Unnatural rhythm. English vowel length in Italian students

Gabriele Azzaro, University of Bologna
E-mail: gazzaro@libero.it

ABSTRACT

After measuring contextual and prosodic vowel (V) length in native speakers, two groups of Italian students were compared to them, to see whether those exposed to naturalistic input fared better than traditionally taught ones. Naturalistic input implied residence abroad for one year (while a control group attended classes in a formal university setting). In native speech Vs were longer when followed by voiced consonants (Cs) and in monosyllabic words. The behaviour of L2 speakers showed some departure from the native model in that Foot-structure was not as influential and following C-voicing was only effective to a lesser degree; in all, the naturalistic input learners fared better. Therefore, mastering these V-lengthening mechanisms, albeit not too awkward for Italians, still needs naturalistic input.

1 INTRODUCTION

Assuming the reported tendency of languages to align with a rhythmical pattern situated between syllable and stress-timed extremes [1, 2], the present study compares the length of English Vs in relation to Foot structure and following Cs in native and foreign speech.

2 ENGLISH AND ITALIAN VOWEL LENGTH

Reportedly, English V-length is influenced by a host of factors, namely tenseness, height, (phrasal) accent and (lexical) stress, speaking rate, voicing, manner and place of articulation of the preceding and following Cs, prosodic structure, syntactic function and phrasal position in the utterance, word frequency. V-length is also an important clue in the perceptual identification of syllable Codas, contributing acoustic information along with other factors like formant transitions, VOT, duration of stop gaps, strength of aspiration burst, F0.

The effect of C-voicing on the length of an immediately preceding V is language-universal, but in some languages it is more marked than in others; in English it is prominent [3], in French significant but smaller [4], in Arabic negligible [5]. This kind of adjustment is nearly phonemic in English (see the minimal pair [bæˑtˀ]/[bæː dˀ]), but barely phonetic in Italian. The average ratio has been found to be 2:3 for American English [6-8], and I am not aware of studies on dialectal variation of this feature. A second type of length adjustment depends on the overall number of segments in a linguistic unit and the number of syllables within each Foot in an utterance. For convenience sake I will here refer to these two types of phenomena as **phonetic** and **prosodic** lengthening respectively.

As opposed to English, Italian V-length is never distinctive, it is the main correlate of lexical stress, it depends on syllable structure (closed syllable Codas nearly half the V-length), it doesn't show a great contextual variability and varies under the effect of following voicing only by 8-17%.

3 LEARNING THE CONTRAST

Due to these different settings in English and Italian, learning the L2 V-timing can be thorny for Italian students of English [9]. The misfit of V-length is a key element in the perception of foreign accent. Foreign accent can be detected by native English speakers through exposure to a speech sample as short as 30 ms [10], and Flege found that the degree of foreign accent perceived by English native speakers decreased as a function of the increase of V duration differences in word pairs such as *beat/bead* [11]. Also, experiments on temporally modified non-native speech show that the relative timing of Vs and Cs can improve the intelligibility of foreign accent [12]. All this bears out the idea that minimal temporal details can often enhance communication or at least the degree of L2 speech acceptability.

Non-native speakers sometimes reduce the amount of mutual adaptation of a V+stop combination away from the English standard towards their L1 norm [13]. Reportedly the same problem arises for Italians, and has sometimes been attributed to the different prosodic makeup of the two languages [9], in that English tends towards stress-timing, Italian towards syllable-timing (although experimental evidence points to other discriminating factors between them, such as different syllable structure, phonetic realisation of stress and a tendency towards V-reduction in English [1, 2]). Finally, segments show greater temporal stability in Italian than English (for instance, unstressed Vs are not as reduced by coarticulation as in English [14]).

4 THE STUDY

The present research is a follow-up of a pilot study carried out on a single group of students, who had no differentiation in kind of tuition and were not compared with native speakers. This study compares Italian students of English to two native speakers in the production of phonetic and prosodic V-length, looking at the possible effect of two types of learning environments. In particular, the investigation aims at discovering whether native speakers of Italian can reproduce the contrast even through a formal learning setting with limited naturalistic exposure to L2.

