THE TIMING OF VOWEL AND CONSONANT GESTURES IN ITALIAN AND JAPANESE

Caroline L. Smith

Yale University and Haskins Laboratories,
New Haven, Connecticut

ABSTRACT

Languages commonly described as syllable-timed, such as Italian, are perceived as having a rhythm in which each vowel is the nucleus of a rhythmic unit. In contrast, in mora-timed languages such as Japanese the basic rhythmic unit, the mora, depends on the durations of both vowels and consonants. It is proposed here that the basis for the contrast between these two types of languages is a correlation between the temporal organization of articulatory gestures for vowels and consonants and the role of vowels in the overall rhythm of a language: in syllable-timed languages, vowels have primacy over consonants, but in mora-timed languages vowels and consonants are of equal importance.

1. INTRODUCTION

The hypothesis being tested is that stress-, syllable- and mora-timed languages are characterized by more or less independence in the timing of vowel and consonant gestures. (The term gesture refers to an abstract, dynamic unit associated with the production of a particular vowel or consonant that controls the spatiotemporal movement of one or more articulators towards a target.) Both of the models of gestural timing that will be compared here assume that the temporally overlapping production of gestures is responsible for their apparent context dependence, but the models differ in their accounts of how gestures are coordinated in time. Both models were proposed to account for English and other languages, but seem to capture characteristics of different types of rhythmic behavior.

The vowel-based timing model [5, 6, 8] claims that gestures for vowels and consonants are coordinated at different levels. Vowel gestures are coordinated with respect to one another, and consonants are coordinated with respect to the vowels. This model predicts that vowel gestures will be unaffected by temporal changes to consonant gestures. Because of the primacy of vowels in determining syllable-timed rhythm, the vowel-based model was expected to apply to Italian.

The vowel-and-consonant timing model [2, 3, 4] claims that vowels and consonants are coordinated at the same level. Since this means that intergestural timing for vowels and for consonants is interdependent, a timing change to any gesture is predicted to cause adjustments in both sets of gestures. This model was expected to apply to Japanese, because mora-timing requires the temporal integration of vowels and consonants. In Japanese, the timing of two vowels relative to each other would be expected to be susceptible to changes in the duration of intervocalic consonants.

Because the two models' predictions differ primarily in the extent to which vowels are affected by changes in the timing of consonants, contrasting utterances that differ only in the length of an intervocalic consonant provide a way to compare the two models. However, since the predictions of the models are couched in terms of abstract gestures, the gestures, in order to be compared experimentally, must be associated with specific articulatory movements. Vowel gestures, for example, can be associated with an appropriate movement of the tongue body (or root), and consonants with the lips or the tongue forming a constriction in the supralaryngeal part of

the vocal tract. Associating gestures with movements in this way makes it possible to compare gestures in different contexts, but it does not differentiate the roles of the various articulators in making the constriction.

2. METHOD

In order to measure the movements of the tongue associated with vowel gestures, as well as the lips and jaw, data were collected at the NIH X-ray microbeam facility at the University of Wisconsin. The microbeam records the movements of the tongue, lips and jaw by means of microscopic X-rays tracking tiny gold pellets attached to the articulators [1]. Pellets were attached to the speakers' nose and upper incisor (to correct for head movement), lower incisor (to measure jaw movement), lower and upper lips, and to four points along the midline of the tongue, starting approximately 1 cm behind the tip of the extended tongue. The microbeam data consist of the horizontal and vertical trajectories of each of these pellets.

The data presented here are a subset of a larger study, in which three native speakers each of Italian and Japanese participated. Data from only one speaker of each language will be discussed in this paper. Each speaker produced, in carrier phrases designed to be comparable across languages, disyllabic utterances of the form "mV_1CV_2", where V_1 and V_2 were /i/ or /ɑ/, and C was one of /p/, /pp/, /t/, /tt/, /m/, /mm/, /n/ or /nn/. In Japanese, utterances with /t/ or /tt/ as the intervocalic consonant and /i/ as the second-syllable vowel were excluded because /t/ palatalizes in this context. The Italian speaker produced 9 to 11 tokens of each utterance, and the Japanese speaker 12 to 16 tokens. The movement trajectories were digitized and smoothed prior to analysis.

