LONG-DISTANCE ANTICIPATORY VOWEL-TO-VOWEL ASSIMILATORY EFFECTS IN FRENCH AND JAPANESE

Faith Chiu[1], Lauren Fromont[2,3], Albert Lee[1], Yi Xu[1]

[1]Speech Hearing and Phonetic Sciences, UCL (University College London), [2]School of Speech Language Pathology and Audiology, University of Montreal, [3]Centre for Research on Brain, Language, and Music, Montreal
faith.chiu.11@ucl.ac.uk, lauren.fromont@umontreal.ca, kwing.lee.10@ucl.ac.uk, yi.xu@ucl.ac.uk

ABSTRACT

This paper examines language-specific differences in anticipatory vowel-to-vowel coarticulation using two non-stress languages. Native speakers of Standard French (n=6) and Tokyo Japanese (n=5) served as subjects to a production study. To investigate possible long-distance effects between and beyond adjacent vowels, linguistic material consisting of /ba.bV/ and /ba.ba.bV/ was embedded within a carrier sentence in each language. The word-final trigger vowel (V) is /a/, /i/ or /u/. Acoustic analysis of continuous F_1 and F_2 trajectories as well as single-point formant measurements revealed opposite patterns in the two languages. Strong anticipatory effects in vowels up to 2 preceding syllables were observed in French. However, Japanese displayed few statistically significant anticipatory effects in any vowel preceding any trigger. We interpret the results as an indication that there are two rather different types of contextual phonetic variability. We also assert not all phonetic assimilatory effects in "coarticulation" are due to articulatory overlap.

Keywords: vowel-to-vowel coarticulation, contextual variation, assimilatory effects, French, Japanese

1. INTRODUCTION

To this day, accounting for anticipatory vowel-to-vowel (V-to-V) coarticulation remains a challenge. A major difficulty stems from understanding the nature of anticipatory coarticulation in general. Among the many proposed accounts [7], a key point of contention is whether there exists a limit to the temporal scope of coarticulation. On one end, the feature-spreading account maintains that a feature is assimilated regressively as long as it is not blocked by a conflicting feature, implying an unlimited temporal scope to anticipatory effects. Supporting evidence can be found in French where anticipatory coarticulation can span up to six segments [3]. The time-locked account, in contrast, argues that the anticipation of a segment has a fixed and limited time window relative to the closure of the syllable-initial consonant [1]. Due to the assertion of a fixed interval, one weakness of this view lies in its inability to model idiosyncratic, conflicting segment-intrinsic properties which can block or delay V-to-V coarticulatory effects. For example, the lack of regressive V-to-V rounding documented in [1] can actually be accounted for by the necessity to complete non labial-spreading of a preceding /i/ vowel [6]. Furthermore, the time-locked account proves inadequate in consideration of long-distance anticipatory coarticulation. In particular, the very mechanism in the time-locked view which explains local coarticulation is a stipulation of a fixed time window. This fixed window also implies a finite temporal scope which problematically does not account for every case of long-distance coarticulation, such as cross-syllable V-to-V assimilatory effects observed in English [9] and French [11].

Long-distance V-to-V coarticulation is known to vary across languages [13]. An existing explanation of such language-specific variability has been attributed the nature of vocalic inventories. The amount of coarticulation depends on the contrastive pressure which is derived from the number of vowels in the language, a more crowded vowel space would lead to less anticipatory coarticulation [10].

The present study is an effort to improve our understanding of long-distance anticipatory V-to-V coarticulation by simultaneously investigating two critical aspects of the effect: language-specificity and temporal scope. For language-specificity, we compared French and Japanese, two languages that differ in terms of density of vowel space, with only five vowels in Tokyo Japanese and fourteen in standard French, including schwa. If contrastive pressure is a key determining mechanism [10], we should expect Japanese to show a greater amount of V-to-V coarticulation than French. In addition, the two languages are similar in terms of a lack of lexical stress and a relatively fast speech rate (number of syllables per second), with Japanese slightly faster than French [14]. Given these similarities and existing findings of a previous study [8], we predicted that Japanese would present similar long-distance anticipatory V-to-V assimilatory effects as in French. To examine the temporal scope of the coarticulatory effect, we adopted a method that compares continuous formant trajectories in controlled disyllabic and trisyllabic /(ba)babV/ minimal pair sequences which varied only in the vowel of the final syllable. The method of comparing continuous trajectories in minimal pairs has been used mainly in studying articulatory movements [1], but is rare in formant analysis. To facilitate the comparison of formant trajectories, we used a recently developed Praat script described below.

