IN SEARCH OF THE PERCEPTUAL CORRELATES OF VOWEL HARMONY

M. Grazia Busà* and John Ohala†
*Università di Bologna, Italy, †University of California, Berkeley, USA

ABSTRACT

This study was designed to investigate universal mechanisms governing patterns of VH. By discovering how listeners deal with the contextual variation of vowels in the context of other vowels, we look for evidence of the phonetic "seeds" of VH in languages which synchronically do not exhibit it. Listeners' criteria for front-back differentiation of V1 when V2 is a high front vowel are tested. The main hypothesis being tested is that a postconsonantal [i] will cause the vowel in the preceding syllable to be perceived as less fronted, due to listeners' perceptual correction of vowel backness in the environment of an adjacent front vowel. The results show that a postconsonantal front vowel affects the perception of a preceding vowel by causing this to be perceived as less fronted. The results, however, could be explained as contrast effects. This requires a deeper investigation of the perceptual mechanisms giving rise to VH.

1. INTRODUCTION

Vowel harmony (VH) is a widespread phonological phenomenon whereby vowels in a word or phrase are modified under the influence of a vowel occurring within the same domain, so that all the vowels in that domain come to share some of their parameters, such as the front/back or the height parameter. Typically, VH is manifested in morphological or morpho-phonemic processes, and the vowels in inflectional endings and functional affixes have variable forms which tend to 'agree' phonetically with the vowels in neighboring nouns, verbs and adjectives. In Hungarian, for example, the vowel of the plural suffix varies according to the vowels in the noun root, e.g., */astal/* "table" ~ */astalok/* "tables"; */ember/* "man" ~ */emberek/* "men". The general understanding of VH captures the intuition that VH is a dynamic process involving the interaction of a sequence of vowels. However, much of the universal mechanisms governing patterns of VH are still unaccounted for [1]. For example, it has not yet been explained why only some properties of vowels, such as fronting or roundness, are subject to harmonic processes, while others, like duration or diphthongization, are not. Likewise, no phonetic account has been provided of the different segments which may block different harmony systems. Such is the case of oral obstruents blocking the spread of nasalization in nasal harmony. The existence of vowels that are neutral to the harmonization process in VH languages is also still unexplained. In Finnish, for example, the front vowels *ä ö y* cannot co-occur with the back vowels *a o u* in the same non-compound word (e.g., */musta/* "black", */vaara/* "danger"; */köyhä/* "poor", */väärä/* "wrong"). The two neutral vowels *e* and *i*, however, can appear in words with either front or back vowels (e.g., */saari/* "island", */sääri/* "leg"; */huone/* "room", */pääte/* "ending").

The possibility of providing an explanation for some of the mechanisms governing patterns of VH arises if VH is considered as the fossilized or phonologized result of between-vowel assimilation. If the view is accepted that before a given phonological variation can develop, it must have phonetic 'seeds', then the examination of phonetic patterns can give us clues about the emergence of phonological processes. This study was designed to investigate some of these phonetic patterns.

It has been shown (e.g., [2]) that, in normal speaker-to-listener interaction, listeners can generally compensate for the distortion in the speech signal by reverting to their linguistic experience and thus parse the influence of contextual effects on speech segments. When listeners fail in their corrective strategy, there is potential for sound change. So, sources of perceptual ambiguity for the listeners are possible sources for sound change. Discovering listeners' mechanisms for dealing with the variation of vowels in the context of other vowels can give us insights into the emergence of VH.

Phonetic studies on V-to-V processes have suggested the possibility of finding phonetic explanations for the patterning of VH in the production and perceptual domains. It has been shown that, due to coarticulation, the formant patterns of V1 in V1CV2 sequences are variable and affected by the quality of V2 [3]. In addition, vowels are perceived in relation to their flanking vocalic context, due to the tight coupling of the listener's production and perceptual strategies [4]. Our hypothesis is that, if we can demonstrate that listeners correct for the predictable distortion of an adjacent palatal vowel, and thus accept as a back vowel a vowel that is more front than usual, it is also possible to predict that a failure in the listeners' perceptual correction mechanism of vowel backness in the environment of an adjacent palatal vowel can lead to palatal VH.

