Italian in the no-man's land between stress-timing and syllable-timing? Speakers are more stress-timed than listeners

Bettina Braun and Sabine Geiselmann

[1] Department of Linguistics, University of Konstanz, Germany

`{bettina.braun,sabine.geiselmann}@uni-konstanz.de`

Abstract

How syllable-timed is Italian? We investigate two contexts for vowel reduction, unstressed syllables and syllables in polysyllabic words. In a production experiment, a large sample of speakers from Tuscany read di- and trisyllabic target words with different stress placement in a sentence context. Results showed vowel reduction in unstressed syllables, both in terms of duration and spectral quality as well as polysyllabic shortening (without spectral reduction). These temporal adjustments are of similar magnitude as reported for stress-timed languages. Results of a two-alternative forced choice task, however, showed little sensitivity to temporal patterns in monosyllabic fragments. Hence, production patterns appear to be more stress-timed than perceptual mechanisms which has implications for duration models in speech synthesis.

Index Terms: Italian, rhythm, stress, polysyllabic shortening, duration models

1. Introduction

Pike suggested to classify languages as either stress-timed (e.g. English or German), where stressed syllables recur in more or less isochronous intervals, or syllable-timed (such as many Romance languages), where syllables recur in more or less isochronous intervals [1].

One way to gain isochrony in stress-timed languages is vowel reduction. In unstressed syllables, vowels are produced considerably shorter than in stressed syllables. Furthermore, stress-timed languages often exhibit polysyllabic shortening: The duration of syllables in a word is inversely related to the length of the word, i.e. the more syllables a word contains (e.g., *mace, mason, masonry*) the shorter its syllables [2,3,4].

Newborns are already able to tell languages apart that belong to different rhythm classes [5,6,7], even though acoustic measures do not support the postulated isochrony (e.g., [8]). Dauer argued that syllable-timed languages have regularly recurring stress beats just like stress-timed languages, with the only difference that those beats seem perceptually less salient than those in stress-timed languages. She therefore does not distinguish between syllable timing (a notion also judged to be psychologically implausible by [9]) and stress-timing but proposes a continuum from more to less stress-based languages.

In this paper we study Italian which has been described as less stress-based [10,11]. Although it has discernible stress beats, it shows considerably less vowel reduction than, e.g. German, a more stress-based language. Word level prominence in Italian is mainly associated with the duration of the vocalic part of the stressed syllable, which is also a property characteristic of stress-timed languages [10,12,13,14]: Vowels in open (non-final) syllables are lengthened when they are stressed, especially in penultimate position [13]. Vowel length is not lexically contrastive in Italian. It has long been debated though whether or not Italian also features polysyllabic shortening. On

the one hand, [15] and [16] did not find consistent evidence for polysyllabic shortening (only for two out of four speakers in [15] and for one out of three speakers in [16]). On the other hand, polysyllabic shortening has been found by [13] and [14]. This paper hence takes a new look at polysyllabic shortening in Italian with a much larger population of speakers (N=24). Furthermore, we compare the durational effect of polysyllabic shortening to the durational difference between stressed and unstressed syllables, a durational difference that is allegedly very salient for the perception of stress in Italian (e.g., [17]). Finally, we investigate the degree of spectral reduction.

In a second experiment, we test whether Italian listeners can make use of durational differences due to stress or polysyllabic shortening in online speech comprehension as the word unfolds over time. While Dutch listeners have been shown to use durational cues for stress perception [18] and polysyllabic shortening [19,20], there are no perception data for Italian. The perception experiments make use of a speeded classification task with resynthesised stimuli varying in duration only (cf. [18]).

2. Experiment 1

Experiment 1 was a production experiment to investigate the effects of stress and syllable number. In the *stress condition*, the first syllable of trisyllabic words, both stressed and unstressed, is compared. In the *syllable condition*, the first syllable in di- and trisyllabic words (stressed in both conditions) is compared.

2.1. Participants

Twenty-four native speakers of Italian from Tuscany, students at Florence University (aged 21 to 38, 20 female) participated for a small fee. They had no self-reported speech or hearing deficits and were unaware of the purpose of the study.

