Filled pauses and prolongations in Roman Italian task-oriented dialogue

Jessica Di Napoli
Institute of English, American, and Romance Studies, RWTH Aachen University
jessica.dinapoli@ifaar.rwth-aachen.de

Abstract

This paper presents work in progress on two markers of hesitation in Roman Italian task-oriented dialogue, namely filled pauses and prolongations. We investigate their form, relative frequency, and distributional characteristics in Italian. Initial results suggest that Italian speakers produce prolongations more frequently than filled pauses, and that the prototypical hesitant prolongation involves a word-final vowel.

1 Introduction

Filled pauses and disfluent segment prolongation signal hesitation in spoken discourse. These two markers of disfluency are similar in that they both signal hesitation through vocalization and duration, which distinguishes them from other markers of disfluency, such as truncations or silences (Eklund, 2001; Betz et al., 2017). However, they are distinct in their specific manifestations. Filled pauses (or fillers) contain non-lexical material and are typically comprised of a central vowel, optionally followed by a nasal, as in *uh* and *um*. Prolongation is instead characterized by the marked lengthening of a segment within a word, prompting a local slow-down in speech rate (Betz & Wagner, 2016).

Previous studies on filled pauses and prolongations in languages such as Swedish (Eklund, 2001) and German (Betz et al., 2017) have shown that filled pauses tend to be longer than prolongations, as well as more frequent (Eklund, 2001). These studies have also pointed to language-specific differences in the distribution of prolongations, both with respect to preferred target segment type (e.g. vowel vs. fricative) and position within the word (i.e., the relative frequency of prolongation in word-initial, word-medial, and word-final position).

In conversation, filled pauses and prolongations play a role in turn-taking. Filled pauses, in particular in task-oriented dialogue such as Map Tasks, help an interlocutor gain time before responding to a question (Lickley, 2001). Prolongation, instead, at the end of a turn can signal that a speaker intends to hold the floor (Savino & Refice, 2000; Gravano & Hirschberg, 2011). A speaker's role in a dialogue can also have an effect on how frequently they produce filled pauses and prolongations. In Map Tasks, instruction *givers*, who describe a route to an instruction *follower*, generally produce more disfluencies because they produce dialogue moves which are more likely to contain disfluencies, such as the instruct move (Lickley, 2001).

With respect to Italian, previous studies suggest that the frequency of filled pauses varies by speaking style, with spontaneous speech displaying more filled pauses than read speech (Magno Caldognetto et al., 1997; see Di Napoli, 2019 for filled pauses in news reading). In addition, Giannini (2003) has shown that within Italian there is regional variation in the vowel quality of filled pauses, with speakers of southern Italian varieties producing a more central vowel than speakers of central Italian varieties.

Savino & Refice (2000) suggest that speakers of Italian may not produce fillers as frequently as speakers of other languages (cf. Eklund, 2001). They suggest that, given the frequency of words which end in open syllables in Italian, word-final vowel prolongation may be more frequent. Savino & Refice (2000) show that disfluent word-final prolongations in Bari Italian (which they refer to as "planning lengthening") are systematically longer than, and thus distinct from, prosodic phrase-final vowels, which undergo default phrase-final lengthening.

A number of open questions remain with respect to filled pauses and prolongations in

Italian. Do Italian speakers consistently produce more filled pauses than prolongations? What are the characteristics of prolongations in Italian? Which segments are most often prolonged, and in what position in the word? Is there any evidence that filled pauses and prolongations function interactionally at the turn level? The present study presents work in progress exploring these questions based on task-oriented dialogue in Roman Italian.

2 Methods

The speech material consists of Map Tasks obtained from the CLIPS corpus. To date, we have completed analysis of two Map Tasks produced by four native speakers of Roman Italian (two female, two male, aged 26-30). Dialogue partners were matched for gender (F-F and M-M). In total, approximately 27 minutes of dialogue have been analyzed, comprised of 9 minutes of dialogue for the F-F pair, and 18 minutes of dialogue for the M-M pair.

Speech annotation was carried out in Praat (Boersma & Weenink, 2018). Speech was first segmented into turns, which we define here as interpausal units (IPUs), that is, stretches of speech from a single speaker bounded by silence longer than 100 ms (Koiso et al., 1998; see also Di Napoli, 2018). Intervals of speech and silence were annotated automatically by script and then verified manually. Filled pauses and prolongations were included within turns. Turn transitions, that is, change or hold were also annotated (see Di Napoli, 2018).

Syllable nuclei were then marked for each turn semi-automatically using a version of the script by de Jong & Wemke (2009), which was modified to allow for manual correction of annotation. In the process of syllable annotation, we marked all filled pauses which occurred, as well as all instances of hesitant prolongation. These tended to be clearly identifiable when comparing segment durations across syllables within a turn or series of turns – prolonged segments displayed markedly longer duration given the surrounding context.

