn, ace gone*, the syllable boundary is between the fricative and the stop, while in *a stone, a scone*, the syllable boundary is before the fricative-stop cluster. Based on Samuel and Pitt's previous work on perceptual

grouping (Samuel & Pitt, 2003), we expect that there will be a significant difference in listeners' responses to the place-of-articulation judging task. Since the fricative's effect varies inversely with its perceptual "affinity"; that is, how closely it binds to its syllable, the typical compensation for coarticulation bias for [s], which is a [k] response, should be much stronger in the /.sC/ condition than the /s.C/ condition. The categorical boundary should shift towards [k] and there should be fewer overall [t] responses for the word pair *a stone, a scone* than for *ace dawn, ace gone*.

3.2 Methods

3.2.1 Participants

Thirty-one native speakers of American English (23 female, 8 male) studying at the University of California, Berkeley, aged 18 to 29 years old, participated as listeners in the study. None of the participants indicated past or present speech or hearing disorders.

3.2.2 Stimuli

The three pairs of stimuli intended to be interpreted as two-word phrases were created using the Klatt synthesizer (Klatt, 1980) at a sampling rate of 22050 Hz with four formants. A continuum of nine steps was created for each pair, with each varying synthesizer parameter evenly spaced. All stimuli were created with the same template: The first vowel lasted 200 ms, followed by the fricative for 200 ms, a period of silence for 75 ms, the onset stop burst for 40 ms, the vowel of the second word for 320 ms (with a 20 ms overlap of burst and vowel), and finally the ending [n] for 110 ms, for a total stimuli length of 925 ms. The fundamental frequency ranged from 120 Hz at the beginning of each stimulus, increasing to 169 Hz at the end of the first word, falling to 148 Hz during the nasal of the second word before ending with a slight increase to 168 Hz.

Frequencies and bandwidths of the stimuli were mostly determined by consulting values in Klatt (1978) together with input from a young San Franciscan female's natural utterances. Nasal formant values were also adjusted based on perception data from Harding and Meyer (2003). The fricative [s] was synthesized with noise centered at 5000 Hz and with peak intensity 60 dB. The fricative [ʃ] was synthesized with noise centered at 3500 Hz and with peak intensity 58 dB. The endpoint /tV/ sequences were created with a stop burst centered at 3500 Hz and F1 starting at 200 Hz, F2 starting at 1650 Hz, F3 at 2900 Hz and F4 at 3600 Hz, before transitioning linearly to the following vowel's

formant values within 90 ms. VOT was 20 ms. The endpoint /kV/ sequences were created with a stop burst centered at 2000 Hz and F1 also starting at 200 Hz, F2 starting at 2000 Hz, F3 at 2400 Hz and F4 at 3400 Hz, before transitioning linearly to following-vowel formants within 90 ms. VOT was 20 ms. In the [ɑn] sequences, the nasal zero was shifted to 400 Hz at 790 ms and held until the end of the stimulus, while in the [on] sequences, the nasal zero was shifted to 400 Hz at 730 ms. Detailed formant values of synthesized segments are shown in Table 3.1. Values in parentheses denote moving formants; for example, the F1 of vowel [e^{j}] begins at 500 Hz but descends to 440 Hz by the end of its duration.

Table 3.1: Summary of segment formants and corresponding bandwidths (Hz).

Vowel	F1	b1	F2	b2	F3	b3	F4	b4
e^{j}	500(–440)	110	2255(–2385)	75	3065(–2875)	130(–110)	3250	200
æ	825(–755)	130	1800(–1885)	100(–75)	2870(–2930)	130	3250	200
ɑ	730(–760)	130	1225	75	3100	110	3800	200
o	500(–421)	130	1017(–1100)	120	3050	110	3800	200
t	200	70	1650	120	2900	130	3600	200
k	200	70	2000	120	2400	130	3400	200
(ɑ) n	600	130	1880	75	2875	110	3800	200
(o) n	300	130	1850	75	2800	110	3800	200

The spectrogram of endpoint stimulus *ace dawn* is shown in Figure 3.1. Spectrograms of the other endpoint stimuli can be found in Appendix A.

3.2.3 Procedure

The experiment consisted of three phases: a teaching phase, a practice phase, and an experimental phase. The listener's task and the stimulus presentation were identical in each phase. The stimulus was played through AKG K240 Studio headphones at a comfortable volume. Participants were then asked to make forced-choice judgments: they pressed [1] on a keyboard if they perceived a word with /t/ or /d/ and pressed [0] if they perceived /k/ or /g/. The categories "t/d" and "k/g" were shown on a computer screen above their corresponding response numbers, [1] and [0]. Participants

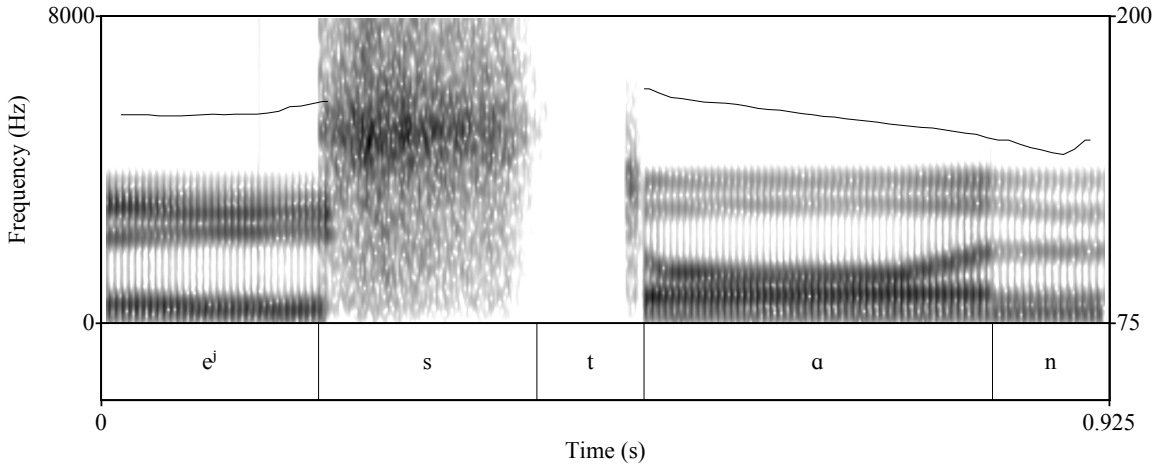

Figure 3.1: Spectrogram and pitch trace of endpoint stimulus *ace dawn*.

were given five seconds to respond, after which a blank screen was shown for one second before the next stimulus was presented.

In the teaching phase, the second half of each stimulus was randomly presented. The tokens were taken from the endpoint stimuli (each stimulus twice) and from the third and seventh stimulus of each continuum once: [tɑn], [kɑn], [to^{w}n], [ko^{w}n] × 2 and the set of four additional tokens, resulting in 12 tokens heard. The purpose of the teaching phase was to familiarize subjects with the two desired categories of sounds, as English orthography does not conform with the relevant IPA symbols in these stimuli.

In the practice phase, all endpoint stimuli of two-word phrases were randomly presented twice; in addition, the third and seventh stimulus of each continuum was also presented: 6 two-word phrases × 2 with 6 additional stimuli, resulting in 18 tokens heard. Participants were told that they would be hearing two-word phrases, and to again press [1] or [0] based on the 't/d' or 'k/g' categories they perceived. The instructions aimed to encourage the listeners to perceive the stimuli as words instead of individual sounds, strengthening the lexical recognition effect and thus changing the position of the syllable break, affecting the subsequent perceptual grouping (or separation) of fricative and stop. The purpose of the practice phase was to familiarize the subjects with the stimuli they would be judging.

The experimental phase was divided into two blocks. All stimuli from all continua were presented: 9 for each pair, with 3 pairs, giving a total of 27 stimuli. This set of stimuli was repeated

thrice in the first block with the entire block of stimuli presented randomly (i.e. all three sets randomly ordered instead of random ordering within a set): 27 × 3, resulting in 81 tokens heard. The participants then had the option of taking a break. In the second block, the set of stimuli was repeated twice and the stimuli presented randomly as before: 27 × 2, resulting in 54 tokens heard. In total, participants judged 135 tokens in this experiment.

3.3 Results

Means of the responses for each fricative and syllabification condition were plotted and data analyzed with ANOVAs. The response curves for /s.C/, /ʃ.C/ and /.sC/ are shown in Figure 3.2.

Comparing the effect of fricative place of articulation; that is, /s.C/ versus /ʃ.C/, the difference between the two curves was found to be significant (response × fricative × step, $F = 9.518$, $p < 0.001$). However, inspection of the /ʃ.C/ curve in Figure 3.2 reveals that the difference is not a categorical boundary shift as Mann and Repp found in 1981; rather, the /ʃ/ curve has a much gentler slope than its /s/ counterpart, indicating that /ʃ/ caused the classification of the following stop to be much less categorical than /s/ does. /ʃ/ also produced slightly more [t] responses on the whole, pooling the responses across all nine steps of the continuum. This is expected following Mann and Repp's (1981) findings, though unlike in the previous study, this difference is insignificant (response × fricative, $F = 0.006$, $p = 0.939$).

Comparing the effect of syllabification; that is, /s.C/ vs /.sC/, the difference between the two curves was found to be significant (response × syllable boundary × step, $F = 8.444$, $p < 0.001$). Figure 3.2 shows that, unlike in the previous condition where the difference was due to the slope of the curve, under these conditions it is the categorical boundary that shifts. Furthermore, inspection of the curves reveals that the 50 % [t] response point lies between steps 5 and 6 for /s.C/, but between steps 6 and 7 for /.sC/: a difference of a whole step. These results are evidently contrary to those of Mann and Repp's study, where they found no significant difference in their investigation of syllabification. Our strongly positive result suggests that their null result was indeed an artifact of merely giving instructions, which do not truly enforce syllabification, whereas the Ganong effect employed in the current experiment was extremely strong and drove changes in the sublexical level of perception. This is in line with expectations of the effects of perceptual grouping (Samuel & Pitt, 2003) and supports the strength of the Ganong effect as previously reported in the literature.

Yet while the categorical boundary shift is unsurprising, its *direction* is. As mentioned above, we

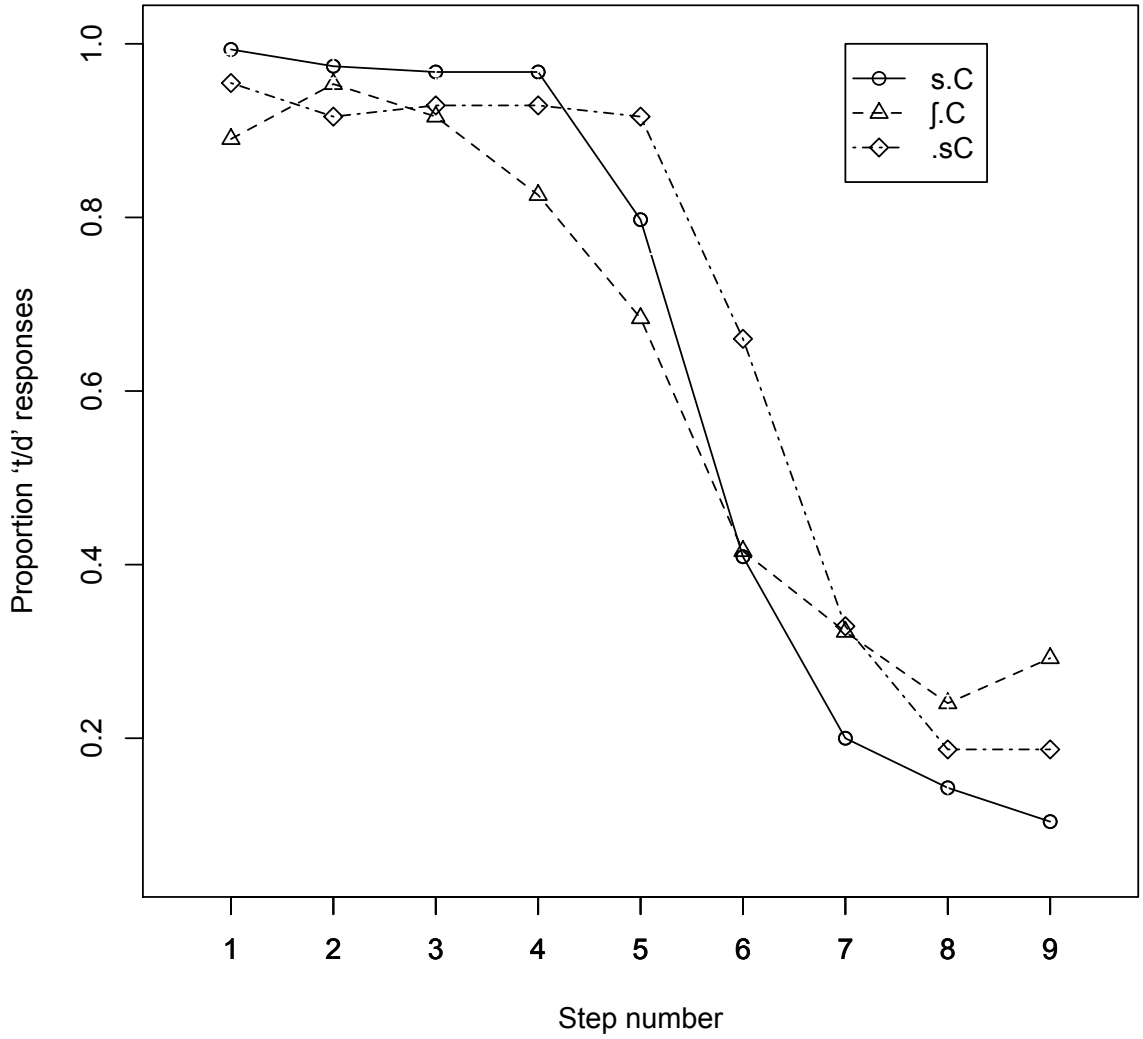

Figure 3.2: Categorical boundary curves of /s.C/, /ʃ.C/ and /.sC/ stimuli.

expected that perceptual grouping enforced by syllabification would cause /s/ to affiliate with the following stop, thereby producing *more* of a compensation for coarticulation effect. We expected that there would be more [k] responses for the /.sC/ condition and that the boundary would shift left. Instead, the boundary shifted to the *right*, indicating that more [t] percepts were heard.

Why, then, would the shift occur in the exact opposite direction as predicted? An explanation could lie in the overall frequency of the words used in the stimuli: *stone* and *scone*. According to the CELEX lexical database of English (Max Planck Institute for Psycholinguistics, 2001), the word

stone is almost 70 times more frequent than the word *scone*. Anecdotally, subjects also reported less familiarity with *scone*: although all subjects knew what the word referred to, most agreed that it appeared much more rarely in childhood (even if it is a regularly encountered word at the UC Berkeley campus), whereas *stone* was quite a frequent lexical item. Given that it is already known that listeners prefer hearing words to nonwords, it is not impossible to imagine that they might prefer to hear higher-frequency words to lower-frequency words as well. Thus, *stone* would be preferred due to listeners' familiarity with that specific lexical item compared to *scone*. If this explanation is true, it implies that lexical frequency effects *override* compensation for coarticulation in rather large significance.