2.2. Materials

Thirty words (3 pairs and 8 triplets) were chosen from the lexical database BADIP [21]. The pairs contained one disyllabic and one trisyllabic word with the same initial (open) syllable: *'capo – 'capita, 'vaso – 'valido, 'vela – 'vedovo*. The second syllable in both words started with the same place of articulation. In the triplets, another trisyllabic word with stress on the penultimate syllable was included: *'caro – 'carico – ca'rota, 'cavo – 'cavolo – ca'vare, 'fila – 'fisica – fi'dare, 'papa – 'papera – pa'pale, 'rapa – 'rapido – ra'pire, 'riso – 'ridere – ri'sata, 'viso – 'visita – vi'sivo, 'vivo – 'vivere – vi'vace*. The three types of words were matched for lexical frequency (disyllabic words 55/million, trisyllabic words with penultimate stress 66/million, trisyllabic words with antepenultimate stress 12/million). Furthermore, we selected 30 filler items with two to four syllables and different stress patterns.

2.3. Procedure

Test and filler words were inserted in a phrasal context to control for phrasal prominence (*Dico la parola ___ tre volte* 'I say the word ___ three times') and printed on a word list in pseudo-randomized order (14pt, 1.5 line spacing). Two members of a pair were separated by at least 3 intervening items. Participants were instructed to read the sentences at normal speed. Recordings took place in a quiet room at Florence University. Their productions were recorded using an m-audio recorder (44.1kHz, 16Bit).

2.4. Results

Some individual productions (less than 1 % of the data) had to be excluded from analysis because of misstressing or disfluencies. The second author annotated the vowels of the first syllable using Praat [22]. Vowel duration, as well as F1 and F2 in the middle of the vowel (in Hz) were automatically extracted. Results for the stress and syllable condition are reported separately.

2.4.1 Stress condition

Fig. 1-3 show the duration, first and second formants of the word-initial stressed or unstressed vowel in trisyllabic words. Data were analyzed using linear mixed-effects regression models with stress and vowel quality as fixed factors and subject and item as crossed random factors [23, 24].
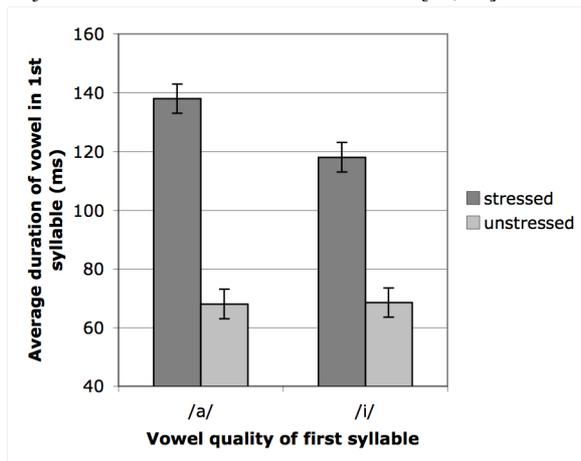

Fig. 1. Duration of vowel in word-initial syllable in trisyllabic words (stressed, unstressed). Whiskers represent standard error
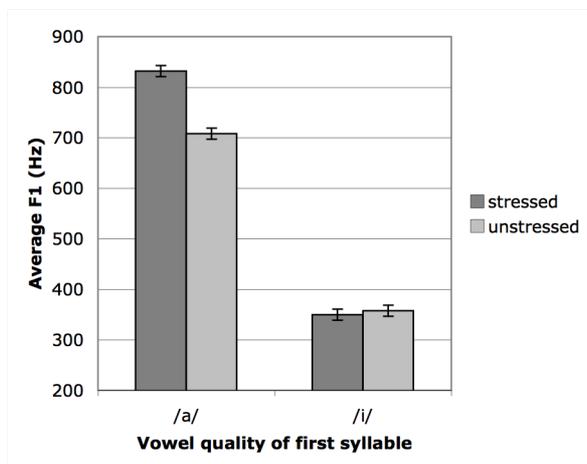

