THE EFFECT OF COGNITIVE LOAD ON TONAL COARTICULATION

Kathryn Franich

University of Chicago
kfranich@uchicago.edu

ABSTRACT

Cognitive load (CL) has been found to influence language perception in many interesting ways, but its role in production has not been explored. In this paper, we look at how CL influences production of tonal coarticulation in Mandarin Chinese. Since coarticulation has been found to involve cognitive planning, this is an especially appropriate domain for investigating the influence of CL. Results indicate that the overall effect of CL on coarticulation was weak, but that anticipatory coarticulation of the dissimilatory variety increased somewhat under CL preceding a tone with a low f_0 onset. These results suggest that anticipatory raising, or 'upstep', may involve a separate cognitive mechanism which is not common to all types of tonal coarticulation.

Keywords: coarticulation, speech production, cognitive load, tone, working memory

1. INTRODUCTION

Humans produce and perceive speech in many different environments from day to day, and are generally adept at normalizing for variation in the communication channel [17, 13]. Contemporary theories of language change posit that the relationship between variation produced by speakers and the ability of listeners to contextualize such variation can be a source of language change [16, 4]. When the listener fails to compensate for variation and 'misperceives' one sound for another, they may interpret the misperceived sound as intended by the speaker, eventually incorporating it into the phonological system. An important objective of current work on sound change is thus to identify possible sources of variation in speech in order to understand how these kinds of variation might be interpreted by the listener.

The goal of the current study is to investigate a new source of language variation: production effects induced by cognitive load. It has long been recognized that the articulation of sounds in speech production involves some level of advance cognitive planning, typically hypothesized to involve the coordination of a detailed motor plan [20] which is stored within working memory [2, 1]. A striking finding concerns processes of coarticulation, or the tendency for one segment to influence a neighboring segment in production: while such effects were previously assumed to occur due to constraints on executing a given motor program (i.e. to physiological constraints on the articulators in reaching distant production targets), it has become clear that at least certain types of coarticulation are planned, likely rooted within the motor program itself [24]. An additional finding is that, while leftward 'anticipatory' coarticulation is robustly found to decrease in conditions where planning is inhibited, rightward 'carryover' coarticulation shows no such effects [24]. One of the goals of the current study is thus to explore these different types of coarticulation under cognitive load.

1.1. The Role of Cognitive Load in Language

Within the linguistic literature, CL has mostly been studied with respect to its role in speech perception. Findings indicate an increase in lexical bias on phoneme identification where cognitive load is higher [14], and a negative impact of cognitive load on word recognition in synthetic speech [8]. Related work examining the influence of attention on language perception shows that attentional distraction impedes listeners' ability to rely on fine-grained phonetic detail in word recognition [23], as well as word reading tasks, where errors reflect more of a word bias in distracted conditions [11]. An fMRI study concluded that attention plays a modulatory role in neuronal activation to speech sounds suggesting attention aids in speech perception [12].

Little work has examined the relationship between CL and language production, and no study that we know of has examined fine-grained phonetic variation in speech as it relates to CL. Since motor plans generated for speech are thought to be housed within working memory, and under the assumption that only a finite number of cognitive resources are available for completing a set of memory-based tasks (such as number recall and speaking) [6, 15, 18], a clear prediction is that performance on one or both tasks will be affected once this limit is reached. The present study investigates precisely this relationship by looking at the influence of cognitive load on the production of tonal coarticulation in Mandarin Chinese.

1.2. Tonal Coarticulation in Mandarin Chinese

The patterning of tonal coarticulation in Mandarin makes it an ideal test case for theories of speech planning and coarticulatory mechanisms. First off, Mandarin exhibits both anticipatory and carryover effects, though carryover effects have been found to be the most robust. Secondly, while carryover effects are assimilatory in nature, anticipatory effects are largely dissimilatory [25]. Recent work has focused on what the precise nature of dissimilatory coarticulation is, since most theories of coarticulation do not predict that it should exist. It has been demonstrated that Mandarin tones tend to dissimilate more where the context tone and target tone differ categorically, possibly due to the existence of an inhibitory mechanism which serves to maintain contrasts between phonological categories [22].

