Chapter 1
Word recognition and phonology: The case of English coronal place assimilation

David W. Gow Jr. and Bob McMurray

Gradient English coronal place assimilation has been shown to produce both progressive and regressive context effects in spoken language perception. In two experiments, the current work examines the timecourse of these effects using the visual world paradigm. In Experiment 1 listeners showed earlier looks to pictures of an object with a noncoronal-initial name (e.g. boat) when it was preceded by an appropriately assimilated item (e.g. *green* pronounced with labial assimilation of the final coronal) than when it was preceded by an unmodified token of the same word (*green*). Experiment 2 employed items that produced potential lexical ambiguity due to assimilation (e.g. assimilation of the /t/ in *cat box* produces a token that resembles *cap box*) to examine regressive and progressive effects concurrently. Progressive effects analogous to those in Experiment 1 were found, albeit somewhat later, suggesting a role of lexical factors in processing. In addition, a regressive effect was shown as listeners favored images depicting cats when the immediate context was labial (e.g. *cat box*) and caps when the context was coronal (*cat drawing*). The regressive effect occurred later than the progressive effect. These results are discussed in the context of evolving lexical activation dynamics.

1. Introduction

The problem of recognizing spoken words has traditionally been couched in terms of how the system recognizes words despite the inherent variability in the acoustic signal. Here we focus on one source of variability: phonological dependencies. As Marslen-Wilson, Nix and Gaskell (1995) point out, listeners must balance the need to discriminate between words that differ by a single feature (e.g. *goat* and *coat*), with the ability to recognize words that undergo systematic phonological modification. How do listeners recognize words if their pronunciation varies from context to context? The true complexity of this problem only becomes clear when one considers the fact that context effects may be bidirectional, and that they operate in the broader context of unfolding lexical activation dynamics.

This paper examines two intersecting issues at the core of this problem. The first is the role of subcategorical variation in spoken word recognition. Many phonological dependencies result in fine-grained, continuous modification to the signal. Moreover, there is evidence that listeners are sensitive to continuous cue variation (cf. Andruski, Blumstein and Burton, 1994; Gow and Gordon, 1995; Utman, Blumstein and Burton, 2000; Gow, 2002; 2003; McMurray, Tanenhaus and Aslin, 2002). However, with few exceptions (c.f.; Gaskell, 2003) current models of spoken word recognition rarely incorporate this sensitivity, or examine how continuous variation interacts with lexical activation dynamics. We examine the role of continuous modification in lexical activation, exploring both how gradient modification influences word recognition, and how lexical factors influence speech processing.

The second issue concerns how phonetic information is integrated over time. The articulatory and acoustic consequences of phonological modification can extend throughout the word-form being modified and neighboring words. This gradient modification is shaped by contingencies between a whole hierarchy of factors ranging from local relationships between proximal segments or cues, to relationships between individual cues and global factors including speaking rate or dialect. The fact that these contingencies operate both forwards and backwards in time and at different timescales makes understanding the temporal dynamics of processing a central challenge for research in spoken language comprehension. This presents a methodological challenge in that the study of the resulting perceptual processes requires experimental paradigms that provide temporally continuous measures of activation.

In this paper we review relevant research on the twin problems of gradient acoustic modification and temporal integration in word recognition. We will then present results from two new experiments that represent the beginning of a systematic program of research assessing these issues.

2. English Coronal Place Assimilation

In this paper we examine assimilation. Chomsky and Halle (1968) describe assimilation as "a process in which two segments are made to agree to a value that is assigned to one or more features" [p.350]. In English coronal place assimilation, segments with coronal place (e.g. /t/, /d/, and /n/) are described as taking the place of a subsequent noncoronal segment (e.g. a velar or labial). Thus, *green boats,* for example, may be pronounced as *greem boats*.

While assimilation may be seen as purely a source of perceptual or phonetic variability in speech perception, it must also be seen as a source of potential lexical ambiguity to the degree that it neutralizes phonetic contrasts in some environments. For example, *cat box* can sound like *cap box* (after assimilation). Thus, the perception of assimilated events may interact with the lexicon in (at least) two ways. First, (as the *cat box* example implies), the phonetic structure of the lexicon (e.g. the presence or absence of minimal pairs) can yield information for the resolution of assimilation-induced ambiguity. Second, by the time assimilation occurs, word recognition processes may have already built significant lexical activation for one or more lexical competitors—other sources of information may prime the system before the assimilated segment is heard. This is evident in a number of studies showing that listeners respond differently depending on whether assimilation does or does not produce a competitor (e.g. Gaskell and Marslen-Wilson, 1996, versus Gaskell and Marslen-Wilson, 2001).

3. Continuous Assimilation and Perception

Chomsky and Halle's (1968) description of assimilation assumes that phonological representation, and thus phonological modification, is discrete. However, evidence from a number of domains suggests that assimilatory modification may be gradient. This debate does not solely concern representation. It has fundamental implications for the sorts of processing mechanisms that may cope with assimilation.

Consider again the phrase *cat box*. By application of the assimilation rule, the coronal /t/ in *cat* should assimilate the labial place of the /b/ in *box* and become a [p]. If assimilatory modification is discrete, *cat box* should sound exactly like *cap box*. This would neutralize a perceptual contrast., and so listeners would have to rely on non-perceptual mechanisms such as contextual constraints (e.g. pragmatic or phonological) to infer the speakers' intent. Gaskell and Marslen-Wilson (1996; 1998; 2001) suggest that listeners use knowledge of place assimilation processes or the regularities they produce to infer the underlying form of assimilated segments based on their phonological context. This type of phonological inference however, may be problematic when one surface form (e.g. *cap box*) is consistent with two lexical interpretations (*cat* and *cap box*), if assimilation is discrete (complete). This might prevent listeners from interpreting the phrase *cap box* at face value when it is intended to describe a box full of caps.

There is however, good evidence that coronal place assimilation is a non-neutralizing process suggesting that listeners may not need to rely on problematic inferences. At an articulatory level, evidence from electropalatographic and articulometer studies shows that the completeness of coronal place assimilation may vary widely, and is modulated by features including speaker, speaking rate and dialect (Barry, 1985; Kerswell, 1985; Nolan, 1992; Byrd, 1996; Kuhnert and Hoole, 2004). In many cases, it is characterized by overlapping coronal and non-coronal closures, with a reduction in the underlying coronal gesture reflected by incomplete closure. Acoustic analyses provide a similar picture. Holst and Nolan (1995) examined the frication envelopes of tokens of place assimilated fricatives that they showed a range of realizations ranging from those associated with canonical tokens of /s/ and /ʃ/, with some showing intermediate realizations with /s/-like onsets and /ʃ/-like offsets. In several studies of place assimilation in stops Gow (2001; 2002; 2003) found that assimilated items have formant characteristics intermediate between those of coronal and non-coronal segments at their offsets. Together, these results suggest that coronal place assimilation is a graded process.

