# HOW PHONOLOGICAL CONTEXT AFFECTS COMPREHENSION: THE CASE OF ASSIMILATED NASALS AND STOPS

*Mercedeh Mohaghegh & Craig Chambers*

University of Toronto, Canada
mercedeh.mohaghegh@mail.utoronto.ca; craig.chambers@utoronto.ca

## ABSTRACT

Four forced-choice identification tasks examined the recognition of words containing sounds that have undergone the process of nasal place assimilation (‘*phone box’*: /n/→[m]) or stop place assimilation (‘*cat box’*: /t/→[p]). Identification scores and response times were measured for words ending in unassimilated or assimilated coronal consonants, which were either presented in isolation or within a carrier sentence that provided the triggering phonological context for place assimilation (i.e., where the word-final coronal consonant is followed by a word-initial labial sound). Identification scores showed that the presence of the context had a positive influence on listeners’ correct identification of the assimilated forms. Furthermore, this effect was comparable across nasal and stop consonants. However, response time measures showed that phonological context speeded the recognition of assimilated nasals but not assimilated stops. This finding is consistent with the idea that compensation for assimilation involves distinct processing mechanisms for nasals versus stop consonants.

**Keywords**: Spoken word recognition, phonological variation, place assimilation, speech perception.

## 1. INTRODUCTION

Lack of invariance in speech signal is one of the major challenges in the recognition of spontaneous speech. This issue becomes even more complicated in cases where production of speech sounds can result in ambiguity at the lexical level. Phonological processes, such as place assimilation, are one common source of such variability. In English, place assimilation is known to affect both coronal nasal and stop consonants, as when they take the place of articulation of a following labial (or velar) sound. For example, ‘*phone box*’ becomes perceptually similar to *‘foam box’* after undergoing place assimilation. A number of studies have suggested that, despite this assimilation, residual acoustic cues reflecting the original coronal place of articulation are detected by listeners and used in perceiving the underlying form of the sound [e.g., 6]. Most studies, however, have focused on the effect of phonological context on recognition of assimilated forms [e.g., 4, 7]. Many of these studies suggest that listeners can compensate for assimilated sounds when the following phonological context in the speech stream is recognized as a trigger for assimilation (i.e., when an assimilated coronal sound is followed by a labial or velar consonant). In the case of strong or complete assimilation that entails lexical ambiguity (e.g., *cat box* might plausibly be *cat box* or *cap box*), some studies have found no effect of phonological context and have suggested higher-level (e.g., syntactic) information to be the main resource used by listeners to overcome effects of assimilation [e.g., 5]. However, in these and other studies, the core comparisons involved cases where assimilated forms were extracted and spliced into contexts that would not trigger assimilation. Thus, it is unclear to what extent phonological context effects might be apparent in studies using naturally produced assimilated forms.

Another aspect of place assimilation that has received little attention in past studies is that although both nasal and stop consonants are known to undergo place assimilation in English, there are known acoustic and probabilistic differences in the nature of place assimilation across the two groups of sounds [3]. In two studies, Mitterer and his colleagues [8, 9] examined whether the viability of the trigger context influenced the recognition of Dutch words ending in assimilated nasals or stops, using an online eye tracking experiment and an offline discrimination task. He concluded that while the contextual information is used at early stages of perceptual processing for assimilated nasals, it is only used at later stages for stop consonants. This brings up the question of whether seemingly similar phonological processes are handled by the same or different underlying mechanisms.

The current study investigates the role of phonological context in the recognition of naturally assimilated words ending in either a nasal or a stop consonant. Results are reported for four forced-choice identification tasks using unassimilated and assimilated forms of real English words, presented either in isolation or preceding a word beginning with a labial consonant (a context that triggers assimilation). Listeners’ recognition of words in isolation was tested to establish the extent to which

residual acoustic cues aid in the perception of the underlying place of articulation. These measures also provide a baseline for analysing the impact of the following phonological context when it is available. Specifically, if phonological context helps with compensation for assimilation, then it should improve listeners' identification of the underlying forms compared to when the words occur in isolation. Finally, a direct comparison of the results for assimilated nasals and assimilated stops should reveal potential differences in the perceptual processing of these sounds.