One possible large confounding factor in this experiment, however, is that the vowels in the syllabification portion of the study are not the same. The /s.C/ stimuli *ace dawn–ace gone* have [eʲ] before the fricative and [ɑ] after the stop, while the /.sC/ stimuli *a stone–a scone* have [ə] before and [o] after. This word paradigm was chosen as it was extremely difficult to find words that fit the requirements of the study, in regards to lexeme status vis-à-vis syllabification condition, that had the same vowels. In 1980, Mann and Repp conducted a study in which they found that the following vowel had effects on the perception of the preceding ambiguous fricative. Although the present study differs in that the ambiguous segments are stops with unambiguous fricatives preceding them, there remains a possibility that the significant results of the syllabification investigation are due to the (un)rounded following vowel rather than the changed syllable boundary. This confounding factor is further explored and discussed in Chapter 5.

4 Experiment 2: Temporal Separation

4.1 Aims

Experiment 2 investigates the effect of gap size between the fricative and the stop on the classification of the stop's place of articulation. Evidently, it is expected that the fricative's effect on the following stop will decrease with time. Mann and Repp's 1981 study showed that there was a sharp decrease in the number of [t] responses for stimuli with the first word ending in /ʃ/ from the stop gap intervals 75 ms to 150 ms, with the other stop intervals (up to 375 ms) having no significant change thereafter. The responses associated with the fricative /s/ did not change significantly.

Like in the previous experiment, the study of the *ace dawn–ace gone* and *ash dawn–ash gone* pairs aims to replicate Mann and Repp's results. Since their 1981 study indicated that the stop was badly perceived at stop gap intervals of less than 70 ms, the intervals used in this study begin at 75 ms and increase to 325 ms. For these word pairs, as gap size increases, there should be less compensation for coarticulation of the following stop, and so the categorical boundary should move to the right (i.e. more overall [t] responses) for /s/ and to the left (more [k] responses) for /ʃ/. In effect, the percentage of [t] responses for the two fricatives should trend towards each other. At 325 ms, a stop gap interval far longer than would be found in any natural utterance, the fricative's effect should fade such that the difference between the responses for /s/ and /ʃ/ should be negligible, just as Mann and Repp initially expected. However, they found that a significant difference remained at their largest interval of 375 ms. Yet the perceptual grouping arising from the Ganong effect in this current experiment may bind the fricative much more tightly to its word; its effect on the preceding stop may thus be so reduced that compensation for coarticulation disappears at large gaps, supporting Mann and Repp's original hypothesis.

More interesting is the part of the experiment that considers the word pairs *ace dawn–ace gone* against *a stone–a scone*. Given that the large significant difference between the two modes of syllab-

ification of the previous experiment was attributed to perceptual grouping effects as described by Samuel and Pitt (2003), it follows that this grouping may be affected by increasing periods of silence within the group. We expect that these additional gaps will erase the grouping effect and perhaps even break up the perceptual group. Thus, subjects should increasingly hear the *a stone, a scone* stimuli as nonwords [əs.to^{w}n] and [əs.ko^{w}n] at larger gap sizes. This should cause the categorical boundary curve to move in the opposite direction as the initial shift due to syllabification; that is, the curve should move to the left (more [k]) with increasing temporal separation. Finally, if listeners really hear only nonwords at the end, the resulting curve at 325 ms should be much closer to that of *ace dawn, ace gone* than to its 75 ms variant, even though the word-nonword "affinity" effect still predicts a slight difference between the two word-pair curves at that gap size.

4.2 Methods

4.2.1 Participants

Participants from Experiment 1 were asked to take part in this follow-up experiment. Those who consented comprised twenty-five native speakers of American English (18 female, 7 male), aged 18 to 29. As before, none of the participants indicated past or present speech or hearing disorders.

4.2.2 Stimuli

The stimuli from Experiment 1 were modified to suit the purposes of Experiment 2. Additional periods of silence were added to the middle of the stimulus (during the silent gap perceived as a stop) in intervals of 50 ms and thereafter 100 ms. The resulting set of stimuli thus had gap sizes of 75 ms, as per Experiment 1, 125 ms, 225 ms and 325 ms. This was repeated for each existing token; that is to say, for each step of each pair of words, yielding 9 steps × 3 pairs × 4 gap sizes for a total of 108 unique tokens to be judged.

4.2.3 Procedure

The experimental procedure was again similar to that of Experiment 1. It consisted of two phases: a practice phase and an experiment phase. The teaching phase was removed as the participants had already been made aware of the task through the previous experiment.

The practice phase consisted of one set of tokens from steps 1, 3, 7 and 9 per two-word phrase, with a random number generator used to select the gap size of each step. In addition, another set of tokens from steps 3 and 7 for each word pair, with gap size again randomly chosen, was presented. This yielded 18 tokens heard in this phase.

Following that, the participants were presented with the experimental phase. This phase was divided into two blocks. In the first block, stimuli of steps 2–9 for each two-word phrase, for gap sizes 125, 225 and 325 ms, were presented. Step 1 stimuli were discarded as the data from Experiment 1 had indicated that the difference in response between steps 1 and 2 was insignificant. Since the data across experiments were matched for each participant, the gap size judgements of this experiment could be compared with participants' previous responses, allowing for the original gap size of 75 ms to be omitted in order to reduce the number of tokens that listeners would have to judge. The set of stimuli was repeated three times in random order, with 8 steps × 3 pairs × 3 gap sizes × 3 repetitions resulting in 216 tokens heard. The second phase presented the set of stimuli twice randomly, resulting in 144 tokens heard. In total, participants judged 360 tokens in this experiment.

4.3 Results

For the sake of consistent comparison, data from the second experiment was analyzed with data from the first experiment only for listeners who had participated in both. Again, the means of the responses for each fricative and syllabification condition were plotted and data analyzed with ANOVAs.

The categorical boundary graph for the /s.C/ condition, across all gap sizes, is shown in Figure 4.1 on page 26. As can be seen from the figure, there appears to have generally been no effect of gap size even though significance was found when the response was interacted with gap size and step number on the continuum (response × gap size × step, $F = 2.089$, $p < 0.01$). But as the graph indicates, the greatest change appears to have occurred from 75 to 125 ms. If the 75 ms token is removed, we obtain an insignificant difference (response × gap size × step, 125–325 ms, $F = 0.552$, $p = 0.903$), which implies that there was no change in listeners' categorical boundaries for fricative-stop separations from 125 to 325 ms.

The categorical boundary curves for the /ʃ.C/ condition (Figure 4.2 on page 27) look very similar to those of the /s.C/ condition, except that at the [k]-end of the continuum (larger step number),

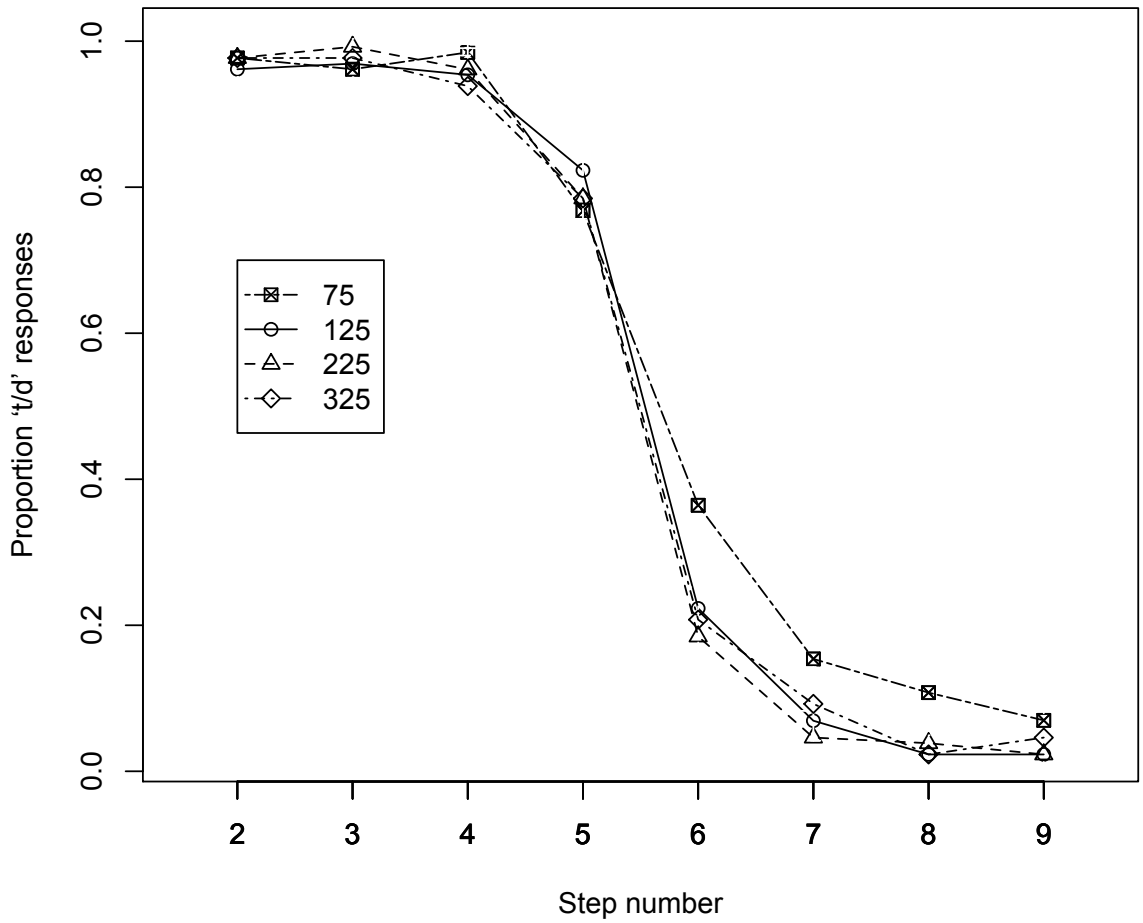

Figure 4.1: Categorical boundary curves of /s.C/ stimuli at varying gap sizes (ms).

the curves are placed higher, indicating that listeners heard more [t] percepts even for tokens with strongly [k]-like values of formant transitions and stop burst. Once again, there is a clear difference between the 75 ms curve and the other gap sizes; overall, there is a significant interaction for gap size (response × gap size × step, $F = 2.570$, $p < 0.001$), but when the 75 ms curve is removed from the interaction, the difference becomes insignificant (response × gap size × step, $F = 0.883$, $p = 0.577$). Based on these results, it appears that the magnitude of compensation for coarticulation, which yields more [t] percepts for /ʃ/, decreases as gap sizes increase. This was expected as larger gap sizes should decrease perceptual grouping, thereby reducing coarticulatory effects.

For the /.sC/ condition, the categorical boundary curves are shown in Figure 4.3 on page 28. Contrary to our expectations, there was no significant difference across the gap sizes ($F = 1.107$,

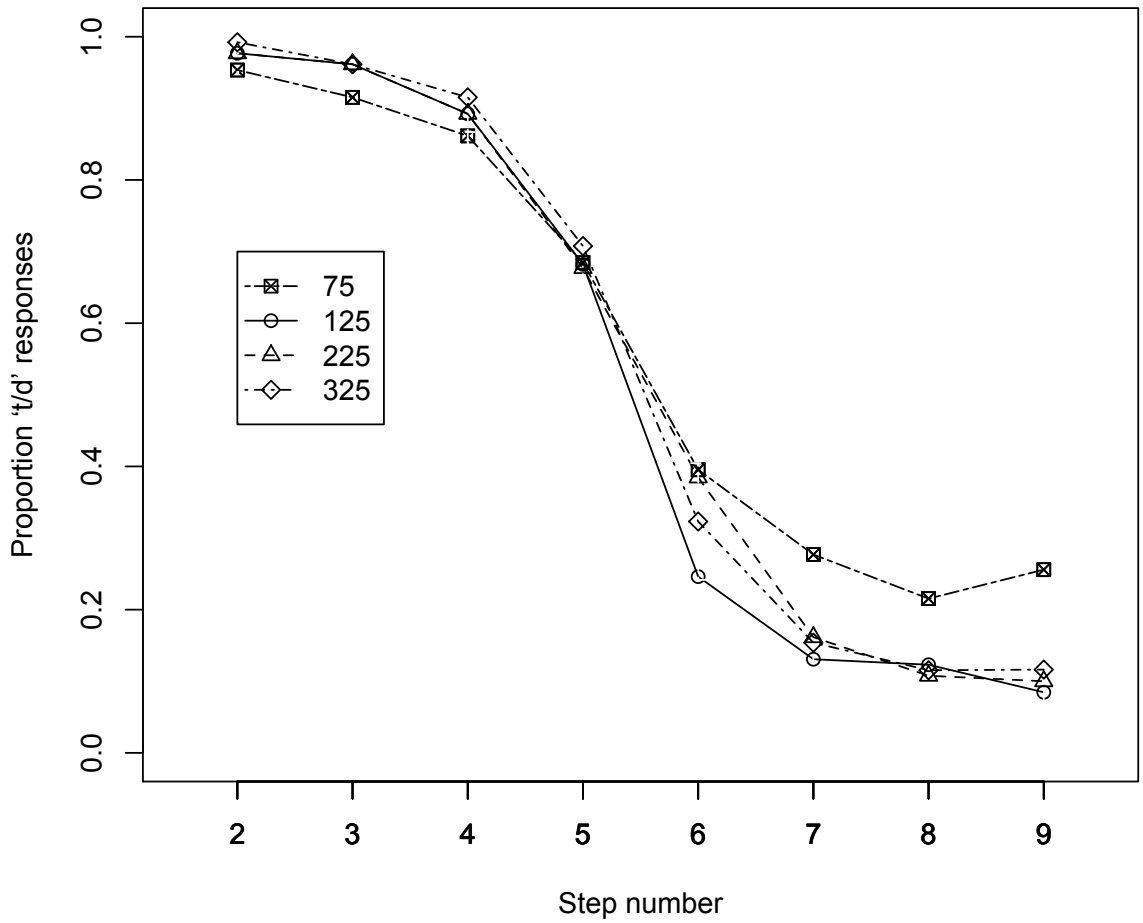

Figure 4.2: Categorical boundary curves of /ʃ.C/ stimuli at varying gap sizes (ms).

p = 0.332). Recall that the perceptual grouping of /s/ and consonant in the second word was expected to disappear with increasing gap silences, as listeners would hear the stimuli as two nonwords instead of real words *a stone* or *a scone*. This would thus cause the word frequency effect as found in Experiment 1 to lessen. That listeners hear nonwords would also imply that the curve of *a stone–a scone* at a 325 ms gap size should closely resemble the 325 ms curve of *ace dawn–ace gone*. Yet the results did not show such behavior (Figure B.1 in Appendix B, with significant difference between the 325 ms curves of /s.C/ and /.sC/: response × syllable boundary × step, F = 44.39, p < 0.001). Instead, the lack of any difference in response behavior between 75 and 125 ms, as was earlier found for /s.C/ and /ʃ.C/ conditions, suggests that the Ganong effect is so strong as to cause listeners to hear real words and thus group the fricative and the consonant together, even at protracted gap sizes.