The gradient nature of assimilation potentially opens the door to more robust word recognition based on other sources of disambiguation. Inferences based on gradient modification would be more powerful than those based on discrete feature change because assimilated forms (e.g. ca^{t}_{p} *box*) may be perceptually distinguishable from unmodified forms (*cat box*). *Graded* perceptual input is also consistent with a probabilistic constraint satisfaction approach, in which multiple sources of partial information are integrated during perception. Under this view, phonemically ambiguous (or partially ambiguous) information may be an *asset* to processing. First, it prevents the system from erroneously committing to an incorrect alternative before potentially disambiguating information becomes available (e.g. McMurray, 2004). Second, gradient modification may reflect recoverable information about the integration of adjacent features, potentially facilitating the perception of both assimilated and assimilating segments.

However, in order to engage in such processing, the system must be sensitive to these subtle acoustic modifications. Early work on the categorical perception of speech sounds (cf. Liberman et al, 1957) suggested that listeners show sharp phonetic category boundaries and little ability to discriminate between members of the same category. However, subsequent research has shown that listeners are sensitive to within category variation, and that category boundaries may not always be clear (c.f. Pisoni and Tash, 1974; Carney et al., 1977). At a purely perceptual level, it seems clear that the system has the necessary sensitivity to the relevant acoustic properties. Moreover, there is now broad evidence that continuous variation affects the dynamics of lexical activation (Andruski, Blumstein and Burton, 1994; Gow and Gordon, 1995; Utman et al., 2000; Gow, 2001; 2002; 2003; Gow and Im, 2004; McMurray et al., 2002; McMurray et al, 2003; McMurray, 2004; Hawkins, 2003). Thus, at both perceptual and lexical levels of processing, the system has the necessary sensitivity to engage in such a probabilistic constraint satisfaction process. We will now turn to the temporal properties of this process.

4. Temporal Processes and Assimilation

The structure of phonetic categories is closely related to the issue of contextual dependency. A great deal of work has shown that feature cue interpretation is context-dependent. Context effects may be internal to a feature of segment. For example speech categories are typically determined by multiple simultaneous acoustic cues that trade (e.g. voicing, by VOT & F1: Summerfield and Haggard, 1977). Moreover, cues to different features of the same phoneme have been shown to interact (c.f. Pisoni and Sawusch, 1977), and visual cues affect categorization of place of articulation continua (McGurk, 1976)

Many context effects act across time. Mann and Repp (1981) have shown that the system compensates for coarticulatory processes affecting adjacent segments—fricatives influence the perception of the subsequent stops. Research has also shown that manner of articulation and voicing are affected by speaking rate, instantiated as the length of the *subsequent* vowel (e.g. Miller and Liberman, 1978; Summerfield, 1981). Ganong (1980) demonstrated that lexical context (a larger temporal domain) can affect phoneme judgments, and the McGurk effect has been shown to be resistant to some temporal asynchronies but not others (Munhall et al., 1996). Thus, not only must phonetic categories be identified across wide temporal domains, but they must take into account the specific temporal properties of multiple cues.

These context effects reflect systematic covariation between acoustic or articulatory phenomena, The temporal asymmetries inherent in such phenomena allow the system to exploit them to improve perception. Coronal place assimilation provides an interesting case as it creates both regressive effects (which resolve ambiguity created by assimilation) and progressive effects (which affect the perception of upcoming material).

A regressive context effect occurs when interpretation of an assimilated segment is influenced by the segment that follows it—essentially, compensation for assimilation. Gaskell and Marslen-Wilson (1996) created fully assimilated token of the form (*wickeb prank*) and measured activation for the initial word (*wicked*) using a cross-modal form priming paradigm. Listeners showed priming for the unmodified form of the target word (*wicked*) in contexts that licensed the observed modification (*wickeb prank*), but not in ones that did not (*wickeb game*). Subsequent studies using both phoneme monitoring and priming paradigms confirmed that post-assimilation context influences the perception of assimilated segments (Gaskell and Marslen-Wilson, 1998; Gow, 2001; 2002; Coenen et al., 2001).

The second context effect (with respect to English coronal place assimilation) is progressive. Gaskell and Marslen-Wilson (1998) found that monitoring for word-initial stop consonants is influenced by the appropriateness of pretarget assimilation. They found that monitoring latencies for an initial noncoronal (such as the labial /b/ in *bearer*) are longer after a contextually unviable modification (e.g. *freight bearer* pronounced *frayk berer*) than they are after unmodified or viably modified tokens of the same word (*freight bearer* or *frayp bearer*). There was no difference in monitoring performance though in comparisons between unmodified and appropriately modified contexts. This may, then, reflect listeners' *inability* to recognize the pretarget lexical context rather than a true progressive context effect. However, the stimuli in this experiment were produced through deliberate mispronunciation, (which induces discrete feature substitution rather than the coarticulation-like modification observed in spontaneously assimilated speech). A number of studies have shown that anticipatory coarticulation facilitates the perception of post coarticulation segments (Kuehn and Moll, 1972; Beddor, Harnsberger and Lindemann, 2002)—thus, the progressive effect may have been diminished by the lack of realistic coarticulation. Indeed, a number of studies on assimilation have revealed progressive facilitation using more natural stimuli. This has been shown in gating (Lahiri and Marslen-Wilson, 1991), phoneme monitoring (Gow, 2001; 2003; Gow and Im, 2004), and inhibitory form-priming (Gow, 2001) paradigms with naturally assimilated stimuli.

It is unclear how or if these progressive and regressive assimilation context effects are related. Gow (2003) showed that the same stimuli can produce both regressive and progressive effects, suggesting that the two effects may occur together. This conclusion is of limited scope, though, since progressive and regressive effects could only be demonstrated in separate experiments using different experimental paradigms. Moreover, mutual bidirectional contingencies raise fundamental questions about the timecourse of processing. Specifically, are the processes applied iteratively or concurrently, and do they reflect one or more underlying process? Very little is known about these questions, because until recently, researchers have lacked the tools to address progressive and regressive effects online, or to examine the timecourse of such effects in the perception of meaningful continuous speech.

5. Methodological Issues in the study of assimilated speech.

The methods that have been used up to this point to examine the processing of assimilated speech have had a limited ability to examine the timecourse of processing. Techniques based on priming allow the experimenter to tap processing at only a couple of timepoints. Moreover, priming itself is a mediated task requiring the perception of two words (the prime and the target, either visual or auditory) as well as a metalinguistic lexical decision. Interpreting the timecourse of these results without a clear linking hypothesis can be difficult.