## 2. PREPARATION OF MATERIALS

The target stimuli used in Experiments 1-4 consisted of 48 monosyllabic words. Half of these words ended in a coronal nasal and the other half ended in a coronal stop consonant (i.e., /n/ in Expts. 1 & 3 and /t/ in Expts. 2 & 4). All words were potentially lexically ambiguous when they underwent place assimilation (e.g. *line-lime*). The words were recorded within a carrier sentence in form of the instruction "Now click on the __ button". A female Canadian English native speaker was recorded producing two versions of the sentence. In the first version, a careful pronunciation of the sentences resulted in the production of words in which the final consonant was not altered by assimilation. In the second version a casual pronunciation of the sentences was adopted to produce assimilated forms. All stimuli were recorded in a sound attenuated booth. Multiple recordings were made for each version of the sentence and the best tokens for unassimilated and assimilated forms were chosen based on the experimenter's judgement. The acoustic properties of the selected tokens (e.g., first and second formant frequency measures and closure duration) were then verified by means of analyses of waveforms and spectrograms in Praat software (Version 5.3.23) [1].

The visual display used for each trial consisted of five orthographically labelled push buttons depicted on a computer screen. For each display there were two auditory stimulus items played (isolated words in Expts. 1 & 2 and full sentences for Expts. 3 & 4). The first auditory stimulus contained words that ended in a coronal sound but were lexically unambiguous, and these will not be discussed here. The second auditory stimulus contained the target item. The labels on the buttons corresponded to the first mentioned item (e.g., *lean*), the second mentioned item (target; e.g., *dine*), a competitor for the target that was the corresponding minimal pair ending in a labial consonant (e.g., *dime*) and two phonologically and semantically unrelated words (e.g., *glove, gulf*).

There were also 72 filler trials to counteract strategic expectations and disguise the manipulation of interest. Stimuli in filler trials consisted of words that did not end in coronal stops and had either semantic or phonological relationships that were not the same as in the critical trials.

## 3. EXPERIMENTS 1 & 2

Experiments 1 and 2 were conducted in order to test listeners' identification of isolated words that ended in assimilated nasal stops (Exp. 1) or assimilated oral stops (Exp. 2).

### 3.1. Method

#### *3.1.1. Participants*

Forty-eight native English listeners were paid for their participation. Participants were female and male adults between 18-40 years of age, who had no history of hearing or speech difficulty and had normal or corrected vision. All participants reported native proficiency in English and that they learned English in early childhood.

#### *3.1.2. Materials*

To create the auditory stimuli for Experiments 1 and 2, target words were clipped from their carrier sentences (e.g., *line* was clipped from "Now click on the *line* button."). Three beep sounds replaced the excised portion of the instruction that originally preceded the test words (duration of beeps: 1422 ms) and the word *button* was omitted.

For critical trials, each participant encountered an equal number of assimilated and unassimilated forms but a given word was presented only once in one of these two conditions. Four versions of each experiment were created, varying the pairing of words to assimilation condition, and each version was assigned to six participants. All stimuli were randomized and were controlled for their lexical frequency based on SUBTLEXus measures [2].

#### *3.1.3. Procedure*

Each participant was seated inside a sound booth in front of a computer screen and at a relatively fixed distance (~109 cm) from two loudspeakers. Participants' task was to click with the computer mouse on the displayed labeled button that best matched with the word they would hear after the three beeps. On each trial, the labelled buttons appeared 3000 ms before the first auditory stimulus

played to allow for enough time for scanning the screen. The second auditory stimulus played automatically after the participant's click response. The trial ended after the participant's second selection. The experiment took approximately 20 minutes in total.

### 3.2. Results

Participants' mouse clicks were analyzed for number of correct responses (i.e., correct selection of the target word, which ended in a coronal nasal or coronal stop) and response time (RT). Figure 1 shows the percentage of correct responses across conditions and Figure 2 shows the corresponding RTs for both experiments.

A paired t-test analysis was performed for each experiment to compare the percentage of correct responses in unassimilated and assimilated conditions. The results showed that the targets were identified significantly more often in unassimilated than assimilated conditions for both nasals ($t(23) = 22.54$ , $p < 0.001$) and stops ($t(23) = 20.97$, $p < 0.001$). A two-sample t-test (Welch's) across experiments revealed that stop targets were correctly identified more often than nasals in both the unassimilated ($t(27) = 6.5$, $p < 0.001$) and assimilated ($t(34) = 3.4$, $p = 0.001$) conditions. RTs (for correct trials) were also analyzed. RTs that were shorter or longer than three times the median absolute deviation (MAD) were identified as outliers. Outliers and missing data points were replaced by the average RT in each condition. A paired t-test analysis comparing the average RTs in unassimilated and assimilated conditions indicated that the RTs were not significantly different for nasals ($t(23) = 1.63$, $p = 0.11$). Surprisingly, for stops, RT was significantly longer in unassimilated condition ($t(23) = 3.01$, $p = 0.006$). This result was unexpected, and might be an unintended by-product of calculating the RT based on very few correct answers that were produced in the assimilated conditions. Finally, a two-sample t-test (Welch's) across experiments showed that the average RT for nasals was significantly longer than for stops in both unassimilated ($t(44) = 3.96$, $p < 0.001$) and assimilated ($t(27) = 4.68$, $p < 0.001$) conditions.