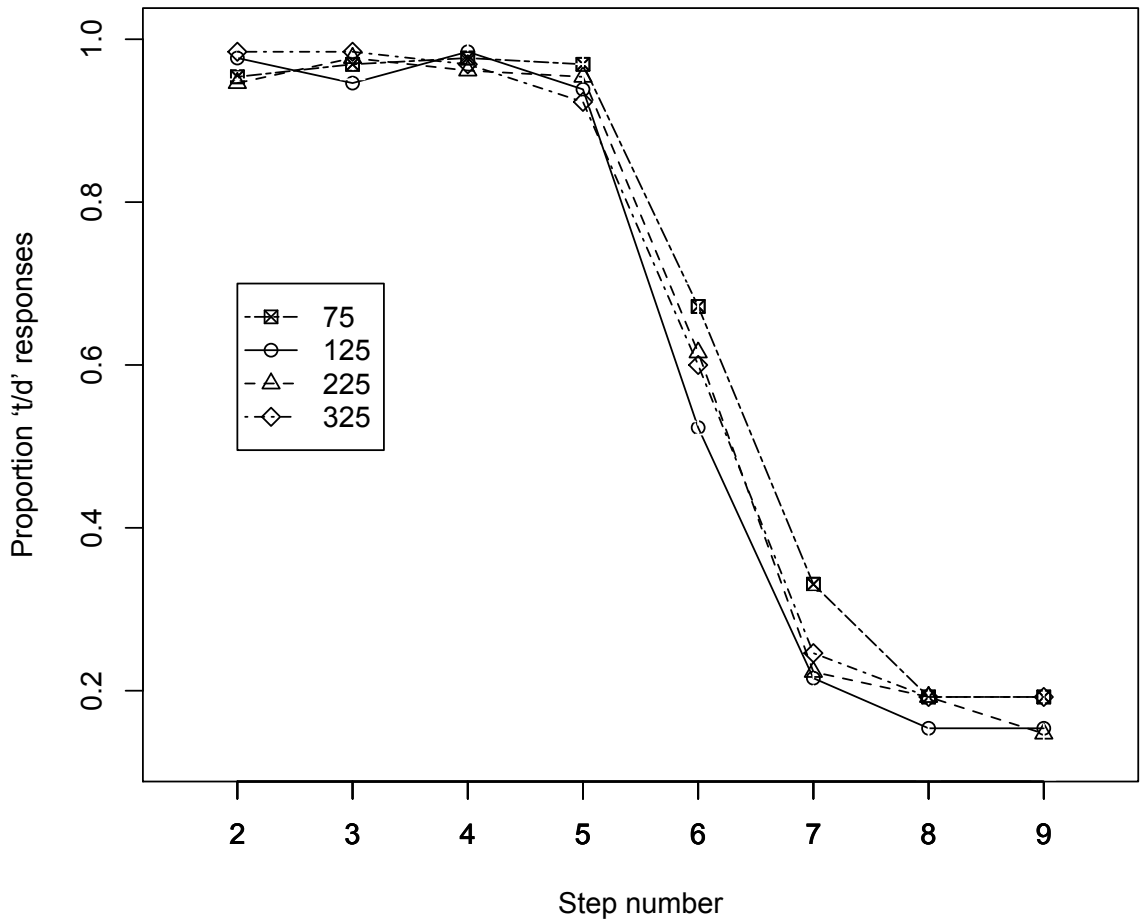

Figure 4.3: Categorical boundary curves of /.sC/ stimuli at varying gap sizes (ms).

The plots of the pooled percentage responses against gap size, for all three fricative and syllabification conditions, are shown in Figure 4.4 on page 29.

Regarding the overall trends seen when comparing /s.C/ and and /ʃ.C/ conditions, it is strange that the both /s/ and /ʃ/ caused a shift in the same direction, even if this shift was statistically insignificant. Both individual plots (Figures 4.1 and 4.2) show that their respective curves for the smallest gap size, 75 ms, produced more [t] percepts than their subsequent curves for larger gap sizes. Why this should be so is unclear. Considering that /s/ causes the following ambiguous stop to be heard more often as [k], if the compensation for coarticulation effect decreases with increasing gap size, the number of [t] percepts should have *increased*. Perhaps the categorical difference in response across both fricatives, as seen in Experiment 1, interacts in some manner with temporal separation.

However, this hypothesis cannot be verified with the data at hand, and must be put aside for now.

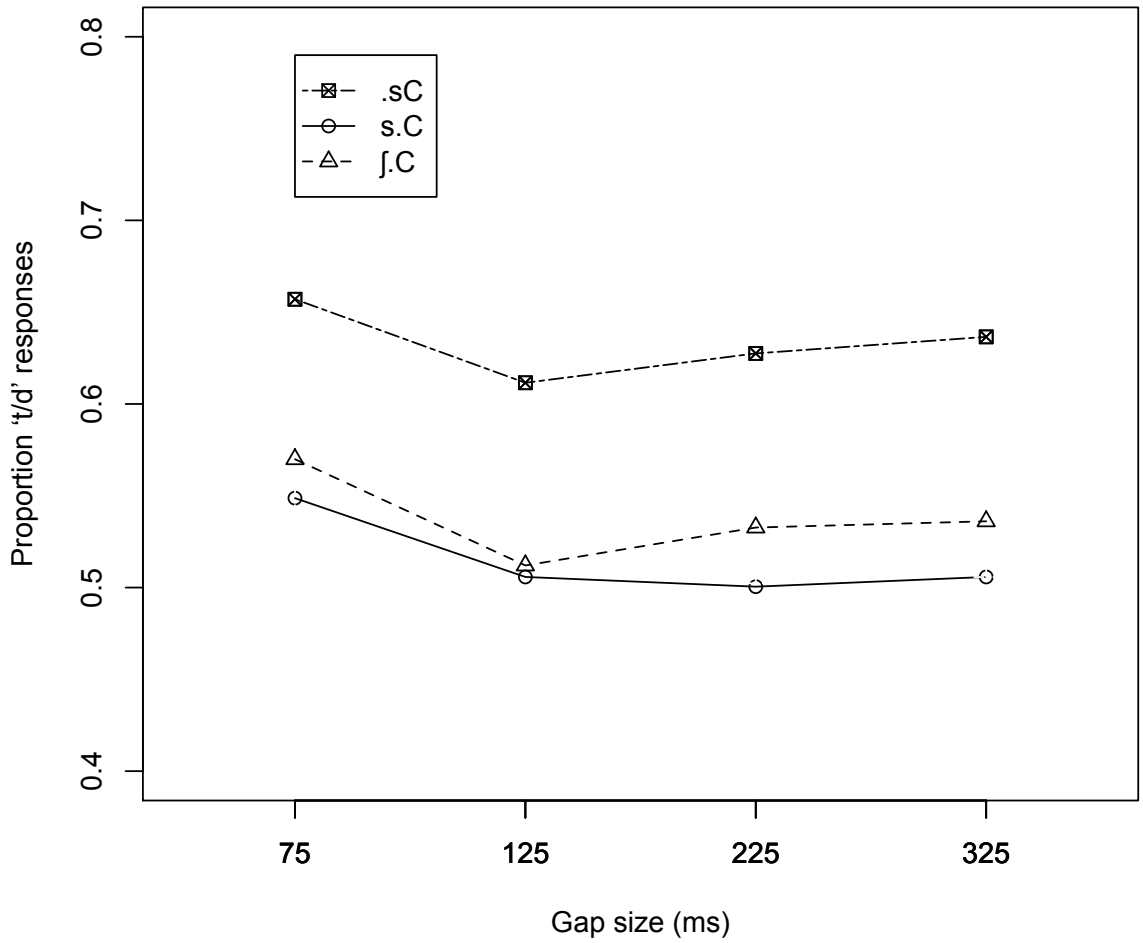

Figure 4.4: Pooled responses of stimuli for fricative/syllabification conditions.

More interestingly, the difference between the responses of the two fricatives, which was large and highly significant (response × fricative, $F = 4.354$, $p < 0.05$), remained constant, even across all gap sizes (response × gap size × fricative, $F = 0.297$, $p = 0.73$). For example, at 325 ms, the difference in proportion of [t] responses was still significantly different (response × fricative, $F = 18.54$, $p < 0.001$; an expanded curve can be seen in Figure B.2 in Appendix B). This implies that whatever effects that /s/ and /ʃ/ may have exerted on the following ambiguous vowel were of the same magnitude (and also in the same direction, as just previously discussed). This is unlike Mann and Repp's original study (1981) in which the difference in response between the two fricatives narrowed after 75 ms, which

suggested that the compensation for coarticulation effect was decreasing with increased separation. In contrast, our results show that the perceptual grouping caused by the Ganong effect seems to *already* separate the fricative from the consonant very distinctly even at small gap sizes.

Regarding the two syllabification conditions with /s/, though there was a clear difference in the proportion of [t] responses when the syllable boundary was placed before or after the fricative (response × fricative, $F = 122.6$, $p < 0.001$), this difference remained constant across all gap sizes (response × gap size × fricative, $F = 0.358$, $p = 0.784$). This is similar to what was found when the fricative was varied; again, the Ganong effect appears to be surprisingly strong. Listeners do not, at all, hear the stimuli as two nonwords with increasing temporal separation (Figure B.1, as discussed above in the context of Figure 4.3).

Therefore, as expected, there was little effect of gap size on the listeners' classifications of the following ambiguous stop for either /s/ or /ʃ/ when the fricative and stop were already perceptually separated. However, we also found, contrary to expectations, that there was *also* little effect for gap size even when the fricative and stop were initially perceptually grouped together. Participants, when asked, reported that they were expecting to hear the words and that the extended silence duration felt more like a strange emphasis on the stop than two separate words. This surprising percept might have perhaps been due to the participants having previously been trained to hear real words. Since English does not have any word /əs/ (given that the word *us* tends to be realized with /ʌ/ being farther back and lower in the vowel space), participants could have been primed, in a sense, to hear a word beginning with /sC/, especially since they already knew that *stone* and *scone* were words in the experiment, which would have then caused unintended fricative-stop perceptual grouping.

5 Experiment 3: Vowel Context

5.1 Aims

Experiment 3 is a follow up to Experiment 1. In that previous experiment, we found significant results; however, a possible confounding factor can be found in the different vowels of *ace dawn–ace gone* versus *a stone–a scone*. As seen in Mann and Repp (1980), vowels have an influence on the segments preceding them. Their study considered ambiguous fricatives between [s] and [ʃ] directly before a vowel (FV sequences), and results showed that there was a strong bias towards /s/ before rounded vowels whereas /ʃ/ was preferred before their unrounded variants. A possible explanation for this can be easily found in the theory of spectral contrast: rounded vowels, given their longer vocal tracts, produce formants that are lower in frequency. Since [s] has a higher center of gravity than [ʃ], it stands to reason that listeners who hear an ambiguous fricative with frequencies higher than the formants of the vowel will classify the fricative as being /s/ rather than /ʃ/.

This effect could extend to stops before vowels as well, which would imply that listeners' judgements of the ambiguous stops are affected both by the preceding fricative and the following vowel. Since the fricative's influence on the stop is modulated by the syllable boundary in the stimulus, the task becomes one of detangling the effects of syllabification from the effects of vowel rounding.

To accomplish this, we attempted to find words of comparable length that could form a stimulus with either [əs.to^{w}]–[əs.ko^{w}] or [e^{j}.stɑ]–[e^{j}.skɑ] that could not be syllabified in the other position. This was a task of monumental proportions that was not feasible. Instead, nonword stimuli are used: *astof–askof*, with the final [n] of the *a stone–a scone* word pairs being changed to an [f]. As Mann and Repp (1981) had already shown that there was no significant change in the results caused by instructing the listeners to differentially syllabify *nonword* stimuli, the location where the listeners place the syllable boundary in the new stimuli should be of little import and the same response should be obtained. Nevertheless, this experiment attempts to persuade listeners to place the syllable

boundary between the fricative and the stop to create a clear contrast between the new stimuli and *a stone–a scone* previously.

If the following vowel has no effect on the place of articulation perception of the stop, the categorical boundary should trace that of *ace dawn–ace gone*. If, however, the significant change ascribed to syllabification as seen in Experiment 1 is actually entirely due to the influence of a rounded vowel, the categorical boundary should trace that of *a stone–a scone*. Yet the strength of Ganong-type lexical effects, as testified by previous studies and as seen in the surprising *a stone* response bias of Experiment 1 as well as in gap size effects (or lack thereof) under varying fricative and syllabification conditions in Experiment 2, suggests that this explanation cannot be easily discarded or overridden by some other phenomenon. We thus expect the results of this present experiment to fall in between the two extremes.

5.2 Methods

5.2.1 Participants

Once again, participants from Experiment 1 and Experiment 2 were asked to take part in this study. In total, there were 19 native speakers of American English (14 female, 5 male) who acted as listeners. None of the participants reported past or present speech or hearing disorders.

5.2.2 Stimuli

The *a stone, a scone* stimuli from Experiment 1 were modified for this experiment. The final [n] was replaced by an [f] and the stimuli recreated from scratch with the Klatt synthesizer (Klatt, 1980). The length of [f] was increased by 50 ms as early consultants reported that the sound was too faint and short to be easily noticeable. Hence, the stimuli for this experiment consisted of [ə] for 200 ms, followed by [s] for 200 ms, silence for 75 ms, the onset stop burst for 40 ms, the vowel [o] for 320 ms (with a 20 ms overlap between burst and vowel) and then [f] for 160 ms, for a total stimuli length of 975 ms. Spectral characteristics of [f] were taken and modified from (Wilde, 1995): the fricative was synthesized with noise dispersed between 1000–6500 Hz, with peak intensity 42 dB. The spectrogram of endpoint stimuli [əsto^{w}f] and [əsko^{w}f] are shown in Figure 5.1.