When considering the contextual effects of coarticulation on the perception of vowels, caution should be taken to rule out the possibility of contrast effects. These are well known phenomena and occur when the identification of a sound is affected by its context in a dissimilative way [5, 6]. It is often difficult to distinguish coarticulatory and contrast effects experimentally. However, evidence that dissimilatory perception occurs in the absence of clear coarticulatory influences should be an indication of dominating contrast effects.

2. EXPERIMENT I

A first experiment was designed to test the following hypotheses: 1) in a V1CV2 sequence where V2 is [i], C is [b] and V1 is part of a continuum spanning from [i] to [u], V2 will cause V1 to be perceived as less fronted due to listeners' perceptual correction of vowel backness in the environment of an adjacent palatal vowel; 2) a shorter intervocalic interval will have a greater effect than a longer one; 3) this perceptual process will be independent of the listener's native language.

2.1. Method

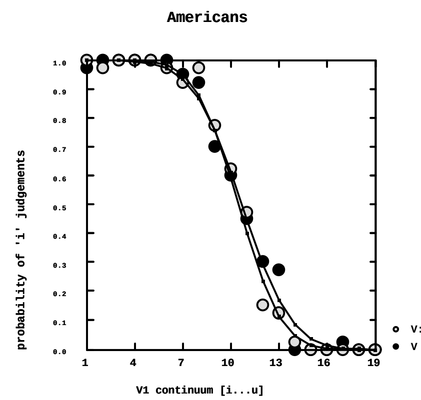
2.1.1. Stimuli. A short (100 ms) and a long (150 ms) 19-point /i–u/ continua were created by manipulating the LPC coefficients of a single /i/ vowel. Four more continua were created by splicing onto the two isolated vowel continua a short and a long /-ˈbi/ sequence (where /b/ was either 60 ms or 90 ms long). The two isolated vowel continua served as control stimuli, while the four combinations of /Vˈbi/ were designed to test the effect of the V-to-V interval duration on the perception of V1. The duration of the vowels and consonants, as well as the formant frequency values of the vowels at the endpoints of the continua, were chosen between possible values of American English and Italian. The F1, F2 and F3 values for [i] were: 300, 2258, 2960 respectively; for [u] they were: 300, 860 and 2260. In the continua F2 was decreased by steps of 45-50 Hz up to 1700 Hz, and then by steps of 140 Hz; F3 was decreased by steps of 20-25 Hz up to 2680 Hz and of 70 Hz afterwards. This particular variation in F2 and F3 was made non-linear to make the crossover between vowel categories closer to the middle of the continuum. A pilot experiment had shown that without this adjustment the categories saturated to /u/ for almost the entire continuum.

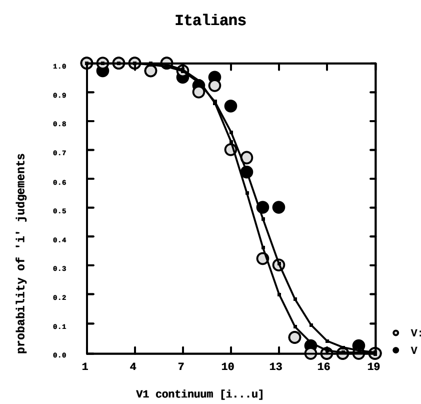
2.1.2. Subjects. The subjects of the experiment were 20 native speakers of American English and 20 native speakers of Italian, non-paid volunteers. English and Italian present relevant differences with regard to vowel and consonant contrasts and to vowel inventory size. English has no vowel or consonant contrasts depending solely on duration, while Italian has consonant but not vowel contrasts depending solely on duration. The difference in vowel inventory size and systems of contrast between the two languages --Italian having fewer vowels than English-- may have consequences on the way the two languages allow for vowel-to-vowel coarticulation [7]. Most importantly, neither language has VH among its current phonological processes. It is appropriate to seek the universal *phonetic* basis of a common cross-language sound pattern in languages which do not at present show the *phonologization* of the given pattern [2].