### 3.3. Discussion

The comparatively low identification scores for words ending in either assimilated nasal or stop consonants suggest that the degree of assimilation for final sounds in the current stimuli was strong. Compared with stops, nasal targets were correctly identified less often and their RTs for correct responses were also longer (even when unassimilated). This indicates that acoustic cues to distinguishing place of articulation (labial vs. coronal) are overall perceptually less salient in nasals than for stop consonants.

## 4. EXPERIMENTS 3 & 4

Experiments 3 and 4 were conducted in order to test the effect of the following phonological context on identification of the words ending in assimilated nasal stops (Exp. 3) or assimilated oral stops (Exp. 4).

### 4.1. Method

#### *4.1.1. Participants*

Sixty-four native English speakers from the University of Toronto, who did not participate in Experiments 1 or 2, were paid for their participation. Participants were recruited based on the same criteria as in Experiments 1 and 2.

#### *4.1.2. Materials*

The source stimuli were the same as those used for Experiments 1 and 2. However, in Experiments 3 and 4 the words were not clipped from their carrier sentences and the sentences were instead played in their entirety. Crucially, in Experiments 3 and 4, the initial sound of the word '*button*' that followed the target word created a context that could trigger assimilation. The provision of this word along with the target word in turn provides the means to test how phonological context affects listeners' ability to compensate for assimilation.

#### *4.1.3. Procedure*

Participants performed a task similar to the task that was described for Experiments 1 and 2. On each trial, participants followed two auditory instructions by clicking on a button (e.g., "Now click on the *dine* button"). Each experiment took approximately 40 minutes in total.

### 4.2. Results

The analysis included participants' mouse clicks for the number of correct responses (identification scores), as well as their RTs for correct trials (Fig. 1 & 2, respectively).

A set of t-test analyses of the identification scores again showed greater accuracy in identifying the target word in the unassimilated condition compared to the assimilated condition for both nasals ($t(31) = 16.46$, $p < 0.001$) and stops ($t(31) = 15.68$, $p < 0.001$). Further, there was significantly greater

success at target identification for stops compared to nasals in both unassimilated ($t(33) = 5.45$, $p < 0.001$) and assimilated ($t(62) = 4.08$, $p < 0.001$) conditions.

Another set of t-test analyses was performed on the RTs for the correct responses after replacing the missing data and outliers with the average RT in each condition. The results showed that the RTs were significantly longer in assimilated condition compared to unassimilated for both nasals ($t(31) = 2$, $p = 0.05$) and stops ($t(31) = 5.15$, $p < 0.001$). However, unlike in Experiments 1 and 2, the average RT for recognizing words with assimilated nasals versus assimilated stops did not turn out to be significantly different ($t(59) = 0.25$, $p = 0.79$).

In a final analysis, the results from Experiments 3 and 4 were compared with the corresponding results from Experiments 1 and 2. Interestingly, a two-sample t-test (Welch's) comparing the identification scores revealed a significant and relatively fixed increase (M = 26%) in the correct identification of assimilated targets for both nasals and stops when the following phonological context was available (nasals: $t(48) = 6.78$, $p < 0.001$; stops: $t(52) = 6.27$, $p < 0.001$). The same test comparing the average RTs across the two sets showed a significant decrease when the context was provided for assimilated nasals ($t(36) = 3.36$, $p = 0.001$) and unassimilated nasals ($t(46) = 6.34$, $p < 0.001$) and unassimilated stops ($t(33) = 7.15$, $p < 0.001$) but no significant difference for assimilated stops ($t(51) = 1.7$, $p = 0.09$).

