# Inhibitory mechanisms in speech planning maintain and maximize contrast

**Sam Tilsen***
**University of Southern California**

Word count: 8,044

*Affiliation: University of Southern California

Mailing address:
Sam Tilsen
University of Southern California
Department of Linguistics
Grace Ford Salvatori 301
Los Angeles, CA 90089-1693
Email: tilsen@usc.edu

Keywords: Contrast; dissimilation; coarticulation; hypocorrection; dispersion; motor planning; inhibition.

## 1. Introduction

This chapter proposes that an inhibitory speech planning mechanism is involved in the maintenance and maximization of phonological contrast. The maintenance of contrast is of central importance to the understanding of phonologization. Generally speaking, assimilatory coarticulation will, unchecked, lead to contrast neutralization. Yet loss of contrast is far from the inevitable consequence of coarticulation; this implies that there exist cognitive mechanisms that oppose the phonologization of coarticulation. A complete theory of phonological change requires an account not only of the mechanisms that lead to loss of contrast, but also the ones that preserve contrast.

Limits on coarticulatory variation are commonly attributed to forces or constraints that maximize the *perceptual* distinctiveness of contrast. Dispersion theories (Liljencrants & Lindblom, 1972; Lindblom, 1986, 1990, 2003; Flemming, 1996, 2004) assert that there exist cognitive mechanisms which function to make speech targets less perceptually similar. The reader should keep in mind that sound systems never literally maximize perceptual differences between sounds, because other things, like coarticulation, often oppose the maximization of perceptual distinctiveness.

Recent experimental work on speech motor planning suggests an alternative view of how contrast is maintained: inhibitory interactions between contemporaneously planned articulatory targets result in dissimilatory effects, and over time these effects can prevent speech targets from becoming perceptually indistinct. For example, experimental observations show that speakers tend to produce an [i] with more peripheral F1 and F2 values when they have very recently planned an [a] (Tilsen, 2009). Likewise, experimental results presented in this chapter show that Mandarin speakers dissimilate tones that are planned in parallel. Findings of this sort suggest that the planning of a speech target is influenced by other simultaneously planned targets. These dissimilatory effects can be understood to arise from inhibitory motor planning mechanisms, and can explain how speakers maintain and maximize contrast.

We will use the phonologization of vowel-to-vowel coarticulation into vowel harmony as a representative example of phonologization processes associated with assimilatory phonetic patterns. This sort of phonologization falls under a general category of sound changes considered to arise from hypocorrection (Ohala, 1981, 1993). Section 2 describes how Ohala's listener-oriented theory of hypocorrective sound change applies to coarticulation, contextualizes this theory in an exemplar-based model of memory, and discusses how dispersion theories model the forces counteracting this process via maximization of perceptual contrast. Section 3 will describe experimental evidence for dissimilation between contemporaneously planned vowels in speech, and will present new experimental evidence that indicates tones in Mandarin exhibit the same

effect. Section 4 discusses these experimental results, argues that they arise from an inhibitory mechanism in the planning of articulatory targets, and explains the importance of this mechanism for understanding phonologization: i.e. inhibition functions to maintain and maximize contrast.

## 2. Background

To exemplify how hypocorrection leads to sound change, and how dispersion theory models the forces opposing this process, we use carryover vowel-to-vowel coarticulation as an example. Vowel-to-vowel coarticulation is an assimilatory influence upon the articulatory movements of one vowel due to the presence of a nearby vowel. Vowel-to-vowel (henceforth V-V) coarticulation is either anticipatory or carryover, and both types have been observed in a variety of languages (Öhman, 1966; Gay, 1974, 1977; Bell-Berti & Harris, 1976; Fowler, 1981; Parush et al., 1983; Recasens, 1984; Recasens et al., 1997; Manuel & Krakow, 1984; Manuel, 1990). Carryover coarticulation in $V_1$-$V_2$ sequences may arise from a combination of several factors. Mechanical constraints on the movement from the articulatory posture for $V_1$ to the posture for $V_2$ may give rise to coarticulation (Recasens, 1984; Recasens et al., 1997). Another potential source of coarticulation is gestural overlap, which in the task dynamic framework of articulatory phonology (Saltzman & Munhall, 1989; Browman & Goldstein, 1986, 1988, 1990), would arise when the gestural activation interval for $V_1$ extends into the time during which $V_2$ is active.

However, mechanical constraints and gestural overlap cannot be the only sources of V-V coarticulation because they are not expected over the observed the temporal range of V-V coarticulation, which can span up to three syllables (Fowler, 1981; Magen, 1997; Grosvald, 2009). A third possibility is that when the articulatory targets for $V_1$ and $V_2$ are planned contemporaneously, those targets may interact, resulting in assimilatory shifts in the target of $V_2$ toward $V_1$, or vice versa (cf. Whalen, 1990). In other words, prior to articulation, there may be variation in the formation of vowel targets that is influenced by other vowel targets in the preceding and subsequent utterance context, which are planned in parallel. Interestingly, the experimental evidence indicates that these interactions are predominantly dissimilatory in nature, and hence tend to oppose the effects of mechanical factors and gestural overlap.

In the highly influential model developed by Ohala (1981, 1993, 1994), V-V coarticulation, and more generally any form of assimilatory coarticulation, can lead to sound change through *hypocorrection*. In this process, sound change begins with a "phonetic perturbation" that frequently occurs in a given linguistic context. The sources of such perturbations can be mechanical, aerodynamic, motoric, and/or perceptual. Carryover V-V coarticulation is one example. The normal functioning of the perceptual apparatus, in this view,

is to compensate for the contextually conditioned perceptual similarity of $V_2$ to $V_1$. In a sense, compensation "corrects" or "normalizes" for the perturbation in $V_2$, undoing its effects on the perception and memory of the sound.

Hypocorrection occurs when the compensatory mechanism under-corrects for phonetic perturbations: “in the vast majority of cases the listener (somehow) parses the signal correctly and infers the speaker’s intended pronunciation. But occasionally a listener may misparse the signal” (Ohala, 1994). The key idea here is that the perturbation is “parsed as independent of the perturbing vowel”. The correction mechanism fails to compensate for coarticulation, and so a subtle phonetic assimilation is reinterpreted as a new pronunciation norm. In the case of V-V coarticulation, hypocorrection leads to vowel harmony, a contrast neutralization in which the vowels in some structural domain (e.g. a root, stem, or word) covary in some of their features (cf. Vernaud, 1980; Rennison, 1990; Krämer, 2001; Finley, 2008).

It is important to note that for phonologization to occur a new “pronunciation norm” must be established both within an individual speaker and across a group of speakers. Exemplar theories (Goldinger, 1992, 1996, 1998; Johnson, 1997, 2006; Pierrehumbert, 2001, 2002) provide a useful way to understand how sound change occurs within a given speaker. In the exemplar model of perception developed in Johnson (1997), every perceived speech sound is stored in memory as a separate exemplar. The exemplars incorporate phonetic details of the particular instantiation of the sound, along with a variety of contextual information and associations to categorial labels. Each exemplar is assumed to have an activation level—its relative salience in memory, which is influenced by its recency and potentially many other contextual factors, such as the word in which it occurred, nearby segments, the listener, speaker, etc. Hence the memory of a sound is not an abstract category, but a large collection of detailed exemplars that include, among other things, spectrotemporal information.