The visual world paradigm (Cooper, 1974; Tanenhaus et al., 1995) provides an alternative task that is able to reliably measure lexical activation with exquisite sensitivity to the temporal dynamics of processing. Moreover, unlike phoneme detection and some uses of priming this methodology uses a *referential* visual context and task. Subjects hear spoken instructions to manipulate one of a small set of visual objects (typically images on a computer screen). The names of these objects are controlled by the experiment to represent possible *interpretations* of the auditory stimulus. Eye movements to each object are monitored during and after the auditory stimulus, and serve as the dependent measure.

With careful control of the verbal instructions and the visual competitor set, fixations reveal unfolding representations at many levels of linguistic processing. For example, Allopenna, et al., (1998), presented subjects with screens containing a target (e.g. *beaker*), a cohort competitor (e.g. *beetle*), a rhyme competitor (e.g. *speaker*) and an unrelated item (e.g. *carrot*). After being instructed to select the target, subjects made more eye-movements to cohort and rhyme competitors than unrelated items. Significantly, fixation probabilities over time reflected the temporal similarity of competitors to the target, replicating and extending previous examinations of lexical competition dynamics. Moreover, when transformed with a simple linking hypothesis, this measure was highly correlated with moment-by-moment activation from the TRACE model of speech perception (McClelland and Elman, 1986), suggesting that eye-movements can yield a detailed picture of lexical activation dynamics. Subsequent work has demonstrated sensitivity to lexical frequency (Dahan et al., 2001), lexical neighborhood (Magnuson et al., 2003) and mismatching coarticulatory information (Dahan et al., 2001b). Thus, these methods provide a detailed picture of lexical processing dynamics.

Recent work applying this technique to phoneme identification reinforces the need for appropriate tasks. McMurray et al. (submitted) compared patterns of fixations in an explicit metalinguistic task (selecting the initial phoneme) with those from a meaning-based task (selecting a picture) for the VOT continua. Results from the lexical task showed more sensitivity to within-category VOT, suggesting that explicit tasks may show less of the sensitivity that is needed to perceive assimilated segments.

The prominent role of context in the perception of assimilated speech suggests that this process is, at its core, one of temporal integration. We have adopted the visual world paradigm because it may allow us to determine *when* processing mechanisms are sensitive to this detail, potentially disambiguating competing theories. Such an approach was taken by McMurray (2004), in determining the relative timing of VOT and vowel-length effects on lexical activation. The following experiments apply this approach to assimilation.

6. Experiment 1: Progressive Effects

The purpose of Experiment 1 was to determine if the the visual world paradigm can provide converging evidence for progressive assimilation effects found with phoneme monitoring and priming paradigms (Gow, 2001; 2003), and can illuminate the timecourse of this effect. Subjects heard assimilated or non-assimilated phrases (e.g. *Select the* $gree^{n}_{m}$ */ green boat*) while eye movements related to possible interpretations of post-assimilation context (*boat*) were monitored. We predicted that listeners would show earlier fixations on the green boat following instructions with the labialized $gree^{n}_{m}$.

6.1. Methods

6.1.1 Procedure

Thirty-four University of Rochester undergraduates were tested in a variant of the visual world paradigm. On each trial subjects saw four pictures. These pictures represented the four possible two-word combinations in which the initial word ended in a coronal or noncoronal and the context word began with a coronal or noncoronal. Thus, one picture contained a coronal followed by a noncoronal (CN: e.g., *green boat*), one picture contained a coronal followed by a coronal (CC: e.g., *green dog*), one picture contained a noncoronal followed by a noncoronal (NN: e.g., *swamp boat*) and one picture contained a noncoronal followed by a coronal (NC: e.g., *swamp dog*). A complete list is provided in Appendix A. Each picture appeared in its own quadrant of a 20 inch computer monitor, with a small gray circle in the center. After 500 ms, this circle turned red, at which point the subject was instructed to click on it to initiate the trial. This initial portion served to orient the subject to the particular pictures and their locations for each trial, minimizing eye-movements due to visual search. After the subject clicked on the red circle, an auditory stimulus was played (of the form "Select the *green boat*"). Subjects then selected the appropriate picture.

Stimuli were blocked by stimulus set, such that subjects saw all of the tokens for two stimulus sets (and two filler sets) as a group. Because a number of the pictures would not be immediately obvious to the subjects (e.g. *swamp dog*), at the beginning of each block the pictures for that block were presented individually, along with their names printed below them.

For each stimulus set, subjects heard all of the six possible auditory stimuli (an initial word ending in a coronal, noncoronal, or assimilated segment followed by a second word beginning with either a coronal or non-coronal). The likelihood of clicking any given picture (or class of pictures) was equated across trials, and additional filler trials were used in each block to minimize strategic responding. Assimilated tokens were heard only once per block. These experimental trials were combined with an equal number of filler items (of similar form) for a total of 288 trials.

Eye-movements were recorded using an SMI Eyelink Eyetracker at 250hz. For each subject, the average probability of fixating each of the four objects (across trials) was computed at every 4ms.

6.1.2. Stimuli

Sixteen stimulus sets were chosen for this experiment. As mentioned, each set of visual stimuli consisted of all of the combinations of a coronal- or noncoronal-initial item with a coronal or non-coronal context item (e.g. *green boat*, a *green dog*, a *swamp boat*, and a *swamp dog).* Pictures were extensively normed to provide maximally iconic representations of depicted items.

Six auditory tokens were constructed for each set (the four items, plus the coronal items produced with assimilation). An experimentally naive male speaker produced each of the four items in the context of "select the" (e.g. *select the green boat)* several times. These productions were recorded digitally and combined to create the six experimental conditions (coronal # noncoronal; assim # noncoronal, noncoronal # noncoronal; coronal # coronal; assim # coronal; and noncoronal # coronal). The initial portion (e.g. *select the green*) was separated from the final item (*boat)* and recombined using cross-splicing techniques to create each condition.

Table 1. Example of splicing for Experiment 1.

<table>
<tr><th>Label</th><th>Example</th><th>Initial Portion Spliced From…</th><th>Final Portion Spliced From…</th></tr>
<tr><td>Assim # Noncoronal</td><td>Select the gree$^{n}_{m}$ boat</td><td rowspan="2">Select the green boat.</td><td>… green boat.</td></tr>
<tr><td>Assim # Coronal</td><td>Select the gree$^{n}_{m}$ dog</td><td>… green dog.</td></tr>
<tr><td>Coronal # Noncoronal</td><td>Select the green boat</td><td>… green dog.</td><td>… green boat.</td></tr>
</table>

Coronal # Coronal	Select the green dog		… green *dog*.
Noncoronal # Noncoronal	Select the swamp boat	… *swamp* dog.	… green *boat.*
Noncoronal # Coronal	Select the swamp dog	… *swamp* dog.	… green *dog.*

As table 1 indicates, initial (unassimilated) coronals were obtained from the coronal # coronal tokens (green dog). Initial noncoronals were obtained from the noncoronal # coronal tokens (green boat). Assimilated coronals were taken from coronal # noncoronal contexts. A combination of subjective comparison by two phonetically-trained listeners and a measurements of the first three formant frequencies prior to closure allowed us to select segments were in fact assimilated. Post-assimilation context was acoustically identical for all tokens (with the same context), and taken from the initial coronal items (as these were more phonetically neutral). These initial items were then spliced onto the post-assimilation contexts to create each of the six conditions. Cuts and splices were made at zero-crossings to avoid artifactual clicks and pops.