**Figure 1**: Percentage of correct responses in Experiments 1-4.

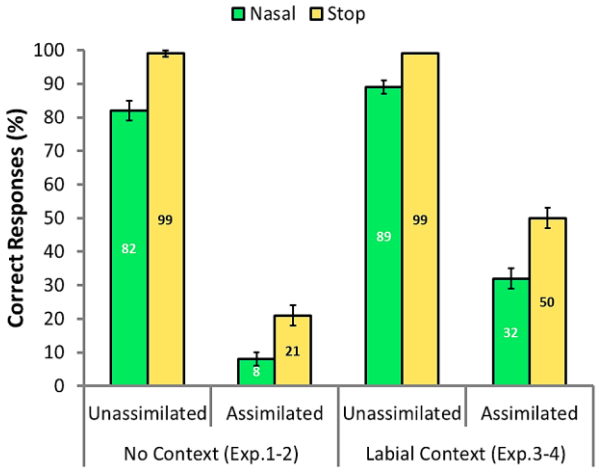
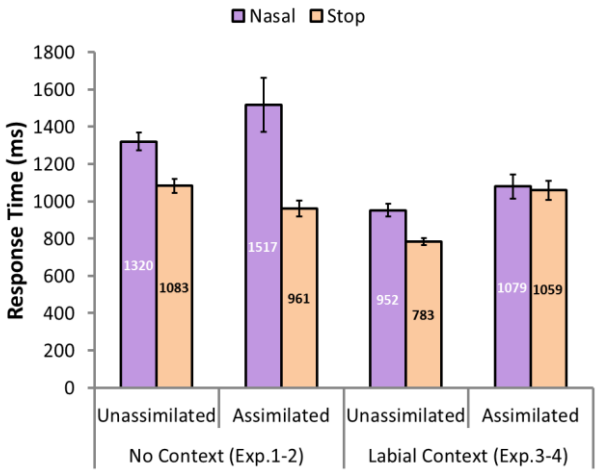



**Figure 2**: Average response time for correct identifications in Experiments 1-4.




### 4.3. Discussion

The results from Experiments 3 and 4 indicate that in fact listeners make significant use of phonological context in recognition of assimilated words. Availability of the following phonological context resulted in an automatic improvement in the recognition of assimilated forms, an effect that was comparable for nasal and stop consonants. In addition, RT results followed the same pattern by showing faster responses for assimilated nasals when the context was available. However, the availability of the phonological context did not significantly change RTs for assimilated stops. This suggests that, even though the provision of the context improved listeners' ability to correctly recognize the words ending in assimilated stops, the place cues were still strongly influencing the perceptual processing, which in turn resulted in showing no change in the processing latency compared to when the context was not provided.

## 5. GENERAL DISCUSSION

The results of the current study show that phonological context influences listeners' comprehension of assimilated forms, independently of the acoustic cues of place of articulation or the type of assimilation process (i.e., nasal or stop place assimilation). However, RT results suggest that there are some processing differences for assimilated nasals vs. stops. For stop consonants, even when phonological context was provided, the strong influence of acoustic cues resulted in no improvement in processing speed compared to when the context was not available. For nasals however, phonological context speeded the identification of assimilated forms even though acoustic cues favoured a labial interpretation. This could reflect the overall low saliency of the acoustic cues to place of articulation in nasals compared to stops, in turn leading to higher dependency on phonological context for correct identification of assimilated forms. Together, the results provide an important extension to earlier claims that the recognition mechanisms involved in dealing with seemingly similar processes might not necessarily be the same [9].

## 6. REFERENCES


[1] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10), 341-345.

[2] Brysbaert, M. and New, B. 2009. Moving beyond Kucera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English. *Behavior Research Methods*, 41(4), 977–990.

[3] Dilley, L., & Pitt, M. 2007. A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *J. Acoust. Soc. Am.*, 122, 2340–2353.

[4] Gaskell, M., & Marslen-Wilson, W. 1998. Mechanisms of phonological inference in speech perception. *J. Exp. Psychol.*: *Human Perception and Performance*, 24, 380‑396.

[5] Gaskell, M., & Marslen-Wilson, W. 2001. Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language*, *44*(3), 325-349.

[6] Gow, D. 2002. Does English coronal place assimilation create lexical ambiguity?. *J. Exp. Psychol.: Human Perception and Performance, 28,* 163–179.

[7] Gow, D., & McMurray, B. 2007. Word recognition and phonology: The case of English coronal place assimilation. In J. Cole, & J. Hualde (Eds.), *Lab. phon. 9*, 173–200. New York: Mouton de Gruyter.

[8] Mitterer, H. 2011. Recognizing reduced forms: Different processing mechanisms for similar reductions. *J. Phon.*, *39*(3), 298-303.

[9] Mitterer, H., & McQueen, J. 2009. Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *J. Exp. Psychol.*: *Human Perception and Performance*, 35, 244–263.