2. METHOD

2.1. Stimuli

The speech material consisted of nonsense words /ba.bV/ and /ba.ba.bV/ where the final vowel was /a/, /i/ or /u/. The segment /b/ was chosen as the intervening consonant since it is labial, not lingual, and therefore should display the least articulatory resistance against lingual manipulations of vowel targets [5, 17]. These nonsense words were embedded in carrier phrases in the two target languages, respectively for French: '*Je veux le jeu* ___' /ʒəvøləʒø ___ /, and for Japanese: 'また ___ です' /mata ___ desu/.

Since all the non-trigger vowels in the stimuli were uniformly set to /a/, it is reasonable to assume that any observable difference in the formants of /a/ is the result of influence from the word-final trigger vowel. Furthermore, we are able to avoid confounds from vowel intrinsic properties as mentioned in [2, 6] by having only one comparable vowel (with multiple triggers). Also, by setting all intervening consonants as /b/, any gestural differences due to CV coarticulation is minimized.

2.2. Subjects

Six native French speakers (three male and three female) and five native Japanese speakers (two male and three female) were recorded. All speakers were monolingual or reported very low proficiency in a second language. To minimize dialectal variability, only Standard French and Tokyo Japanese speakers were chosen. All subjects were between 22 to 33 years old and reported normal to corrected-to-normal vision and no history of hearing or language impairment.

2.3. Procedure

The recording took place in sound-treated booths at the University of Nantes and University College London for French and Japanese, respectively. The speech material in the carrier sentence was randomized and consecutively presented in written form on a computer screen. Japanese materials were given in *kana* syllabaries with no pitch accent specified, subjects defaulted to an unaccented LHHH sequence. Each subject read aloud the stimuli at normal speed with 10 repetitions per token, for a total of 60 stimuli per subject per language. Recordings were made with a microphone (French: Sennheiser e234 dynamic cardioid; Japanese: RØDE NT1-A condenser cardioid), an audio interface recorder (French: Presonus Audiobox USB; Japanese: RME Fireface UC) into software (French: Praat [4]; Japanese: Cool Edit) at a sampling rate of 44.1 kHz.

2.4. Measurements

Continuous trajectories of the first two formants (F1 and F2) were obtained from segmented recordings using the Praat script FormantPro [18], with 20 time-normalized points per segment. The utterances were segmented and labelled manually, first by syllables in the carrier sentence and then tokens containing target vowels were further segmented by phonemes. The beginning of each vowel was demarcated by the first formant onset [16]. The formant values were transformed into the Bark scale [19]. The measurements were then averaged across repetitions. For graphical analysis, the measurements were further averaged across speakers to generate continuous formant trajectories as those shown in Figs. 1 and 2. For statistical analysis, extreme values of F1 and F2 in each vowel were taken by FormantPro.

3. RESULTS

3.1 Comparison of formant trajectories

In Figs. 1 and 2, continuous mean F1 and F2 trajectories (in bark) are plotted in normalized time for the tri-syllabic sequences with /a/, /i/ and /u/ as the final vowel. In the French data (Fig. 1), the F1 contours of /a/ when preceding final triggers /i/ and /u/ are considerably lower than when preceding /a/; the F2 contours of /a/ when preceding the final trigger /i/ are substantially higher than when preceding /a/ and /u/. The overall formant contours are therefore influenced by the final vowel, the trigger. This leftward coarticulation occurs in all preceding syllables, including even those in the carrier ('*Je veux le jeu'*). At the same time, however, both formants tend to converge at all syllable boundaries, which suggests that it is only the vowel targets that are affected by an assimilatory effect rather than a continuous leftward spreading of either the height or backness feature to also affect consonants.

In Fig. 2, in contrast, all the Japanese non-final /a/ vowels mostly converge in their F1 and F2 trajectories, regardless of the final vowel. This suggests the final trigger asserted little assimilatory influence.

Figure 1: French formant contours for /bababa/ (blue), /bababi/ (green) and /bababu/ (red): mean F1 and F2 values (bark) as a function of time.

