DURATION COMPENSATION IN NON-ADJACENT CONSONANT AND TEMPORAL REGULARITY

Hee-Sun Kim

Department of Linguistics, Stanford University
Stanford, CA 94305 U.S.A.
heesun@csli.stanford.edu

ABSTRACT

This study reports a case of consonant-induced duration compensation within a higher phonological unit in Korean. The main finding of this study is that longer duration in a consonant is partly compensated by shorter duration even in non-adjacent consonants. Results of experiments revealed that the shortening in non-adjacent consonants is a subsidiary compensation process to maintain constant duration at the larger domain than the syllable. Based on these observations, an attempt at modeling speech timing is presented with the perspective that speech production system is a process having multiple simultaneous tasks and the optimal output is obtained by a compromise between their objectives.

1. INTRODUCTION

A number of studies on the temporal structure of speech have shown that there is a tendency to keep relatively constant intervals between phonological units, e.g., syllable, stress, or morae, in speech production. This temporal regularity of speech, however, has been a controversial issue in speech production modeling and synthesis areas since timing adjustments, such as duration compensation, occur between phonological units, but they are often insufficient to achieve isochrony.

Cross-linguistic studies have shown a wide range of compensatory timing relationships at many different phonological levels during speech production. Within the syllable domain, vowels are normally shorter in closed than in open syllables [15] and before voiceless than voiced consonants [2]. At the foot level, a stressed syllable is shortened when unstressed syllables follow it in the same metrical foot [5]. In some languages, the word tends to remain relatively constant by decreasing of the duration of the individual segmental sounds when the more sounds the word contains [12]. The literature generally offers accounts in view of temporal regularity, i.e., syllable, stress, or possibly word timing tendency.

However, most duration compensation processes are not sufficient to maintain strict isochrony within the domains. For instance, the amount of vowel shortening due to a following voiceless long consonant is not equal to the differences of durations between voiced and voiceless consonants in the syllable domain. Shortening of stressed syllable by a following unstressed syllable is substantially less than the duration of the unstressed syllable itself, and so on.

Studies employing synthetic speech have shown that deviant durational patterns and fundamental frequency contours decrease intelligibility and naturalness [1]. Proper durational cues are effective in disambiguating sentences via their prosodic structures [13]. Most duration compensation effects could serve as perceptual cues for identifying phonological structure, e.g., to signal the stress interval or to mark syllable/word boundaries. Thus, it is necessary for a speech timing model to provide a general account of duration compensation effects with respect to temporal regularity.

The present study reports a finding of a relationship of duration compensation between non-adjacent consonants: long duration of a consonant shortens non-adjacent consonants. Most previous studies on duration compensation have more or less neglected such effects between non-adjacent consonants, since the sensitivity of duration change by the context variations is more on vowels and consonants. Given the existence of duration compensation, we can further assume that it is not necessary that compensatory relationships should be restricted to adjacent vowels or segments. This study explores such possibility of duration compensation in a larger phonological domain with Korean stops, which are characterized with specific durational relationships depending on their phonemic distinctions, lax, tense, and aspirated.

Generally, Korean tense stop closures are more than twice as long as lax stop closures in intervocalic position and the closure duration has been reported to be a strong perceptual cue for differentiating lax stops from tense stops [6]. On the other hand, VOT serves as a strong cue distinguishing aspirated stops from either lax or tense. The sum of closure duration and VOT, consonant duration, is longest in aspirated stops; lax stop duration is the shortest; tense stop duration is intermediate. The durational relationship of the three types of Korean stops is summarized in Figure 1.

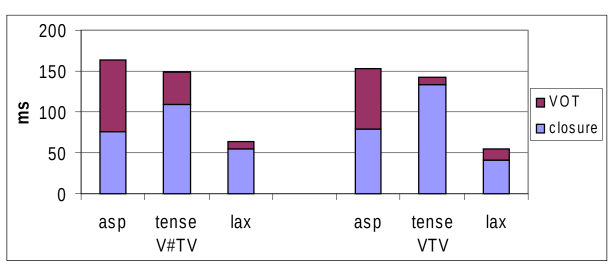

Figure 1: Durational relationship of Korean lax, tense, and aspirated alveolar stop closures and VOTs in word initial (V#TV, Kim [8]) and word medial position (VTV, Shin [16]). The VOT of lax stop in word medial position is reconstructed from Han [6]). T stands for alveolar stop.