We then analyzed the frequency of occurrence of filled pauses and prolongations, both with respect to total dialogue time (including silences), and the amount of speech produced, quantified either in terms of the number of syllables or number of turns produced by a speaker. As speakers in the Italian Map Tasks took turns as instruction giver and follower, frequencies are compared for speakers in their two dialogue roles.

For each filled pause and prolongation, we then analyzed the duration, the segmental composition, and the position of the filled pause or word presenting a prolonged segment with respect to the turn (initial, medial, or final). Additionally, for prolongations, we marked the part of speech of words with prolonged segments, as well as the position of the prolonged segments with respect to the word (again, initial, medial, or final).

3 Results

In total, we identified 151 filled pauses and prolongations across the two Map Task dialogues. Table 1 presents the results for the relative numbers of filled pauses and prolongations produced by each speaker according to dialogue role (instruction giver vs. follower). As is evident in Table 1, all speakers produce prolongation much more frequently than filled pauses; prolongations are approximately four times more frequent across speakers. Additionally, both markers of hesitation are produced more frequently by speakers when they act as instruction giver than as instruction follower. In particular, filled pauses are produced only rarely by instruction followers.

A degree of individual variability is also evident in Table 1 – speaker M1 produces many more filled pauses and prolongations than the other three speakers, both as instruction giver and follower. However, each dialogue had a different duration, and each speaker produced a different amount of speech, so it is important to consider the frequency of filled pauses and prolongations with respect to speech actually produced.

Table 2 shows results for the frequency of filled

Speaker	Giver		Follower		Total
	FP	PR	FP	PR	
F1	3	20	0	3	26
F2	4	16	0	6	26
M1	20	44	0	6	70
M2	2	14	1	12	29
Total	29	94	1	27	151

Table 1: Number of filled pauses (FP) and prolongations (PR) by speaker according to dialogue role (instruction giver or follower).

Speaker	#/min		#/100syll		% turns	
	FP	PR	FP	PR	FP	PR
F1	0.32	2.43	0.22	1.69	1.7	13.3
F2	0.42	2.32	0.38	2.08	2.4	13.4
M1	1.61	4.02	0.78	1.96	5.9	14.7
M2	0.24	2.09	0.12	1.05	1.0	8.4
Mean	0.65	2.72	0.38	1.70	2.8	12.4

Table 2: Frequency of filled pauses and prolongations by speaker per minute of dialogue and per 100 syllables of speech. The final column shows the percent of turns with filled pauses or prolongations.

pauses and prolongations. It is clear from the table that the frequency differs somewhat according to the measure adopted. On average, approximately 12 percent of turns display prolongation, and 3 percent contain filled pauses. There are 1.7 prolongations every 100 syllables (or roughly 3.4 per 100 words, if we estimate an average of 2 syllables per word).

In terms of duration, the mean duration of filled pauses is 568 ms (*SD* = 325 ms), while the mean duration of prolongations is 396 ms (*SD* = 229 ms). Filled pauses thus appear to be longer than prolongations, but since our sample size is very small, and the standard deviations are very large, we can only view this as a possible tendency.

Table 3 presents the results for the position of filled pauses and prolongations with respect to the turn (or IPU). All speakers are combined here. Filled pauses occur most frequently in turn-initial position, while prolongations occur most frequently in turn-medial and turn-final position. In total, approximately 80 percent of filled pauses and 60 percent of prolongations occur together with a silent pause.

We now report results for the segmental composition of filled pauses and prolongations. All filled pauses were either entirely vocalic, as in *eeh*, or entirely nasal, as in *mmh*. The vocalic filled pauses were most frequent, and accounted for 2/3 of the total. In terms of prolongation, vowel prolongation was much more frequent than prolongation of a consonant. Details are provided in the following section.

	turn-initial	turn-medial	turn-final
FP	66.7 %	20.0 %	13.3 %
PR	18.2 %	43.0 %	38.8 %

Table 3: Frequency of filled pauses and prolongations according to position in the turn.

3.1 Additional features of prolongations

In this section, we zoom in to look at prolongations in more detail. Beginning with the type of segment subject to prolongation, Table 4 presents the relative frequencies of vowel and consonant phones found to undergo prolongation. As mentioned above, and as is clear in the table, the vast majority of prolongations are cases of vowel prolongation (at approximately 90 percent of the total).

V phones	% total	C phones	% total
o	28.1	**n, l, m**	6.6
e	24.8	**s**	2.5
a	18.2	**t**	0.8
i, u	9.1		
diphthong	9.9		
Total % V	90.1	**Total % C**	10.0

Table 4: Prolongations according to phone type.

In particular, /o/, /e/, and /a/ are the most frequently prolonged segments, and in this order. The least frequent vowel phoneme is /u/ at only 1 percent. Diphthongs such as /ai/ and /oi/ were prolonged somewhat more often. In terms of consonants that undergo prolongation, /n/ is the single most frequent, at 4 percent.