Our first prediction, in line with previous work, is that carryover coarticulation will be stronger than anticipatory coarticulation overall. We also predict that carryover coarticulation will be assimilatory and anticipatory coarticulation dissimilatory. For the role of CL, we anticipate one of four possible outcomes. Under the hypothesis that both carryover assimilation and anticipatory dissimilation are planned, we predict that both types of coarticulation will be altered under CL. If, however, we hypothesize that only anticipatory coarticulation is planned, we predict that only this variety will show effects of CL. A third possibility is that both carryover and anticipatory coarticulation are planned, but through different mechanisms. In this case, we might see an impact of CL on carryover coarticulation, while anticipatory dissimilation remains intact. The fourth possible outcome, of course, is that neither type of coarticulation is influenced by CL.

2. METHOD

2.1. Participants

Eleven native Mandarin speakers (7 female) from Mainland China participated in this study for either $10 or course credit. Ages ranged from 19 to 45 and all were affiliated with the University of Chicago.

2.2. Design

Stimuli were drawn from [25] and consisted of disyllabic forms containing the sequence /mama/ where each syllable bore one of the four lexical tones of Mandarin, the High-Level tone (T1), the Low-Rising tone (T2) the Falling-Rising tone–or, as it is also known for its overall low f_0, the Low tone–(T3), and the High-Falling tone (T4). All logically possible combinations of tones were used, for a total of 16 disyllabic forms. Though only one of these forms was a real word (māmā means 'mother') participants were asked to read all forms as naturally as possible.

Each disyllabic form was embedded four times in three conditions. In the first condition (Control), participants simply read the forms on the computer screen. In the second condition (Mid CL), participants recalled three two-digit numbers while reading the target word. Two-digit numbers were presented one at a time, and displayed for 1000 ms each before the next number was presented. Immediately after the numbers were presented, the target word appeared in Simplified Chinese. After reading the word on the screen, participants were asked to type the numbers horizontally. The third condition (Severe CL) was identical to the Mid condition, but participants had to recall five two-digit numbers. Participants encountered each combination of tones four times per condition, for a total of 192 trials. The study was carried out in four blocks using E-Prime software, with one instance of each tonal combination per CL condition displayed per block in randomized order. Words were recorded as uncompressed .wav files using a Marantz PMD670 Professional Solid-State Recorder, digitized at a sampling rate of 44.1K. Extraction of f_0 contours was done using ProsodyPro, a Praat script [5] designed by [26]. Vocal pulses were corrected manually.

2.3. Coarticulation Measures

Since coarticulation is difficult to measure by simply comparing f_0 trajectories, we adopt a method for measuring coarticulation via tonal 'deviation scores' as inspired by [7]. This can be described as the distance of the z-score log f_0 of a syllable utterance from the center of a speaker's f_0 space. As f_0 cues are most acoustically salient during the vocalic portion of the syllable, we took measurements at 5 equidistant points from each vowel of a given /mama/ sequence. For each position i under investigation (σ_1 or σ_2), we computed the centroid f_0 of the speaker's tonal space. A similar calculation was used to compute standard deviation. We then calculated z-scores based on the log-transformed values of centroid and standard deviation values (1).

$$\text{(1)} \quad \frac{log f0^i - mean(log f0^i)}{sd(log f0^i)}$$

3. RESULTS

Only those trials in which participants accurately recalled all digits in the cognitive load task were included. Since over 60% of trials in the Severe condition resulted in inaccurate responses, this condition was deemed too difficult and was excluded from analysis. An additional 9% from the Mid condition was also excluded due to inaccurate recall. Trials in which tone deviation scores reached more than 2 standard deviations from the mean f_0 for more than 3 consecutive time points were excluded from analysis, amounting to an additional 4% of the total data.

3.1. Model and Predictors

For each tone condition, deviation scores were analyzed using linear mixed effects models (`lme4` package [3] for *R* statistical software [19]). Five fixed variables were examined: TIME (normalized), DURATION (of target syllable), TRIAL, CONTEXT, and LOAD. TIME, DURATION and TRIAL were all treated as continuous variables, and were centered to avoid collinearity effects. CONTEXT, which refers to the the f_0 range of the context tone, was coded as a categorical variable with two levels, High and Low. For carryover coarticulation, High contexts encompassed T1 and T2, which both have high f_0 offsets, and Low contexts encompassed T3 and T4. For anticipatory coarticulation, High contexts included T1 and T4, which both have high onsets, and Low contexts included T2 and T3. LOAD was treated as a categorical variable with two levels, Control and Cognitive Load (CL–formerly 'Mid'). Both CONTEXT and LOAD were sum coded.