4.2. Results and Discussion

Our hypothesis was that variability due to assimilation would improve processing by facilitating activation for the post-assimilation context (e.g. *boat,* in *gree*$^{n}_{m}$ *boat*). Since fixations to objects in the visual world paradigm reflect lexical activation, we compared fixations to the non-coronal target (e.g., the *green boat*) when the auditory stimulus (*green*) was assimilated or not. Our hypothesis was that there would be a higher probability of fixations after an assimilated context.

We first assessed the degree to which subjects selected the correct picture. This was quite high, averaging 99% correct. However, even a single incorrect trial could yield significant noise, as in this case the target would become the competitor (or a filler) and high fixation probabilities (to the target) would be replaced with low (to the new competitor). Thus, to guard against this, subjects performing less than -.5 z-scores from the mean were excluded from the analysis. This left 28 subjects in the analysis of the original 34.

Figure 1a shows the probability of fixating the target (the *green boat*) as a function of time for assimilated and non-assimilated coronals. There was a reliable effect of assimilation ($t(27)=2.71$, $p=.011$), with targets receiving more fixations after assimilated consonants. A complementary facilitory effect was also seen in looks to the coronal competitor objects (e.g. *green duck*)—assimilated stimuli resulted in marginally fewer looks to the competitor (figure 1b; t(27)=2.01, p=.054). Thus assimilatory modification to coronals can help facilitate activation of upcoming targets, and also rule out competing objects faster.

In order to determine when the effect could first be seen, we used a technique similar to McMurray (2004) to analyze the temporal pattern of fixations. For each subject, the probability of looking at each of the four classes of visual targets was computed for each 20ms time slice, for each of the six conditions. These were then smoothed using a triangular smoothing window (with a width of 100 ms). From this data, the "size" of the

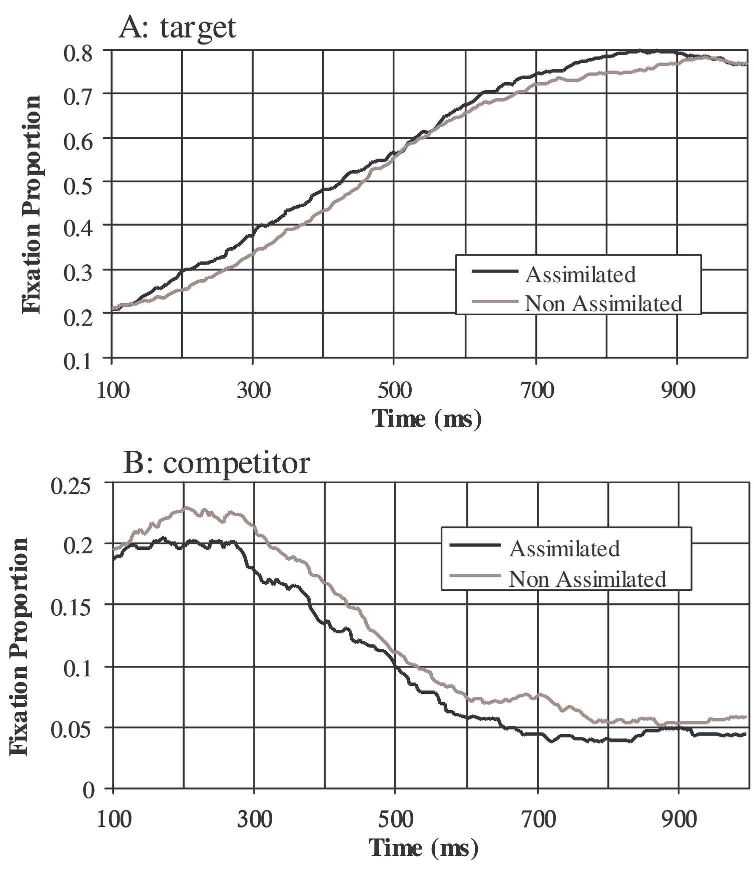

Figure 1: Looks to the target (panel A) or competitor (panel B) as a function of time given an assimilated or non assimilated auditory stimulus.

progressive effect was computed at each 20ms time-slice. First we computed overall bias to the non-coronal object by subtracting probability of fixating either of the coronal competitors (the *green dog* or the *swamp dog*) froo the probability of fixating either of the non-coronals (e.g., either of the *boats*) in the assimilated condition from the same probability in the unassimilated (coronal) condition. Thus, where the average effect size was zero, these probabilities were the same, while when the effect size was positive this would indicate a progressive effect.

The results of this analysis yielded a progressive effect that was marginally significant at 120 ms (t(27)=2.03, p=.053) after the onset of the context word (e.g. *boat*). It was fully significant at 140 ms (t(27)=2.33, p=.027) and lasted until 340 ms.. At this point, looks to the target and competitor were equal across the two conditions. Since it takes roughly 200ms to plan and launch an eye-movement, this places the onset of this progressive effect sometime during the last portion of the assimilated item, and extending throughout the post assimilation context.

Experiment 1 demonstrates a progressive effect of assimilatory modification—an assimilated coronal facilitates activation for upcoming contexts that are consistent with assimilation (e.g noncoronals), and inhibits activation for upcoming competitors (coronals).

The next experiment examines the processing of assimilation when it produces lexical ambiguity. This will extend the findings of Experiment 1 in three ways. First, comparison between any progressive effects found in the two experiments may illuminate any role of lexical factors in the timecourse of processing. Second, the creation of lexical ambiguity makes it possible to examine regressive context effects that may play a role in disambiguation. This makes it possible to look for evidence of both regressive and progressive effects in a single task, and to compre their timecourses, as filtered through the dynamics of lexical activation.