On the production side, the exemplar model described in Pierrehumbert (2001, 2002) uses the collection of stored exemplars to form a production target in the following way. First, an exemplar is randomly selected, then a weighted average of the phonetic values of similar exemplars is taken in order to form a production target. The activation level is a factor in the weighting, and hence more recent exemplars will play a greater role in target formation. The phonetic values are considered to be perceptually or articulatorily relevant variables, which for vowels includes formant values. Moreover, the categorial labels and phonetic values can be used to define a similarity metric, allowing for a notion of "similar" exemplars.

In the context of this model, regularly present phonetic perturbations can gradually shift the distribution of exemplars in phonetic space. For example, frequent carryover V-V coarticulation will tend to assimilate the target of $V_2$ to $V_1$ in that context. This happens because each time a production target is formed, previously stored exemplars influenced the weighted

averaging. Furthermore, the exemplar memory of a given speaker is part of a network of interacting agents, each with their own exemplar memory. If the phonetic perturbations occur with sufficient frequency across the population, then memories of both self-generated and other-generated sounds will feed into the sound change (cf. Oudeyer, 2006; Pierrehumbert, 2004; Wedel, 2004). Left unchecked, this will lead to partial contrast neutralization, and in the present example, vowel harmony. What, then, opposes these tendencies?

Dispersion theories describe a formal approach to understanding the maintenance and maximization of contrast, but these approaches do not explain how speakers accomplish these things. There are two prominent dispersion theories we consider here. The adaptive dispersion theory of Liljencrants & Lindblom (1972)—cf. also Lindblom (1986, 1990, 2003)—models vowels as mutually repelling objects in a perceptual space (e.g. a 2-D F1,F2 space), and models vowel system organization as an optimization problem. In contrast, the constraint-based approach of Flemming (1996, 2004) employs three goals, implemented as constraints: minimize articulatory effort, maximize the number of contrasts, and maximize the perceptual distinctiveness of contrasts.

Both approaches have in common an appeal to a cognitive mechanism which functions to make perceptual contrasts maximally distinct, and both require that this mechanism coexists with factors that indirectly reduce perceptual distinctiveness. In the case of V-V coarticulation, both theories correctly predict that in languages with more vowels, those vowels will exhibit a lower degree of V-V coarticulation because there is more pressure to maximize perceptual contrast (cf. Manuel & Krakow, 1984; Manuel, 1990, 1999; Magen, 1989). However, adaptive dispersion and constraint-based dispersion do not explain, nor purport to explain, how speakers implement the repulsive forces or constraints in real time; rather, they describe patterns that are fairly removed from individual speakers and utterances. In that regard, dispersion theories fall short of describing *how* contrast is maintained. Experimental evidence presented in the next section points to an alternative understanding of contrast maintenance and maximization, one that utilizes a well-motivated motor planning mechanism.

## 3. Experimental evidence of dissimilation in motor planning

Recent experimental work indicates that contemporaneously planned vowel and tone targets are dissimilated. It is argued that these dissimilatory effects arise from an inhibitory motor planning mechanism. The experimental methodology reported on here, as well as the theoretical analysis of results, was inspired by studies of reaching and oculomotor control which have probed the interaction between movements planned simultaneously. In short, with numerous variations, the nonspeech studies show the following: when movement A to one target

location is prepared in the context of planning a distractor movement B to a sufficiently different target location, then the executed trajectory of movement A deviates *away from* the target of movement B (cf. Sheliga et al., 1994; Doyle & Walker, 2001; Van der Stigchel et al., 2005, 2006; Welsh & Elliot, 2005; Houghton & Tipper, 1996; Ghez et al., 1997). In addition, more salient distractors induce greater deviations away (Tipper et al., 2000). As we will see, these experiments are relevant to understanding analogous effects observed in speech.

*3.1 Dissimilation between vowels in a primed shadowing task*

Tilsen (2009) reports dissimilation between the vowels /a/ and /i/ in a *primed vowel-shadowing* task. In this paradigm, the subject hears a *prime* vowel, then after a delay of several hundred milliseconds, the subject hears a *target* stimulus, which is either a vowel or a beep. There are three types of trials: *concordant* trials, in which the prime and target vowels belong to the same phonemic category; *discordant* trials, in which the prime and target vowels belong to different phonemes, and *no-target* trials, in which the target is a beep. On the concordant and discordant trials, the speaker shadows (repeats) the target vowel as quickly as possible. On the no-target trials, the speaker produces the prime vowel as quickly as possible. In order to respond quickly, the speaker must pre-plan the prime vowel on every trial. Hence on all trials, the speaker first plans to produce either /a/ or /i/, but on 1/3 of the trials (the discordant ones), the speaker subsequently produces the other vowel. Importantly, the paradigm allows one to investigate speech target planning interactions in a $V_1$-$V_2$ sequence without the mechanical and motoric confounds associated with articulation of $V_1$.

Acoustic analyses comparing response vowel F1 and F2 on concordant and discordant trials revealed quasi-*dissimilatory* effects: /a/ responses after /i/ primes were acoustically less similar to /i/ than were /a/ responses after /a/ primes, and vice versa for /i/ responses. In other words, on discordant trials, /a/ and /i/ responses were more peripheral in F1,F2 vowel space, as if dissimilated from the /i/ and /a/ primes, respectively. Figure 1 shows normalized bivariate mean F1 and F2 95% confidence regions for productions on concordant and discordant trials. Formant trajectories were obtained using a Matlab implementation of a robust LPC algorithm, and a dynamic formant tracking algorithm developed at the University of California, Berkeley Phonology Laboratory. Formants were averaged over the middle 1/3 of each vowel, and normalized within subjects. Each subject produced approximately 80-120 vowels in each of the conditions. The figure shows normalized values combined across all 12 native speakers of American English (6 male, 6 female) who participated in the experiment.