Fig 2. Mean first formant (with standard error) of vowel in word-initial syllable in trisyllabic words (stressed, unstressed)
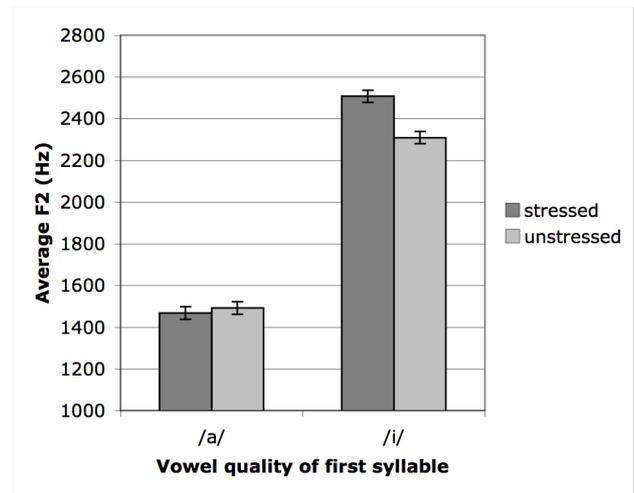

Fig. 3: Mean second formant of vowel in word-initial syllable in trisyllabic words (stressed and unstressed)

For the measures vowel duration, F1 and F2, there were significant effects of stress (all p's < 0.001), vowel quality (all p's < 0.01) and interactions between the two (all p's < 0.0001).

2.4.2 Syllable condition

Fig. 4 shows the duration of the word-initial vowel in the syllable condition (e.g. *'papa* vs. *'papera*) split by vowel quality. Results showed significant main effects of syllable number $(F(1,515)=128.2, p < 0.0001)$, of vowel quality $(F(1,515)= 6.2, p < 0.005)$ and a significant interaction between the two $(F(1,515=11.3, p < 0.0001)$. Further analysis showed that the effect of syllable number was significant for all vowel qualities (all p's < 0.05), 24.7ms difference for /a/, 12.8ms for /e/ and 9.0ms for /i/. Syllable number had no effect on the first formant (p > 0.1), the second formant (p > 0.5), or on spectral tilt (p > 0.1).
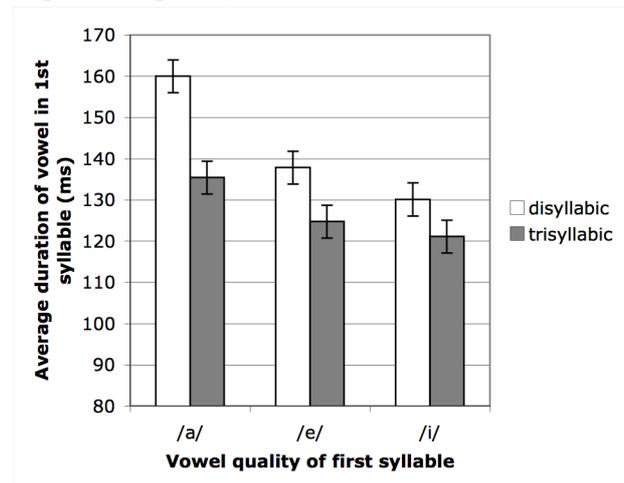

Fig. 4: Duration of the vowel in word-initial (stressed) syllable in di- and trisyllabic words. Whiskers represent standard error

2.5. Discussion

The results support the claim that word stress is signalled by vowel duration in Italian [17]. Stressed vowels were almost twice as long as unstressed vowels but the difference was larger for low than for high vowels. In addition, unstressed vowels were produced more centralized than stressed vowels. In the syllable condition, Italian speakers produced word-initial syllables in trisyllabic words with shorter duration than

word-initial syllables in disyllabic words, especially for the vowel /a/. Effects for the vowels /a/ and /e/ are comparable to the polysyllabic shortening effect reported for Italian [13] and English [4], a prototypical stress-timed language. In contrast to the reduction in unstressed syllables (Fig. 2 & 3), vowels were not centralized due to polysyllabic shortening.