Individual models were run for each target tone beginning with a maximal model consisting of fixed effects TRIAL, DURATION, CONTEXT, TIME, and LOAD, with two-way interactions between CONTEXT and TIME, between LOAD and TIME, and between LOAD and CONTEXT, as well as a three-way interaction between LOAD, CONTEXT, and TIME. TIME was analyzed using growth-curve analysis; the effect of such a model is to describe the change in either slope (if linear) or curvature (if quadratic) in the trend attributable to one level or another of a categorial variable. If a parameter increases significantly with time, its linear coefficient will be positive. Where a parameter exhibits significant uniform curvature with time, its quadratic coefficient will be positive if concave, and negative if convex. Only one of these two parameters (that with the larger absolute coefficient) will be reported for each tone. Random effects structure of the models consisted of a random intercept for Subject, and random slopes for TRIAL, DURATION, CONTEXT, TIME and LOAD.

Comparisons were done between nested models using the `car` package for *R* to evaluate model fit; results from the optimal model for each target tone are reported. Calculations of *p*-values are based on the Kenward-Roger degrees of freedom approximation, calculated using the *R* package `pbkrtest` [10]. For lack of space, we will discuss terms of interest collectively for the four carryover models, and then the four anticipatory models.

3.2. Carryover Coarticulation

As expected, carryover coarticulation was robust, as indicated by strong effects of CONTEXT for all target tones. Larger effects were found for T2 (β=0.59373, t=10.398, p<0.001) and T4 (β=0.56417, t=7.752, p<0.001), the two contour tones, and smaller effects found for T1 (β=0.43575, t=5.846, p<0.001) and T3 (β=0.47206, t=5.264, p<0.001). For all target tones, higher deviation scores were found in High contexts, and lower deviation scores were found in Low contexts, indicating assimilation (Figure 1). An effect of TIME was also found for all target tones, with stronger effects of the quadratic parameter found for T2 (β=1.06197, t=16.080, p<0.001) and T4 (β=-0.97292, t=-10.985, p<0.001) and stronger effects of the linear parameter found for T1 (β=0.46322, t=3.316, p<0.001) and T3 (β=-3.53930, t=-15.845, p<0.001). Interactions between CONTEXT and TIME were also found for all target tones, with larger effects found for the linear parameter for all. Effects were strongest for T1 (β=-0.62576, t=-4.506, p<0.001) and T2 (β=-0.65563, t=-5.571, p<0.001), and smaller for T3 (β=-0.63723, t=-2.884, p<0.01) and T4 (β=-0.45237, t=-2.885, p<0.01). Generally, assimilatory effects of both contexts were found to dissipate over time. No effect of LOAD was found for any target tone. None of the two-way interactions nor the three-way interaction between LOAD, CONTEXT and TIME showed any significant effects.

3.3. Anticipatory Coarticulation

Effects of anticipatory coarticulation were overall much weaker, as expected. In fact, a main effect of CONTEXT was not found for any target tone. Significant effects of TIME were found for the linear parameter for T3 (β=-1.41770, t=-7.695, p<0.001) and for the quadratic parameter for both T2 (β=0.83408, t=17.136, p<0.001) and T4 (β=-0.69512, t=-13.066, p<0.001), but no effects were found for T1. No effect was found for the interaction between CONTEXT and TIME. A main effect of LOAD was found for T2 (β=-0.07555, t=-2.348, p<0.05), but no other tone. Of primary interest, a significant effect emerged for the interaction between LOAD and CONTEXT for T1 (β=-0.03039, t=-2.204, p<0.05), and a marginal effect for T3 (β=0.03721, t=1.731, p=0.08). In both cases, anticipatory raising before a Low context was increased under CL (Figure 2). No effect for the two-way interaction between LOAD and TIME nor the three-way interaction between LOAD, CONTEXT and TIME was found.

