Perceptual reality of phonetically-driven phonology: Place assimilation and consonant cluster simplification at different prosodic boundaries

Taehong Cho & James McQueen
Max Planck Institute for Psycholinguistics

It has been proposed that some phonological patterns are best explained by the acoustic/perceptual characteristics of segments (see [1]). For example, alveolars tend to undergo place assimilation more often than labials, which are in turn more susceptible to place assimilation than velars. This asymmetry in place assimilation is arguably accounted for by the so-called Production Hypothesis: speakers make a greater effort to preserve perceptually stronger segments (velars) than perceptually weaker segments (alveolars), from which listeners benefit ([2,3]; cf. [4]). Such a phenomenon is also seen in tri-consonantal cluster simplification in Korean: for a C1C2C3 sequence, when C2 is optionally deleted, speakers delete /t/ most often and /k/ least often. This has been interpreted as implying that /k/ is kept more often because its perceptual salience is greater than /t/ ([5]). However, the perceptual reality of such interpretations have not yet been corroborated by speech perception data.

We evaluated the perceptually-grounded Production Hypothesis by investigating how the asymmetry in place assimilation and cluster simplification in Korean would influence phoneme recognition. 104 Seoul Koreans participated in a series of phoneme monitoring-restoration experiments in which they monitored for an underlying coda target in either an assimilated (/t/ or /p/ -> [k]/_k) or an unassimilated real word form (Exp.1); in either an illegally assimilated (/p/ or /k/->[t]/_t) or an unassimilated real word form (Exp.2). In another experiment (Exp.3), they had to monitor for C2 when it was present in a C2-kept word ({C1}C2C3, ‘{}’=a deleted phoneme) or restore C2 when it was absent in a C2-deleted word (C1{C2}C3). In order to examine whether and how the phoneme recognition process is constrained by prosodic structure, words with target phonemes were placed Accentual Phrase (AP) -initially and AP-medially (cf. [6,7])

Several findings have emerged. First, the speed and accuracy of recognition of the unassimilated targets were in general greatest for velars, intermediate for labials and smallest for alveolars in line with the hypothesized perceptual salience ranking: velars > labials > alveolars. More specifically, comparing the unassimilated targets with assimilated ones, the recognition of the unassimilated targets were faster and more accurate than assimilated ones only when the targets were labials (but not when the targets were alveolars) (Exp.1).

Second, comparing the illegally assimilated targets with the unassimilated ones, the recognition of the illegally assimilated targets (e.g., restoring /k/ when /k/-> [t]/_t) in real word forms was significantly slower and less accurate than that of the unassimilated targets. However, there was a significant interaction with the consonant type: the effect was significantly far greater for velars than for labials (Exp.2). This supports the hypothesis that velars are perceptually more salient than labials, such that the loss of (more salient) velars due to illegal assimilation is more harmful than that of labials, making the restoration of the former harder than the latter.

Third, for the cluster C1C2C3 which is simplified by deleting either C1 or C2, when the C2 targets in words were velars, the recognition of C2 was significantly faster for {C1}C2C3 items (with the velar target present) than for C1{C2}C3 items (with the velar target absent). Crucially, however, when the C2 targets in words were labials no such effect was found, suggesting that listeners benefit from the preservation of velars, but not from that of labials (Exp.3).

Finally, as for prosodic boundary effects, the differences in phoneme recognition/restoration due to segment type were significantly more robust AP-initially than AP-medially, indicating that listeners are more sensitive to phonological changes at a stronger prosodic boundary.
Overall, the results suggest that asymmetries in place assimilation and consonant cluster simplification in Korean are indeed perceptually grounded, supporting the view that phonetics can be a source of explanation for some phonological patterns. Moreover, we infer that the human speech recognition system is sensitive to the acoustic/perceptual properties of individual segments to a different degree, which is further modulated by the prosodic structure of a given language.

References
Prosodic and phonotactic influences on fricative voicing assimilation in German

Claudia Kuzla & Taehong Cho
Max Planck Institute for Psycholinguistics

Recent phonetic studies have shown that the phonetic realization of segments varies with the prosodic position in which they occur. For example, domain-final segments are longer than domain-initial ones; segments are produced more strongly domain-initially than medially; these effects increase in magnitude as the prosodic domain becomes larger, i.e. phrase-initial segments are stronger than word-initial (but phrase-medial) ones. (Edwards, Beckman & Fletcher, 1991, JASA; Fougeron & Keating, 1997, JASA; Cho & Keating, 2001, JPhon; Cho, 2003, 15th IPhS).