The hypothesis is tested that long consonant duration, e.g., tense or aspirated, can induce shortening of non-adjacent consonants across a vowel. Along with positive results, this study attempts to outline a model accounting for duration compensation effects observed in this study. The model can be regarded as a development of a 'parallel' model [3], which assumes that the phonetic output is optimized with respect to conflicting, violable constraints rather than derived by sets of rules. In the model, duration compensation is considered as the consequence of a compromise between multiple constraints on timing for each phonological unit, e.g., segment, syllable, etc. While further study is necessary to discover how the proposed model is related to processes of articulation, the model should serve as an adequate account of the relationship between duration compensation and temporal regularity as a whole.

2. EXPERIMENT 1

Procedure

The first experiment examined duration compensation between non-adjacent consonants with disyllabic real and nonsense words ((C_1)VC_2V+ particle '-ka'). Specifically, it was to investigate the degree to which a preceding consonant (C_1) affects the closure duration and the VOT of the consonant (C_2) in the following syllables. Only alveolar stops were chosen in this experiment since velar and bilabial lax stops in intervocalic position are easily spirantized in casual speech.

Four Seoul Korean speakers, two males and two females, took part in the present investigation. The stimuli consisted of twelve disyllabic forms with consonants ranging from lax to aspirated stops, *ata, tata, t'ata, t^{h}ata* (group 1): *at'a, tat'a. t'at'a, t^{h}at'a* (group 2): *at^{h}a, tat^{h}a, t'at^{h}a, t^{h}at^{h}a* (group 3). The words were placed in the sentence frame, ______*+ka yəki+e is'+ə* ('______ is here'). Subjects practiced the tokens in order to pronounce them like real words before recording. Sentences were repeated 6 times in randomized order. All of the subjects reported that the two forms of tokens, *t^{h}at'a* and *t'at^{h}a*, with the carrier sentence, were hard to pronounce as real words and made many mistakes in actual recording. Therefore, the series of two tokens are exempted from the analysis. The closure duration and the VOT of C_2 were measured.

Results

All four speakers show considerable and consistent shortening of C_2 when the C_1 stop has a longer consonant duration (backward shortening). Figure 2 shows the compensatory relationships between C_1 and C_2 in terms of the mean durations of C_2. The figure shows that C_2's durations decrease (lax stop /t/) as C1's closure duration increases (no onset < lax < aspirated < tense). The mean closure duration of C_2 is slightly shorter when C_1 is an aspirated stop than a tense stop (*t'ata*: *t^{h}ata*) even though they did not show statistical differences.

VOTs of tense and lax stops (groups 1 and 2) do not show any shortening due to the different preceding consonantal environments. The VOT ranges of tense and lax stops in /ata/ and /at'a/ overlap each other in word medial position as other previous observations on Korean stop. On the other hand, VOTs of the aspirated stops (C_2) decrease in two speakers as the consonant duration of C_1 increases (no onset < lax < aspirated). But the VOT shortening of aspirated stops was not consistent across speakers.

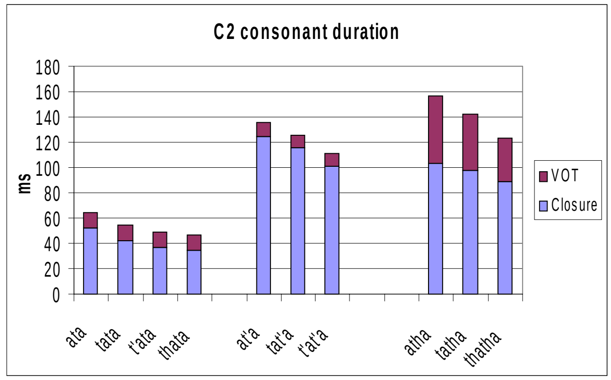

Figure 2: Shortening pattern of the consonant duration of C_2 depending on the type of C_1, separated by closure duration and VOT. Data were pooled across speakers.