4.1 HYPOTHESES

Three hypotheses were verified: **1.** native and foreign speech differ in phonetic V-length in that L2 speakers do not lengthen Vs preceding voiced Cs as much as natives do,

ISBN 1-876346-48-5 © 2003 UAB

given the higher temporal regularity of Italian Vs [15]; **2.** native and foreign speech differ prosodically, namely Vs are pronounced with different timing compression in one-syllable vs. two-syllable words in native and L2 speech, given a greater amount of V reduction in English [16]; **3.** exposure to native input is essential in approximating the target behaviour.

4.2 METHOD

Two female native speakers of standard British English and fifteen female Italian students of English read target sentences after having seen them and checked that they understood them; the sentences consisted of the carrier phrase "He said …" plus a keyword, and were uttered at normal speed. The Italian subjects had studied English from age 11, and had recently been subjected to language input of different kind: three had lived abroad as part of an exchange project. There were thus three groups of speakers according to quality of received input: native speakers (N), non-native students with experience abroad (NN1) and without (NN2).

The independent variables controlled for were voicing of the C immediately following the stressed V in the keywords (2 levels: voiceless and voiced C), Foot type (2 levels: one or two-syllable words) and L2 input quality (3 levels: native (N), non-native exchange students (NN1), non-native students with formal tuition (NN2)). All the standard British English Vs /i ɪ e æ ʌ ɑ ɒ o ʊ u/ were employed (/ɜ/ was excluded for lack of high frequency lexical items to administer), even though the V-type effect was not controlled. The Vs were all stressed. Four words for each V were used, satisfying the criteria exemplified in Table 1.

Following voicing	Syllables	
	ONE	TWO
+VOICED C	/si:d kɪd bed/…	/'si:dɪŋ 'kɪdɪŋ 'bedɪŋ/…
-VOICED C	/si:t kɪt bet/…	/'si:tɪŋ 'kɪtən 'betɪŋ/…

Table 1: Test conditions and example keywords.

A battery of 40 sentences was thus created, which was then read by each subject, obtaining a total of 680 items. V-lengths were then measured both in absolute (ms.) and in proportional terms (relative to individual speech rate, by dividing each V-length by its respective word length).

Whenever a speaker wanted to repeat an item, s/he was allowed to do so; when present, the repetition only was included in the analysis. Only on five occasions repetition was deemed necessary by non-native readers. The data were recorded in a quiet room on a Sony DAT tape recorder at 44KHz-16bit (mono) with a Sony ECM-F8 Electret Condenser microphone. The microphone was placed on a soft cushion at about 20 cm from the speaker's mouth, sideways. Unfortunately two of the subjects had to be edited out due to airflow interference with the microphone. The recordings were then transferred to a PC at 22kHz, 16 bit, and each keyword was normalised at -.5 db and analysed with MultiSpeech 3700 ver. 2.2. V-lengths were measured with reference both to waveforms (onset of regular waveform and volume increase) and spectrograms (F0 through F4 presence, exclusion of bursts, frications and possible final glottal stops). Word lengths were measured from inception of initial stop burst or fricative to complete end of signal. About 10% of measurements were random checked by a second linguist, with no signifiant discrepancies. Statistical analyses were performed with the ViSta 6 package and Corel Quattro Pro (ver. 10).

5 RESULTS AND DISCUSSION

First I examined how consistent was the production of the two native speakers. Correlation between V-lengths was .93 (p<.0001) for proportional values (V-length/word length) and .94 (p<.0001) for absolute values. A two-tailed t-test was not significant (p<.986 for proportions and <.337 for absolute lengths). The performance of the two speakers was then extremely even. Yet, since a two-tailed t-test for the difference between their word lengths was of relative significance (p<.039), comparison between proportions as well as absolute lengths is crucial.

Foot	Monosyllabic	*SD*	Bisyllabic	*SD*
	145 ***33%***	*8%*	101 ***20%***	*11%*
Voicing	Voiced		Voiceless	
	135 ***30%***	*12%*	112 ***24%***	*11%*
L1	Native (N)		Non-native (NN)	
	130 ***30%***	*14%*	122 ***26%***	*11%*

Table 2: Mean V-lengths in ms. for all subjects (with percentage of V/Word-length in italics) under three conditions: Foot-type, following voicing and L1 of speaker.