In this study, we examined the acoustic realization of the initial German fricatives /f, v, z/ in different prosodic positions. In particular, we investigated whether the assimilatory devoicing of voiced fricatives /v, z/ after a voiceless obstruent (e.g., /z/ -> /s/ t_) is a gradient process sensitive to prosodic structure, i.e. varying with the size of intervening prosodic boundaries. It was hypothesized that a smaller prosodic boundary between segments increases the cohesion between them such that the degree of voicing assimilation of /z, v/to voiceless /t/ is greater across a smaller prosodic boundary. In addition, we examined how the hypothesized variation in voicing assimilation due to prosodic position is further conditioned by a phonotactic constraint: In German, while both voiced and voiceless labiodental fricatives /v, f/ occur word-initially, only the voiced alveolar fricative /z/ (not /s/) is allowed in that position. Excessive assimilatory devoicing of voiced /v/ would confound the phonological contrast with its voiceless counterpart /f/, which would constrain the voicing assimilation of /v/. In contrast, no such restriction would apply to /z/, allowing for more voicing assimilation.

Ten speakers read various sentence types intended to elicit different prosodic boundaries. Two versions of each sentence type were used: one containing a stop-fricative sequence (e.g., /…t#v…/, # = prosodic boundaries, hat[t]#[v]?lder ‘has forests’), the assimilation environment, and one containing a vowel-fricative sequence (e.g., /…E#v…/, hatte#[v]?lder ‘had forests’), as a control condition. The elicited utterances were grouped into three prosodic categories (Major, Minor, and Word), with regard to the size of the prosodic boundary in the test sequences. The Major boundary was defined as having both a pause and an intonational marker (“boundary tone”); the Minor boundary as having no pause, but an intonational marker; and the Word boundary as having neither a pause nor an intonational marker. Two speakers were excluded from the analysis because they did not produce any tokens belonging to the Minor boundary. Acoustic measurements included preboundary syllable duration, fricative duration, and the amount of voicing as measured by percent (%) of voicing during the fricative.

Results are as follows. First, the three-way prosodic grouping was supported by the preboundary (final) lengthening pattern of Major > Minor > Word. Second, there was a main effect of boundary size on the fricative duration, but the effect was not completely in accordance with the domain-initial strengthening patterns, showing only a pattern of Minor > Word. Third, the relative amount of voicing showed a voicing assimilation effect: both /v/ and /z/ were realized with less voicing after /t/ than after a vowel. Considering /z/, the amount of voicing assimilation varied with prosodic boundary size, showing a general pattern of Major > Minor > Word. This supports the hypothesis of voicing assimilation as a gradient process sensitive to prosodic structure. However, when /v/ was considered, no such effect was found. This asymmetry between /z/ and /v/ suggests that the effect of prosodic boundary is constrained by the
language-specific phonotactics, such that when there is a need to maintain a phonological contrast (/v/ vs. /f/), the boundary effect is suppressed.

The results of this study suggest that phonetic and phonological processes, as found in the fricative duration and voicing assimilation in German, are bounded by the interplay between prosodic structure and phonotactics of the language.
Tracking the timecourse of multiple context effects in assimilated speech

Bob McMurray & David Gow
University of Rochester

Recent research has demonstrated the sensitivity of lexical activation to fine-grained phonetic detail (Andruski, Blumstein & Burton, 1994; Gow & Gordon, 1995; McMurray, Tanenhaus & Aslin, 2002). However, few studies have systematically examined how such detail is integrated over time in word recognition. Place assimilation is an ideal context in which to examine these issues. In English, segments with underlying coronal place often show partial assimilation to the place of a subsequent labial or velar (e.g. /m/ or /k/). This produces a surface form with acoustic and articulatory properties that are partially consistent with both coronal and non-coronal place (e.g. Green Boat => Green\textsubscript{m} Boat). Listeners can use the resulting acoustic/phonetic cues to anticipate upcoming material and use post assimilation context to resolve the ambiguous consonant (Gow, 2001; Gaskell & Marslen-Wilson, 1998). However, there has been no detailed examination of the timecourse of these effects to probe their relationship to one another or to lexical activation. Thus, we used an eye-tracking task in which subjects respond to spoken instructions to manipulate visual objects whose names have linguistic properties of interest (Tanenhaus, Spivey-Knowlton, Sedivy & Eberhart, 1995). Fixation probabilities in this task mirror the activation of lexical items as speech unfolds over time (Allopenna, Magnuson & Tanenhaus, 1998).