3. Experiment 2

Experiment 2 was a speeded categorization experiment to investigate whether or not Italian listeners use vowel duration as a cue to stress and number of syllables.

3.1. Participants

The same 24 participants as in Experiment 1 took part in Experiment 2 (in the opposite order). All were right-handed with normal or corrected-to-normal hearing and eyesight.

3.2. Materials

A female Tuscan speaker produced the triplets *'papa – 'papera – pa'pale* and *'vivo – 'vivere – vi'vace* in a phrasal context (*Adesso udirete la parola ___ tre volte,* 'Now you will hear the word _____ three times'). Recordings took place in a sound-attenuated cabin at Konstanz University (44.1kHz, 16bit). These triplets were chosen because they allowed for easy segmentation and because durational differences were most extreme for these two vowels. We used the three-syllabic words with antepenultimate stress (*'papera, 'vivere*) as basis for manipulation because they already had an intermediate duration. The initial syllables were extracted and lengthened or shortened (PSOLA) to obtain an 8step continuum with 20ms steps – ranging from the duration of the initial unstressed vowel (*pa'pale, vi'vace*) to a duration 20 ms longer than that of the stressed initial vowel in the disyllabic member of the triplet (*'papa, 'vivo*). This was done for each syllable separately (for /pa/: 60 – 200ms; starting point *'papera* with 140ms, for /vi/: 52 – 192ms; starting point *'vivere*). Then, pitch was flattened to ascend only 10 Hz in 250 ms, which was a pitch movement lying in-between the two extreme f0-contours for the stressed and unstressed members of each pair. The manipulated syllables were spliced into another recording of the context (*Adesso udirete la parola + vi / pa*). For the stress condition, steps 1-5 were used (/pa/: 60-140 ms, /vi/: 52-132 ms). For the syllable condition, steps 4-8 were used (/pa/: 120-200 ms, /vi/: 112-192 ms).

3.3. Procedure

Four experimental lists were created to which participants were randomly assigned. Each list contained one of the stimuli in the stress condition (e.g. *'papera – pa'pale*) and the other one in the syllable condition (e.g. *'vivo – 'vivere*). Two lists contained the syllables "pa" in the stress condition and "vi" in the syllable condition, one of them starting with the stress condition, the other one starting with the syllable condition. In the other two lists the assignment of syllables to condition was reversed. The two alternatives to choose from (e.g., *vivo-vivere*) were presented visually on a laptop screen (Arial, font 28). Following [18] the alternatives were always presented at the same side of the screen, so that the two-syllable word always stood on the left side in the syllable condition and the word with antepenultimate stress always stood on the left in the stress condition. There were ten blocks of ten stimuli each. Each trial started with a fixation cross presented for 500 ms centred on screen. Participants' responses were collected via a button box, time-locked to the end of the word fragment with

a timeout of 2000 ms. Participants were told to react as quickly as possible. The experiment was controlled with presentation. The alternatives were read out before the start of the experiment and participants were asked to repeat them.

3.4. Results

Responses were analyzed separately for the stress and the syllable condition using binomial logistic regression models with the factor in question (stress or syllable number) and syllable (*pa* vs. *vi*) as fixed factors as well as participant as random factor. Fig. 5 shows responses in the stress condition, Fig. 6 responses in the syllable condition.
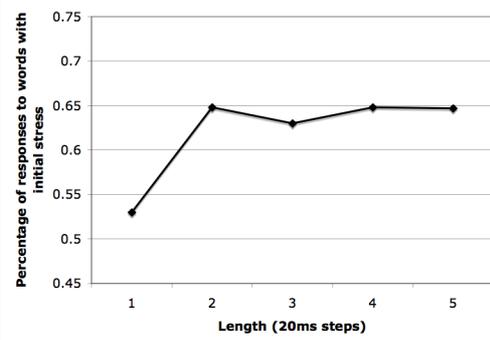