To sum up, the duration of C_2 is shorter as the whole consonant duration of C_1 increases. The statistical analysis supports the shortening effect on C_2 due to a longer consonant C_1: in pairs with no-onset and lax ($F(1, 142)= 27.28$ $p<.0001$), lax and tense ($F(1, 94)= 17.34$ $p<.0001$), and lax and aspirated ($F(1, 94)= 25.28$ $p<.0001$), but in the pair of tense and aspirated, no significant difference was found ($F(1, 46)= 0.17$ $p=0.68$). As can be seen from the results, however, the shortening of C_2 is not exactly proportional to the durational differences of C_1. It implies that there are shortening of other segments, e.g. intervening V, and some limit of stop duration compression to preserve their phonemic distinctions.

3. EXPERIMENT 2

Procedure

The second experiment was designed to explore further questions raised from Experiment 1: (i) In what domain does consonant shortening occur?, (ii) Are there any other segments showing shortening due to a long consonant? If so, which segments are shortened and how much?, and (iv) Is the shortening a duration compensation process to maintain some temporal regularity in a larger domain than the syllable? For these questions, a multi-syllabic string, /t_1*a*.t_2*a*.C_3*a*.t_4*a*+*ka*/, was used, where C3 ranges from alveolar lax to aspirated stops, *tatatata* (type 1), *tatat'ata* (type 2), *tatat^{h}ata* (type 3). Two male and two female native speakers participated in the recording. The carrier sentence form used in Experiment 2 is, _______*+ka tʃəki+e is'+ə* ('______ is there').

Even though the words used in this experiment are 4 syllable nonsense words, they are phonotactically permissible forms in Korean and subjects did not report any special difficulties in speaking the words in the carrier sentence. Before the recording, the tokens with a following particle *–ka* are controlled to have one accentual phrase, defined by Jun [7], in order that subjects read the tokens like a word as in Experiment 1. If the speaker used a different phonological phrasing while they were practicing the sentences before recording, the experimenter (the author) presented two or three sentences as samples to induce subjects to read *tataC$_3$ata+ka* as one accentual phrase. Subjects were asked to read the sentences in a manner. The list was also repeated 6 times in randomized order. The durations of each syllable separated by each consonant and vowel in the *tataC$_3$ata+ka* string were measured. The first syllable duration does not include the closure duration of consonant /t/ since the closure duration in sentence initial position is impossible to measure with spectrogram and waveforms. Thus, the first syllable duration is measured from the release of the first stop to the end of the first vowel.

Results

Both anticipatory (in t_2) and backward (in t_4) shortening of consonant durations are observed when the non-adjacent consonant C_3 is long, tense or aspirated. Overall, the mean duration of the third syllable is significantly longer when the laryngeal type of C_3 is tense or aspirated (35-50 ms). However, the whole duration of the phonological phrase remains relatively constant regardless of the laryngeal types of C_3. The statistical analyses show no statistical differences between the sums of the five syllable durations from Type 1 to 3 both individually and across speakers. In addition, the first syllable and the fifth syllable also keep similar durations in all types (Type 1, 2, and 3) for every speaker: statistics show no duration differences in the first syllable and in the fifth syllable with respect to different types. That is, the sum of the 2nd, 3rd, and 4th syllables are relatively consistent in duration. Figure 3 shows the mean duration of consonants and vowels in second, third, fourth syllable as percentages of the interval from C_2 to V_4.

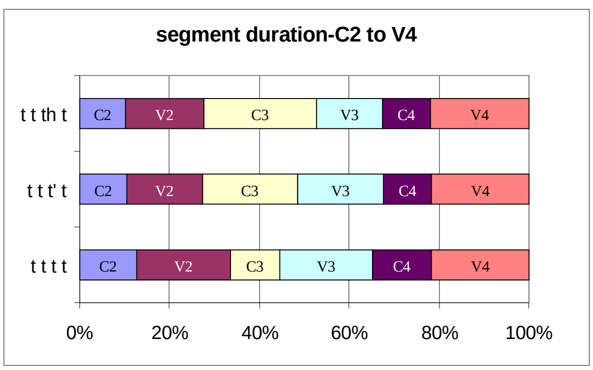

Figure 3: The relationship of the second, third, and fourth syllable duration in percentile separated by each segment.