With respect to the position of prolonged segments in a word, Table 5 shows that prolongations are word-final in just over 90 percent of cases. Prolonged segments in Italian are thus typically vowels, and typically in word-final position.

Regarding part of speech and word class, 53 percent of the words exhibiting prolongation were open class words (in particular verbs, nouns, and adverbs). Forty-five percent of words were closed class words (most frequently prepositions, conjunctions, and determiners). For the remaining 2 percent of words, part of speech could not be determined due to speech error.

word-initial	word-medial	word-final
5.0 %	4.1 %	90.9 %

Table 5: Frequency of prolongations according to position in the word.

4 Discussion and Conclusion

Preliminary results presented in this paper reveal some important characteristics of filled pauses and prolongations in Italian. First of all, filled pauses appear to be much less frequent than prolongations, for all speakers. This stands in contrast with what Eklund (2001) found for Swedish, but is in line with Savino & Refice's (2000) claim that speakers of Italian do not need to insert segments to indicate hesitation; this same goal can be achieved through word-final vowel prolongation.

In fact, the present study has shown a strong preference for prolongation of vowels in word-final position – 87 percent of prolonged segments in the two Map Task dialogues are word-final vowels. Overall, the distribution of prolongation in word-initial, word-medial, and word-final position determined here (5–4–91%) differs considerably from languages such as American English and Swedish (as reported in Betz et al., 2017).

Our results also provide support for an interactional function of filled pauses and prolongations in Italian. Both markers of hesitation are produced more frequently by instruction givers than instruction followers, in line with Lickley (2001). Filled pauses tend to occur turn-initially in our speech material, and when they do, they sometimes (30 percent of the time) occur immediately after a question. They could serve the function here of giving additional time to an interlocutor to answer. Prolongations instead occur more often turn-finally, and turn-final prolongations are systematically followed by holds (rather than speaker changes) in the dialogues presented here, 86 percent of the time.

Overall, these findings contribute to ongoing work on disfluency and provide insights into language-specific characteristics of hesitation.

References

Betz, S., Eklund, R., and Wagner, P. Prolongation in German. 2017. In R. Eklund and R. Rose (eds.), *Proceedings of DiSS 2017, Disfluency in Spontaneous Speech*, 18–19 August 2017, Stockholm, Sweden, pp. 13–16.

Betz, S. and Wagner, P. 2016. Disfluent lengthening in spontaneous speech. In O. Jokisch (ed.), *Elektronische Sprachsignalverarbeitung (ESSV) 2016*, *Studientexte zur Sprachkommunikation (81)*, Dresden: TUD Press, pp. 135–144.

Boersma, P. and Weenink, D. 2018. *Praat: doing phonetics by computer*. Computer program, version 6.0.43.

CLIPS corpus (*Corpora e Lessici di Italiano Parlato e Scritto*). http://www.clips.unina.it.

de Jong, N. H. and Wemke, T. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behaviour Research Methods* 41(2): 385–390.

Di Napoli, J. 2018. Overlapping speech and turn-taking in Roman Italian. In *Proceedings of P&P13: Conference on Phonetics and Phonology in German-speaking countries*, 28–29 September 2017, Berlin, Germany, pp. 37–40.

Di Napoli, J. 2019. Functions of pause in Italian television news broadcasts. In *Proceedings of the 1st International Seminar on the Foundations of Speech – Pausing, Breathing and Voice*, 1–3 December 2019, Sønderborg, Denmark, pp. 109–111.

Eklund, R. 2001. Prolongations: A dark horse in the disfluency stable. In *Proceedings of DiSS 2001, Disfluency in Spontaneous Speech*, 29–31 August 2001, Edinburgh, UK, pp. 5–8.

Giannini, A. 2003. Hesitation phenomena in spontaneous Italian. In M. J. Solé, D. Recasens, and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 3–9 August 2003, Barcelona, Spain, pp. 2653–2656.

Gravano, A. and Hirschberg, J. 2011. Turn-taking cues in task-oriented dialogue. *Computer Speech and Language* 25(3): 601–634.

Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., and Den, Y. 1998. An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogs. *Language and Speech* 41: 295–321.

Lickley, R. J. 2001. Dialogue moves and disfluency rates. In *Proceedings of DiSS 2001, Disfluency in Spontaneous Speech*, 29–31 August 2001, Edinburgh, UK, pp. 93–96.

Magno Caldognetto, E., Zmarich, C., and Ferrero, F. 1997. A comparative acoustic study of spontaneous and read Italian speech. In G. Kokkinakis, N. Fakotakis, and E. Dermatas (eds.), *Proceedings of EUROSPEECH*, Rhodes: ISCA, pp. 779-782.

Savino, M. and Refice, M. 2000. Acoustic cues for classifying communicative intentions in dialogue systems. In P. Sojka, I. Kopeček, and K. Pala (eds.), *Text, Speech and Dialogue*, Berlin: Springer, pp. 421–426.