2.1.3. Presentation. The stimuli were presented to the listeners through headphones over a computer. The first part of the test consisted of the four /Vˈbi/ continua, each token repeated twice and randomized, for a total of 19 x 4 x 2 = 152 trials. The second part consisted of the two isolated vowel continua, each token presented twice and randomized, for a total of 19 x 2 x 2 = 76 trials. In each part, the subjects were to make a forced choice judgement of the target vowel by clicking on an [i] or an [u] box on the computer screen. The subjects were told that we were trying to assess vowel quality for a speech recognition system and their task was to help us decide how to classify vowels. The whole session lasted 10 minutes.

3. RESULTS

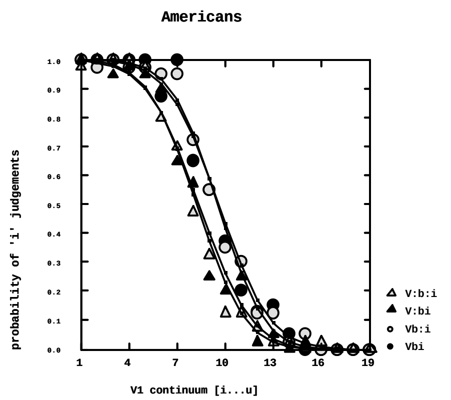
Figures 1-4 present the results of the experiment, in the form of probability of 'i' judgements. In all the figures, a smooth curve calculated with Probit analysis is superimposed on the raw data. Each data point represents 40 judgements (20 subjects x 2 token repetitions). In the figures, the numbers on the x axis correspond to the stimulus number, so that 1 and 19 represent the first and last endpoint stimuli, i.e., [i] and [u] respectively; the values on the y-axis indicate the probability of 'i' judgements, from 0 to 1 (a decrease in number of 'i' judgments indicates a shift in perception in the expected direction, i.e., that more tokens have been identified as 'u'). The statistical analyses (ANOVA's), both between and within group, were performed on the raw data.

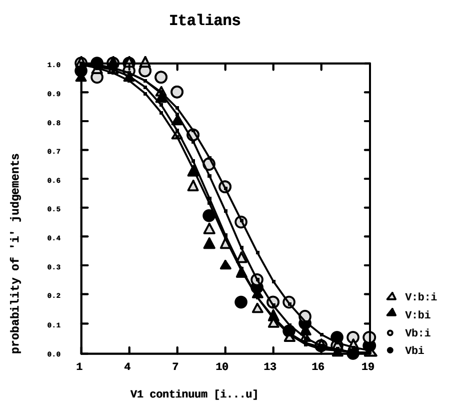




Figs. 1 (Americans) and 2 (Italians): Fitted curves to the data showing identification as /i/ of the stimuli from the /i-u/ continuum in isolation.

In Figures 1-2, showing the results for the isolated vowels, both language groups show saturated judgements at 0 and 1 endpoints, which indicates that the subjects had no trouble identifying the endpoint vowels [i] and [u]. The two language groups differ in where they have the crossover between [i] and [u]. For the Americans it is around token 11 for both continua, while for the Italians it is around 11-12 for the short vowel continuum and around 12 for the long vowel continuum. The difference in identification of short vs long isolated vowels does not reach significance. The language effect, **L**, is significant (F(1,76) = 76.05, p=0.004).





Figs. 3 (Americans) and 4 (Italians): Fitted curves to the data showing identification as /i/ of the stimuli from the /i-u/ continuum in the context /Vˈbi/ (conditions: long vowel + long /b/; long vowel + short /b/; short vowel + long /b/; short vowel + short /b/).

Figures 3-4 show the results for the vowels in the /Vˈbi/ continua. Here too, both endpoints of the continua for both groups show saturation. Compared to the data of the vowels in isolation, these data show that there is a clear shift in the probability of 'i' judgement due to context. For the Americans, the crossover between [i] and [u] has shifted to token 10 for the short V1 continua and between token 8 and 9 for the long V1 continua. For the Italians, the crossover between [i] and [u] is around token 9 for both the [V:bi] and the [V:b:i] continua, around 10 for the [Vbi] continuum and around 11 for the [Vb:i] continuum. The language as well as the context effect, **C**, are significant (for **L**, F (1, 152) = 6.44, p=0.01; for **C**, F (3, 152) = 4.40, p=0.005). The differences in the two experimental conditions (i.e, isolated vowels vs four /Vˈbi/ continua) are significant both for the language and the context factors (for **L**, F (1, 335) = 24.74, p=0.000; for **C**, F (1, 390) = 28.84, p=0.000). The effect of the four /Vˈbi/ conditions is different for the two language groups. The Americans show two distinct trends, one for the two /Vˈbi/ continua with short V1 and the other for the two continua with long V1 (for **C**, F (3, 76) = 5.09, p=0.003) – this latter causing a greater shift in perception than the former. The Italians too show a greater perceptual effect for the two long V1 continua, but of the two short V1 continua, the one where the vowel is followed by a short consonant has a greater effect than the one where it is followed by a long consonant, which reflects a sensitivity to the long-short consonant distinction (but the difference between the conditions is not significant). Unexpectedly, for both languages a longer vowel-to-vowel interval appears to cause a greater shift in identification threshold than a shorter one.