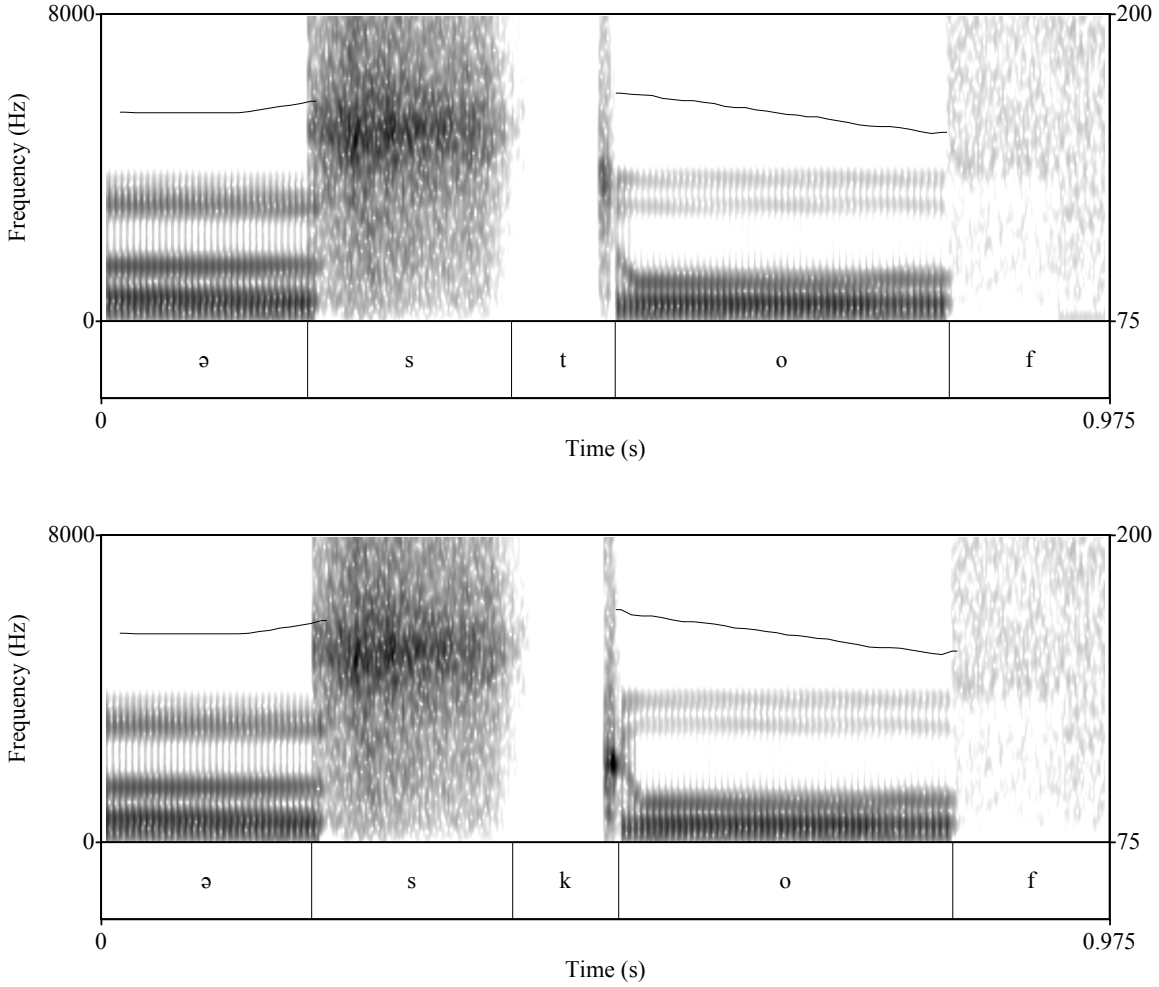

Figure 5.1: Spectrogram and pitch trace of endpoint stimuli [əsto^{w}f] and [əsko^{w}f].

5.2.3 Procedure

The procedure was similar to that used in Experiments 1 and 2. There were two phases: a practice phase and an experimental phase. The practice phase was intended to be a mix between previous 'teaching' and 'practice' phases. 6 tokens of the second half of a stimulus ([to^{w}f] or [ko^{w}f]) were used: two each of the endpoint stimulus (steps 1 and 9), one from step 3 and one from step 7. These six tokens were randomly presented. Following this, another set of twelve tokens were randomly presented: the six previous tokens, 1 each of two-word tokens of original gap size 75 ms from steps 1,

3, 7 and 9, as well as the endpoint two-word tokens of gap size 225 ms. In total, participants heard 18 tokens in this phase. Participants were told that the stimuli would be nonwords: either "uhs doef" or "uhs goef". Instead of seeing categories "t/d" and "k/g", they were instructed to categorize the utterances as "d" or "g". These differences from the previous experiments, coupled with the additional tokens consisting of only the second half of the stimulus, were meant to psychologically reinforce the syllable boundary between fricative and stop, much like in Mann and Repp's (1981) study.

The experimental phase was conducted in one block. Stimuli from steps 2–9, with gap sizes 75 and 225 ms, were presented; again, step 1 of each continuum was discarded as the response difference between steps 1 and 2 was insignificant. This set of stimuli was repeated five times in random order, with 8 steps × 2 gap sizes × 5 repetitions yielding 80 tokens judged by the participants in this experiment.

5.3 Results

Means of the responses for each fricative and syllabification condition were plotted and data analyzed with ANOVAs.

The categorical boundary curves for all three conditions: /.sC/, /s.C/ and nonword at a gap size of 225 ms are shown in Figure 5.2 on page 35. The figure shows that at 225 ms, the longer gap size, the categorical boundary for the nonwords, which listeners were encouraged to hear under /s.C/ syllabification, in fact appears to have traced the curve for /.sC/ real words (response × lexical status /.sC/ × step, $F = 1.447$, $p = 0.182$). Statistically, the curve for nonwords was also significantly different from that of /s.C/ real words (response × lexical status /s.C/ × step, $F = 14.99$, $p < 0.001$). The similarity of the nonword curve to the /.sC/ real word curve seems to indicate that the original Experiment 1 results were entirely due to vowel rounding affecting the stop percept rather than the placement of the syllable boundary, as previously explained in the aims of this experiment.

However, for the shorter gap size of 75 ms (Figure 5.3 on page 36), the perceptual boundary for nonwords fell neatly between those of /s.C/ and /.sC/ real words. The nonword curve was significantly different both from the /s.C/ real word curve (response × lexical status × step, $F = 4.642$, $p < 0.001$) and from the /.sC/ real word curve as well (response × lexical status × step, $F = 5.611$, $p < 0.001$).

Close inspection of the individual boundary curves reveals that the nonword curve traces the

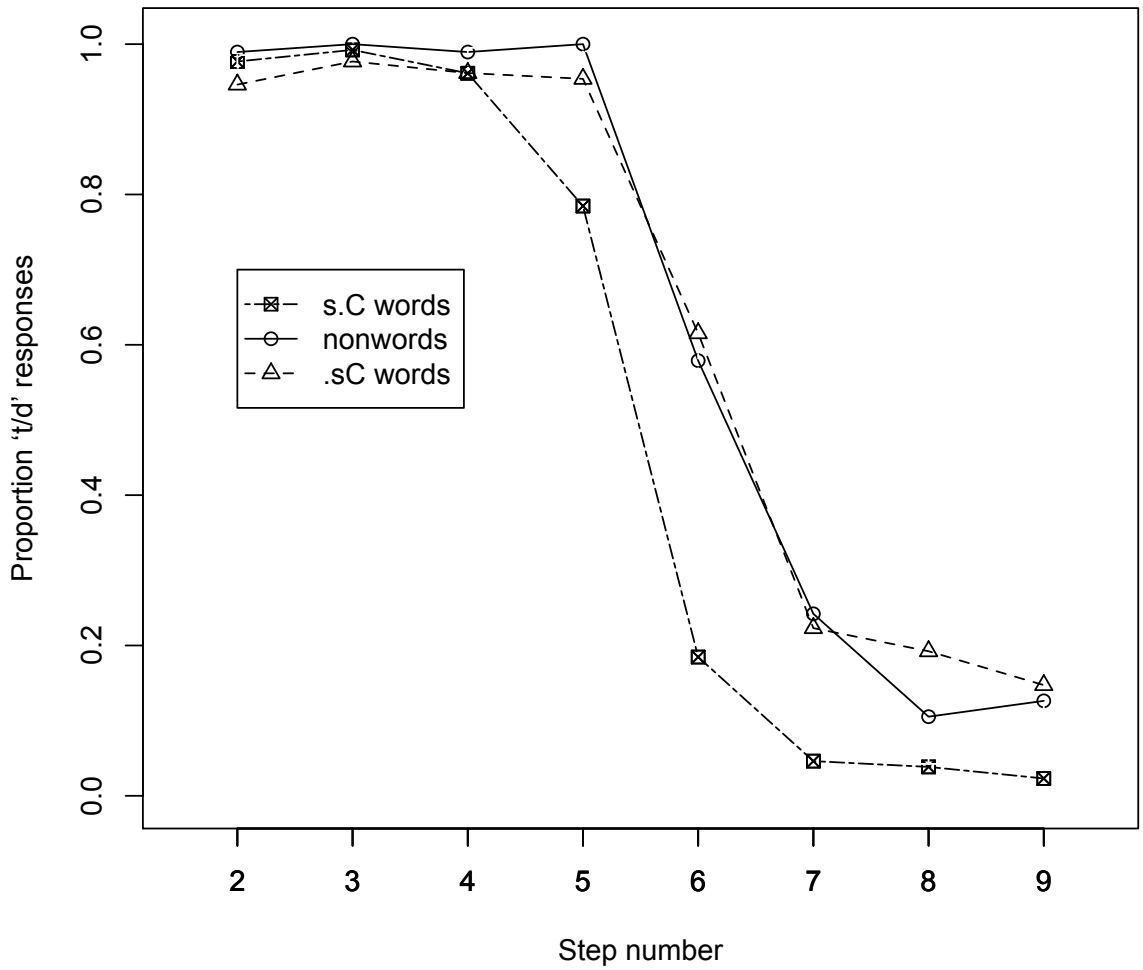

Figure 5.2: Categorical boundaries of /s.C/, nonwords and /.sC/ at 225 ms.

/.sC/ real word curve until step 5, after which the proportion of [t] responses drops dramatically, almost to the level of /s.C/ real words. By step 7 of the continuum, the [t] responses of the nonwords are even lower than their /s.C/ equivalents. This shows that close to the [t] end of the continuum, the nonwords are parsed much like *a stone–a scone*, whereas at the [k] end, they are perceived more like *ace dawn–ace gone*.

Although this mixed result, which appears to fall in the middle of real-word /.sC/ and /s.C/ curves, was expected, what roles do syllabification and vowel roundedness play, respectively, and how do these interact with gap size to produce the differing behavior found? The key notable finding of this experiment is that the classification of the following stop changes significantly with gap size for

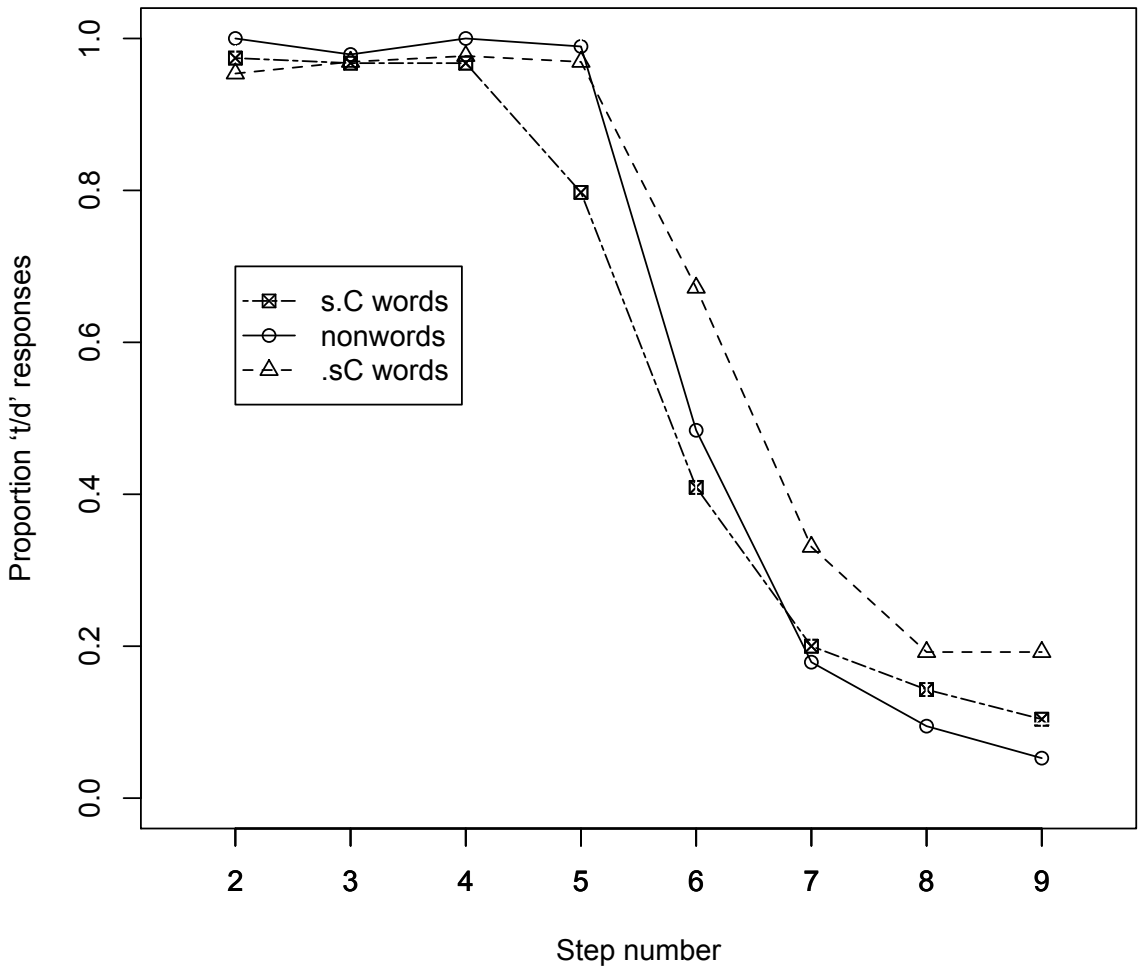

Figure 5.3: Categorical boundaries of /s.C/, nonwords and /.sC/ at 75 ms.

nonwords (response × gap size, $F = 5.915$, $p < 0.05$), but in the *opposite* direction as the response change found for the /s.C/ and /.sC/ real word conditions in the previous two experiments. The pooled results show this rather dramatically (Figure 5.4 on page 37).

This suggests that the nonword criteria may perhaps be showing a pure vowel-on-stop effect isolated from any other phenomenon, especially given that it is uncertain how listeners actually syllabified the nonword stimuli. Thus, perceptual boundaries must be mediated by both syllabification and vowel context. At 75 ms, the contribution of the two factors are almost equal. For the /s.C/ condition, the decreased perceptual grouping between fricative and stop enforced by the intervening syllable boundary biases the listener to hear more [k] percepts to compensate for coarticulation.