Because of the very high correlations among the tongue pellets (as high as .95 between the x-dimensions of two pellets), a factor analysis was performed on the x and y positions of the pellets at successive 5 ms frames, with the intention of extracting factors that would reflect the positioning of the tongue for the different vowels. Examination of the movement trajectories had suggested that the frontmost tongue pellet showed primarily movement associated with the alveolar consonants, so it was excluded from the factor analysis, leaving 6 dimensions, from which 2 factors were extracted.[1] The first of these was primarily associated with horizontal movement, and the second with vertical movement. Pellet trajectories were also measured that were expected to show movement typical of specific gestures. The trajectories that were measured are shown below.

Table 1. Trajectories measured.

Trajectory	Associated Gesture
Lower Lip vertical	initial /m/, bilabial intervocalic consonants
Tongue Tip vertical	alveolar intervocalic consonants
Tongue Dorsum horizontal	vowels
Tongue Body Rear vertical	vowels
Horizontal tongue factor	vowels
Vertical tongue factor	vowels only in Italian

The utterances measured were those in which the two vowels were different, as these permitted the identification of movements from the first vowel to the second. Five time points, defined as the edges of periods of zero velocity, were located in each of the trajectories associated with vowel gestures: the onset of movement towards the first vowel, the time at which the movement for the first vowel reached its target, the end of the plateau region for the first vowel, the time at which the movement for the second vowel reached its target, and the end of the plateau region for the second vowel.

3. RESULTS

Different intervals between the labelled time points were measured in order to determine whether the time between the two vowels was changing when the length of the intervocalic consonant changed. ANOVAs were run for each subject separately, with the intervals between the labelled points as dependent variables and grouping factors Length (of intervocalic consonant), Place of articulation, Consonant Identity, and Vowel quality.

Figures 1 and 2 illustrate tokens of /mipa/ (solid line) and /mippa/ (dotted line) from Italian and Japanese, aligned at the release of the initial /m/. In Italian (Figure 1), the large humps associated with the production of /i/ in the top three articulatory trajectories are virtually identical in /mipa/ and /mippa/; the rear and downward movements for /a/ also coincide. The two utterances differ in the positioning of the central hump in the Lower Lip trajectory, which corresponds to the intervocalic consonant, relative to the other movements. The raising of the lower lip for /pp/ occurs earlier relative to the preceding lip movement ($p<.001$ for the effect of Length)[2] and to the tongue movement for the /i/ than does the raising for /p/ ($p<.001$), resulting in the preceding vowel being shorter acoustically before the geminate ($p<.001$), a well-known characteristic of Italian [7].

Figure 2, for Japanese, shows the raising of the lower lip for the intervocalic consonant occurring at about the same time relative to the preceding lip and tongue movements, with the preceding vowel slightly longer acoustically preceding the geminate ($p<.001$). Although the tongue raising and fronting begins at approximately the same time in both utterances, the lowering and backing for /a/ is significantly delayed when following /pp/ ($p<.001$ for the effect of Length).

This impressionistic pattern is borne out by measurements of the interval between the times at which the two vowels reach their targets, whose approximate locations are indicated by arrows on the figures. This interval was consistently longer in Japanese ($p<.001$ for all trajectories). The statistical results for Italian were more variable, but with negligible numerical differences found in the contexts of the two consonant lengths. These results support the hypothesis that the second vowel is delayed relative to the first in Japanese but not in Italian.

Although preliminary, the results shown here do suggest that the timing of the two vowels relative to each other is controlled independently of the consonants in Italian but in conjunction with them in Japanese. The most immediate implication of this is that neither model of timing organization can claim to be the most insightful for both types of languages. The apparent relation between the form of temporal coordination between vowel and consonant gestures and the corresponding differences in linguistic rhythm suggests that the organization of gestural timing may be a source for the differing rhythmic behavior between syllable- and mora-timed languages.

4. REFERENCES

[1] Abbs, J., & Nadler, R. (1987), "User's Manual for the University of Wisconsin X-ray microbeam".

[2] Browman, C.P., & Goldstein, L. (1987), "Tiers in articulatory phonology, with some implications for casual speech", *Haskins Laboratories Status Report, 92*, 1-30.

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

[4] Browman, C.P., & Goldstein, L. (1990), "Gestural specification using dynamically-defined articulatory structures", *Journal of Phonetics, 18*, 299-320.

[5] Fowler, C. (1981), "A relationship between coarticulation and compensatory shortening", *Phonetica, 38*, 35-50.