4. DISCUSSION

The results of experiment I point to the existence of a vowel-to-vowel perceptual effect affecting V1. For both language groups, the probability of 'i' judgements for V1 decreases when the stimuli are in the environment of an [i] in the following syllable, i.e., when they occur in the /Vˈbi/ context as compared to when they are found in isolation. In other words, the listeners identify the first vowel as ‘u’ a greater number of times. This finding could be explained as due to listeners’ compensation for the effect of the neighboring vowel by accepting as a back vowel /u/ some vowels that previously had been identified as an /i/. This would mean that listeners have learnt to expect that a high back vowel followed by a high front vowel becomes fronted and that this phonetic effect (back vowel fronting) occurs in their language. The failure to normalize this effect could be seen as a possible explanation for how VH arises.

However, the pattern observed might also be due to contrast effects. According to [4], a contrast effect on a V-to-V sequence would work by erasing the coarticulatory influence of the neighboring vocalic context. Contrast effects would also work in the opposite direction, and be similar in magnitude, as coarticulatory effects. Accepting accounts that explain the observed perceptual patterns as due to contrast effects would disregard the tight coupling between the listener’s production and perception strategies, which is undesirable. To examine if the results obtained with our experiment are due to contrast effects, which would yield similar results, we decided to run a second experiment.

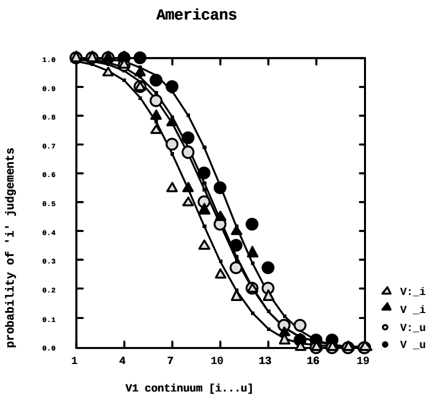
5. EXPERIMENT II

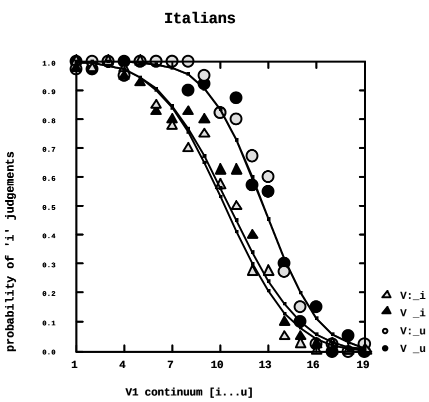
The hypothesis to verify with this experiment was: the existence of contrast effects would be proven if, in a sequence of two vowels separated by a certain time interval, the second vowel influences the perception of the first vowel, and this effect is comparable in magnitude to the effect observed in experiment I.

4.1. Method

4.1.1. Stimuli. Using the same short (100 ms) and long (150 ms) 19-point /i–u/ continua as in experiment I, four more continua were created, in which the short or long vowel of the continuum was followed, after a 650 ms time interval, by either [i] or [u]. The duration of the time interval was designed to be long enough to avoid sounding like an actual utterance. The duration of the second vowel was the same as the duration of V2 in the /Vˈbi/ sequence.

4.1.2. Subjects and presentation. The subjects of the experiment were, again, 20 Americans and 20 Italians, all native speakers, non-paid volunteers. The stimuli presentation was the same as that described in 2.1.3. above.