Figure 1: Carryover coart. by context and tone.

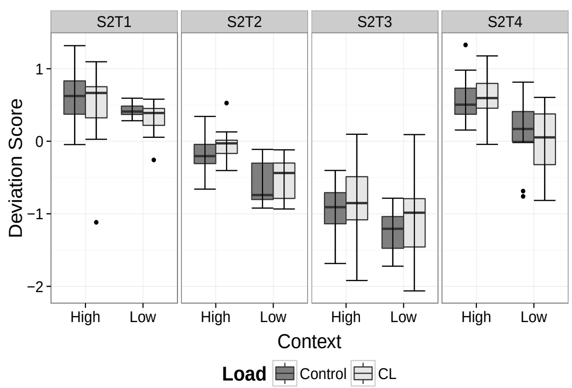
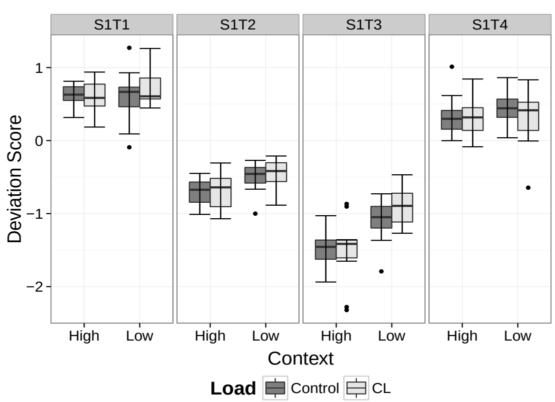


Figure 2: Anticipatory coart. by context and tone.



4. DISCUSSION

Our first finding, that carryover coarticulation showed no effect of CL, is in line with previous work indicating a lack of advance planning for carryover coarticulation [24], and is also consistent with Target Approximation models which see carryover effects as resulting from articulatory inertia in approximating tonal targets [27]. For anticipatory coarticulation, we found increased dissimilation on T1 and T3 where they preceded tones with low f_0 onsets. That these melodies were the most affected is striking given that such patterns of 'upstep'–or 'downstep', depending on one's perspective–are common crosslinguistically, possibly attributed to speakers' overcompensation for declination, which serves to lower f_0 over the course of an utterance [25]. While downstep typically raises a high tone before a low tone, low tone downstep is also attested [21]. Mandarin also has a sandhi process of pre-low raising where T3 is produced with a rise before another T3.

It is interesting that CL served to increase anticipation, as previous work has suggested that interfering with speech planning should lead to a decrease in coarticulation. This leads us to posit a link between the commonness of processes such as downstep and the observed influence of cognitive load: if such processes serve a specific linguistic function (e.g. to enhance perceptibility of tonal contrasts), it may be that a dedicated cognitive mechanism serves to bolster them even when CL is high. For example, it could be that the articulatory apparatus is directed to selectively overcompensate when cognitive resources for speech production are depleted and fine-grained phonetic distinctions are vulnerable. A recent model which may serve to explain this mechanism is Tilsen's 'field model' of speech production which posits that contrasts are maintained through the targeted excitation and inhibition of regions of the speech planning space (e.g. vowel space, tone space, etc.) [22]. Though the model was originally proposed to explain dissimilation patterns between contrastive vowels, a similar mechanism could be extended to other contrast maximization processes.

Effects of CL on contour Tones 2 and 4 were more mixed. For T2, a main effect of LOAD was found in the anticipatory environment only, suggesting that CL may serve to influence tone production in more general ways beyond coarticulation. It is possible that the more complicated gestural composition of this tone led to its being more affected by cognitive load [9]. It is unclear to us why this tone should have been more affected in σ_2 position, though we note the effect of LOAD also approached significance for T2 in the carryover context (p=0.13). As for T4, previous work has shown it to be the most resistant to coarticulatory effects[25]; thus, it is not entirely surprising that no contextual effects (CL-related or otherwise) were found of this tone.

5. CONCLUSION

We have shown that CL shows a modest influence on some types of tonal coarticulation, namely anticipatory raising before Low f_0 contexts. Such findings suggest that the process of upstep/downstep may serve an important linguistic function crosslinguistically in the maintenance of tonal contrasts.