Fig 5. Stress condition: Percentage of responses to word with initial stressed syllable

Results showed a main effect of stress ($p < 0.05$), but no effect of syllable type and no interaction. There were significantly more responses to the word with word-initial stress (e.g. *'papera*) in length conditions 2-5 compared to length condition 1 ($\beta = 0.5$, $z = 2.3$, $p < 0.05$), see Fig. 5. Responses to length condition 1 did not differ from chance ($t(18) = 0.2$, $p > 0.8$) but responses to all other conditions were significantly different from chance (all p's < 0.05; one-sample t-test). Analysis of reaction times showed a significant effect of length: the longer the vowel in the initial syllable, the faster participants' responses ($\beta = -0.34$, $p < 0.05$).
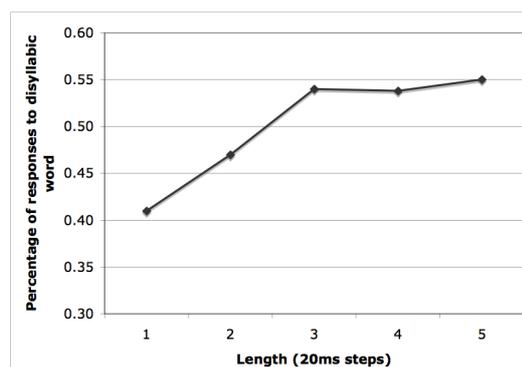

Fig 6. Syllable condition: Percentage of responses to disyllabic word

In the syllable condition, responses did not differ from chance level (all p's > 0.13), for none of the length steps. Not surprisingly, therefore, there was no effect of length on reaction times ($p > 0.6$). Nevertheless, there was a significant effect of length on participants' responses ($\beta = 0.14$, $z = 3.0$, $p < 0.0001$), see Fig. 6. There were significantly more responses to the disyllabic word in length conditions 3–5 compared to length condition 1 (all $\beta > 0.3$, $z > 1.9$ $p < 0.05$). Length condition 2 was intermediate and did not differ from the other length conditions ($p > 0.1$).

3.5. Discussion

The results of the Italian subjects in the stress condition corroborate that duration is a cue towards word stress in Italian, all else being equal. However, the effect of duration is not very strong. Very likely, the controlled factors pitch and intensity also play an important role (or a more important role) in natural stress perception. On the other hand, Italian listeners were not able to use durational patterns of word fragments to reliably predict the structure of the word (in terms of its number of syllables).

General Discussion

Our results clearly show that Italian speakers shorten syllables when they occur in longer words compared to when they occur in shorter words. This effect was larger for low than for high vowels and comparable to the effects previously reported for Italian [13] and English [4]. Not surprisingly, durational differences due to polysyllabic shortening were smaller than durational differences due to stress. What is more, unstressed vowels are also spectrally reduced, which is not the case for reduction due to polysyllabic shortening.

A speeded 2AFC task showed that the duration of a monosyllabic word fragment influenced responses in the stress condition, all else being equal. The longer the monosyllabic fragment, the more responses to the word starting with a stressed syllable. However, the identification function is not as steep as reported for Dutch in a comparable experiment [18]. This suggests that Italian listeners are less sensitive to durational cues to word stress than listeners of more stress-based languages. Regarding the perception of polysyllabic shortening, Italian listeners performed at chance level. In other words, they were not able to use the durational differences for identification (deciding whether the fragment stems from a disyllabic or trisyllabic word). Such an asymmetry between production and perception has also been discussed by [25]. Conceivably, Italian listeners' reduced sensitivity to vowel length differences may stem from the fact that vowel length is not phonemic in Italian as it is in Dutch. Therefore, Italian listeners might not process vowel duration for lexical activation. On the other hand, in non-lexical tasks, Italian listeners scored as high as German listeners who have a phonemic length distinction for some vowels. This task involved same-different judgments on vocalic length contrasts in CVC and CVːC syllables [26]. Therefore, we hypothesise that Italian listeners *are* sensitive to vowel duration but do not use it for prelexical or lexical processing.