5. Experiment 2: Bidirectional Context Effects

Assimilation can potentially create lexical ambiguity. For example, a labially assimilated /t/ in *cat* results in a potentially lexically ambiguous surface form, which is partially consistent with both *cat* and *cap*. While lexical biases might have allowed listeners in Experiment 1 to determine that the assimilated form $gree^{n}{}_{m}$ is best interpreted as the familiar word *green*, such biases would not help listeners discriminate between modified instances of *cat* and *cap*. Nonetheless, subsequent context may provide partial cues to the underlying lexical form. That is, since coronals undergo assimilation, but not noncoronals (in English), an assimilated segment followed by a noncoronal context is more likely to be due to assimilation (and hence, an underlying coronal), than if it is followed by a coronal (which would not create assimilatory cues). A probabilistic constraint satisfaction approach could evaluate these likelihoods, in conjunction with the likelihood of each interpretation given the gradient, bottom-up signal and other contextual factors, to yield the most likely interpretation.

Experiment 2 examined the resolution of this ambiguity by using stimulus sets in which assimilation created lexical ambiguity, such as *cat box*, *cat drawing, cap box,* and *cap drawing.*

5.1. Methods

Thirty-four University of Rochester undergraduates were run in an identical procedure to Experiment 1. Subjects were shown four pictures for a brief period and then heard a phrase instructing them to select one of them. Again, stimulus sets were presented in blocks and subjects were shown each picture and name at the beginning of the block.

Like Experiment 1, subjects heard each item once per stimulus set, with the frequency of each selectrion balanced across the experiment. Again, an equal number of filler items (non lexically ambiguous) were also employed, to yield a total of 288 trials. Eye movements were monitored thoughout the experiment and analyzed in the same way as Experiment 1.

5.1.3. Stimuli

Stimuli were created using the cross-splicing technique used in Experiment 1 with the exception that items were chosen that would be lexically ambiguous after undergoing assimilation. A complete list of items is given in appendix B.

5.2. Results and Discussion

The mouse-click responses were analyzed first. Subjects were not as accurate as in Experiment 1, averaging 90% correct. This was likely due to the ambiguity in the initial word. As before, subjects performing at less than -.5 standard deviations from the mean (on the mouseclick response) were excluded from analysis. This left 19 of 34 subjects in the analysis.

The pattern of preferred interpretations in coronal and noncoronal contexts is consistent with the pattern of priming found by Gow (2003). Figure 2 displays the probability of fixating one of the initial items (either

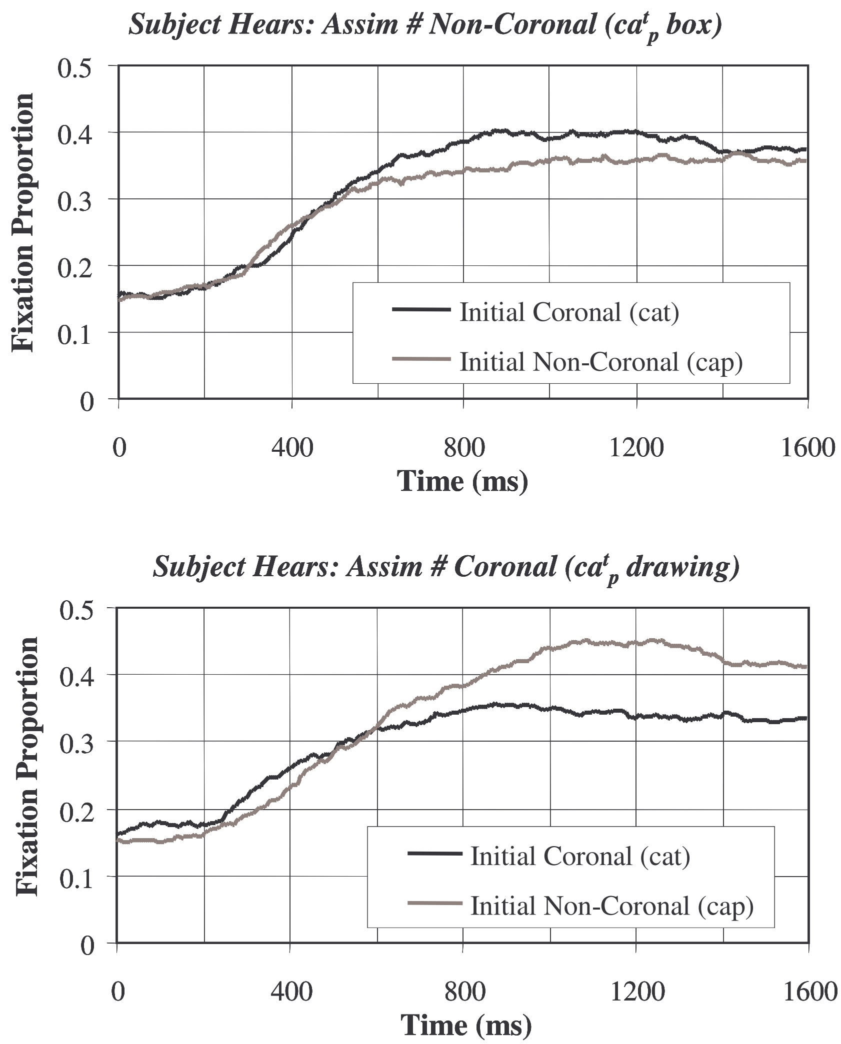

Figure 2: Looks to the initial-coronal or initial-noncoronal objects (e.g. cat or cap) as a function of time after hearing either Assim # Coronal (panel A) or Assim # Noncoronal (panel B)

the coronal interpretation, *cat*, or the noncoronal, *cap*) as a function of time. The top panel (A) shows the pattern of fixations after hearing an assimilated token followed by a noncoronal (e.g. *cat/p box*). It is clear that early on (before about 500 ms after the onset of the post-assimilation context word) there is a small bias to prefer a coronal interpretation.

However, at around 500 ms, the bias reverses towards the coronal interpretation (given a non-coronal context). The bottom panel (B) shows the fixations after hearing the same assimilated token followed by a coronal (e.g. *cat/p drawing*). Here, both interpretations are maintained initially, with a later bias emerging towards the non-coronal interpretation. This difference in interpretation between the conditions demonstrates that for *the same assimilated token*, interpretation is dramatically modified by post-assimilation context.

This regressive effect was tested in a 2x2 ANOVA using fixations to potential interpretations as a dependent measure, and the interpretation (coronal or noncoronal) and post-assimilation context as independent factors. There was a significant effect of interpretation ($F(1,18)=4.5$, $p=.048$)—with noncoronal interpretations receiving slightly more overall fixations. There was no main effect of context ($F<1$), in that there were not more fixations (overall) in one context than the other. However, as predicted, a significant interaction was found between context and interpretation ($F(1,18)=81.93$, $p=.01$).