In order to check how the production of each Italian student correlated to the native behaviour as a whole, I merged the data from the natives and performed a Monotonic Univariate Multiple Regression analysis with each Italian subject as a predictor (both on percentages and on absolute lengths). High and significant correlations (above .78) with the bundled native data were obtained for 8 of the 13 Italian subjects (ranging from .86 to .93 for proportional data and from .78 to .91 for absolute lengths; p values ranged from .0076 to <.0001 with both Monotonic and Robust procedures). The r^2 values went from .74 to .86, indicating a fairly large amount of shared variance with the native speakers. This points to a high level of consistency between the students' overall performance and the target; the phonetic and prosodic lengths of these words were in general handled well. Yet the best correlations were with the three students who had taken part in the student exchange. The lowest correlation with the native data was .53 for proportional length, and .37 for raw ms. (p<.001), which indicates some kind of approximation even for the L2 speaker who fared worst. Despite obvious data size limitations, PCA and Factor Analysis were attempted to see whether the same components could explain subjects' variance across speakers. Although five factors emerged able to account for over 91% of the native speakers' variance, none showed up in the Italians.

Turning to an overall comparison of average V-lengths in the four test conditions examined (pre-voiced and pre-voiceless

monosyllabic, pre-voiced and pre-voiceless bisyllabic), the data are summarised in Table 2 and 3. Table 2 shows how overall the tendency to lengthen Vs was strong in one-syllable words, as expected (see first row). Following voicing too was generally effective (Vs before voiced Cs were 6% longer), and it is also interesting that native speakers' Vs took up 30% of their average word-length, while non-native speakers were on average shorter by 4%. Despite the above mentioned correlations there was therefore some discrepancy between the two groups. This must be looked into with greater detail.

Foot		Monosyllabic	Bisyllabic	**Difference**
Following Voicing	[-voice]	130 *29%*	93 *18%*	***11%***
	[+voice]	160 *38%*	110 *23%*	***15%***
L1	NN	142 *32%*	103 *20%*	***12%***
	N	167 *40%*	93 *20%*	***20%***
Voicing		[-voice]	[+voice]	
L1	NN	110 *23%*	134 *29%*	***6%***
	N	120 *26%*	140 *35%*	***9%***

Table 3: Mean V-lengths in ms. and percentages of word-length under various interactions of factors (N=native; NN=Non-native).

A quick glance at the interactions in Table 3 shows us that the prosodic compression typical of English Foot structure is evident in the native speakers more than in the Italians: whereas English speakers (N) half V-length in bisyllabic words, Italians (NN) do not (on average a length disparity of 20 vs 8%, reported in the third row on Table 3). This reflects both the longer nature of Vs in native speech and the smaller compression in bisyllabic Feet. Besides, the phonetic expansion of Vs before voiced Cs is less significant in Italian speakers overall: they expand Vs from 23 to 29% of their word-length (+6%), while English speakers climb from 26 to 35% (+9%; cf. the last row in Table 3). Possibly, the smaller discrepancy between short and long segments is a mark of non-native speech. All this calls for further analysis.

First, a 2-way ANOVA for Foot and Voice (2x2) was performed for the native speakers alone (twice, first on percentages, then on absolute lengths with no significant discrepancy); Foot and Voice were both significant (p<.0001) with no interaction. The two subjects showed very cohesive performance (between-subjects p=.21). This means that both prosodic and phonetic adaptations emerge and both subjects follow the same trends.

A 3-way ANOVA for Voice, Foot and Subjects on Italians alone yielded high significance (p<.0001) for all three conditions. Therefore Italians lengthened Vs before voiced Cs and their monosyllabic words had longer Vs (somewhat mirroring the N model); but they also varied considerably amongst themselves, with little coherence. Besides, a slight interaction emerged between the Subject and Foot factors (p=0.03), suggesting that V-length does not change with consistent criteria under the effect of prosodic structure, or at least that it varies unevenly in different subjects since some counteracted the main effect of Foot structure.