Figs. 5 (Americans) and 6 (Italians): Fitted curves to the data showing identification as /i/ of the stimuli from the /i-u/ continuum in the context /V/, 650 ms time interval, and /i/~/u/ (conditions: long vowel + /i/; short vowel + /i/; long vowel + /u/; short vowel + /u/).

6. RESULTS

Figures 5-6 show the results of experiment II. The data are represented as explained in 3 above. The two languages have clearly different crossover points for the four continua. Both the language and the context factor are significant (for **L**, $F(1,152) = 765.9$, $p=0.00$; for **C**, $F(3, 152) = 158.47$, $p=0.00$). The Americans have crossover points around tokens 8-9 for the /V:_i/ continuum; tokens 9-10 for the /V_i/ and the /V:_u/ continua; and token 11 for the /V_u/ continuum. The differences due to context are significant (for **C**, $F(3,76) = 61.61$, $p=0.04$). The Italians have crossover points around token 10 for both the /V:_i/ and the /V_i/ continua, and at token 13 for the /V:_u/ and the /V_u/ continua. These differences are significant (for **C**, $F(3,76) = 138.41$, $p=0.00$).

Comparing these results with those from experiment I, the Italians show evidence of a contrast effect for all four context conditions. This is shown by the fact that in the /V:_i/ and /V_i/ conditions the number of ‘i’ judgements has decreased when compared to the judgements of the vowel in isolation, which shows that the percept has shifted to ‘u’ a greater number of times for contrast with the second /i/. In the /V:_u/ and /V_u/ conditions the number of ‘i’ judgements has increased as compared to the judgements of the vowel in isolation, which shows that the percept has shifted to ‘i’ a greater number of times for constrast with the second /u/. The effect is similar in magnitude to the effect obtained in the four /Vˈbi/ continua. The Americans show evidence of a contrast effect for the /V:_i/ and the /V_i/ conditions: here too the number of ‘i’ judgements has decreased when compared to the judgements of the vowel in isolation, and the effect is similar in magnitude to the effect obtained in the four /Vˈbi/ continua. For the two /V:_u/ and /V_u/ conditions the data are less clearly interpretable, since the former shows a decreased number of ‘i’ judgements, and the latter an about equal number of ‘i’ as when the vowel is in isolation.

7. FINAL DISCUSSION AND CONCLUSIONS

We presumed that listeners adjust their criteria for vowel interpretation in environments where they expect some coarticulatory distortion of the vowel quality. In the search of perceptual correlates for VH, we though that the failure to implement this strategy could be a plausible source for the emergence of VH in languages. When we set out to prove this hypothesis, some data in our experiment led us to suppose that contrast effects may be responsible for the results obtained. A second experiment showed us that, in fact, the results are in the direction of contrast. V-to-V contrast effects appeared significant in spite of the long duration of the time interval between the two vowels. The two languages examined have appeared to be different in the way in which the four contextual conditions affect the shift in perception. This suggests that universal, as well as language-specific factors may be involved. However, the main question remains unanswered: if the V-to-V effect found is due to contrast, how does VH come about? How can we explain VH if there are forces resisting it? At some point listeners have to learn to expect that a vowel is colored by some other vowel in order for VH to take place. More research is needed to better understand V-to-V perceptual mechanisms so that a phonetic account of VH can be given after the contrast effect has been ruled out.

REFERENCES

[1] Ohala, J. 1994. Towards a universal, phonetically-based theory of vowel harmony. In *Proceedings International Conference on Spoken Language Processing ’94*. 2, 491-94.
[2] Ohala, J. 1993. The phonetics of sound change. Jones, C. (ed.), *Historical linguistics: Problems and perspectives.* London: Longman, 237-78.
[3] Öhman, S. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *JASA,* 39, 1, 98-104.
[4] Fowler, K. 1981. Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research,* 46, 127-139.
[5] Eimas, P. 1963. The relation between identification and discrimination along speech and non-speech continua. *Language and Speech,* 6, 206-17.
[6] Fry, D., Abramson, A., Eimas, P. and Liberman, A. 1962. The identification and discrimination of synthetic vowels. *Language and Speech,* 5, 171-89.
[7] Manuel, S. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *JASA,* 88, 3, 1286-98.