In order to better understand the timing of the regressive effect, a temporal analysis similar to Experiment 1 was conducted. As before, the average probability of fixating each of the four alternatives was computed at each 20 ms time-bin and smoothed. Next, the regressive effect was computed at each time-bin, for each subject. This effect consisted of the overall bias towards a coronal interpretation of the assimilated segment (CN - NN: looks to the *cat box* minus looks to the *cap box*) subtracted from the same bias in the assim # coronal condition (CC – NC: looks to the *cat drawing* minus looks to the *cap drawing*). This is the statistical interaction computed above (at each timeslice): if the coronal bias as the same in each condition it would be zero.

A series of one-sample t-tests were conducted at each time-slice to determine when the effect was reliably non-zero. It was found to be marginally significant ($t(18)=2.02$, $p=.058$) at 560 ms and significant at 580 ms ($t(18)=2.30$, $p=.033$). It lasted for quite some time, becoming non-significant at 1300 ms ($t(18)=2.058$, $p=.054$). Given the known oculomotor delay of 200 ms, it would thus appear that the regressive effect can be seen from approximately 360 ms after the onset of the post-assimilation context, (close to the mean offset of 403 ms). This quite late and later than the progressive effect found in Experiment 1.

To test for a progressive effect, the assimilated # noncoronal context (*$ca^t{}_p$ box*) was compared to the coronal # noncoronal context (*cat box*). The progressive effect, operationalized as increased looks to items with noncoronal contexts (e.g. *box* as opposed to *drawing*) was found to be significant at 380 ms (t(18)=2.11, p=.049), lasting until 480 ms (t(18)=1.99, p=.06). It is notable that this progressive effect becomes significant 240 ms later than the comparable effect found in Experiment 1 (and is 100 ms shorter). This likely reflects the fact that assimilation created lexical ambiguity in Experiment 2, but not in Experiment 1—such ambiguity slows the dynamics of processing, resulting in a later progressive effect. This is consistent with a finding that assimilation context effects are stronger when assimilation does not create lexical ambiguity (c.f. Gaskell and Marslen-Wilson, 2001).

In summary, the results of Experiment 2 show that progressive and regressive assimilation effects may be shown within the same trials using the same task. This finding raises the possibility that these effects may be related, either through a common source, or may be dependent on one another. Furthermore, they suggest that lexical factors play some role in such effects.

6. Continuous Input – Graded Behavior

Results from both experiments suggest that progressive and regressive perceptual effects are not discrete. Viewed through a sensitive measure of lexical activation they are shown to vary in timecourse and strength. In this section we examine whether these gradations are related to variations in the acoustic manifestation of assimilation. Evidence from several cross-linguistic studies (Beddor, Harnsberger and Lindemann, 2002; Gow and Im, 2004) suggests that the strength of assimilation context effects may be a function of the degree of acoustic modification an assimilation produces.

To examine this relationship we conducted an acoustic analysis on each of the stimuli used in the previous two studies to extract their continuous features. The frequencies of F1, F2, F3 were measured for each item and then related across conditions (assimilated vs. non-assimilated) and to behavior in a series of statistical analyses. The first analysis examined whether modification due to assimilation is in fact graded in this experiment, or whether it is in the form of a discrete, feature-changing process. Three analyses compared F1, F2 and F3 for coronal, noncoronal or assimilated consonants. In addition to these measurements, labiality/velarity (of the noncoronal) was included as a factor (since the formant cues for a labial differ from those for a velar). If assimilation were complete, assimilated tokens should look the same as noncoronal tokens. This is not what was found. The analysis of F1 (figure 3, panel A) yielded

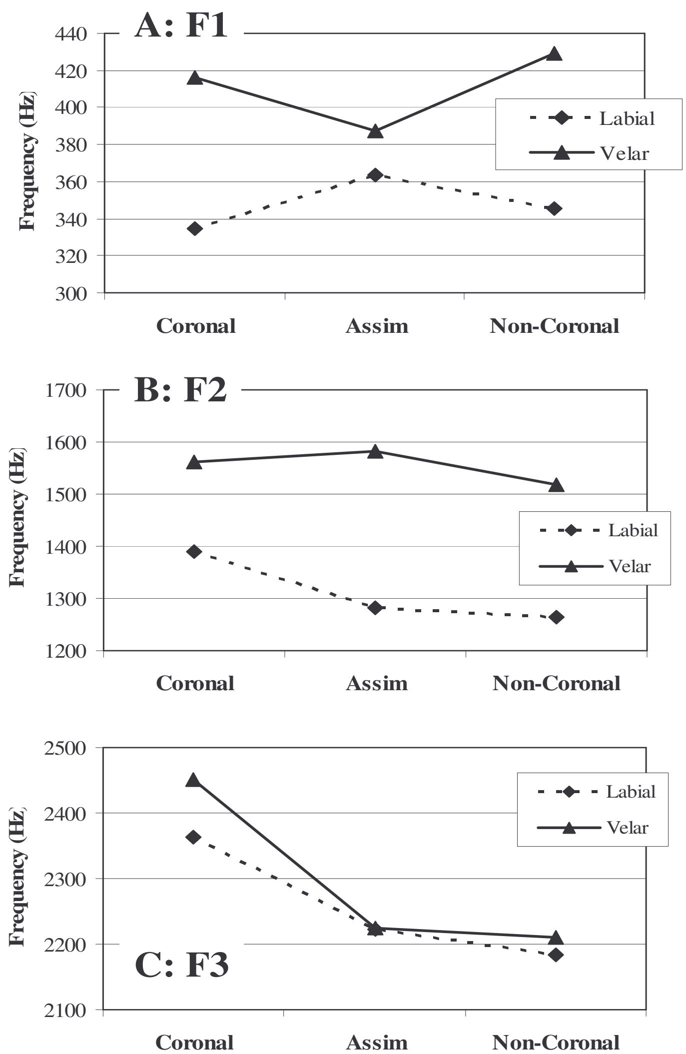

Figure 3: Formant frequencies for the first three formants of coronal, assimilated and non-coronal consonants used in Experiments 1 and 2.

no overall effect of consonant type ($F<1$), but a marginally significant interaction of consonant with labiality ($F(2,60)=2.9$, $p=.058$). This was due to the fact that assimilated consonants had more extreme values than either of the non-assimilated tokens, but this value was lower for velars and higher for labials ($F(1,30)=4.7$, $p=.039$). Analysis of F2 (figure 3, panel B) showed a pure linear trend—coronals had the highest frequencies, noncoronals the lowest and assimilated tokens were in between ($F(1,30)=4.6$, $p=.039$). This did not interact with labiality ($F(1,30)=1.1$, $p>.2$). Finally analysis of F3 (figure 3, panel C) yielded a significant linear trend ($F(1,30)=16.1$, $p<.001$) that did not interact with labiality ($F<1$).