In conclusion, our results show that Italian speakers alter the durational structure of words in a similar way as speakers of a stress-timed language do (vowel reduction, polysyllabic shortening). On the other hand, Italian listeners do not use durational cues during online speech recognition. This suggests that speech synthesis applications for Italian need not be as concerned with precise timing of sounds as in other languages (e.g., [27]). Future research will have to investigate other cues to word stress such as intensity and pitch, as well as their respective weighting with respect to duration – both in the perception of natural and synthesized speech.

References

[1] Pike, K.L. (1945). *The Intonation of American English*, Ann Arbor: University of Michigan Press.

[2] Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *JASA* 51, 2018–2024.

[3] Port, R.F. (1981). Linguistic timing factors in combination. *JASA* 69, 262-274.

[4] White, L. & Turk, A. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics* 38, 459–471.

[5] Nazzi, T., Bertoncini, J. & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm, *JEP:* 24, 756-766.

[6] Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal, *Cognition* 73, 265-292.

[7] Nazzi, T. & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants, *Speech Communication* 41, 233-243.

[8] Dauer, R. M. (1983): Stress-timing and syllable-timing reanalyzed, *Journal of Phonetics* 11, 51-62.

[9] Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm, *Phonetica* 66, 46-63.

[10] Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives, *Proceedings of Speech Prosody*, Aix-en-Provence, 115-120.

[11] Bertinetto, P.M. (1981): *Strutture Prosodiche dell'Italiano*, Firenze: Accademia della Crusca.

[12] Bertinetto, P.M. (1980). The perception of stress by Italian speakers, *Journal of Phonetics* 8, 385-395.

[13] D'Imperio, M. & Rosenthall, S. (1999). Phonetics and phonology of Italian main stress, *Phonology* 16, 1-28.

[14] Hajek, J., Stevens, M. & Webster, G. (2007). Vowel duration, compression and lengthening in stressed syllables in Italian, *Proceedings ICPhS* 16, 1057-1060.

[15] Marotta, G. (1985). *Modelli e Misure Ritmiche: la Durata Vocalica in Italiano*, Bologna: Zanichelli.

[16] Vayra, M., Fowler, C.A. & Avesani, C. (1987). Word-level coarticulation and shortening in Italian and English speech, *Haskins Laboratories Status Report on Speech Research* SR/91, 75-89.

[17] Bertinetto, P.M. (1977). A contrastive study on the production and perception of stress by English and Italian speakers" In Conte, M., Giacalone R.A. & Ramat, P. (eds.). *Sprache im Kontext. Akten des 12. Linguistischen Kolloquiums*, Tübingen: Niemeyer, 79-92.

[18] Reinisch, E., Jesse, A. & McQueen, J.M. (2011). Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue, *L&S* 54, 147-165.

[19] Shatzman, K. B. & McQueen, J. M. (2006). Prosodic knowledge affects the recognition of newly acquired words, *Psychological Science* 17, 372-77.

[20] Salverda, A.P., Dahan, D. & McQueen, J.M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension, *Cogn.* 90, 51–89.

[21] De Mauro, T., Mancini, F., Vedovelli, M. & Voghera, M. (1993). BADIP – Banca Dati dell'Italiano Parlato, [badip.uni-graz.at].

[22] Boersma, P. & Weenik, D. (2009): Praat: Doing phonetics by computer.

[23] Baayen, H.R. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. CUP

[24] Baayen, H.R., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items, *JML* 59, 390-412.

[25] Labov, W., Karen, M. & Miller, C. (1991). Near-mergers and the suspension of phonemic contrast. Language Variation and Change, 3, 33 – 74.

[26] Berger, I., Braun, B. & Altman, H. (2011). L1 influence on the perception of consonantal and vocalic length contrasts. Phonetics & Phonology in Iberia. Tarragona.

[27] Möbius B., & Santen P. H. J. (1996). Modeling Segmental duration in German Text-to-Speech Synthesis. *Proceedings of ICSLP*, Vol. 4, pp 2395-2398.