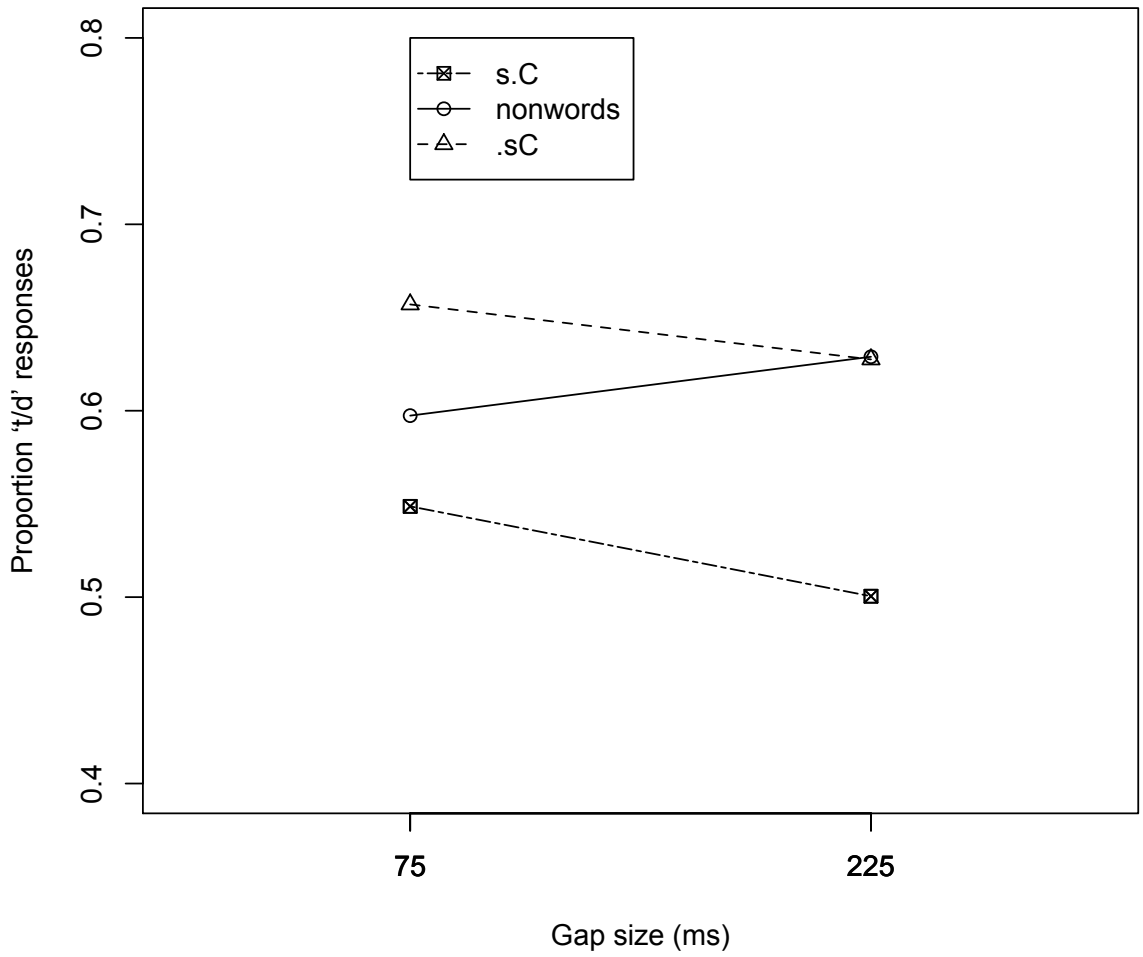

Figure 5.4: Pooled responses of stimuli for fricative/lexical status conditions.

The nonword stimuli show a neutral syllabification effect, in that the fricative and stop are neither separated, as would be in the /s.C/ condition, nor bound together, as would be in the /.sC/ condition, but produce a greater percentage of [t] responses, which must then be due to the effect of vowel rounding. Further up the scale, the /.sC/ condition yields the highest number of [t] responses, being a combination of the tendency towards [t] with vowel rounding *and* the lexical frequency bias caused by Ganong-enforced fricative-stop perceptual grouping.

At 225 ms, however, the response behavior of /.sC/ could be entirely accounted for by the vowel rounding effect, which strengthens as gap size increases. It is not readily discernible why the vowel rounding effect would change proportionally with temporal separation; however, a possible

explanation involves the nonword stimuli themselves, which were derived from the *a stone–a scone* stimuli. It is not inconceivable that the listeners might realize this and thus be primed to place the syllable boundary before the stop, despite instructions to the contrary; much like in Experiment 2. Their experience from the previous experiment might then have produced a phantom lexical frequency effect triggered by subconscious memory of the "emphasized" stop; section 6.2 provides an in-depth discussion of the consequences of this perceived extended stop emphasis.

If this account of the effects of gap size on the nonword stimuli is true, then the response behavior of /.sC/ must also be a combination of syllabification and vowel effects, much like that of the 75 ms condition. In addition, the strong lexical frequency bias arising from the Ganong effect probably has a large influence on listeners' responses as well. We must thus surmise that, in general, the respective contributions of syllabification and vowel context are difficult to detangle, especially when lexical effects are also involved.

6 General Discussion

6.1 Recapitulation

After an extensive review of the literature, we suggested that the original Mann and Repp study (1981), in which was studied the effect of the fricative, the fricative-stop temporal separation, and the syllabification condition on the perception of an ambiguous stop in /VFCV/ stimuli, could be modified by enforcing syllable boundaries with real words instead of mere instructions. In particular, the Ganong effect (Ganong, 1980) and previous investigations into perceptual grouping (Samuel & Pitt, 2003) suggested that compensation for coarticulation was highly permeable to lexical effects (Elman & McClelland, 1988). The present study thus used real words to investigate compensation for coarticulation under varying syllabification and temporal separation conditions.

The first experiment studied syllabification effects. Three pairs of stimuli were created: *ace dawn, ace gone*: [eʲs.tɑn]–[eʲs.kɑn]; *ash dawn, ash gone*: [æʃ.tɑn]–[æʃ.kɑn]; and *a stone, a scone*: [ə.stoʷn]–[ə.skoʷn]. A nine-step continuum was created between the alveolar and velar stops for each pair. Contrary to Mann and Repp's 1981 findings, percent alveolar response did not increase for /ʃ/: instead, the response was simply less categorical. Comparing /s.C/ to /.sC/ conditions, more [k] responses were expected for /.sC/ due to its tight affiliation and hence greater compensation, but the opposite effect was found: the perceptual boundary moved to the right, i.e. more [t] responses. This was hypothesized to be due to the higher frequency of the word *stone* in the English language compared to *scone*; that is, listeners were not only biased towards real words, but also towards higher-frequency lexical items.

The second experiment studied temporal separation effects. Gap sizes ranged from 75 to 325 ms for all three pairs of stimuli. It was expected that larger gap size would reduce the magnitude of compensation effects, particularly for the /.sC/ condition as the word association effects would disappear. The results replicated Mann and Repp's 1981 findings for /s.C/ and /ʃ.C/ conditions,

with the [t] response rate lowering from 75 to 125 ms and staying constant thereafter, though the magnitude of difference between the conditions itself did not change significantly. The /.sC/ condition, on the other hand, showed no effect for gap size at all: it produced percent alveolar responses consistently higher than those of /s.C/ or /ʃ.C/ conditions throughout. It appears that the word recognition Ganong phenomenon is so strong that it overcomes weak compensation effects.

The third experiment was designed to separate the effects of syllabification from the potential confounding factor of vowel roundedness, whose effect on preceding ambiguous fricatives, but not stops, had been previously characterized by Mann and Repp in 1980. A stimuli continuum *astof–askof* [əsto^{w}f]–[əsko^{w}f] was created and experiment instructions attempted to persuade listeners to perceptually separate the fricative and the stop into two different syllables. Two gap sizes were studied: 75 and 225 ms. At the longer gap size, the perceptual boundary for the nonwords followed the curve for /.sC/ real words, seemingly indicating that the original Experiment 1 results were entirely due to vowel rounding affecting the stop percept rather than syllabification. However, for the shorter gap size of 75 ms, the perceptual boundary for nonwords fell neatly between those of /s.C/ and /.sC/ real words. This suggests that perceptual boundaries are mediated by both syllabification and vowel context.

6.2 Some contentious issues

6.2.1 The lexical frequency effect

In the first experiment, the surprising result that the boundary curve shifted towards [t] under /.sC/ conditions, instead of [k] as expected, was explained by a lexical bias created by the much higher frequency of the word *stone* compared to *scone*. This then implies that this lexical frequency effect *overrides* compensation for coarticulation in rather large significance. But is this explanation at all valid? Furthermore, in this experiment, the lexical effect does not simply compensate for compensation for coarticulation, which would yield a null result as both forces cancel one another. Instead, can the effect be so strong as to also cause a change in the opposite direction?

In fact, evidence from the literature supports this hypothesis. First, to examine the validity of the lexical effect, we see that Ganong himself examined his data from the **tash–dash* and *task-*dask* continua, cross-referencing it to "the sum of the logarithm of the frequency of occurrence" and finding that the correlation coefficient was "not reliably different from zero" (Ganong, 1980). Yet

he cautioned that this result may be of little import as the experiment had not been specifically designed to investigate word frequency effects. Indeed, in 1984, Fox studied a word-word continuum: *bad–dad* [bæd]–[dæd], and found that its categorical boundary curve showed a bias towards /b/ as compared to that of a neutral nonword-nonword continuum [bæ]–[dæ]. This matches lexical frequency data showing that *bad* is more frequent than *dad* (Carroll, Davies, & Richman, 1971). It is likely that word-medial stops are affected in the same way as well, which would extend this result to our *stone–scone* stimuli.

Next, building on Mann and Repp's study (1980) of vowel-on-fricative coarticulation while varying the roundedness of the vowel, Nguyen juxtaposed Ganong-type lexical effects onto this set of coarticulatory biases (Nguyen, 2001); that is, each effect would produce opposite results, thus pitting them directly against each other to see which was the stronger effect. As mentioned before, Mann and Repp showed that the probability of /s/ response increases before rounded vowels than that of /ʃ/, and vice versa for unrounded vowels. Nguyen constructed a paradigm of word-nonword pairs with ambiguous fricatives in French: *choucroute–*soucroute* [ʃukʀut–sukʀut] and **chalier–salier* [ʃalje–salje], in which *choucroute* and *salier* are real words while their counterparts on the opposite end of the fricative continuum are not. If vowel-on-fricative coarticulation were the stronger effect, listeners would hear **soucroute* and **chalier*; however, Nguyen hypothesized that the Ganong effect would take precedence and that instead, listeners would prefer the real words. This hypothesis was found to be true: French listeners showed a significant bias for real words despite the rounded and unrounded vowel conditions being reversed. In effect, Nguyen showed that the Ganong effect could override the smaller, sublexical effect of compensation for coarticulation. Accordingly, if the Ganong effect can do this, it stands to reason that word frequency, another lexical-level effect, could produce a similar consequence as well.

These two studies thus lend credence to the hypothesis of the present study. We can conclude that the words *stone* and *scone* are preferred differently by listeners based on their respective frequencies in the language, and that this lexical-level bias can indeed be much stronger than sublexical compensation for coarticulation.

6.2.2 Coarticulation and gap size

A recurring effect found in the experimental data is that an increased gap size between the fricative and the stop in /.sC/ stimuli, that is, in stimuli that are perceptually grouped together due to lexical

status, results in higher percent alveolar response than in /s.C/ stimuli, whereas the hypotheses of the perceptual grouping account (Samuel & Pitt, 2003) would predict that the less tightly bound cluster would reduce the Ganong-enforced lexical frequency bias towards [t] and therefore result in fewer [t] percepts. In Experiment 2, this tendency was manifest in the null result for /.sC/ stimuli when interacted with gap size: the categorical boundary curves did not shift significantly even though their /s.C/ and /ʃ.C/ counterparts showed a decrease in [t] response. In Experiment 3, we found a straightforward increase in percent alveolar response for /sC/ nonwords (with uncertain syllabification, although listeners were encouraged to parse the nonword stimuli as /s.C/) compared to /s.C/ real words; in fact, the nonword categorical boundary curve traced the /.sC/ real word curve at a gap size of 225 ms.

If the curves are moving in the opposite direction of what was expected, there must be a reason for this. One possible explanation that was briefly touched upon in the results sections of Experiments 2 and 3 involves a priming or memory effect. In Experiment 2, the lack of movement of the boundary as gap size increased suggests that listeners continued to group the fricative and stop together to hear the real words *stone* and *scone*. How much of this is due to a strong Ganong effect versus a priming effect is unclear. Priming has been shown to be an extremely resilient phenomenon, sometimes even reversing lexical frequency effects if the lexeme of low-frequency were presented often enough before the judging task (Scarborough, Cortese, & Scarborough, 1977). Yet such a strong reversing bias is not necessary to explain our data as we did not attempt to influence lexical frequency; we only require the simple memory portion of the priming effect. Considering that English has no words /don/ or /gon/, it follows that listeners would be strongly primed to hear real words, especially words that they have heard before and that they remember judging. Hence listeners would treat the large-gap stimuli as if they were normal, small-gap *a stone–a scone* stimuli, leading to the same response behavior in the classification of the following ambiguous stop.

The same reasoning can be applied to the results of Experiment 3 for the nonword stimuli. Even though the stimuli were not real words, only the last phoneme was changed to create them; the first four-fifths of the stimuli remained the same. It was found that in the visual mode, when subjects were first shown an incomplete sketch and then later shown more of the picture until it was recognized, they later identified the same sketch much earlier than they previously did (Kolb & Whishaw, 2009). Our auditory stimuli could have paralleled this finding: with the subjects already primed from hearing *stone* and *scone* in the earlier experiments, the simple sequence /sC/, which was followed by the same vowel, could have led to an association with the real word continuum and

thus a similar response behavior.

However, this explanation is not enough to explain all the variation in Experiment 3. We would expect that the priming effect would apply equally to small or large gap sizes. Yet the Experiment 3 results show that priming appears to be less effective at 75 ms, as the curve quickly falls away from /.sC/ to /s.C/. Alternatively, this could be explained by another effect that specifically targets large gap sizes, that invokes low-level on-line speech perception instead of off-line memory echoes.

This possibility reduces the role of lexical-level Ganong effects and instead posits that the increased [t] response at long gap sizes is an artifact of speech perception peculiarities. It relies on the perception, as testified by some subjects, that increased temporal separation lengthened the anticipation of the following stop, which would have then caused compensation for coarticulation in the opposite direction. As the listeners hear an [s], a motor theory account of speech perception (Liberman & Mattingly, 1985) or a gesture perception account (Fowler, 2006) would posit that they perceive the tongue position at the alveolar ridge. In the absence of other cues, listeners assume that the tongue stays in the same position, leading to the extended stop being perceived at that place of articulation as well.

Indeed, a quick electropalatography investigation lends credence to this assumption. A male native speaker of American English, who had not taken part in any of the prior experiments and who was unaware of previous results or experiment hypotheses, was asked to produce the phrases *ace dawn* and *a stone*, first quickly, and then exaggeratedly slowly as though his speech were being played back at a slower speed. The electropalatography system WinEPG from Articulate Instruments was then used to record his articulatory movements. Figure 6.1 shows the averaged electropalatograms during the segment [t] under both conditions. As is immediately obvious, the tongue remains in contact with the alveolar ridge for *a stone* whereas it pulls away for *ace dawn*. Figures B.3 and B.4 (in Appendix B) show a larger portion of the time evolution of the speaker's gestures, which demonstrate that for *ace dawn*, alveolar contact is near zero during the gap of silence, and for *a stone*, not only is there tongue contact during the stop in *a stone*, but tongue contact also remains throughout the entire duration of [t].