A third analysis was performed for all the speakers together: significant were the main effects of Foot, Voice and L1 (p<.001), indicating the expected V effects and also strong differentiation between speakers. A significant interaction between Foot and L1 (p=.0004), not found in the English subjects, confirmed that it was the Italians to contrast the main effect of prosodic structure. Only some of them but not all then, followed the target language tendency to reduce V-length appropriately in two-syllable words, as noted above (cf. Table 3). On the other hand speakers did not vary phonetically (p=.22), which suggests that lengthening before voiced Cs is fairly constant across all subjects (Italians too).

The statistics were finally computed for comparing the natives with the two subgroups of NNs identified earlier. The correlation with the N speakers was best for the NN1 group (the exchange students) (Table 4).

	NN1	NN2
N	.88 *(.83)*	.77 *(.51)*
NN1		.77 *(.67)*

Table 4: Correlations of proportional V-length (and absolute length in brackets) between the three groups(p<.001).

But how did the three groups compare? A three-way ANOVA for Foot, Voice and L1 (2x2x3) with interactions showed that: **1.** for prosodic length NN1 did not differ from N significantly (p=.65) but NN2 did (p=.0001); **2.** also for phonetic length NN1 did not differ from N (p=.73) but NN2 did (p=.0003) (Table 5). It must be added that for Voice by L1 no interaction was significant for both pairs of groups (p=.34 and .38), while there was a significant interaction for Foot by L1 between N and NN2 (p<.0001, and .65 between N and NN1). This means that N and NN2 contribute contradictorily to the main effect of prosodic structure: NN2 was not therefore consistent with the target behaviour, while NN1 was. This can be seen in Figure 1, where the insufficient length by NN2 in monosyllabic Feet stands out.

	Monosyllabic	SD	Bisyllabic	SD	**Diff.**
L1 Input x Foot	N = 40%	10	N = 20%	9	**20%**
	NN1 = 38%	10	NN1 = 21%	9	**17%**
	NN2 = 31%	10	NN2 = 20%	8	**11%**
	[-voice]		[+voice]		
L1 Input x C-voicing	N = 26%	12	N = 35%	14	**9%**
	NN1 = 27%	12	NN1 = 32%	13	**5%**
	NN2 = 22%	10	NN2 = 28%	10	**6%**

Table 5: Mean proportional V-length in relation to the factors: input received, Foot, voicing of following C.

It is true that proportional gain of V-length with C-voicing was better approximated by NN2 (6% increase) than NN1 (5% increase), but NN2 Vs stayed too short altogether (Figure 2). Figure 2 shows that the slopes are very similar for NN1 and NN2, and both groups offer a lesser degree of lengthening than the native group. A tentative explanation for NN2's global production of shorter Vs may lie in the above mentioned prosodic and syllabic structural differences between the two languages, particularly the fact that Italian Vs are more stable. Besides, Italian Vs, even if generally longer than English Vs, nearly half their length in closed syllables; since all the words analysed had closed syllables the

ISBN 1-876346-48-5 © 2003 UAB

effect should have applied here.

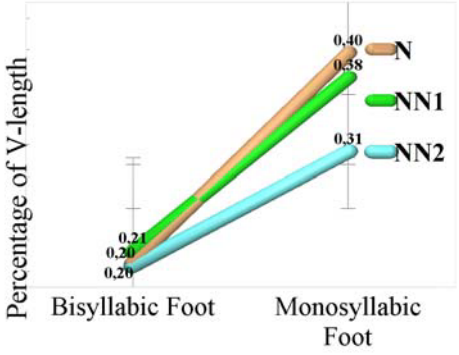

Figure 1: Plot of proportional V-lengths in bisyllabic vs monosyllabic words for the three groups.

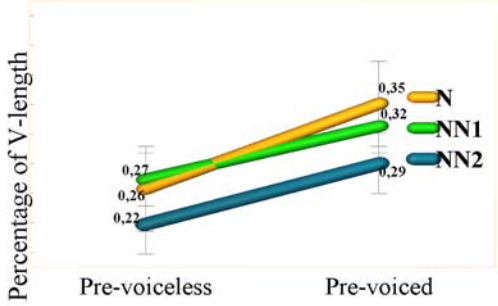

Figure 2: Plot of proportional V-lengths before voiced and voiceless Cs for the three groups.