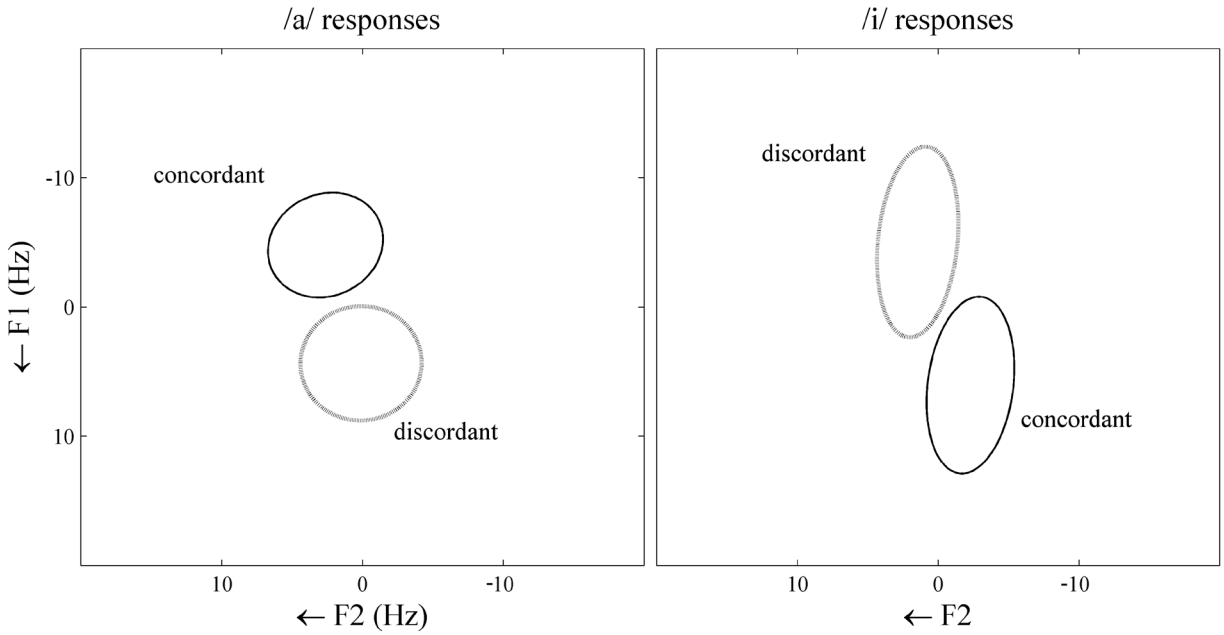


Figure 1. Comparisons of primed vowel shadowing responses on concordant and discordant trials. Ellipses represent 95% confidence regions for within-speaker normalized F1, F2 bivariates averaged over the middle 1/3 of each response.

Figure 1 shows that discordant trial productions of /a/ had significantly higher F1 than concordant trial productions. Discordant productions of /i/ had significantly lower F1 and higher F2 than concordant trial productions. It should be noted that, although not all subjects exhibited these patterns, dissimilation was the predominant trend across the population. For more detailed information on the design, analysis, and subject-specific variation, the reader should consult Tilsen (2009). The "dissimilation" observed here should be understood in a literal, phonetic sense, entailing less similarity. These dissimilatory effects, although relatively subtle, are fairly remarkable in that they point to a mechanism that subtly alters a vowel target as a function of other targets that are planned in parallel.

*3.2 Dissimilation between Mandarin tones*

*3.2.1 Methodology*

In most respects, the experimental design of the primed tone-shadowing task reported here is identical to the primed vowel-shadowing design described above and in Tilsen (2009), with the following important differences. Stimuli were the vowel [ai] with Mandarin Tone 1 (55 high), Tone 2 (35 rising), and Tone 4 (53 falling). To construct stimuli, 100 samples of each tone were recorded by a female native speaker of Mandarin. These tokens were subjected to automated F0 analysis (described below), and the ones most similar in F0 to the mean contours for each set were selected as the experimental stimuli. The stimuli were windowed to 250 ms and amplitude-normalized. Participants were 12 native speakers of Mandarin Chinese, ages 18-30. Each speaker participated in 2 one-hour sessions, and only produced two of the three tones. There were 4 speakers for each combination of tones. In instruction and practice phases, it was emphasized that subjects should produce the correct tone, should avoid starting the response with one tone then switching to the other, and should avoid producing the tones too rapidly.

In processing the data for analysis, responses were excluded which were initiated early (i.e. with an RT of less than 150 ms after the onset of the target tone), initiated late (with an RT more than 2.2 s.d. greater than the mean for each subject), or which were duration outliers (more than ± 2.2 s.d. from the mean for each subject). F0 analysis was conducted using a robust automated pitch tracking algorithm implemented in the Voicebox Speech Processing Toolbox (Mike Brookes, Department of Electrical & Electronic Engineering, Imperial College) for Matlab. Analysis frames of 10 ms were used. For each subject and tone, F0 contours were normalized by linear interpolation or compression to the median number of frames, and then unweighted moving-average smoothing with a five-frame window was applied. Because subjects occasionally produce incorrect tones, or switch from one to another during the response, it is necessary to identify such occurrences and exclude them from the analysis. To accomplish this, average contours and first-difference ($\Delta$F0) contours were calculated for each target tone. Then for each frame of each response, the number of standard deviations of F0 and $\Delta$F0 from the target and non-target averages were calculated. If more than 15% of the frames in a response were outliers (F0 or $\Delta$F0 > than 2 s.d. away from the target mean), or if there were more outliers relative to the target than the non-target, the response was considered an errorful production or mis-analysis, and was excluded. The total number of excluded responses was about 9.5% of the total number of responses.

*3.2.2 Results*

8 of the 12 subjects exhibited significant or marginally significant dissimilation on discordant trials compared to concordant trials. However, the interpretation of dissimilation is sometimes ambiguous due to the dynamic nature of F0 in contour tones. Figure 2 shows within-

speaker comparisons of F0- and duration-normalized tone contours for each of the three tone combinations. Average concordant trial contours are shown with a solid line, average discordant trial contours with a dotted line. Both contours are accompanied by 95% confidence standard error regions. Statistical tests comparing F0 on concordant and discordant trials were conducted for the first, middle, and last third of each tone. Significant differences ($p < 0.05$) are indicated with "*", marginally significant differences ($p < 0.15$) are indicated with a "+".

Figure 2a shows results for subjects who produced Tone 1 (high) and Tone 2 (rising). Subjects s05, s15, and s06 show dissimilation in one or both tones, i.e. the discordant contour for a given tone is less similar to the other tone than the concordant contour. Subject s11 exhibits an anomalous average discordant trial contour, in which the high tone responses appear to initially assimilate to the non-target rising tone (which begins lower), and then subsequently dissimilate from the rising tone. Since the non-target tone rises toward the end, it is possible to see the dissimilation in Tone 1 as a form of assimilation to the rising pattern of Tone 2. In other words, the similarity between tones can be assessed on the basis of relative F0 values, or on the basis of a pattern of change in F0. However, this latter form of assimilation does not appear to occur generally across the subject population.

Figure 2b shows results for Tone 1 (high) and Tone 4 (falling). Subjects s10 and s12 exhibit dissimilatory patterns, while s08 and s14 exhibit assimilatory patterns. Note that s14, who had the largest assimilatory pattern in the experiment, produced anomalously short tones. The interpretation of dissimilation in s12 is based upon the observation that the F0 in the final third of the falling discordant trial contour is further away from the high tone contour than the concordant trial one. This is more suggestive of dissimilation than the pattern produced by s08, for whom the discordant falling tone both begins and ends lower than concordant one. In the s08 case, the contour is most readily viewed as the consequence of an assimilatory contour-wide lowering of F0; in the s12 case, the relative fall in F0 in the final third of the falling tone is more straightforwardly interpreted as a propensity to exaggerate the fall in F0.