Since assimilation yields graded modification, we next assessed the relationship between continuous acoustic variation and the strength of the progressive and regressive effects. The first analysis examined the progressive effect in Experiment 1. We computed the progressive effect size for each item, by pooling across subjects and subtracting the proportion of fixations in the coronal condition from the assimilated condition. This was the dependent variable in a regression analysis. Independent variables consisted of each of the difference between each formants' frequency for a coronal or assimilated item. This was used (as opposed to the raw frequency) because it better matched the DV (since it too was the difference between these two conditions) and because it takes into account differences in absolute F2 due to place of articulation or nasality. Additionally (as before) labiality and its interaction with each formant was added to account for place-specific effects. The formant frequencies together with labiality accounted for 23% of the variance, but were not significantly more than zero ($F<1$). However, when the interactions were included, the model accounted for 72% of the variance ($F(3,8)=4.7$, $p=.035$). Thus, continuous acoustic cues to assimilation are systematically related to the size of the progressive effect, but they must be interpreted with respect to place of articulation.

The same model was applied to the regressive effect in Experiment 2. Regressive effect size was computed in the same way as our temporal analyses of Experiment 2, but grouped across time and within items. With this as a dependent variable, the model accounted for a total of 52% of the variance, a substantial R^2, but was not significant due to the small number of data points (there were only 16 items). Thus, this analysis supports a relationship between continuous acoustics and the regressive effect, but cannot conclusively demonstrate it. While it is clear that increasing the power of the design with more items would help, it may also be the result of the fact that this analysis did not include acoustic cues from the post-assimilation context, which may be crucial for the perceptual processes that compensate for assimilation.

7. Discussion

The present experiments yield important findings on two aspects of the perceptual problem created by assimilation: the continuous nature of the input and the perceptual response, and the complex timing of the effects.

Our acoustic analyses of our naturally produced stimuli replicated a prior studies that demonstrated that the acoustic correlates of assimilation are continuous and gradient, not discrete and feature-changing. Assimilated segments have some features of coronals and non-coronals, some features of both, and some features that may be unique to assimilation. Similarly the present experiments show both progressive and regressive effects to be continuous as well. The progressive effect found in Experiment 1 took the form of a facilitation in the rise-time of lexical activation for the target. The regressive effect found in Experiment 2 was not a complete perceptual decision—context seemed to provide only a probabilistic bias to interpret an ambiguous sound as either a coronal or non-coronal. Finally, analyses relating the these continuous perceptual processes to continuous acoustic parameters yielded a compelling relationship (particularly Experiment 1). These results show that coping with phonological variability is a gradient problem, from signal to perception.

Analyses of temporal dynamics were focused primarily on determining the relative onsets and durations of the two effects. We found a very early progressive effect in Experiment 1 that appeared during the offset of the assimilated segment. This is just prior to the onset of the context, and thus is more consistent with a model in which progressive effects are not simultaneous with regressive effects. Experiment 2 also broadly support this class of models: the regressive effect appeared later than the brief progressive effect.

Complicating this explanation is the fact that the progressive effect in Experiment 2 was weaker and delayed relative to Experiment 1. We offer two (non-mutually-exclusive) explanations for these enigmatic results. First, our measure of activation for the post-assimilation context was confounded with our measure of activation for the assimilated word. That is, looks to the *cat box* reflect progressively heightened activation for *box* but also reduced activation for *cat* (due to assimilatory modification). This clearly played a role in the weakness of the progressive effect seen in Experiment 2, but it is unclear if it can account for its lateness. Second, the assimilated items in Experiment 2 were lexically ambiguous (or close to it). This could have slowed lexical competition dynamics. This would account for the lateness of the regressive effect, but also for the late progressive (relative to Experiment 1). This implies that perceptual inferences made in the process of resolving assimilation may be intimately bound up with the dynamics of lexical competition.

8. Conclusion

Here we have argued that phonological, phonotactic and phonetic processes create lawful regular variability that is distributed across time and provides a potentially rich information source. Two experiments examining the assimilation of place of articulation support this point by showing that continuous variability in the properties of coronals can facilitate recognition of upcoming items, and that phonetic context can help cope with variability caused by assimilation. Moreover it is clear that this project is richly temporal. Progressive effects appear early and before regressive effects. More interesting, the changes in temporal dynamics between Experiment 1 and 2 imply that dynamics of online lexical competition may play an important role in this process—when such dynamics are slowed by lexical ambiguity or competition, the temporal properties of cue integration change too.

This work suggests a complex three-way interaction between perceptual, phonological and lexical processes. Continuous lawful variability in the speech signal cannot be ignored by word recognition—it represents useful information. Likewise, the perceptual processes that cope with such information are affected by lexical dynamics. This work lies on the tripartite interface of these three domains of language science and represents only a small portion of the potential relationships and interactions between them. Empirical and theoretical work that bridges these domains will ultimately provide a richer set of questions and answers concerning fundamental speech processes.

References

Allopenna, P., Magnuson, J.S. and Tanenhaus, M.K. (1998) Tracking the time course of spoken word recognition using eye-movements: evidence for continuous mapping models. *Journal of Memory and Language*, **38(4**), 419-439.

Andruski, J.E., Blumstein, S.E. and Burton, M.W. (1994) The effect of subphonetic differences on lexical access. *Cognition*, **52**, 163-187.

Barry, M.C. (1985). A palatographic study of connected speech processes. *Cambridge papers in phonetics and experimental linguistics, vol. 4*. Department of Linguistics, University of Cambridge.

Beddor, P. S., Harnsberger, J. and Lindemann, S. (2002). Acoustic and perceptual characteristics of vowel-to-vowel coarticulation in Shona and English. *Journal of Phonetics, 30,* 591-627.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, **24**, 209-244..

Carney, A.E., Widin, G.P. and Viemeister, N.F. (1977) Non categorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, **62**, 961-970.

Chomsky, N., and Halle, M. (1968). *The sound pattern of English*. Cambridge, MA: MIT Press..

Cooper, R. (1974) The control of eye fixation by the meaning of spoken language. *Cognitive Psychology*, **6**, 84-107.

Dahan, D., Magnuson, J.S. and Tanenhaus, M.K. (2001) Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, **42**, 317-367.

Dahan, D., Magnuson, J.S., Tanenhaus, M.K. and Hogan, E. (2001b) Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, **16 (5/6)**, 507-534.

Darcy, I. (2002). Online processing of phonological variation in speech comprehension: The case of assimilation. In S. Hawkins and N. Nguyen (eds), *Temporal integration in the perception of speech* (p. 32). ISCA.

Darcy, I., and Dupoux, E. (this volume). Perceptual learning and plasticity in L2 learners: building a new system for phonological processes.

Gaskell, M.G. and Marslen-Wilson, W.D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, **22**, 144-158.

Gaskell, M.G. and Marslen-Wilson, W.D. (1998). Mechanisms of phonological inference in speech perception. Journal of Experimental Psychology: Human Perception and Performance, 24, 380-396.