Spectrally (Lotto & Kluender, 1998), movement of the tongue closure to the velum would have been accompanied by auditory flux; specifically, the center of gravity of frication should have decreased as is typical for frication arising from the places of articulation further back in the mouth. In contrast, a flat spectrum would be expected if the tongue remains in the same position. Given that [s] was synthesized as a steady-state fricative, it is unsurprising that, by both the gesture and

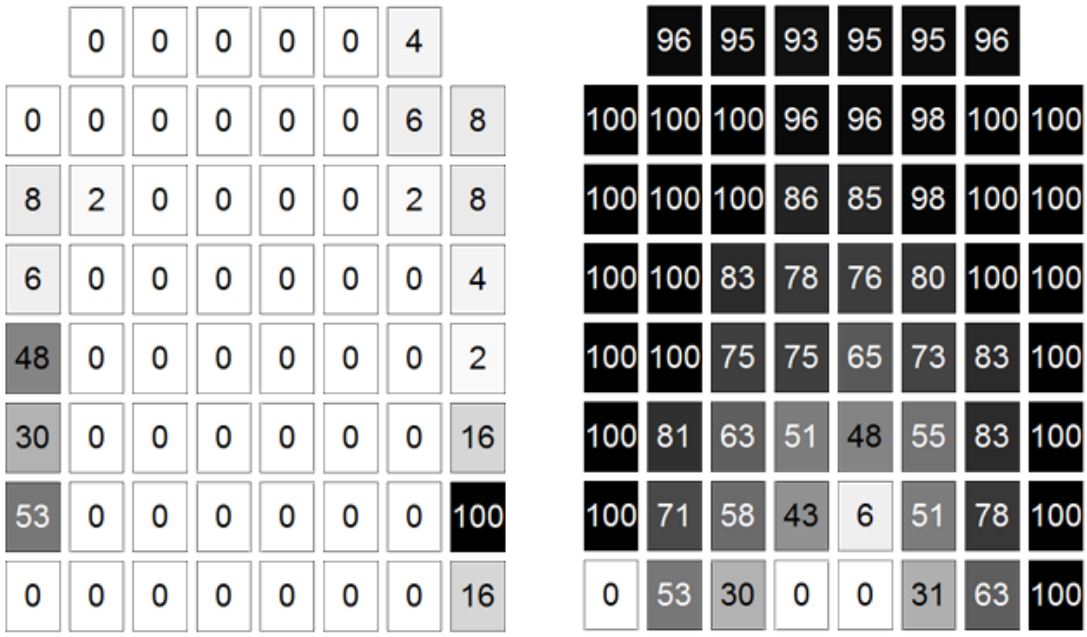

Figure 6.1: Averaged electropalatograms during the gap of silence of *ace dawn* (left) and *a stone* (right). Maximal contact of a cell across all frames is marked with "100"; no contact is marked with "0".

spectral accounts, listeners were more likely to hear the stop as [t] initially, with extended silence only confirming this percept as there is nothing to immediately contradict it.

Evidently, it is difficult to pinpoint exactly which explanation is correct, especially since the experiments of the current investigation were not designed to answer these questions. However, a number of generalizations can be drawn from the data. First, the Ganong effect is stronger than expected and also interacts with lexical frequency effects. Second, syllabification *does* affect compensation for coarticulation, in line with expectations from previous literature on perceptual grouping. Third, vowel context and auditory speech perception artifacts also contribute to listeners' response behavior. These generalizations provide insights into the ongoing debate on the nature of speech perception, as was briefly mentioned in the literature review for this study (Chapter 2).

6.3 Implications for models of speech perception

When does lexical knowledge get inserted into the perception process? Most models of speech perception are split by this question and fall into one of two categories: those that posit the existence of top-down effects; that is, that there are lexical effects on sublexical perception, and those that assert that perception is entirely bottom-up, with any higher-level effects taking place in post-perceptual processing. Examples of successful and influential models on both sides of the question are described below.

In Merge (Norris, 1994), lexical effects are post-perceptual rather than pre-perceptual. Norris contends that the streams of lexical knowledge and auditory information are separate. Only after the output of auditory processing has been generated is it then merged with lexical knowledge, after which a final decision is made. In this model, there is no overlap from one system to another: that is, lexical information cannot change the fundamental output of the auditory perception system; it cannot encourage or discourage the perception of specific segments. It is a wholly feedforward model of speech perception. Indeed, Norris, McQueen, and Cutler (2003) contend that perhaps the strongest argument for Merge is precisely its lack of feedback, as due to perceptual learning taking place only off-line, "feedback could not possibly help in recognizing words": it would thus be more parsimonious to exclude such a feature in a model of speech perception.

In contrast, there are two prominent theories that oppose this view. One is TRACE, developed by Elman and McClelland (1986); the other is Adaptive Resonance Theory (Grossberg, 1980). TRACE posits an interactive process between the auditory perception and lexical knowledge levels. The activation of a unit at any fixed level (phonemic, morphological, lexical, et cetera) causes the activation of units at other levels that correspond to the possibilities at those levels *given* the initially activated unit. This activation can proceed in any direction. For example, the fragment *exting…* can activate, at the lexical level, the whole word *extinguish*, which then in turn increases activation of a phonemic-level final /ʃ/ (Samuel, 2011). Adaptive Resonance Theory (ART), a general cognitive processing model, uses the same principles of cross-level activation when applied to speech perception (Loritz, 1991); however, instead of having fixed levels like TRACE, ART uses "chunks" of varying sizes that the listener hears: they may be as small as fleeting phonetic transitions or as large as multi-word phrases. This model avoids the problems associated with positing the existence of specific units such as phonemes that appear from time to time. But disregarding the smaller details of the models at present, in general, the distinguishing feature of TRACE and ART is that they are

models that possess both feedforward *and* feedback processes.

Given the large mass of literature that shows indirect effects of top-down processing, such as lexeme-driven compensation for coarticulation, it seems as though models that incorporate feedback into on-line processing must be valid. Consider Elman and McClelland's (1988) study using *Christmas–*Christmash* and **foolis–foolish* continua. Since the subjects were not asked to identify the final sound of the first word, it stands to reason that higher-level decision processing remains uninvolved. Instead, if the first word had any effect on the categorical boundary of the initial sound of the second word, this must have been due to some form of lexical recognition exerting an effect on subsequent auditory information; that is to say that feedback from lexical knowledge influenced on-line perceptual processing.

However, Norris et al. (2003) present an argument against this view. In support of Merge, they explain that lexeme-driven compensation for coarticulation could be, in fact, a result of post-perceptual processing. They contend that the higher transitional probabilities of /əs/ and /ɪʃ/ sequences account for the percept of /s/ in *Christmas* and /ʃ/ in *foolish* without lexical input. Purely bottom-up models of speech perception such as Merge learn, through repeated exemplars, that [t] after [s] is not the same as [t] after [ʃ]; listeners then use these specific cues to create the compensation for coarticulation effect at the post-perceptual level. In this manner, Merge keeps the sublexical and lexical stages of processing separate by assigning the former to the perception stage and the latter to the decision-making stage.

Our results, on the other hand, cannot easily be reconciled with the two-stage process of bottom-up models such as Merge. Our key finding, which produced results that differed from the original Mann and Repp 1981 study, is that the Ganong effect is much stronger than expected: it was able to enforce syllabification of the stimuli in the perception of the listeners. This then also enforced perceptual grouping, which affected the magnitude of compensation for coarticulation in the following stop. Previous studies had investigated one of these two phenomena, but never both at once.

Under Merge, both phenomena operate at the decision-making stage: that is, listeners hear the stimuli first before then making a decision on two issues: where the syllable boundary should be placed, and what place of articulation to assign the stop to. But which comes before the other? If the place of articulation decision were made first, then we would expect the same categorical boundary curves for both /s.C/ and /.sC/ stimuli, given that syllabification has not yet played a role. But this implies that the percept would then be subsequently changed by the application of lexical effects resulting from the placement of the syllable boundary; for example, a [k] percept would be passed

on by the sublexical level of processing, remain a [k] based on previous experience with /sk/ clusters, but then changed to [t] by the recognition than *stone* is more common than *scone* in the English lexicon. Given the strength of the lexical frequency effect (the categorical boundary was shifted an entire continuum step for /.sC/ versus /s.C/ stimuli), it is difficult to believe that it applies only at the end of a long chain of processing.

Yet if the decision of syllabification were made first, we would then expect no difference in response behavior across gap sizes in Experiment 2, as two words that are separated by 75 ms of silence should be no different than two words separated by 325 ms of silence. It is true that for /s.C/ and /ʃ.C/ real words, there is no significant change after 125 ms, but the drop in percent alveolar response from 75 to 125 ms must be accounted for. Furthermore, the mixed behavior of nonword stimuli from Experiment 3 would be highly irregular. The fact that the percent alveolar response is in between the /s.C/ and /.sC/ real word curves can be explained by variation in the way listeners syllabify the stimuli. However, we would then expect this response to be in the same direction as the real word responses as temporal separation increases, instead of seeing the change in the opposite direction as observed in Figure 5.4. We should also not have found that the nonword stimuli syllabified entirely like /.sC/ at a larger gap size when the response was between the two syllabification conditions at 75 ms, as there should have been no reason for any further lexical processing applied to the nonword stimuli. A caveat here is that the auditory processing explanation of the [t] bias at increased temporal separation, as described in the previous section, could be a working possibility for Merge, since this explanation posits that the preference for [t] arises from sublexical cues that result from the perception of the lengthened stop.

Still, it is unlikely that this effect could alone account for the exact tracing of the /.sC/ real word curve that the nonword curve exhibited. Therefore, a review of this paradigm of results across the three experiments reveals that the two decisions of syllable boundary and stop place of articulation must be dependent on one another to a certain extent, especially when adding the novel effect of lexical frequency on compensation for coarticulation that has not been previously considered.

In contrast to Merge, top-down models such as TRACE and ART easily incorporate all the effects discussed, with no particular requirement for the order in which they are applied. Instead, as these effects are activated, they filter the probabilities of auditory units of any size at the sublexical level on an on-line basis, thereby directly affecting low-level perception. This comfortably accounts for the results that were found in our experiments. A model like TRACE posits that as the stimuli is presented, segment by segment, possibilities are continuously being reviewed and subsequently

discarded or promoted. This allows any number of effects to be active and does not oblige any of them to *always* be active, such that certain effects can appear or be prominent in certain stimuli and not in others.

Let us consider the perception of a particular nonword stimulus at a gap size of 225 ms in Experiment 3, that lies near the middle of the continuum but whose stop will eventually be identified as [t]. As the phonemes /ə/ and /s/ are perceived, the Ganong effect deactivates the possibility that the stimuli is syllabified after the fricative. At the same time, the priming effect from the listeners' memories of previous trials activates the words *stone*, *scone* and other words beginning with /s/. The lexical frequency effect further variably modulate the activation of those words. As the stimuli continues to be presented, with the listener hearing the ambiguous stop and /o/, the other words beginning with /s/ are discarded. Since the word *stone* is much more activated than *scone*, the listener expects to hear [t] more often, and as a result, the cues of the highly ambiguous stop, which are not extremely different enough to be classified as [k], are interpreted as probably belonging to [t] instead; therefore *scone* becomes less activated. Consequently, [n] is also activated and the listener expects to hear it next. When an [f] is heard instead, since the [t] percept for the ambiguous stop is already so much more activated than [k], the listener reports [t], as though he had been hearing *stone* rather than the nonwords **stof* or **skof* or their counterparts if the syllable boundary had been placed after the fricative, a possibility which was deactivated early in the process. This is exactly what TRACE and ART predict through their cyclical, continual on-line processes in both directions, using feedforward and feedback mechanisms to whittle away the possibilities.

This example, simply one of the many stimuli that were presented in our experiments, illustrates how the bidirectional activation systems of models like TRACE and ART in fact provide for more flexibility in terms of the effects that may influence the perception of an ambiguous stop. One of the criticisms leveled by Norris et al. (2003) against models of speech perception that include top-down feedback is that said models are not parsimonious, as the consistent flow of information is extremely loose and unstructured. Yet from the scrutiny of both models applied to the experimental data, it seems that it is Merge that is more unstructured. It must be acknowledged that the linear information flow of Merge is perhaps more parsimonious when there are few effects to be considered, as there are fewer decisions made and the percept choice filters through them very easily. However, when more effects are added and when their order appears to be fluid and sometime simultaneous, a model like Merge becomes extremely disorderly as it generates a huge number of available choices at post-perceptual processing, given the many decisions that have to be made. From a single sublexical

percept are derived an overwhelming number of perceptual possibilities that must then be narrowed into one final judgement.

On the contrary, TRACE and ART propose a process that has less redundancy. With the addition of a feedback mechanism, the constant refining of the perceptual signal creates a system in which information is copious at the very bottom, but is gradually reduced through the activation and deactivation cycles. When the model must deal with a large number of effects, only as many of them that are necessary become activated and feed into the model. This appears to be a much more parsimonious model of speech perception. Processing capability is conserved by integrating the various effects at as low a level as possible; that is, sublexically. While Merge posits that all processing occurs post-perceptually when the various streams of information are integrated at once, TRACE spreads out the cognitive load early in the perceptual process and thus avoids the chaos near the point of judgement that may occur with Merge.

In the case of the present study, we consider a great number of effects that influence the percept of one single unit, the ambiguous stop. The effects are not always present, nor do they always operate at the same strength. Accordingly, given the varied outcomes of these effects on listeners' judgements, we infer that our experimental data is more consistent with feedback-incorporating models.

7 Conclusions

This study began as a response to Mann and Repp's (1981) conclusions that syllabification and fricative-stop temporal separation do not have effects on compensation for coarticulation of the ambiguous stop. Based on the large body of existing literature on factors that mediate compensation for coarticulation, it was suggested that their methodology contained a number of confounding factors that cast doubt on the null result that was obtained. In light of the extensive review, the present study revised the original method of enforcing syllabification (that is, by instruction), and instead exploited the Ganong effect (Ganong, 1980) to ensure the placement of syllable boundaries either before the fricative or between the fricative and the ambiguous stop. Continua of real words *ace dawn–ace gone, ash dawn–ash gone*, or *a stone–a scone* were heard by subjects and the ambiguous stop judged as being alveolar or velar. In an effort to separate lexical and prelexical effects, such as vowel rounding, a continuum of nonwords **astof–*askof* was also presented to and judged by the listeners.

Given the original goals of the study, what has this investigation accomplished? All three experiments have established, quite definitively, that syllabification does have a role in the perception of an ambiguous stop, even if its consequent magnitude of change might not be enough to overcome the prelexical effect of vowel rounding, as found in Experiment 3 (Chapter 5). On the other hand, we also found that supposing a certain perceptual grouping arising from syllabification, the typical and expected compensation for coarticulation shift may itself be overwritten by higher-order lexical effects, such as word frequency. This is a significant finding, as previous studies on this phenomenon treated only purely coarticulatory effects (Nguyen, 2001), whereas our study demonstrated this phenomenon on *compensation* for coarticulatory effects as well.

In addition, the experiments have shown that higher-level lexical effects interact with lower-level prelexical effects to change listeners' judgements of the ambiguous stops. These prelexical effects

range from the temporal separation between fricative and stop, as initially predicted, to the vowel context around the cluster. The overall result of the sum of these effects causes listeners' percepts to trend in different directions. A specific example, that is on its own a novel finding of the study, is that temporal separation also affects categorization response differently based on perceptual grouping, another higher-level effect. When the gap was increased between two perceptual groups, such as between the words *ace* and *dawn*, there was change in listeners' response. On the contrary, when the gap was increased within a perceptual group, such as within the word *stone*, there was no response change, indicating that the lexically-favored perceptual group remained intact.