As can be seen in Figure 3, a long stop in C3 (tense or aspirated), shortens the durations of C2, V2 to the left and V_3, C_4 and V_4 to the right. The shortening effect is greater in the second syllable (20-35ms) than in the fourth syllable (7-10ms). In this experiment, however, limited evidence for backward shortening is found in the fourth syllable. In particular, V_4 does not show a significant shortening across speakers. In addition, YJ and DO, show neither lengthening nor shortening in V_3 when C_3 is a tense stop whereas HY and HJ show shortening of V_3 in the same context. When C_3 is an aspirated stop, V_3 is significantly shortened across speakers. All subjects show shortening of C_4, but shortenings are smaller (7-10ms) than in Experiment 1 (10-15ms). For the variations of the length of V_3 in Type 2 depending on subjects, more data with the extensive number of subjects will be required in future research.

The results of Experiment 2 indicate that shortening effects observed in Experiment 1 are a subsidiary timing compensation process to hold some regular temporal cycle within a large phonological domain than the syllable. And the duration compensation is locally restricted to a subpart of an accentual phrase.

4. DISCUSSION: A PARALLEL MODEL

There have been two main theories in modeling of speech timing: Intrinsic timing theory [4,8] and Extrinsic timing theory [10, 11, 14]. In view of intrinsic timing theory, speech timing is intrinsic in coordinated speech motor plan and there is no neural functioning in the timing of speech actions. A number of phenomena in speech timing, however, are difficult to explain without reference to absolute time which may be controlled by certain neural structures directly involved in speech production (Extrinsic timing theory). Within the extrinsic timing theory, two different models have been proposed [10, 11]. Klatt's model can be categorized as a bottom-up model while Kohler's model as a top-down model. Flemming [3] points out that both models have problems to account for the relationship between duration compensation and the tendency to maintain temporal regularity. As an alternative, Flemming [3] proposes a parallel model, which incorporates both top-down and bottom-up effects.

The observations on duration compensation in Experiment 1 and 2 are to be accounted for by adopting the parallel model. The model is an optimization model, in which durations are assigned to best satisfy multiple constraints for producing the optimal output. In other words, there are inherent timing targets for both segments and larger constituents such as the syllable. These constraints conflict when the target duration of the larger phonological constituent is different from the sum of the targets for sub-components. Then, the optimal results will be obtained by compromising between segmental and higher targets, so that it yields compensatory effects between segments. To explain the qualitative properties of duration compensation between phonological units, the cost function is proposed to calculate the extent of compromise of each target. Specifically, the cost function imposes the sum of costs for the deviation from each target, i.e., the difference between the achieved duration and the inherent target duration. The output requiring the minimal cost will be the optimal one.

For the results of Experiment 2, suppose that the duration of the third syllable in Type 1 has the inherent syllable target duration. When the segment, C_3, is a tense stop which has a longer target duration, it conflicts with the syllable target duration because the sum of the target segment duration of C_3 and V_3 will be significantly longer than the syllable target duration. For this reason, C_3 and V_3 will be shortened in the extent of yielding bearable costs at the segment level and maintaining the phonemic distinction from a lax stop. But they are still too long with respect to the target duration of the syllable domain and ultimately it cannot satisfy the target duration of the accentual phrase. Thus, to avoid additional costs by the deviation from the target phrase duration, other segments and syllables existing in the domain will be shortened to minimize the final cost. Consequently, the differences in the third syllable duration due to a longer consonant, as shown in Experiment 2, are completely compensated within the accentual phrase while the compensation is incomplete with respect to the target duration of the syllable. This implies that the tendency for the temporal regularity may exist in a larger domain than the syllable in Korean.

For a robust model of duration compensation, however, we should note the locality of duration compensation observed in Experiment 2: a longer consonant duration is compensated in a subpart of the phonological phrase domain. Perhaps in the process of speech production, the sounds are stored in various sizes of concatenated buffers and there is a tendency for speakers to accommodate the extra duration by small adjustments of timing in the domain. If not, the duration compensation in Experiment 2 may occur in all five syllables in small steps without requiring further costs on the contrary to our observations. And the minimal local domain conditioned by a syllable having the deviant duration from the target duration will be determined by the amount of duration difference. In Korean, it seems that the minimal domain for duration compensation triggered by a longer consonant consists of the preceding and the following syllables as well as the syllable having the consonant.