Figure 2c shows results for Tone 2 (rising) and Tone 4 (falling). Subjects s03, s07, and s09 exhibit a dissimilatory pattern in one of the tones. Subject s13 exhibited no differences between the discordant and concordant conditions for either tone. s07 and s09 tended to dissimilate Tone 2 from Tone 4 on discordant trials by lowering F0; the effect was highly significant for s07, but marginally significant for s09 and localized to the middle third of the contour. The dissimilation observed in s03 is of the sort identified in s12, where the final third of the falling contour falls lower on discordant trials, making it less similar to the rising pattern of the non-target rising Tone 2.

Table 1 shows mean duration and RT data by subject, for each tone-concordance condition. There were no significant differences in duration or RT between concordant and

discordant trials. One subject, s07, appears to have responded anomalously slow compared to the others. The absence of any effects of discordance on duration or RT indicates that the dissimilatory F0 patterns cannot be interpreted as indirect consequences of differences in reaction time or tone duration between the two conditions.

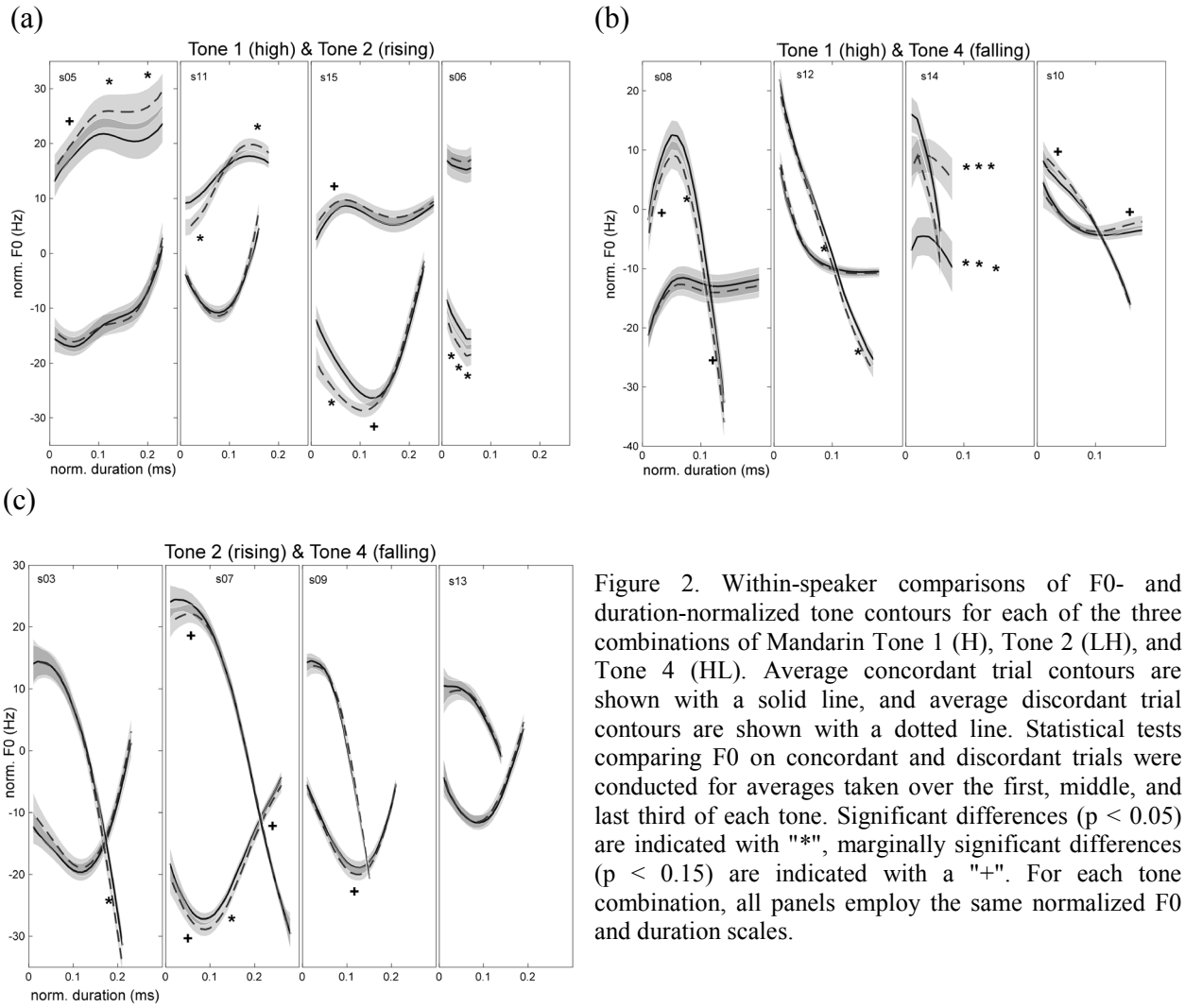


Figure 2. Within-speaker comparisons of F0- and duration-normalized tone contours for each of the three combinations of Mandarin Tone 1 (H), Tone 2 (LH), and Tone 4 (HL). Average concordant trial contours are shown with a solid line, and average discordant trial contours are shown with a dotted line. Statistical tests comparing F0 on concordant and discordant trials were conducted for averages taken over the first, middle, and last third of each tone. Significant differences ($p < 0.05$) are indicated with "*", marginally significant differences ($p < 0.15$) are indicated with a "+". For each tone combination, all panels employ the same normalized F0 and duration scales.

Table 1. Mean durations and RTs for each tone and concordance condition.