In summary, this study has considered the following factors that affect the perception of an alveolar-velar ambiguous stop, which can themselves be grouped into a number of categories. Those which were expected to play a role were syllabification and lexical status, which were linked in this study as lexical status was used to enforce syllabification. Indeed, these factors had perhaps the greatest influence on subjects' responses throughout the entire study. There were also factors which appeared unexpectedly but had been previously investigated: these were lexical frequency and vowel context, which played large roles in the categorization of stops in specific circumstances; that is to say, their influence was marked but highly localized in these experiments. Finally, the finding that large gap sizes did not break up perceptual groups was surprising and thus far not investigated in the literature. This factor could be ascribed to the strength of the Ganong effect, to the presence of auditory cues (such as the flat spectrum of [s]), or to a motor theory account of perception, in which the lack of movement of the tongue suggests a certain perceptual grouping. Clearly, this finding has wide-ranging theoretical implications in local segment recognition and categorization, whether by spectral or gesture accounts.

Globally, our results strongly support a top-down approach to speech perception, with multiple layers of feedforward and feedback processes activating and deactivating possibilities even at the lowest level of perception. Given the large number of effects that interact with each other, as well as the pervasive influence of the Ganong-type lexical bias across all three experiments, it seems less parsimonious to posit a purely bottom-up model to process the massive amount of variation that may occur in the process of making one final judgement. Instead, top-down models scale better and are able to handle larger numbers of interactions with more efficiency.

Of course, the investigation still leaves some questions unanswered. Exactly which effect dominates another? In which context? In addition, why are certain effects dominant under some conditions and not others? These issues remain contentious and have roots both in experimental data

and in theoretical models. In the present study, although Experiment 3 was an attempt to study the strength of lexical versus prelexical effects, our results instead showed that the influence of these effects was hard to separate with any great conviction. Furthermore, unintended priming was a possible confounding factor in the later experiments. Theoretically, while current models of speech perception provide excellent frameworks on which the numerous perceptual phenomena can be scaffolded, they do not yet possess mechanisms to explain the dominance of certain biases under specific conditions other than simply stating that it must be so. In order to formulate such explanations, further detailed experiments must be conducted, where as many factors are controlled as possible and individual effects studied.

Such an undertaking is necessarily difficult. The current investigation aimed to examine two or three individual phenomena and yet, a mere step from the start, had to posit and invoke multiple other possible effects to explain the results. Still, we have drawn a number of conclusions from this study, both expected and unexpected. These findings build upon Mann and Repp's original work and extend the literature on compensation for coarticulation in novel ways, with implications for models of speech perception.

References

Carroll, J. B., Davies, P., & Richman, B. (1971). *Word frequency book*. Boston, MA: Houghton Mifflin.

Dehaene-Lambert, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *Journal of Cognitive Neuroscience*, *12*(4), 635–647.

Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes*, *16*(5/6), 491–505.

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

Elman, J. L. & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, *27*, 143–165.

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics*, *68*(2), 161–177.

Fox, R. A. (1984). Effect on lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(4), 526–540.

Frauenfelder, U. & Segui, J. (1989). Phoneme monitoring and lexical processing: evidence for associative context effects. *Memory and Cognition*, *17*, 134–140.

Ganong, I., William F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychology Review*, *87*, 1–51.

Hallé, P. A., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: a case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 592–608.

Harding, S. & Meyer, G. (2003). Changes in the perception of synthetic nasal consonants as a result of vowel formant manipulations. *Speech Communication*, *39*, 173–189.

Kingston, J., Kawahara, S., Mash, D., & Chambless, D. (2007). *Auditory contrast and autonomy in context effects: data from Japanese and English listeners*. Unpublished manuscript. Retrieved from http://people.umass.edu/jkingstn/web%20page/teaching/ling_820/website/28%20mar%2007/jsshtk%20ms%20submitted.pdf

Kirk, C. (2001). *Phonological constraints on the segmentation of continuous speech* (Doctoral dissertation, University of Massachusetts Amherst).

Klatt, D. H. (1978). Analysis and synthesis of CV syllables in english. Unpublished manuscript. Cambridge, MA: Massachusetts Institute of Technology.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*, 971.

Kolb, B. & Whishaw, I. Q. (2009). *Fundamentals of human neuropsychology*. New York, NY: Worth Publishers.

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Loritz, D. (1991). Cerebral and cerebellar models of language learning. *Applied Linguistics*, *12*(3), 299–318.

Lotto, A. J. & Kluender, K. R. (1998). General contrast effects in speech perception: effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, *60*(4), 602–619.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003a). Lexical effects on compensation for coarticulation: a tale of two systems? *Cognitive Science*, *27*, 801–805.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003b). Lexical effects on compensation for coarticulation: the ghost of Christmash past. *Cognitive Science*, *27*, 285–298.

Malécot, A. & Chermak, A. (1966). Place cues for /ptk/ in lower cutoff frequency shifts of contiguous /s/. *Language Speech*, *9*, 162–169.

Mann, V. A. & Repp, B. H. (1980). Influence of vocalic context on perception of the [s]-[s] distinction. *Perception & Psychophysics*, *28*(3), 213–228.

Mann, V. A. & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustic Society of America*, *69*(2), 548–558.

Massaro, D. W. & Cohen, M. M. (1981). Phonological context in speech perception. *Perception & Psychophysics*, *34*, 338–348.

Max Planck Institute for Psycholinguistics. (2001). WebCELEX. Retrieved from http://celex.mpi.nl/

Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition*, *84*, 55–71.

Nguyen, N. (2001). Rôle de la coarticulation dans la reconnaissance des mots. *L'Année psychologique*, *101*, 125–154.

Norris, D. (1994). SHORTLIST: a connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238.

Pitt, M. A. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics*, *60*(6), 941–951.

Pitt, M. A. & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, *39*, 347–370.

Pitt, M. A. & Samuel, A. G. (2006). Word length and lexical activation: longer is better. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1120–1135.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 621–637.

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, *125*(1), 28–51.

Samuel, A. G. (2011). Speech perception. *Annual Review of Psychology*, *62*, 49–72.

Samuel, A. G. & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, *48*, 416–434.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 1–17.

Schwartz, M. F. (1967). Transitions in American English /s/ as cues to the identity of adjacent stop consonants. *Journal of the Acoustic Society of America*, *42*, 897–899.

Segui, J., Frauenfelder, U., & Hallé, P. A. (2001). Language, brain and cognitive development. In E. Dupoux (Ed.), (Chap. Phonotactic constraints shape speech perception: Implications for sublexical and lexical processing). Cambridge, MA: The MIT Press.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167* (3917), 392–393.

Wilde, L. L. (1995). *Analysis and synthesis of fricative consonants* (Doctoral dissertation, Massachusetts Institute of Technology).

Appendices

A Spectrograms of endpoint stimuli

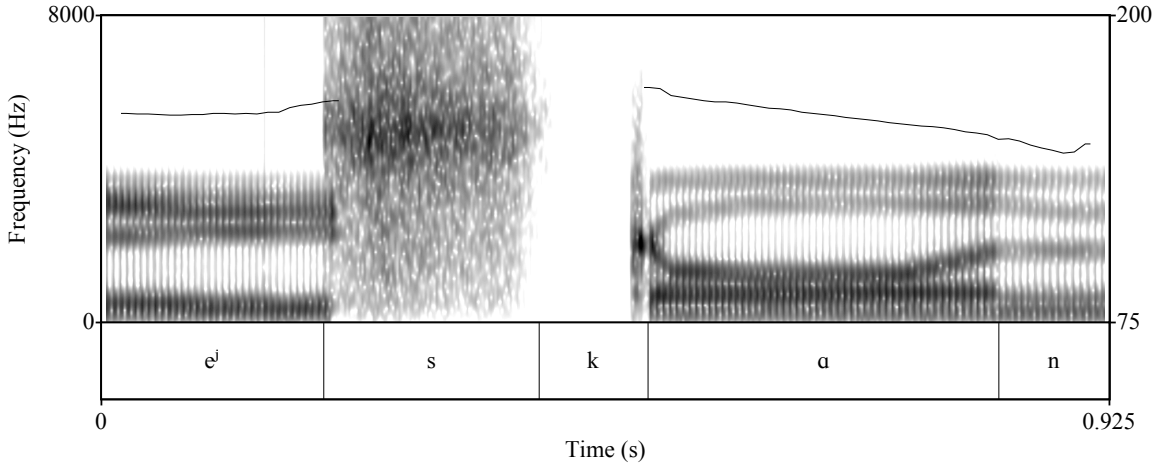

Figure A.1: Spectrogram and pitch trace of endpoint stimulus *ace gone*.

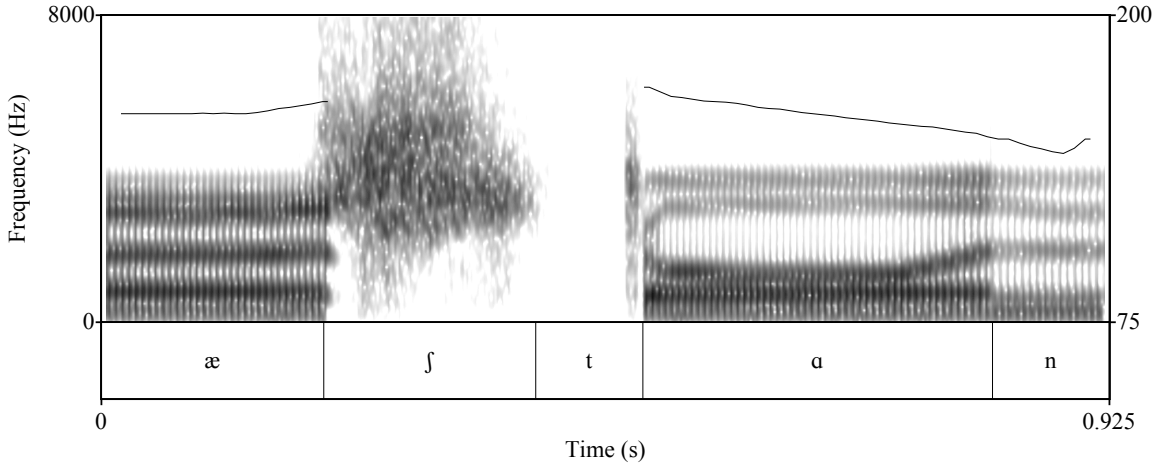

Figure A.2: Spectrogram and pitch trace of endpoint stimulus *ash dawn*.

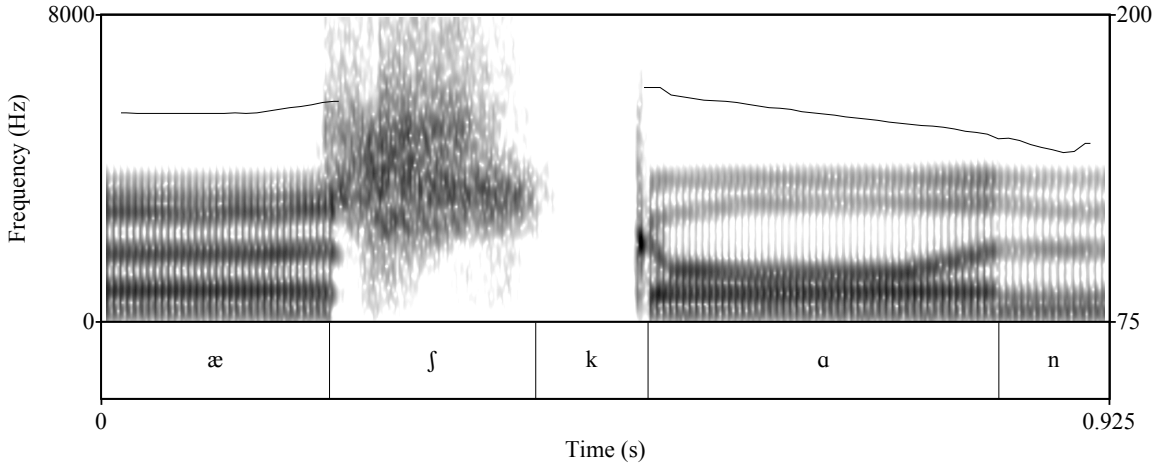

Figure A.3: Spectrogram and pitch trace of endpoint stimulus *ash gone*.

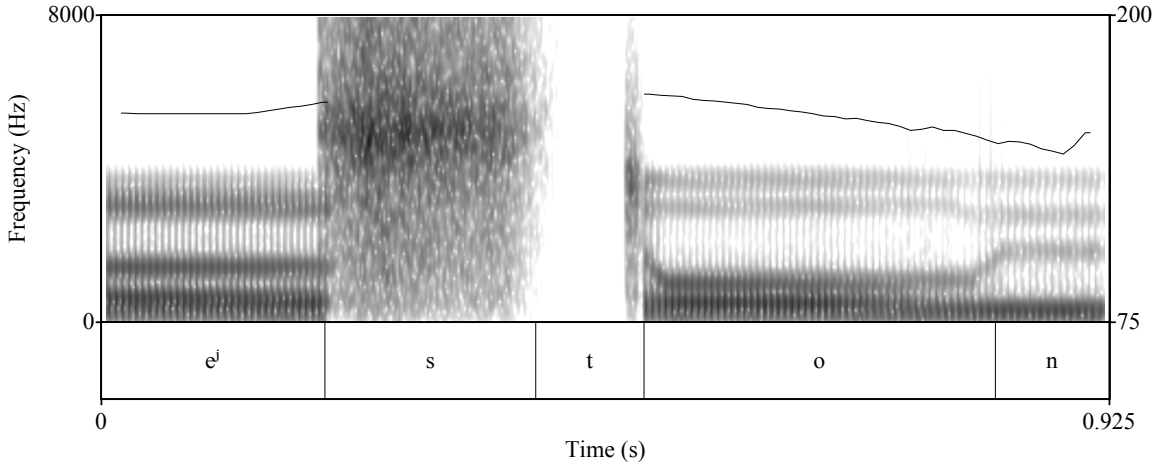

Figure A.4: Spectrogram and pitch trace of endpoint stimulus *a stone*.