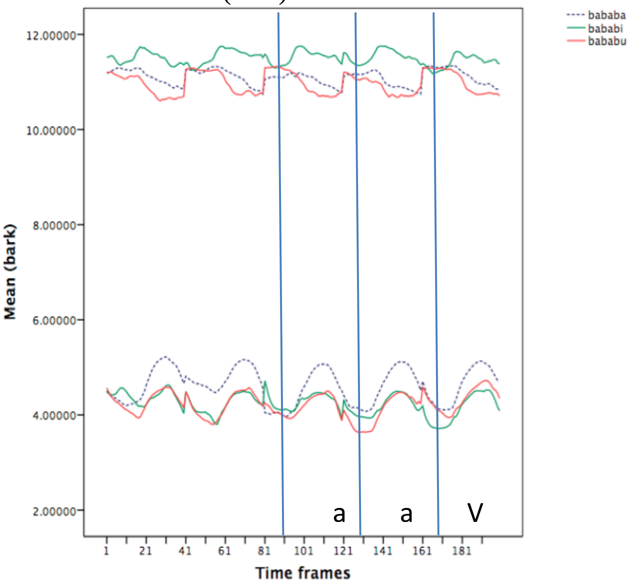
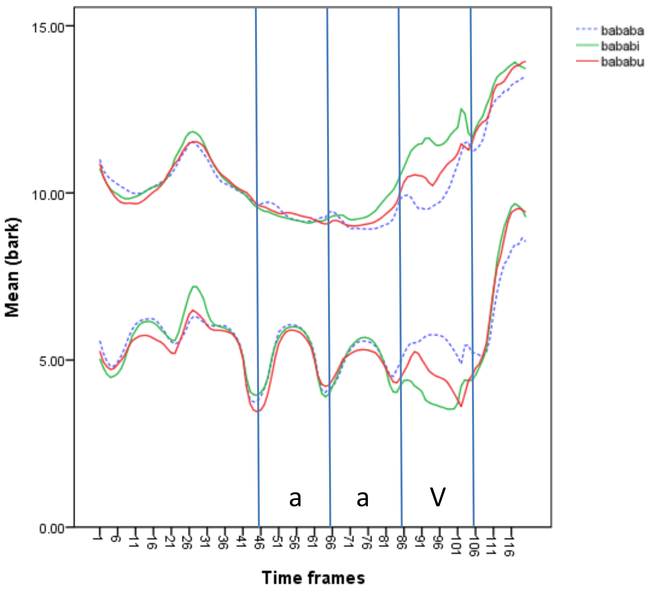


Figure 2: Japanese formant contours /bababa/ (blue), /bababi/ (green) and /bababu/ (red): mean F1 and F2 values (bark) as a function of time.



3.2 Quantitative analysis

The extreme F1 and F2 values for each vowel, which are obtained at the spectral turning point, were extracted in each language. These values were then averaged for each speaker into three pools according to the trigger /a/, /i/ and /u/ in the 2-syllable and 3-syllable conditions. Table 1 and Table 2 display the mean extreme formant values, averaged across repetitions and speakers, in 2-syllable and 3-syllable conditions respectively. As can be seen in both tables, the formants of /ba/ show clear differences depending on the trigger vowel V in French, but not in Japanese.

Table 1: Mean extreme vowel values for F1 and F2 (bark) in the 2-syllable condition.

Language		French			Japanese		
Stimuli	**V**		**ba**	**bV**		**ba**	**bV**
F1 (bark)	a		6.12	5.19		5.92	5.86
	i		4.36	3.15		5.91	3.92
	u		4.11	3.15		5.99	4.57
F2 (bark)	a		10.90	10.82		9.25	9.27
	i		11.64	13.10		9.35	12.07
	u		10.5	9.22		9.13	9.38

Table 2: Mean extreme vowel values for F1 and F2 (bark) in the 3-syllable condition.

Language		French			Japanese		
Stimuli	**V**	**ba**	**ba**	**bV**	**ba**	**ba**	**bV**
F1 (bark)	a	5.22	5.16	5.13	6.05	5.56	5.75
	i	4.63	4.49	4.38	5.99	5.68	4.38
	u	4.58	4.57	4.71	5.89	5.31	5.25
F2 (bark)	a	10.87	10.81	10.94	9.12	8.91	9.50
	i	11.47	11.71	11.64	9.10	9.14	11.85
	u	10.67	10.74	10.74	9.25	9.01	9.37

Results from a 3×1 repeated measures ANOVA for each language support the observations from both Figs. 1–2 and Tables 1–2. For French, the main effect of trigger vowel is significant for F1 ($F(1,24) = 109.29$, $p < 0.001$) as well as F2 ($F(1,24) = 348.95$, $p < 0.001$). This shows that the vowel /a/ varied depending on the trigger vowel of the sequence. For Japanese, there is a marginally significant effect on F1 ($F(1,24) = 4.25$, $p = 0.050$). But from Fig. 2 we can see that the effect is not assimilatory, since the highest F1 peak is when the trigger vowel is /i/ rather than /a/. There is no significant effect on F2.