| | | | Tone A | | | | Tone B | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | concordant | | discordant | | concordant | | discordant | |
| | | Tone A-B | mean | (s.d.) | | | | | | |
| DUR. (ms) | s05 | 1-2 | 0.321 | (0.029) | 0.316 | (0.029) | 0.314 | (0.027) | 0.311 | (0.028) |
| | s06 | 1-2 | 0.137 | (0.019) | 0.135 | (0.020) | 0.142 | (0.019) | 0.138 | (0.018) |
| | s11 | 1-2 | 0.272 | (0.027) | 0.269 | (0.032) | 0.241 | (0.028) | 0.247 | (0.023) |
| | s15 | 1-2 | 0.341 | (0.027) | 0.342 | (0.030) | 0.327 | (0.025) | 0.334 | (0.029) |
| | s03 | 1-4 | 0.325 | (0.028) | 0.332 | (0.028) | 0.310 | (0.037) | 0.312 | (0.040) |
| | s07 | 1-4 | 0.351 | (0.029) | 0.373 | (0.034) | 0.409 | (0.042) | 0.394 | (0.040) |
| | s09 | 1-4 | 0.289 | (0.023) | 0.296 | (0.023) | 0.271 | (0.035) | 0.273 | (0.034) |
| | s13 | 1-4 | 0.276 | (0.030) | 0.277 | (0.031) | 0.218 | (0.031) | 0.226 | (0.025) |
| | s08 | 2-4 | 0.278 | (0.063) | 0.287 | (0.062) | 0.245 | (0.057) | 0.242 | (0.056) |
| | s10 | 2-4 | 0.260 | (0.032) | 0.274 | (0.033) | 0.239 | (0.021) | 0.238 | (0.025) |
| | s12 | 2-4 | 0.270 | (0.024) | 0.275 | (0.022) | 0.263 | (0.024) | 0.255 | (0.021) |
| | s14 | 2-4 | 0.165 | (0.017) | 0.160 | (0.013) | 0.152 | (0.012) | 0.146 | (0.011) |
| RT (ms) | s05 | 1-2 | 0.434 | (0.065) | 0.417 | (0.064) | 0.400 | (0.058) | 0.428 | (0.062) |
| | s06 | 1-2 | 0.304 | (0.062) | 0.295 | (0.069) | 0.298 | (0.068) | 0.307 | (0.067) |
| | s11 | 1-2 | 0.231 | (0.074) | 0.230 | (0.076) | 0.228 | (0.071) | 0.221 | (0.072) |
| | s15 | 1-2 | 0.245 | (0.066) | 0.249 | (0.062) | 0.247 | (0.056) | 0.253 | (0.064) |
| | s03 | 1-4 | 0.262 | (0.088) | 0.268 | (0.088) | 0.266 | (0.082) | 0.269 | (0.089) |
| | s07 | 1-4 | 0.513 | (0.052) | 0.510 | (0.051) | 0.505 | (0.055) | 0.521 | (0.049) |
| | s09 | 1-4 | 0.284 | (0.072) | 0.281 | (0.061) | 0.285 | (0.065) | 0.284 | (0.067) |
| | s13 | 1-4 | 0.294 | (0.077) | 0.299 | (0.081) | 0.288 | (0.085) | 0.283 | (0.074) |
| | s08 | 2-4 | 0.423 | (0.053) | 0.412 | (0.048) | 0.394 | (0.046) | 0.410 | (0.050) |
| | s10 | 2-4 | 0.304 | (0.083) | 0.307 | (0.080) | 0.292 | (0.078) | 0.286 | (0.071) |
| | s12 | 2-4 | 0.307 | (0.044) | 0.308 | (0.043) | 0.304 | (0.047) | 0.308 | (0.042) |
| | s14 | 2-4 | 0.346 | (0.059) | 0.344 | (0.059) | 0.340 | (0.058) | 0.343 | (0.071) |

## 4. Discussion

To summarize, a majority of subjects exhibited dissimilation on discordant trials, in at least one of the tones. However, substantial inter-subject variability was observed in this regard, along with instances of assimilatory patterns. Section 4.1 will address the potential sources of this variability. Section 4.2 will argue that the dissimilatory patterns arise from an inhibitory motor planning mechanism, and section 4.3 will explain how this inhibitory mechanism may be responsible for the maintenance and maximization of contrast.

### *4.1 Intersubject variability*

Not all subjects exhibited dissimilation in both responses of the tone-shadowing task, and some of them produced assimilatory patterns. This variation is consistent with the results of primed vowel-shadowing in Tilsen (2009), and does not negate the importance of the dissimilatory behavior. If one views the mechanism of dissimilation as learned, or perhaps, as innate but modulated by context and experience/learning, then one should expect speaker-specific variation in its effects. The mere presence of the dissimilation in some speakers—here the majority—begs for an explanation. Moreover, there are a number of factors which may mask the output of the dissimilatory mechanism in some cases.

For one, there may be ceiling or floor effects attributable to F0 register. Some speakers may not normally produce F0 above or below a certain range; thus where a dissimilatory pattern would raise or lower F0 beyond that range, no dissimilation is produced. This could account for the near-absence of dissimilation in the initial third of Tone 4 (falling), since this tone tends to begin at the top of the normal F0 range. Stimuli and speaker gender may also have an influence on dissimilatory behavior, although the current design was not well-suited to analysis of such effects. It is also possible that variation results from differences in attention to the task. In Tilsen (2009), subjects who produced assimilatory patterns either responded abnormally slowly or with high error rates, indicative of a lack of focus—here, however, no such correlation was observed.

It is important to consider why dissimilatory patterns are not generally observed in paradigms where speakers execute both elements of a sequence. For example, in studies of articulated VCV sequences or tonal coarticulation (e.g. Shen, 1990; Xu, 1997; Gandour et al., 1994), assimilatory patterns are by far the predominant ones. This is presumably because mechanical factors, gestural overlap, or other sources of assimilatory coarticulation *tend* to overwhelm the dissimilatory effects of contemporaneous target planning. The primed-shadowing task circumvents these effects by inducing the speaker to plan, but not articulate, the first element of the sequence. The assimilatory patterns in fully articulated sequences are, like dissimilation in primed vowel-shadowing, only tendencies. There is, indeed, one study that has reported a dissimilatory effect between articulated vowels: Fletcher (2004) found a slight dissimilation between /a/ and /i/ in Southern British English /ə'kaki/ sequences, particularly for one subject. Furthermore, on an token-by-token basis, dissimilation is still observed in articulated sequences, and the extent to which assimilation or dissimilation occurs in natural speech, where various phonological, prosodic, semantic, and discourse factors are uncontrolled, is not well known. One should not conclude that just because assimilation is the tendency observed in the lab, that only assimilation occurs outside the lab.

*4.2 Dissimilation is caused by inhibitory mechanisms*

Inhibitory mechanisms have been broadly implicated in the control of sequential movement, and are necessary for understanding how action sequences are performed when actions are planned contemporaneously. Lashley (1951), on the basis of observations of anticipatory and perseveratory errors in movement sequences, argued that plans associated with each element in a sequence must be activated in parallel. Parallel activation has found experimental support in studies of prepared movement sequences, for example in a series of experiments conducted by Sternberg et al. (1978, 1988). They showed that the number of syllables and number of interstress units (or feet) in an utterance have independent, additive effects on the latency to initiate the utterance. Similar results were obtained for typing, and in a related speech paradigm by Wheeldon & Lahiri (1997, 2002). The theoretical interpretation Sternberg and colleagues offer for these findings is as follows. Prior to the initiation of an utterance, all action units are active in working memory. All but the first unit must be suppressed to initiate the utterance. Hence the more units there are, the longer it takes to inhibit the non-initial ones and begin the first one. The concept of competition between movement plans activated in parallel has been modeled as competitive queuing in neural networks (Grossberg, 1978; Bullock & Rhodes, 2003; Bullock, 2004).

The dissimilation of movement targets and trajectories from competing ones has been theoretically related to inhibition. Houghton & Tipper (1996) and Tipper et al. (2000) argue that deviation away from distractors is the result of *selective inhibition* of motor plans associated with the distractor. In this view, the trajectories and targets of movements are represented by activity patterns in overlapping populations of neurons. In order to move to one target, other movement plans that are simultaneously active in working memory must be selectively inhibited. Moreover, because the neural populations encoding for these plans overlap to some degree, inhibition of one population can have an effect on the population encoding the target movement.