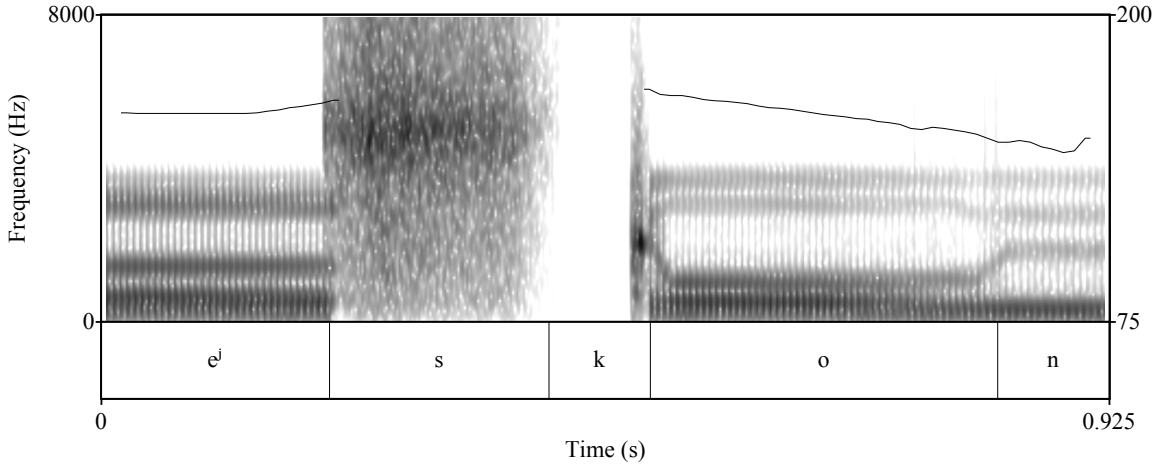

Figure A.5: Spectrogram and pitch trace of endpoint stimulus *a scone*.

B Supplementary figures

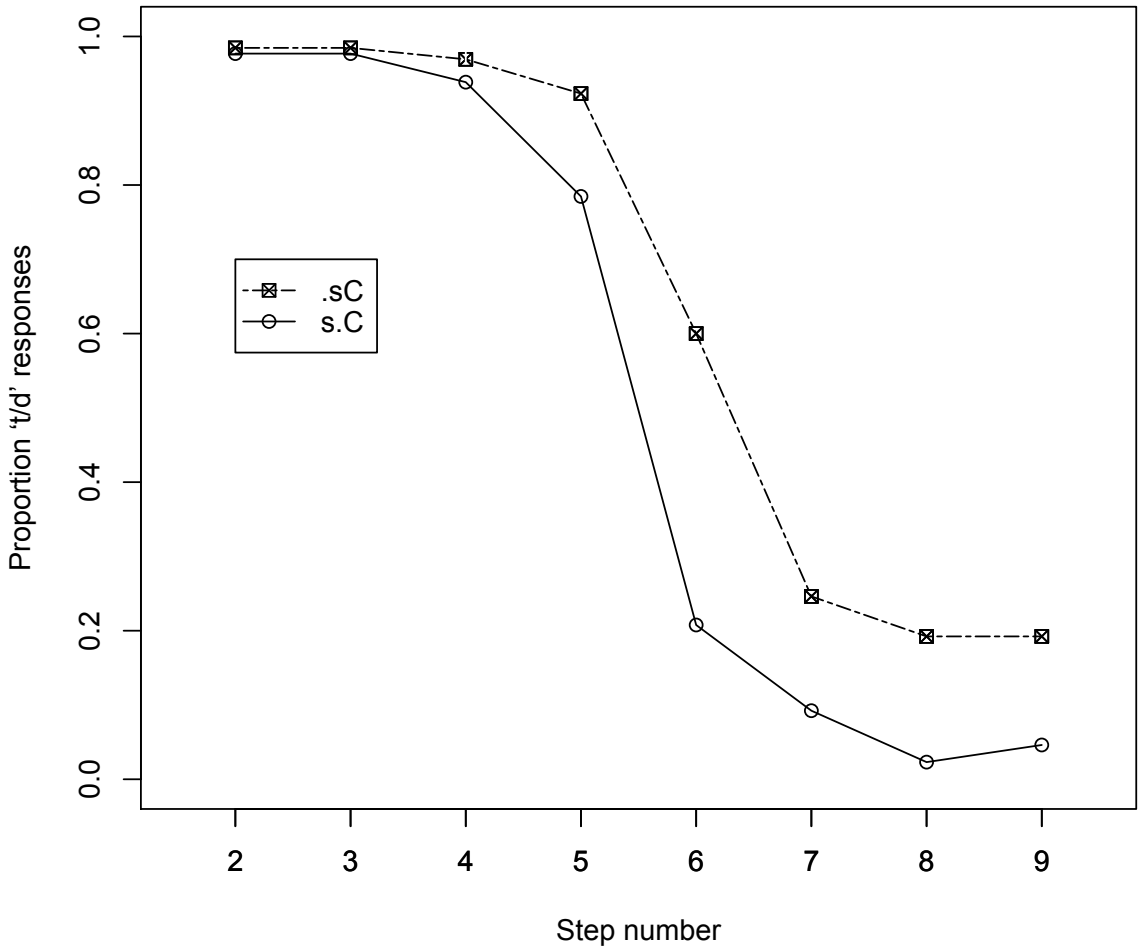

Figure B.1: Categorical curves at 325 ms for /.sC/ and /s.C/.

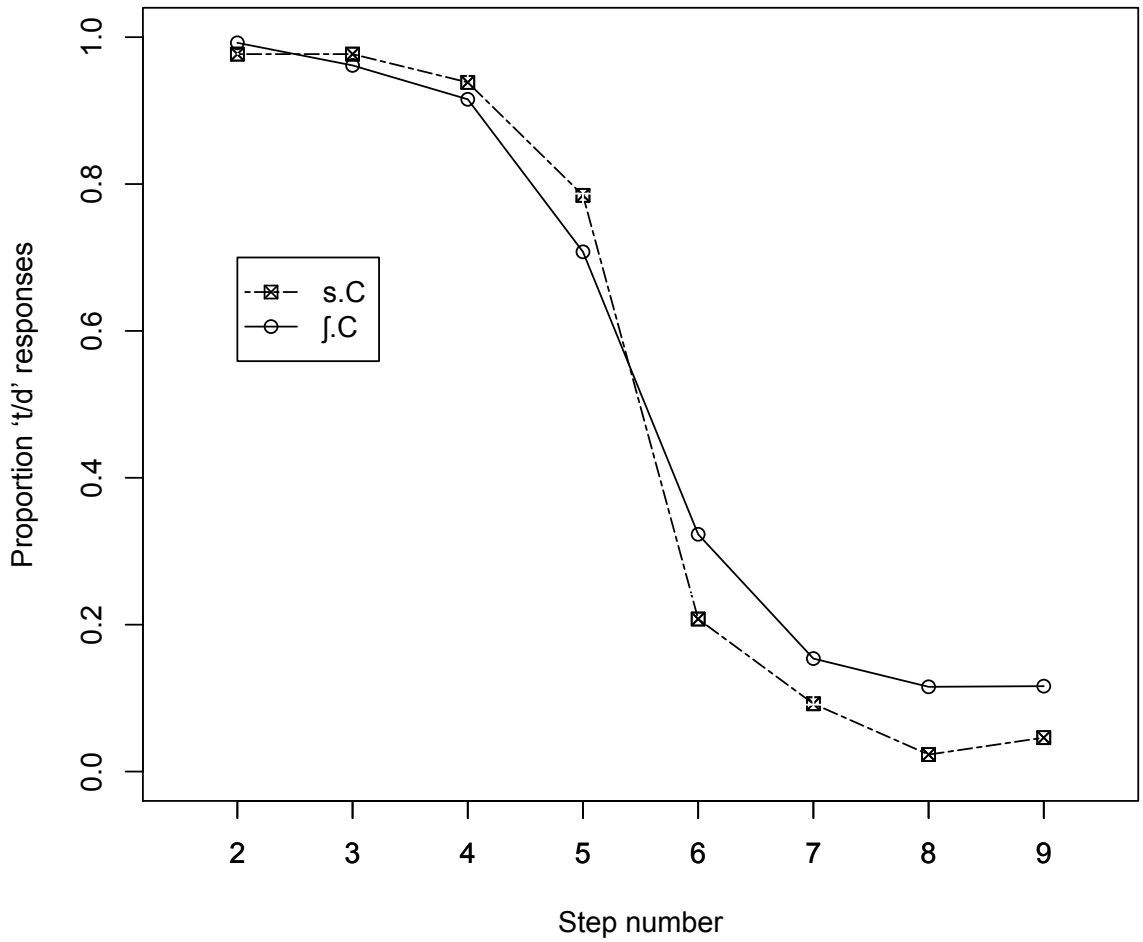

Figure B.2: Categorical curves at 325 ms for /s.C/ and /ʃ.C/.

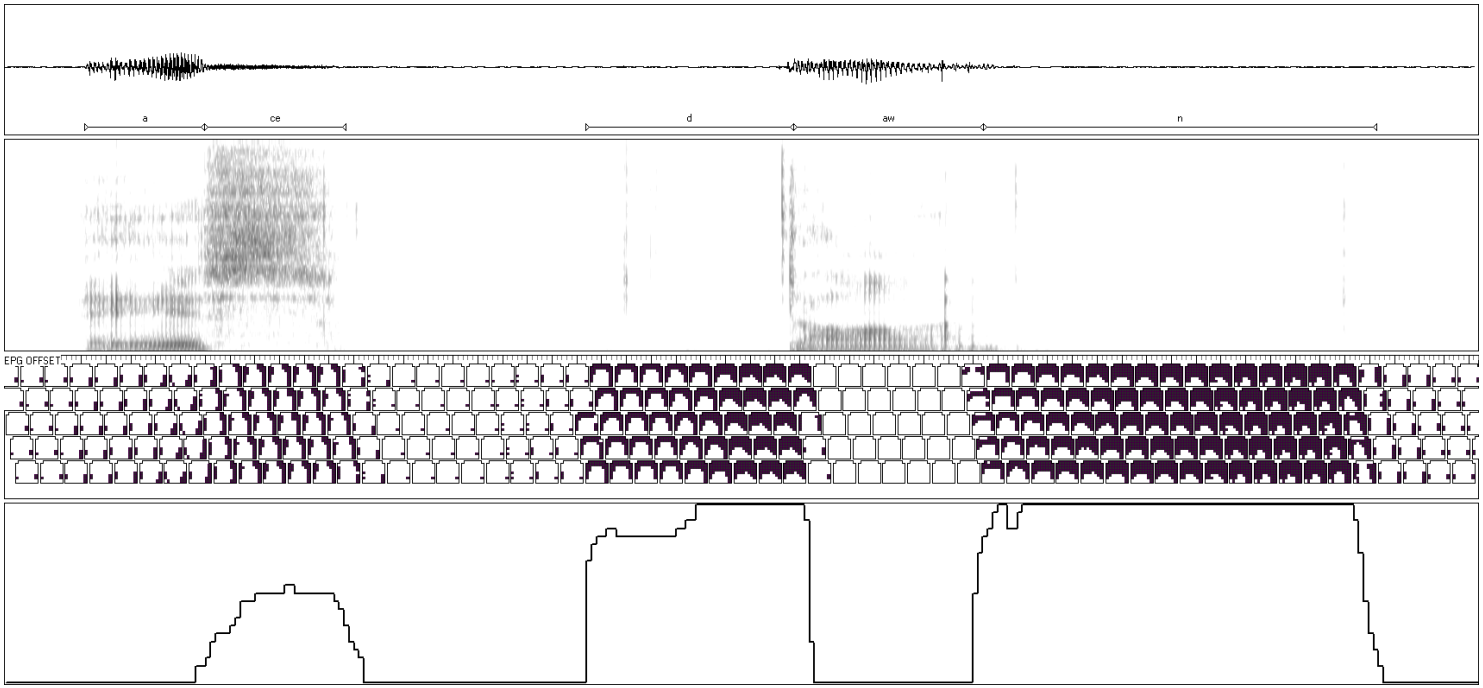

Figure B.3: Time evolution of articulatory gestures of *ace dawn*. The top panel shows the waveform of the recording. The middle panels show the spectrogram and the corresponding electropalatogram frames. and The bottom panel shows a graph of percent alveolar contact (top two rows of electropalatogram cells) over time.

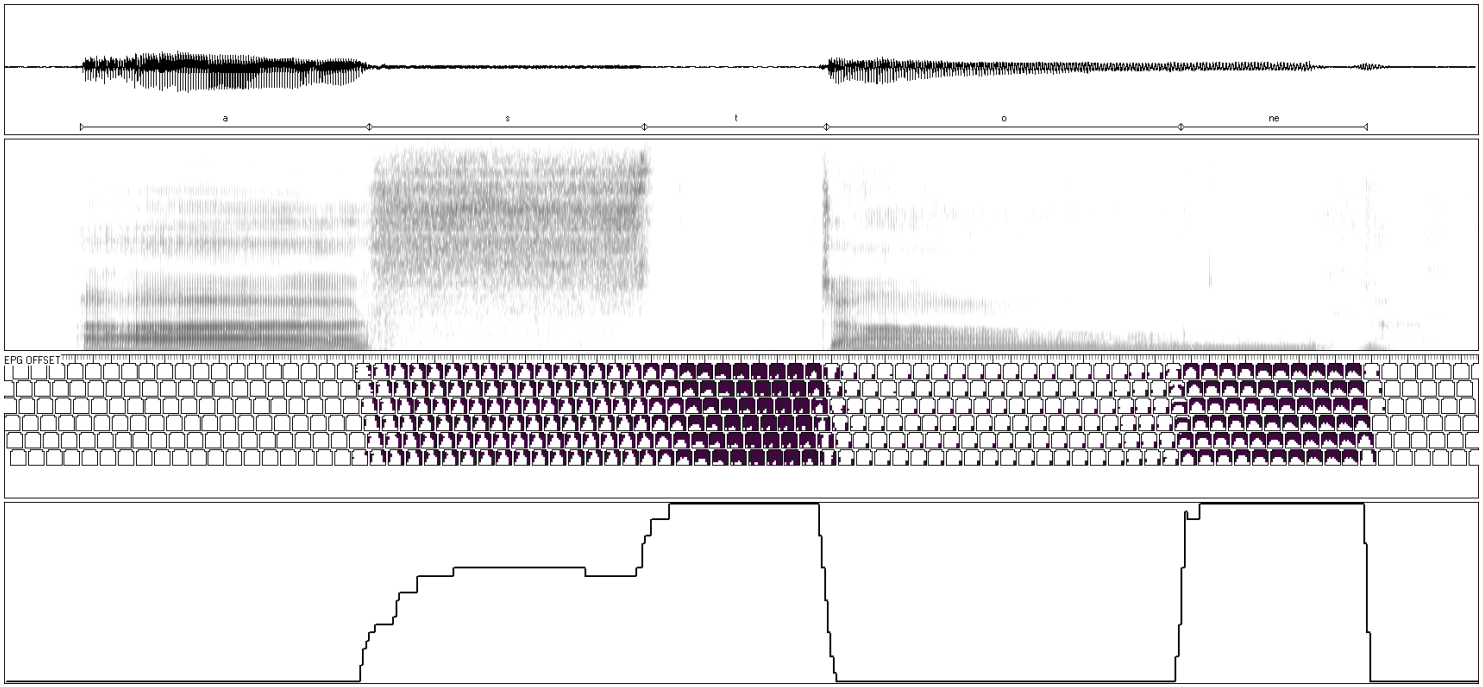

Figure B.4: Time evolution of articulatory gestures of *a stone*. The top panel shows the waveform of the recording. The middle panels show the spectrogram and the corresponding electropalatogram frames. The bottom panel shows a graph of percent alveolar contact (top two rows of electropalatogram cells) over time.