As a further comparison, three paired comparisons were conducted for the French data of the three-syllable condition. Following the method in [16], five measurement points were selected for each pool of combined pre-trigger /a/. These measurements were then averaged between repetitions and speakers. Given that, respectively, F1 and F2 values are inversely correlated to the height of the tongue body and its backness, the obtained values were subtracted between the groups: [a-u], [a-i], [i-u]. When comparing differences in F1, paired samples t-tests revealed significant differences between [a-i] ($t(19) = 25.975$, $p < 0.001$) and [a-u] ($t(19) = 36.228$, $p < 0.001$) after Bonferroni correction. In terms of F2, significant differences were established between [a-i] ($t(19) = 25.979$, $p < 0.001$) and [i-u] ($t(19) = 36.228$, $p < 0.001$). Respectively, this suggests adjustments were in tongue height (F1) in vowels preceding /i/ or /u/ (and not /a/), and the backness of the tongue is different when anticipating a final /i/ versus /a/, and the position of backness also varies between vowels preceding /i/ versus /u/.

For Japanese, the results of a repeated measures ANOVA on the F1 and F2 values for each of the 20 timepoints of the vowel indicate significant differences only in F1, and only at the last 4 timepoints adjacent to the trigger vowel. Significant differences were found for F1 at point 16 ($F(1,24) = 4.50$, $p = 0.044$), 17 ($F(1,24) = 4.66$, $p = 0.041$), 18 ($F(1,24) = 6.18$, $p = 0.020$), 19 ($F(1,24) = 6.39$, $p = 0.018$) and 20 ($F(1,24) = 5.39$, $p = 0.029$).

4. DISCUSSION

Overall, the results revealed strong long-distance anticipatory V-to-V coarticulation in French, at both the word-level and the sentence-level. However, Japanese showed little anticipatory coarticulation, except in F1 in the last 4 points of the vowel (final 20% of the duration) in 2-syllable sequences.

4.2 Long-distance coarticulation in French

The continuous formant trajectories in Fig. 1 enable easy visualisation of the time course of the anticipatory assimilation. Such time-varying patterns pose challenges to all major existing accounts of coarticulation. The time-locked model [1] cannot account for the large time domain of long-distance V-to-V assimilatory coarticulation in French. Neither the gestural overlap [5] nor the hybrid [15] account can explain why the magnitude of the assimilation had not reduced as the distance between the trigger and the affected vowel increased. For the feature-spreading account, unless it is stipulated that this phonetic feature-spreading can only affect vowels, it is difficult to explain how the assimilatory effect can cross syllable boundaries where both F1 and F2 have shown convergence at the consonants or onsets, and hence should have blocked both the height (associated with /a/) and frontness (/i/) features of the trigger vowel.

In contrast to the major accounts of coarticulation, a vowel harmony model [11] seems to better account for the French data found here at least at the descriptive level. In vowel harmony, a certain feature (e.g., height, frontness, rounding, etc.), is shared by all the vowels in a certain time domain, usually within a word [11, 12]. From an articulatory perspective, the effect of such phonological feature-sharing would now be to reset all the vowel targets within the time domain. Once reset, there will be no further readjustment of the targets based on distance from the trigger vowel. During articulation, the speaker just tries to linearly realize both the consonant and vowel targets in order. As a result, both the vocalic features (assimilated) and consonantal features (unassimilated) are achieved in this model, as in Fig. 1.

The vowel harmony account has been used to model the data found in [11]. However, in that specific study, the account was used instead to explain assimilatory effects in the vowel immediately preceding the trigger. Our study has shown the surprising result of vowel harmony-like assimilation that was long-distance and even extended all the way to the beginning of the carrier sentence. Such a long-distance sentential effect can be partially explained by the fact that most syllables in the carrier sentence had the vowel /ə/, which has low coarticulatory resistance [17]. Such segments with low coarticulatory resistance are said to be receptive to segmental influence. It remains to be seen whether other vowels would have been affected just as much as the schwa on a sentential level. Future studies can thus consider minimal pairs that differ not only in the final trigger vowel but also in early target vowels. It also awaits further research to establish the reasons as to why a whole sentence can become the domain of vowel harmony in a language.