Gaskell, M.G. (2001). Phonological variation and its consequences for the word recognition system. *Language and Cognitive Processes*, **16(5/6),** 723-729.

Gow, D. (2001) Assimilation and Anticipation in continuous spoken word recognition. *Journal of Memory and Language*, **45**, 133-139.

Gow, D. (2002) Does English Coronal Place Assimilation Create Lexical Ambiguity?. *Journal of Experimental Psychology: Human Perception and Performance*, **28(1),** 163-179.

Gow, D. (2003) Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, **65(4),** 575-590.

Gow, D. and Gordon, P. (1995) Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, **21(2),** 344-359.

Gow, D.W. and Holcomb, P. (November, 2002). Electrophysiological correlates of place assimilation context effects. Paper presented at the meeting of the Psychonomic Society, Kansas City, MO.

Gow, D.W., and Im, A. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language*, **51**, 279-296.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, **31**, 373-405.

Holst, T., and Nolan, F. (1995). The influence of syntactic structure on [s] to [ß] assimilation. In B. Connell and A. Arvanti (Eds.), *Phonology and phonetic evidence: Papers in laboratory phonology IV*. (pp. 315-333). Cambridge: Cambridge University Press.

Kerswill, P.E. (1985). A sociophonetic study of connected speech processes in Cambridge English: An outline and some results. *Cambridge Papers in Phonetics and Experimental Linguistics,* **4**, 1-39.

Kuhnert, B., and Hoole, P., (2004). Speaker-specific kinematic properties of alveolar reductions in English and German.

Liberman, A.M., Harris, K.S., Hoffman, H.S. and Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, **54(5),** 358-368.

Magnuson, J.S., McMurray, B., Tanenhaus, M.K. and Aslin, R.N. (2003). Lexical effects on compensation for coarticulation: the ghost of Christmash past. *Cognitive Science*, **27(2)**, 285-298.

Magnuson, J.S., Tanenhaus, M.K., Aslin, R.N. and Dahan, D. (2003). The microstructure of spoken word recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, **133(2)**, 202-227.

Marslen-Wilson, W.D., Nix, A. and Gaskell, M.G. (1995). Phonological variation in lexical access: Abstractness, inference and English place assimilation. *Language and Cognitive Processes*, **10**, 285-308.

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

McMurray, B., Aslin, R., Tanenhaus, M., Spivey, M., and Subik, D. (submitted). Two “B” or not 2 /b/: categorical perception in lexical and nonlexical tasks.

McMurray, B., Tanenhaus, M.K. and Aslin, R.N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, **86(2)**, B33-B42.

McMurray, B., Tanenhaus, M.K., Aslin, R.N. and Spivey, M.J. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research*, **32(1),** 77-97.

McMurray, B. (2004). Within-Category Variation is Used in Spoken Word Recognition: Temporal Integration at Two Time Scales. Doctoral Dissertation. University of Rochester.

Munhall, K.G., Gribble, P., Sacco, L., and Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception and Psychophysics*, **58**, 351-362.

Nolan, F.(1992). The descriptive role of segments: Evidence from assimilation. in Doherty Ladd, *Papers in laboratory phonology II: Gesture, segment, prosody*. pp. 261-280, Cambridge MA, Cambridge University Press.

Pisoni, D.B. and Sawusch, J.R. (1974) On the identification of place and voicing features in synthetic stop consonants. *Journal of Phonetics, 2,* 181-194.

Pisoni, D.B. and Tash, J. (1974) Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, **15(2),** 285-290.

Summerfield, Q. (1981) Articulatory rate and perceptual constancy in phonetic perception. Journal of the Acoustical Society of America, 7(5), 1074-1095.

Summerfield, Q. and Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, **62(2),** 435-448.

Tanenhaus, M.K., Spivey, M.J., Eberhard, K.M. and Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, **268**, 1632-1634.

Utman, J.A., Blumstein, S.E. and Burton, M.W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics*, **62(6),** 1297-1311.

Weber, A. (2001). Help or hindrance: How violation of different assimilation rules affects spoken-language processing. *Language and Speech*, **44(1),** 95-118.

Appendix A: Stimuli Used in Experiment 1

Cor_Non	**Cor_Cor**	**Non_Non**	**Non_Cor**
Green Boat	Green Dog	Swamp Boat	Swamp Dog
Brown Pig	Brown Tree	Slim Pig	Slim Tree
Red Monkey	Red Notebook	Plump Monkey	Plump Notebook
Maroon Goose	Maroon Duck	Patriotic Goose	Patriotic Duck
Gold Medal	Gold Necktie	Platinum Medal	Platinum Necktie
Tan Pants	Tan tie	Denim Pants	Denim tie
Salmon Blouse	Salmon Dress	Plum Blouse	Plum Dress
Violet Bikini	Violet Door	Cream Bikini	Cream Door
Crimson Clock	Crimson Tire	Pink Clock	Pink Tire
Scarlet Curtain	Scarlet Train	Black Curtain	Black Train
Chocolate Pie	Chocolate Tart	Grape Pie	Grape Tart
One Ball	One Daisy	Aluminum Ball	Aluminum Daisy
Seven Bandaids	Seven Darts	Drab Bandaids	Drab Darts
Ten Bats	Ten Dollars	Damp Bats	Damp Dollars
Nine Bees	Nine Dice	Dim Bees	Dim Dice
Eleven Goldfish	Eleven Dwarves	Dark Goldfish	Dark Dwarves

Appendix B: Stimuli Used in Experiment 2

Cor_Non	**Cor_Cor**	**Non_Non**	**Non_Cor**
Eight Babies	Eight Dolls	Ape Babies	Ape Dolls

Mat Book	Mat Drawer	Map Book	Map Drawer
Bud Case	Bud Tie	Bug Case	Bug Tie
Cat Box	Cat Drawing	Cap Box	Cap Drawing
Cone Bending	Cone Dropping	Comb Bending	Comb Dropping
Gun Popping	Gun Twins	Gum Popping	Gum Twins
Heart Breaker	Heart Doctor	Harp Breaker	Harp Doctor
Road Patch	Road Twisting	Robe Patch	Robe Twisting
Rat Gleaming	Rat Dangling	Rack Gleaming	Rack Dangling
Phone Box	Phone Donor	Foam Box	Foam Donor
Street Cleaner	Street Train	Streak Cleaner	Streak Train
Line Peeler	Line Tag	Lime Peeler	Lime Tag
Mast Creator	Mast Trophy	Mask Creator	Mask Trophy
Bite Guard	Bite Damage	Bike Guard	Bike Damage
Net Cushion	Net Trap	Neck Cushion	Neck Trap
Mud Gear	Mud Drinker	Mug Gear	Mug Drinker