6 CONCLUSIONS

Under the apparent overall similarity between N and NN, the latter group did not offer a uniform performance: individuals strayed from the target model. Also, there was a difference in performance between NN1 and NN2, with the former group approximating the native speakers' timing effect more closely. We saw that in general there was less discrepancy in the Italian subjects between longer and shorter Vs, specially for NN2, and the contrast triggered by following voicing was better replicated than the prosodic one: while English speakers tend to compress segments in two-syllable words, Italians seemed less affected by this strategy and only those who had been abroad approximated the target disparity of length, having altogether longer Vs in one-Foot words.

Finally, NN2 had shorter Vs in all conditions, which renders appropriate their slope of V-lengthening before voiced Cs but not their proportional length (Figure 2). Why all their Vs were shorter remains to be investigated.

Although the above post hoc discrepancies were found, further research is needed to confirm these findings with larger samples, also to investigate the nature of various other factors affecting the performance of non-native speakers, such as perception of the contrast, intrinsic V-length, kind of tuition received, indirect exposure to native input (via the media for instance) and effects of persistent L1 use, amongst others.

REFERENCES

[1] Dauer, R.M., "Stress-timing and syllable-timing reanalyzed", *Journal of Phonetics*, vol. 11, pp. 51-62, 1983.

[2] den Os, E.A., *Rhythm and Tempo of Dutch and Italian; a Contrastive Study*, Utrecht: Drukkerij Elinkwijk, 1988.

[3] Flege, J.E., "The production and perception of foreign language speech sounds", in *Human Communication and Its Disorders. A Review*, H. Winitz, Ed. Norwood, NJ: Ablex, pp. 224-401, 1988.

[4] Mack, M., "Voicing-dependent vowel duration in English and French: Monolingual and bilingual production", *Journal of the Acoustical Society of America*, vol. 71, pp. 173-8, 1982.

[5] Port, R.F., S. Al-Ani, and S. Maeda, "Temporal compensation and universal phonetics", *Phonetica*, vol. 37, pp. 235-52, 1980.

[6] Peterson, G.E. and I. Lehiste, "Duration of syllabic nuclei in English", *Journal of the Acoustical Society of America*, vol. 32, pp. 693-703, 1960.

[7] Kluender, K.J., R.L. Diehl, and B.A. Wright, "Vowel-length differences before voiced and voiceless consonants. An auditory explanation", *Journal of Phonetics*, vol. 16, pp. 153-69, 1988.

[8] Laeufer, C., "Patterns of voicing-conditioned vowel duration in French and English", *Journal of Phonetics*, vol. 20, pp. 411-40, 1992.

[9] Busà, M.G., *L'inglese degli italiani*, Padova: Unipress, 1995.

[10] Flege, J.E., "The detection of French accent by American speakers", *Journal of the Acoustical Society of America*, vol. 76, pp. 692-707, 1984.

[11] Flege, J.E., "Production and perception of a novel, second-language phonetic contrast", *Journal of the Acoustical Society of America*, vol. 93, pp. 1589-608, 1993.

[12] Tajima, K., R. Port, and J. Dalby, "Effects of temporal correction on intelligibility of foreign-accented English", *Journal of Phonetics*, vol. 25, pp. 1-24, 1997.

[13] Flege, J.E. and R.F. Port, "Cross-language phonetic interference: Arabic to English", *Language and Speech*, vol. 24, pp. 125-146, 1981.

[14] Vayra, M., C. Avesani, and C.A. Fowler, "Patterns of temporal compression in spoken Italian", in *Proceedings of the X International Congress of the Phonetic Sciences*. Dordrecht: Foris Publications, pp. 541-6, 1984.

[15] Ferrero, F., et al., "Some acoustic characteristics of the Italian vowels", *Journal of Italian Linguistics*, vol. 3, pp. 87-95, 1978.

[16] Nooteboom, S., "The prosody of speech: melody and rhythm", in *The Handbook of Phonetic Sciences*, W. Hardcastle and J. Laver, Eds. Oxford: Blackwell, pp. 640-73, 1997.