Dissimilatory patterns observed in primed vowel- and tone- shadowing can be understood to arise from *intergestural inhibition* in the context of an exemplar theory of production. Figure 3 uses schematized planning stages to model the effect of inhibition in a vowel-shadowing task. The figure compares /i/-exemplar activation in F1,F2 space on a concordant (left) and discordant trial (right). Since each vowel is associated with many exemplars, it is reasonable to approximate the excitation of exemplars having any particular pair of F1 and F2 values with a smooth function. In this case a bivariate Gaussian is used, though the concept generalizes to any relatively smooth function. The excitation function, minus any inhibition, constitutes a net activation function which can be viewed as a probability that a given F1, F2 bivariate target will be produced. In Stage 1, after the prime vowel stimulus, Figure 3 shows that the excitation function is substantially greater for prime vowel exemplars than for non-prime exemplars. Since

the probability of producing the prime is 2/3, the summed excitation of prime vowel exemplars is twice the excitation of the non-prime exemplars.

In Stage 2, when the target is known, intergestural inhibition is applied and the target exemplars are fully excited. The inhibition function, shown to the right of the Stage 2 excitation function, is modeled as a bivariate Gaussian located on the center of mass/activation of the non-target excitation function for /a/-exemplars. There are two important aspects of the inhibition. First, inhibition of the non-target /a/-exemplars is greater on the discordant trial than on the concordant trial. This is justified by the observation that more salient distractors produce stronger dissimilatory effects (Tipper et al., 2000). In other words, more inhibition is necessary on the discordant trials because the non-target prime was more highly excited. Second, the inhibition function is non-zero throughout the region of F1,F2 space where the target /i/-exemplars are located, and crucially, the inhibition is greater on the side of that region closer to the non-target /a/-exemplars. From these two characteristics, it follows that the center of mass of the activation function (excitation minus inhibition, shown in stage 3) is shifted further away from the non-target on the discordant trial, compared to the concordant trial. In Stage 3, the concordant (white ●) and discordant (white o) centers of activation are shown in both activation functions, for purposes of comparison. The F1,F2 difference between discordant and concordant trials is about [-30, 55] Hz. Model equations and further details of implementation can be found in Tilsen (2007).

A more complicated version of this model would treat a larger number of phonetic variables, as well as dynamical aspects of speech targets. After all, vowel formants are often dynamic, and Mandarin tones exhibit substantial change over time; this must be incorporated into target planning and should therefore be subject to dissimilation. Exemplar theory allows for modeling time as an additional dimension of exemplar space (cf. Johnson, 1997), so that memories incorporate spectrotemporal information. Hence the model proposed above should be generalizable to higher-dimensional exemplar spaces with a temporal dimension. It is also noteworthy that the model does not require one to commit to representation in either perceptual or motoric coordinate space. Acoustic coordinates were used here for expository purposes only.

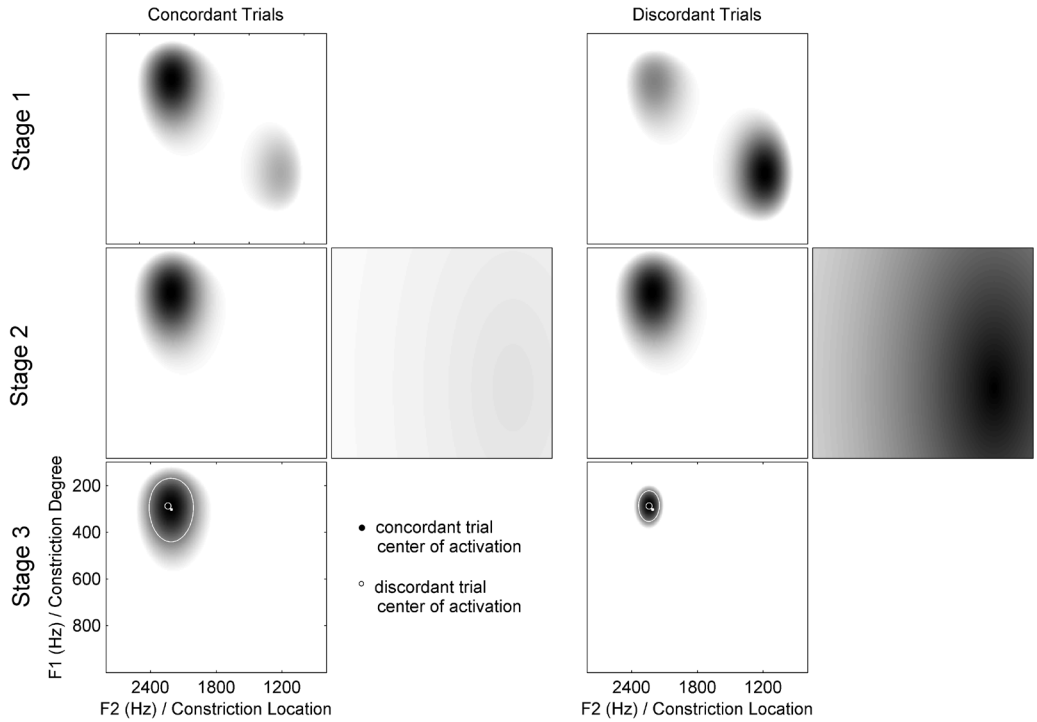


Figure 3. Simulation of the effects of intergestural inhibition on concordant and discordant trials with an /i/ target. Stage 1 shows excitation functions after the prime vowel. Stage 2 shows excitation and intergestural inhibition functions after the target stimulus. Stage 3 shows the activation function from which a production target is derived. For comparison, the concordant (white ●) and discordant (white o) centers of activation are shown in both activation functions.

*4.3 Intergestural inhibition, coarticulation, and contrast*

In the context of an exemplar model, intergestural inhibition can function to maintain contrast and maximize the use of a phonetic space. Consider once more the phonologization of V-to-V coarticulation to vowel harmony. First, the dissimilatory effect of intergestural inhibition to some extent opposes assimilatory coarticulation between $V_1$ and $V_2$ by subtly dissimilating the target of $V_2$ from $V_1$, and perhaps vice versa. On average, the tendency in VCV sequences

appears to be assimilatory coarticulation, due perhaps to some combination of mechanical factors and gestural overlap. This suggests that these factors tend to outweigh the effects of intergestural inhibition. Over time, if unconstrained, this situation could lead to loss of contrast, i.e. phonologization of vowel harmony.

However, the inhibition model also predicts that as $V_1$ and $V_2$ exemplar distributions shift closer in phonetic space, the strength of intergestural inhibition will become greater on the region of $V_2$ exemplar space (this follows as long as the inhibition function remains constant over time). In other words, closer targets are more strongly dissimilated. In some cases, this stronger inhibition will not dissimilate the target of $V_2$ enough to prevent loss of contrast, but in other cases, the dissimilation may be strong enough to do so. The exemplar distribution in the latter case will come to reflect a balance between the assimilation from coarticulatory forces and the dissimilation from inhibitory ones. This balance is precisely what is described by dispersion theories. Indeed, intergestural inhibition can be seen as a mechanism through which the speaker attempts to maximize contrast on an utterance-by-utterance basis. Whether or not a relatively stable balance occurs in any given language is likely to depend on many factors, particularly on vowel and consonant inventories of a language and co-occurrence frequencies of the units in VXV sequences. Ultimately, what intergestural inhibition provides is a real-time, utterance-anchored mechanism for maintaining and maximizing contrast. Contrast is never *fully* maximized because highly variable coarticulatory forces are always influencing the exemplar distribution, but dispersion theories likewise do not predict that a phonetic space is actually maximally used—they only posit a tendency toward this.