6. REFERENCES

[1] Baddeley, A. 1986. *Working Memory*. Oxford: Oxford University Press.

[2] Baddeley, A., Hitch, G. 1974. Working memory. Bower, G., (ed), *The psychology of learning and motivation* volume 8 New York. Academic Press 47–90.

[3] Bates, D., Maechler, M., Bolker, B., Walker, S. 2014. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7.

[4] Blevins, J. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.

[5] Boersma, P., Weenink, D. 2014. PRAAT (version 5.3.71). http://www.praat.org.

[6] Broadbent, D. E. 1958. *Perception and Communication*. Pergamon Press.

[7] DiCanio, C. 2014. Triqui tonal coarticulation and contrast preservation in tonal phonology. *Presentation at SSMCA 2014*.

[8] Francis, A., Nusbaum, H. 2009. Effects of intelligibility on working memory demand for speech perception. *Attention, Perception & Psychophysics* 71(6), 1360–1374.

[9] Gao, M. 2008. Gestural coordination among vowel, consonant and tone gestures in Mandarin Chinese. *Presentation at the 8th Phoneic Conference of China and the International Symposium on Phonetic Frontiers*.

[10] Halekoh, U., Højsgaard, S. 2014. A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models – the R package pbkrtest. *Journal of Statistical Software* 59(9), 1–30.

[11] Hartsuiker, R., Corley, M., Martensen, H. 1975. The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related reply to baars et al. (1975). *Journal of Memory and Language* 52, 58–70.

[12] Hugdahl, K., Thornsen, T., Ersland, L., Rimol, L. M., Niemi, J. April 2003. The effects of attention on speech perception: An fMRI study. *Brain and Language* 85(1), 37–48.

[13] Johnson, K. 2005. Speaker normalization in speech perception. Pisoni, D., Remez, R., (eds), *The Handbook of Speech Perception* Oxford. Blackwell Publishers 363–389.

[14] Mattys, S., Wiget, L. 2011. Effects of cognitive load on speech recognition. *Journal of Memory and Language* 65, 145–160.

[15] Norman, D., Bobrow, D. 1975. On data-limited and resource-limited processes. *Cognitive Psychology* 7, 44–64.

[16] Ohala, J. 1993. The phonetics of sound change. Jones, C., (ed), *Historical Linguistics: Problems and Perspectives* London. Longman.

[17] Pisoni, D. 1997. Some thoughts on "normalization" in speech perception. Johnson, K., Mullennix, J., (eds), *Talker Variability in Speech Processing* San Diego. Academic Press 9–32.

[18] Posner, e. a., M.I. 1980. Attention and the detection of signals. *Journal of Experimental Psychology Gen* 109, 160–174.

[19] R Development Core Team, 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. ISBN 3-900051-07-0.

[20] Sternberg, S., Monsell, S., Knoll, R., Wright, C. 1978. The latency and duration of rapid movement sequences: comparisons of speech and typewriting. Stelmach, G., (ed), *Information Processing in Motor Control and Learning* New York. Academic Press 469–505.

[21] Tadadjeu, M. 1974. Floating tones, shifting rules, and downstep in Dschang-Bamileke. *Studies in African Linguistics* Supplement 5, 283.

[22] Tilsen, S. 2013. Inhibitory mechanisms in speech planning maintain and maximize contrast. Yu, A. C., (ed), *Origins of Sound Patterns: Approaches to Phonologization* 112–127.

[23] Vitevitch, M. 2003. The facilitative influence of phonological similarity and neighborhood frequency in speech production. *Memory & Cognition* 31, 491–504.

[24] Whalen, D. 1990. Coarticulation is largely planned. *Haskins Laboratories Status Report on Speech Research* SR-101/102, 149–176.

[25] Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25, 61–83.

[26] Xu, Y. 2013. ProsodyPro – a tool for large-scale systematic prosody analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP)*. Aix-en-Provence, France. 7-10.

[27] Xu, Y., Liu, F. 2012. Intrinsic coherence of prosodic and segmental aspects of speech. Niebuhr, O., Pfitzinger, H., (eds), *Understanding Prosody – The Role of Context, Function, and Communication*. Walter de Gruyter 1–26.