Experiment 1 examined anticipation. 21 subjects heard phrases such as “Select the maroon goose” in which “maroon” was either naturally assimilated or unassimilated. Subjects selected (with a mouse) the picture from a screen containing all four combinations of word-final coronal/non and word-initial coronal/non: a maroon goose, a maroon duck, a patriotic goose and a patriotic duck. More eye-movements to the target (maroon goose) were observed immediately after an assimilated than an unassimilated coronal (t(20)=2.13, p=.046) and there were fewer eye-movements to the competitor, maroon duck (t(20)=2.43, p=.024). These results suggest that assimilation can help prepare the system by increasing priors for potential non-coronal targets, and decreasing them for potential coronal competitors.

Assimilation can often create lexical ambiguity: “cat” is ambiguous between “cat” and “cap”, when it assimilates bilabial place (Cat Box => Cat\textsubscript{p} Box). This ambiguity can be resolved when the post-assimilation context is integrated with the subphonemic cues for assimilation in the ambiguous consonant. In Experiment 2 subjects heard phrases like “Select the cat box”, in which cat was either assimilated or not. Screens contained pictures of “cat box” (coronal-non), “cat drawing” (coronal-coronal), “cap box” (non-non) and “cap drawing” (non-coronal). Eye-movements were analyzed for looks to the item corresponding to the initial coronal (cat vs. cap). Late in the trial, more looks were seen to the initial coronal (cat) than the non-coronal (cap) when the second item was non-coronal (box). The reverse was found (F(1,20)=11.2, p=.003) when the second item was a coronal (drawing). This supports a regressive ambiguity resolution mechanism consistent with feature-parsing (Gow, 2003). Additionally, anticipatory effects on early eye-movements were marginally significant for the correct target, e.g. “box” (t(20)=1.84, p=.081), and significant for looks to the competitor (t(20)=2.45, p=.024), suggesting both anticipatory and regressive mechanisms may be active for the same items (though not necessarily simultaneously).
These experiments support the on-line use of subphonemic detail for anticipating upcoming material and resolving ambiguity and suggest that the eye movement paradigm may be ideal for studying the integration of acoustic detail over time. We argue for active recognition process that is sensitive to phonological regularities in the signal and in which information integrated over a relatively large temporal range.

References
The Role of Gestural Overlap in Perceptual Place Assimilation: Evidence from Korean

Minjung Son*‡, Alexei Kochetov*† & Marianne PouplierΘ*
Yale University‡, Haskins Laboratories*, Simon Fraser University† and University of MarylandΘ*

The question of the origins of place assimilation has been a subject of increasing interest and controversy [1, 2, 4, 7, 8, 9, 11]. While researchers disagree on specific factors that cause perceptual place assimilation, it is commonly regarded as an important source of phonological assimilation [1, 8]. In particular, there are opposing views on whether perceptual assimilation is exclusively attributable to gestural reduction [5, 6] or can be triggered by gestural overlap as well [1, 3, 12]. The present study investigates whether perceptual place assimilation is uniquely attributable to either gestural reduction or overlap.

As an argument for the gestural reduction hypothesis, Jun [5, 6] presented data on regressive place assimilation in Korean /pk/ clusters. Investigating changes of oral pressure during the production of /pk/, Jun found that some tokens displayed gradient gestural reduction but no overlap, while other tokens showed gestural overlap without any reduction. In a follow-up perceptual study with Korean and English listeners he found that partially reduced tokens of /pk/ were overwhelmingly perceived as assimilated [kk], while overlapped tokens did not undergo perceptual assimilation. However, Jun's methodology did not allow him to distinguish between a partially and a fully reduced lip gesture. His assumption that Korean assimilation is a gradient, postlexical process was based largely on theoretical grounds. Additionally, all the tokens he identified to be partially reduced contained the sequence /pkw/. The small changes in oral pressure observed during these tokens can thus not uniquely be attributed to a potentially reduced /p/ or coarticulation of the velar labial-glide sequence.