As further application, the model can include many different timing targets besides segments and syllables: languages showing the tendency of stress timing like English may have the foot target, and mora timing languages like Japanese may have the mora target in the cost function. And the dependency of maintaining each target duration will be determined by language-specific weight of the target in the cost function. If a language has a high weight at the foot domain, then greater duration compensation effects will occur at the foot level in that language since the cost function will penalize deviation at the foot level more than at other levels.

5. CONCLUSIONS

So far, we have seen that a long consonant can trigger a wide range of duration compensation in a domain, which is larger than the syllable but smaller than the accentual phrase. This locally restricted duration compensation suggests that in speech production, there is a tendency to minimize the number of units participating in the timing adjustment process for the temporal regularity.

The duration compensation observed in this study is explained within the parallel model. The parallel model is suitable for providing a generalized account for the relationship between incomplete duration compensation and temporal regularity since this model fundamentally assumes multiple timing targets and their compromise. Obviously, we need extensive studies using a large body of speech data to assess its validity, and to specify principles of temporal organization of speech. However, this study would offer a more explicit model accounting for the relationship between duration compensation and temporal regularity in general.

6. REFERENCES

1. Carlson, R., B. Granström, and D. Klatt. “Some notes on the perception of temporal patterns in speech”. In B. Lindblom and S. Öhman (eds.) *Frontiers in Speech Communication Research*, Academic Press, New York, 233-243, 1979.
2. Chen, M. “Vowel length variation as a function of the voicing of the consonant environment”, *Phonetica* 22: 129-159, 1970.
3. Flemming, E. “Phonetic optimization: Compromise in speech production”, *University of Maryland Working Papers in Linguistics* 5, 1997.
4. Fowler, C. “Coarticulation and theories of extrinsic timing”, *Journal of Phonetics* 8:113-133, 1980.
5. Fowler, C. “A relationship between coarticulation and compensatory shortening”, *Phonetica* 38: 35-50, 1981.
6. Han, J.-I. “The Phonetics and Phonology of Tense and Plain Consonants in Korean”, Ph.D. dissertation, Cornell University, 1996
7. Jun, S.-A. “The Phonetics and Phonology in Korean Prosody”, Ph.D. dissertation, Ohio State University, 1993.
8. Kelso, J. and Tuller, B. “Intrinsic time in speech production: Theory, methodology, and preliminary observations”, In Keller, E and M. Gopnik (eds*.)*, *Sensory and Motor Progress in Language*. Hillsdale, NJ: Erlbaum, 203-222.
9. Kim, M.-R. “Acoustic Characteristics of Korean Stops and Perception of English Stop Consonants”, Ph.D. dissertation, University of Wisconsin, 1994.
10. Klatt, D. “Synthesis by rule of segmental durations in English sentences”, In B. Lindblom and S. Öhman (eds.), *Frontiers in Speech Communication Research*, Academic Press, New York, 1979.
11. Kohler, K. “Invariance and variability in speech timing: from utterance to segment in German”, In S. Perkell and D. Klatt (eds.), *Invariance and Variability in Speech Processes*, Hillsdale NJ: Erlbaum Associates, Inc., 268-299, 1986.
12. Lehiste, I. “Temporal organization of spoken language”, In L. Hammerich et al (eds*.)*, *Form and Substance*. Copenhagen: Akademisk, 159-169, 1971.
13. Lehiste, I, Olive and Streeter 1976. Lehiste, I., J. Olive and L. Streeter. (1976) “The role of duration in disambiguating syntactically ambiguous sentences”, *JASA* 60, 1199-1202.
14. Lindblom, B. et al. “The Concept of Target and Speech Timing”, In R. Channon and L. Shockey (eds.), *Honor of Ilse Lehiste/Ilse Lehiste Puhendusteos*. Foris Dordrecht, 161-181,1987.
15. Maddieson, I. “Phonetic cues to syllabification”, in V. Fromkin (ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, Orland: Academic Press, 203-221, 1985.
16. Shin, J.Y. “Consonantal production and Coarticulation in Korean”, Ph.D. dissertation.University of London, 1997.