Hence intergestural inhibition is not a priori mutually exclusive with perceptual dispersion or perceptual correction. It can be seen in two ways, either as operating alongside perceptual mechanisms, or as the underlying basis for them. It is also reasonable to see inhibition both as an intrinsic aspect of how working memory operates and as something modulated by experience. Whenever articulatory plans are brought into working memory, the serial ordering of those plans is accomplished by interacting excitatory and inhibitory processes; the production of one articulation requires the simultaneous suppression of others, yet the extent to which inhibition is exerted between plans is inferred and learned from the linguistic experience of a speaker.

One problem with dispersion theories is they lack an account of how articulatory targets are planned so as to maximize perceptual contrast. These theories hold that the speaker, for functional reasons, produces sounds that maximize perceptual contrast. However, there is limited evidence for a real-time perceptual dispersion mechanism. The most suggestive evidence to date is the hyperspace effect reported in Johnson, Flemming, & Wright (1993) and in Johnson (2000). In Johnson et al. (1993), listeners identified the "best" examples of a range of synthetic vowel

stimuli as the ones that were more peripheral than their own productions. The source of this difference can interpreted as a consequence of target undershoot in production, or as the result of an active perceptual process. An alternative account of the hyperspace effect is suggested by mounting evidence that the perception of a sound involves simulation of the corresponding motor activity that speakers would use to produce the sound themselves. It is well-established that activity in cortical premotor and motor regions, via the mirror system, accompanies the perception of actions (including the production of speech sounds), and that this motor activity plays an important role in accurate and quick perception (D'Ausilio et al., 2009; Watkins et al., 2003; Fadiga et al., 2002; Pulvermüller et al., 2006; Rizzolatti & Craighero, 2004; Gallantucci et al., 2006; Gallese et al., 1996). Hence, the "best" example of a speech sound may correspond to the motor simulation which involves maximal inhibition of related speech targets. In other words, the best /i/ target would be the one formed when other vowel sounds are maximally suppressed, and hence, the most dissimilated /i/ target. This reasoning could extend to the selection of "as you say it" examples, and to the stimuli which were used to avoid consonantal context and talker-unfamiliarity confounds in Johnson (2000). In sum, the hyperspace effect could very well involve a perceptuo-motor mechanism which relies heavily on intergestural inhibition in the motor simulation of sensory stimuli.

Worthy of mention is an alternative account of perceptual correction that involves lower-level perception, advocated by Holt, Lotto, & Kluender (2000). They suggest that on very short timescales, a general mechanism of neural adaption to a perceptual stimulus results in a subsequently diminished perceptual response to acoustically similar stimuli. It is likely that both low-level perceptual adaptation of this sort, and higher-level inhibitory interactions associated with more categorial speech percepts, are involved in perceptual compensation.

In sum, intergestural inhibition in motor planning is important for understanding what limits assimilatory coarticulation and its phonologization. The effects of inhibition are manifest on two timescales: in the real-time planning of speech targets, and indirectly on a diachronic timescale by virtue of dissimilating exemplar distributions. Perceptual dispersion can be seen as a pattern emerging from intergestural inhibition, exemplar memories, and interacting agents—as opposed to a cognitive mechanism in and of itself. It is of course possible to see the production-perception interaction in a causal loop, whereby the diachronic selection of more contrastive sounds reinforces the extent of motor planning dissimilation. Ultimately, we must conclude that there is another domain of constraints on phonological change that is neither strictly perceptual nor strictly articulatory. Such constraints arise not from perceptual discriminability, nor from physical forces or temporal overlap of articulations, but rather, from cognitive mechanisms governing the planning, suppression, and execution of sequential movement.

**Acknowledgments**

Thanks to Keith Johnson for discussions of this research. Two anonymous reviewers contributed to the improvement of this manuscript. Thanks to Yao Yao and Ron Sprouse for assistance in the University of California, Berkeley Phonology Lab. This work was supported by the National Science Foundation under Award No. 0817767.

**References**

Bell-Berti, F. & Harris, K. (1976). Some aspects of coarticulation, *Haskins Laboratories Status Report on Speech Research,* SR45/46, 197-204.

Browman, C. & Goldstein, L. (1986). Towards an articulatory phonology. In C. Ewen and J. Anderson (Eds.), *Phonology Yearbook 3*, 219-252, Cambridge: Cambridge University Press.

Browman, C. & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica 45,* 140-155.

Browman, C. & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In M. Beckman & J. Kingston (Eds.), *Papers in laboratory phonology I*: *Between the grammar and the physics of speech*, 341-376. Cambridge: Cambridge University Press.

Bullock, D. & Rhodes, B. (2003). Competitive queuing for serial planning and performance. In Arbib, M. (Ed.), *Handbook of Brain Theory And Neural Networks*, 241–244. MIT Press: Cambridge, MA.

Bullock, D. (2004). Adaptive neural models of queuing and timing in fluent action. *Trends in Cognitive Sciences*, 8(9): 426-433.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The Motor Somatotopy of Speech Perception. *Current Biology*, 19, 381–385

Doyle, M. & Walker, R. (2001). Curved saccade trajectories: Voluntary and reflexive saccades curve away from irrelevant distractors. *Experimental Brain Research*, 139: 333–344.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, 15, 399–402.

Finley, S. (2008). *Formal and cognitive restrictions on vowel harmony*. Doctoral dissertation, Johns Hopkins University.

Flemming, E. (1996). Evidence for constraints on contrast: The dispersion theory of contrast. Chai-Shune K. Hsu (Ed.), *UCLA Working Papers in Phonology* 1, 86-106.

Flemming, E. (2004). Contrast and perceptual distinctiveness. In Hayes, B., Kirchner, R., & Steriade, D. (Eds.), *The Phonetic Bases of Markedness*. Cambridge University Press, Cambridge.
Fletcher, J. (2004). An EMA/EPG study of vowel-to-vowel articulation across velars in Southern British English. *Clinical Linguistics & Phonetics*, 18(6): 577-592.
Fowler, C. A. (1981). A relationship between coarticulation and compensatory shortening. *Phonetica,* 38: 35-50.
Galantucci, B., Fowler, C.A., & Turvey, M.T. (2006). The motor theory of speech perception reviewed. *Psychonomics Bulletin Review*, 13, 361–377.
Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 593-609.
Gandour, J., Potisuk, S., & Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics*, 22, 4: 477-492.
Gay, T. (1974). A cinefluorographic study of vowel production. *Journal of Phonetics,* 2: 255-266.
Gay, T. (1977). Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America,* 62: 183-193.
Ghez, C., Favilla, M., Ghilardi, M., Gordon, J., Bermejo, R., & Pullman, S. (1997). Discrete and continuous planning of hand movements and isometric force trajectories. *Experimental Brain Research,* 115: 217-233.
Goldinger, S. (1992). *Words and voices: Implicit and explicit memory for spoken words.* PhD dissertation, Indiana University.
Goldinger, S. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22(5), 1166-1183.
Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 2: 251-278.
Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In Rosen, R. and Snell, F. (Eds.), *Progress in Theoretical Biology*, 5, 233– 374. Academic Press.
Grosvald, M. (2009). Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics*, 37, 2: 173-188.
Holt, L., Lotto, A., & Kluender, K. (2000). Neighboring spectral context influences vowel identification. *Journal of the Acoustical Society of America,* 108(2), 710-722.