4.3 Lack of anticipatory coarticulation in Japanese

Contrary to our prediction based on contrastive pressure from vowel inventory [10], speech rate [14], and a prior articulatory study [8], the current results did not show evidence of anticipatory V-to-V assimilation in any vowel trigger context, except only in a very local effect (4 timepoints before the trigger) on F1 and limited to the 2-syllable condition. While somewhat surprising, the finding can be viewed as consistent with a vowel harmony account of the French results. That is, given that vowel harmony is language-specific rather than universal, it may be the case that certain diachronic developments have been moving French in the direction of evolving full vowel harmony, a process proposed by [12], while Japanese happens not to have experienced such similar pressures. This is certainly highly hypothetical, and further research is needed to look into the issue.

4. CONCLUSION

Through a systematic comparison of continuous formant trajectories, this study surprisingly reveals long-distance anticipatory V-to-V assimilatory effects across several syllables in French. Just as surprisingly, Japanese was found to show little anticipatory assimilation, despite having lower contrastive pressure in vowel space than French and possessing an overall speech rate just as fast as French [14]. We interpret these findings as offering support for a vowel-harmony account of V-to-V assimilation in French [11]. This interpretation, while contentious, may motivate future investigations of coarticulation to take vowel harmony and its possible diachronic origin [12] as a serious source of explanation when encountering French-like assimilatory effects.

5. REFERENCES

[1] Bell-Berti, F. and Harris, K. S. 1981. A temporal model of speech production. *Phonetica* 38: 9-20.

[2] Bell-Berti, F. and Krakow, R. A. 1991. Anticipatory velar lowering: A coproduction account. *Journal of the Acoustical Society of America* 90: 112-123.

[3] Benguerel, A.-P., and Cowan, H. A. 1974. Coarticulation of upper lip protrusion in French. *Phonetica* 30: 41-55.

[4] Boersma, P. and D. Weenink. 2014. Praat: doing phonetics by computer.

[5] Fowler, C. and Brancazio, L. 2000. Coarticulation Resistance of American English Consonants and its Effects on Transconsonantal Vowel-to-Vowel Coarticulation. *Language and Speech* 43: 1-41.

[6] Gelfer, C. E., Bell-Berti, F. and Harris, K. S. 1989. Determining the extent of coarticulation: effects of experimental design. *Journal of the Acoustical Society of America* 86: 2443-2445.

[7] Hardcastle, W. J. and Hewlett, N. Eds. 1999. *Coarticulation: Theory, Data and Techniques*. Cambridge, Cambridge University Press.

[8] Löfqvist, A. 2009. Vowel-to-vowel coarticulation in Japanese: The effect of consonant duration. *Journal of the Acoustical Society of America* 125: 636-639.

[9] Magen, H.S. 1997. The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics* 25: 187-205.

[10] Manuel, S. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different langauges. *Journal of the Acoustical Soaciety of America* 88: 1286-1298.

[11] Nguyen, N. and Fagyal, Z. 2008. Acoustic aspects of vowel harmony in French. *Journal of Phonetics* 36: 1-27.

[12] Ohala, J. J. 1994. Towards a universal, phonetically-based, theory of vowel harmony. In *Proceedings of ICSLP*, Yokohama. 491-494.

[13] Öhman, S.E.G. 1966. Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* 39: 151-168.

[14] Pellegrino, F., Coupé, C. and Marsico, E. 2011. Across-language perspective on speech information rate. *Language* 87: 539-558.

[15] Perkell, J.S. and Matthies, M.L. 1992. Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability. *Journal of the Acoustical Society of America* 91: 2911-2925.

[16] Recasens, D. 1999. Acoustic Analysis. In *Coarticulation: Theory, Data and Techniques*, W. J. Hardcastle and N. Hewlett (Eds.). Cambridge, Cambridge University Press. 322-336.

[17] Recasens, D. and Espinosa, A. 2009. An Articulatory Investigation of Lingual Coarticulatory Resistance and Aggressiveness for Consonants and Vowels in Catalan. *Journal of the Acoustical Society of America* 125: 2288-2298.

[18] Xu, Y. 2007-2014. FormantPro.praat. From: http://www.phon.ucl.ac.uk/home/yi/FormantPro/.

[19] Zwicker, E. 1961 Subdivision of the audible frequency range into critical bands. *Journal of the Acoustical Society of America* 33: 258.