[6] Fowler, C. (1983), "Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet", *J. of Exp. Psych.: General, 112*, 386-412.

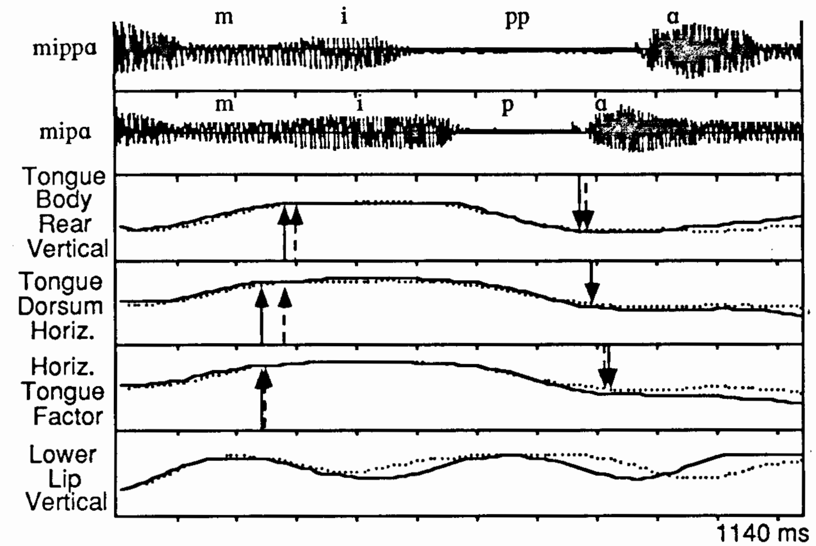
[7] Maddieson, I. (1985), "Phonetic cues to syllabification", in Fromkin, V., (ed.), *Phonetic Linguistics*, 203-221, Orlando: Academic.

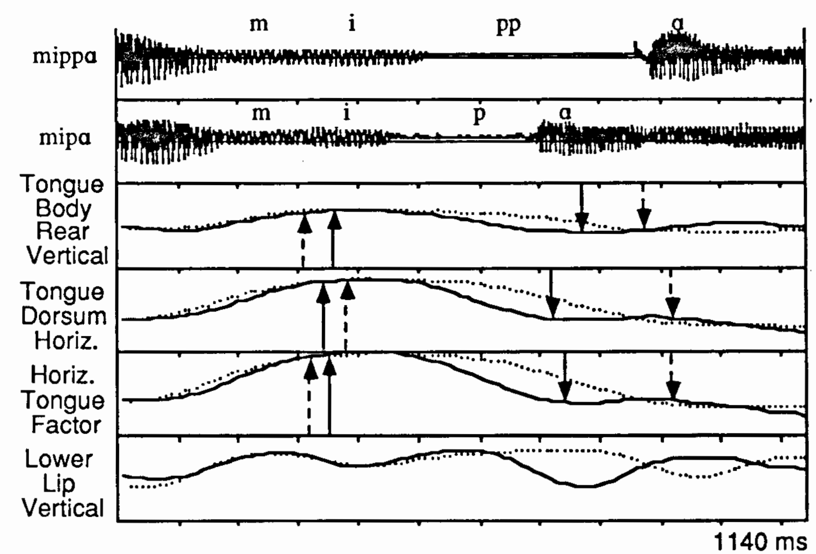
[8] Öhman, S.E.G. (1966), "Coarticulation in VCV utterances: spectrographic measurements", *J. Acoust. Soc. Am., 39*, 151-168.

Work supported by NSF grant BNS-8820099 and NIH grant DC-00121 to Haskins Laboratories.

[1] The factor analysis was a principal components analysis using BMDP 4M, with a VARIMAX rotation. The factor scores were then calculated for each frame of data, providing trajectories similar in form to the pellet trajectories.

[2] Results reported here for the effect of Length are based on 1,145 degrees of freedom for Italian and 1,195 for Japanese.





Figures 1 and 2. Productions of, at the top, Italian "Dica **mipa** molto" (solid lines) and "Dica **mippa** molto" (dotted lines), and at the bottom, Japanese "Boku wa **mipa** mo aru" (solid lines) and "Boku wa **mippa** mo aru" (dotted lines). Time is along the horizontal axis; each tick mark indicates 100 ms. The utterances within each picture were aligned at the release of the initial /m/ in the target word. The times at which the vowels reached their targets are shown by arrows (solid for the single consonant, dotted for the geminate).