The present study reports articulatory movement data recorded with an EMMA system [10] during the production of /pk/ clusters by a single Seoul Korean speaker. Stimuli consisted of real words with /pk/ clusters word-medially and at word boundaries, produced at two speaking rates; nonwords with word-medial /pk/ clusters were also collected. The results show that in word-medial /pk/ clusters the lip gesture was either fully present (with varying degrees of overlap) or completely absent, regardless of speech rate. Notably /pk/ clusters across word boundaries and word-medially in nonwords exhibited substantial overlap, but no reduction of C1. Contrary to Jun's hypothesis, the results indicate that word-medial lip reduction during /pk/ in Korean is a categorical rather than a gradient, postlexical process (though its occurrence is optional, or stochastic). Crucially, Jun's listeners failed to identify a /p/ in reduced /pk/ clusters not because of perceptual place assimilation, but likely because it was not articulated by his speakers.

Using the production data obtained in the EMMA experiment, we conducted a phoneme identification experiment with Korean and English listeners, presenting them with VC and VCCV stimuli that exhibited either overlap or categorical reduction of the lip gesture. Data from 10 subjects show that listeners often failed to recover the lip gesture in overlapped tokens and perceived [kk]. This was observed in up to 39% of the cases for Korean listeners and in up to 36% for English listeners. Unsurprisingly, a fully reduced lip gesture in /pk/ was not recovered by listeners, but perceived as [kk]. While the results allow us to conclude that overlap does play a role in perceptual place assimilation independent of reduction, any unique contribution of reduction cannot be investigated on the basis of Korean. The results suggest that gestural overlap
can be regarded as an important factor in the evolution of language-particular patterns of assimilation [3, 8].

References
Why Some Things are Um-bearable: Investigating Perceptual and Articulatory Influences on Place Assimilation Processes

Steve Winters
Indiana University

Several cross-linguistic surveys have revealed that place assimilation processes target nasals more often than stops in the languages of the world. These surveys have also revealed that there is an implicational relationship between nasals and stops as targets of place assimilation: stops never undergo this process in any given language unless nasals do so, as well. Some researchers (e.g., Jun 1995, Boersma 1998) have hypothesized that there are perceptual reasons behind the asymmetric behavior of nasals and stops in place assimilation, suggesting that nasals are more susceptible to this process because they have weaker acoustic cues to their place of articulation.

This hypothesis was tested by investigating the ability of Dutch and English listeners to discriminate between nasals and stops of varying places of articulation in an AX discrimination experiment. Speakers of these two languages were chosen because Dutch and English exhibit different assimilatory patterns with respect to nasals and stops in connected speech. In Dutch, only coronal nasals are susceptible to place assimilation (Booij 1995), while in English, both coronal nasals and stops may undergo this process (Barry 1991). The listeners in this experiment were asked to discriminate between consecutive VCCV sequences; the consonants in these sequences were either nasal-stop or stop-stop clusters. These nasals and stops were all produced at either the labial, coronal, or dorsal points of articulation. The analysis of listeners’ ability to discriminate between these VCCV sequences focused on VCCV pairs which differed in the place of articulation of the initial consonant, since these consonants are the most likely targets of place assimilation in English and Dutch. Listeners more easily discriminated stops than nasals in this context when the stops had audible release bursts; without these cues, stops did not have a perceptual advantage over nasals. The language of the listener had no significant effect on the discriminability of place in either nasals or stops. These results suggest that the hypothesized perceptual advantage of stops over nasals is mediated by the likelihood of speakers to produce them with audible release burst cues.

A subsequent experiment investigated the possibility that the difficulty of articulating consonant clusters consistently and accurately might account for nasals’ comparative susceptibility to place assimilation. To test this hypothesis, Dutch and English speakers participated in a production experiment wherein they listened to and attempted to imitate the VCCV stimuli that were used in the AX discrimination experiment. The results of this study showed that speakers of both languages exhibited less accuracy in reproducing nasals than stops. Examining the acoustic variability of subjects’ spoken repetitions also revealed an interaction between manner of articulation and the language of the speaker: nasal-stop sequences had greater durational variability than stop-stop sequences for only Dutch speakers—for English speakers, the durational variability of these sequences was the same. This articulatory variability seems to correspond to the different assimilatory patterns in the two languages. Together, these results suggest that the cross-linguistic asymmetry between nasals and stops as targets of place assimilation may be motivated more by the difficulty of articulating nasals in consonant clusters than by listeners’ relative inability to perceive their place of articulation correctly.
References Cited