Houghton, G. & Tipper, S. (1996). Inhibitory Mechanisms of Neural and Cognitive Control: Applications to Selective Attention and Sequential Action, *Brain and Cognition*, 30: 20-43.
Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In Johnson, K. & Mullennix, J. (Eds.), *Talker Variability in Speech Processing*, 145-166, Academic Press: New York.
Johnson, K. (2000). Adaptive Dispersion in Vowel Perception. *Phonetica*, 57: 181-188.
Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34: 485-499.
Johnson, K., Flemming, E. & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, 69: 505-528.
Krämer, M. (2001). *Vowel harmony and correspondence theor*y. Doctoral dissertation, University of Düsseldorf.
Lashley, K.S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral Mechanisms in Behavior*. New York: Wiley.
Liljencrants, J. & Lindblom, B. (1972). Numerical simulations of vowel quality systems: the role of perceptual contrast. *Language,* 48: 839-862.
Lindblom, B. (1986). Phonetic universals in vowel systems. In: Ohala, J.J. and Jaeger, J.J. (Eds.), *Experimental Phonology*. Academic Press, New-York, 13–44.
Lindblom, B. (1990). On the notion of possible speech sound. *Journal of Phonetics* 18, 135–152.
Lindblom, B. (2003). Patterns of phonetic contrast: towards a unified explanatory framework. *Proceedings of the 15th International Congress of Phonetic Sciences*, 39–42.
Magen, H. (1989). *An acoustic study of vowel-to-vowel coarticulation in English*, Ph. D. dissertation, Yale University.
Magen, H. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics,* 25: 187-205.
Manuel, S. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America,* 88, 1286–1298.
Manuel, S. (1999). Cross-language studies: relating language-particular coarticulation patterns to other language-particular facts. In Hardcastle, W. & Hewlett, N. (Eds.), *Coarticulation: Theory, Data and Techniques*, 179-198. Cambridge: Cambridge University Press.
Manuel, S. & Krakow, R. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research,* SR77/78: 69-78.
Ohala, J. (1981). The listener as a source of sound change. In Masek, C., Hendrick, R., & Miller, M. F. (Eds.), *Papers from the Parasession on Language and Behavior, Chicago Linguistic Society*, 178-203.

Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In MacNeilage, P. (Ed.), *The production of speech*. New York: Springer-Verlag. 189-216.
Ohala, J. (1993). The phonetics of sound change. In Jones, C. (Ed.), *Historical Linguistics: Problems and Perspectives*. London: Longman.
Ohala, J. (1994). Towards a universal, phonetically-based, theory of vowel harmony. *3rd International Conference on Spoken Language Processing*. Yokohoma, Japan. 491-494.
Öhman, S. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America,* 39: 151-168.
Oudeyer, P. (2006). *Self-organization in the evolution of speech: Studies in the evolution of language*. Oxford: Oxford University Press.
Parush, A., Ostry, D., & Munhall, K. (1983). A kinematic study of lingual coarticulation in VCV sequences. *Journal of the Acoustical Society of America,* 74: 1115-1123.
Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee and P. Hopper (eds.) *Frequency effects and the emergence of lexical structure*. John Benjamins, Amsterdam. 137-157.
Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory Phonology VII*, Mouton de Gruyter, Berlin, 101-139.
Pierrehumbert, J. (2004). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 115–154.
Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, USA*, 103, 7865–7870.
Recasens, D. (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. *Journal of the Acoustical Society of America*, 76(6): 1624-1635.
Recasens, D., Pallares, M., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America*, 102(1): 544-561.
Rennison, J. (1990). On the elements of phonological representations: The evidence from vowel systems and vowel processes. *Folia Linguistica* 24, 175-244.
Rizzolatti, G. & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192.
Saltzman, E. & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1: 333-382.
Sheliga, B., Riggio, L., & Rizzolatti, G. (1994). Orienting of attention and eye movements. *Experimental Brain Research*, 98: 507-522.
Shen, X. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics*, 18, 2: 281-295.

Sternberg, S., Knoll, R. Monsell, S. & Wright, C. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica,* 45, 175-197.
Sternberg, S., Monsell, S, Knoll, R., & Wright, C. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typing. In G. E. Stelmach (Ed.) *Information Processing in Motor Control and Learning.* New York: Academic Press. 117-152.
Tilsen, S. (2007). Vowel-to-vowel coarticulation and dissimilation in phonemic-response priming. *UC-Berkeley Phonology Lab Annual Report*, 416-458.
Tilsen, S. (2009). Subphonemic and cross-phonemic priming in vowel shadowing: evidence for the involvement of exemplars in production. *Journal of Phonetics*, 37: 3, 276-296.
Tipper, S., Howard, L. & Houghton, G. (2000), Behavioral consequences of selection from neural population codes. In Monsell, S. & Driver, J. (Eds.), *Control of Cognitive Processes*, 225-245. MIT Press.
Van der Stigchel, S. & Theeuwes, J. (2005). The influence of attending to multiple locations on eye movements. *Vision Research.*
Van der Stigchel, S., Meeter, M., & Theeuwes, J. (2006). Eye movement trajectories and what they tell us. *Neuroscience Biobehavioral Review*.
Vergnaud, J. (1980). A formal theory of vowel harmony. In Vago (Ed.), *Issues in vowel harmony*. Amsterdam: Benjamins. 49-62.
Watkins, K., Strafella, A., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41, 989–994.
Wedel, A. (2004). Category competition drives contrast maintenance within an exemplar-based production/perception loop. In *Proceedings of the seventh meeting of the ACL special interest group in computational phonology*, 1–10. Barcelona, Spain: Association for Computational Linguistics, July 2004.
Welsh, T. & Elliot, D. (2005). The effects of response priming on the planning and execution of goal-directed movements in the presence of a distracting stimulus. *Acta Psychologica,* 119: 123-142.
Whalen, D. (1990). Coarticulation is largely planned. *Journal of Phonetics*, 18: 3-35.
Wheeldon, L. & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, 37, 356–381.
Wheeldon, L. & Lahiri, A. (2002). The minimal unit of prosodic encoding: prosodic or lexical word. *Cognition*, 85(2): B31-B41.
Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25, 1: 65-83.