Disfluency in Spontaneous Speech (DiSS) Workshop 2023
28-30 August 2023, Bielefeld, Germany

Prolongation in Italian

Loredana Schettino[1], *Robert Eklund*[2]

[1]Interdepartmental Research Center Urban/Eco, Università di Napoli Federico II, Italy
[2]Department of Culture and Communication, Linköping University, Sweden

loredana.schettino@unina.it, robert@roberteklund.info

Abstract

This study follows the framework set by the strand of corpus studies aimed at investigating and unveiling the role of prolongations as linguistic elements by comparing their use in different languages, ranging from Germanic languages (Swedish, American English, German) to Tok Pisin, Chinese Mandarin, Hungarian and Hebrew, and provides evidence on the use of prolongation in a Romance language, namely Italian. The analysis is conducted on different speech styles, i.e., descriptive informal dialogic speech as well as informative monologic speech, and concerns the distributional characterisations of prolongation phenomena, their segmental and durational traits also considering the comparison with another type of voiced speech management phenomena, that is non-verbal vocalisations. The main results show that in the considered Italian data speakers use prolongations more frequently than non-verbal vocalisations and the latter are generally longer, which argues for the fact that these two voiced phenomena are differently involved in speech management. Then, the distributional and segmental features of prolongations in Italian as compared to other languages support the idea that prolongations, as linguistic elements, are subjected, to a certain extent, to the phonotactic constraints of languages.[1]

Index Terms: prolongations, filled pauses, Italian, semi-spontaneous speech

1. Introduction

Phenomena such as segment prolongations and non-verbal vocalizations have been described as belonging to the heterogeneous class of linguistic elements that speakers may use to manage the complex online speech production and perception processes [1, 2, 3]. More specifically, speakers may suspend their speech and gain valuable time by lengthening segmental material or by producing non-verbal vocalizations or nasalizations, i.e., *eeh*, *ehm*. The occurrence and surface realisations of these phenomena may vary due to extra-linguistic contextual factors, like situational or individual-dependent factors (see [4] for an overview), but cross-linguistic variation has also been documented, supporting the idea that it may be subjected to language-specific phonological, syntactic, and semantic constraints as well as exhibit common properties [5].

In particular, various (and counting) cross-linguistic studies by Eklund and colleagues have investigated the characteristics of prolongations (PRs) showing common patterns as well as language-specific traits. The evidence concerning the distribution in words, initial–medial–final, that emerged from the first of these studies [6] for Swedish and American English and from [7, 4] for Swedish and Tok Pisin showed an almost identical distribution for Swedish and American English (both Germanic languages), i.e., 30–20–50% for initial–medial–final position, and a very different distribution for Tok Pisin, 18–0–82%. Hence, Eklund suggested the "morphology matters" hypothesis which posited PR distribution as a function of morphological complexity of the language in question, with Swedish and English being of similar complexity and Tok Pisin being much less complex. This hypothesis was given support in ensuing work on Japanese, 10–5–85% [8], and Mandarin, 4–5–85% [9], also two languages with a less complex morphology than Swedish and English. However, ensuing investigations made this picture more complicated. While studies of PRs in Hungarian presented a distribution of 18–19–63% [10, 11] and more or less supported the "morphology matters" hypothesis – then renamed to the perhaps more accurate "phonotactics matters" hypothesis – studies of German [12] provided a clear counter-example with a 7–15–78% distribution, being more similar to Tok Pisin, Japanese and Mandarin than to Swedish, English and Hungarian, despite exhibiting morphology/phonotactics more similar to Swedish, English and Hungarian than to Tok Pisin, Japanese and Mandarin. This new picture was further supported by a study of Hebrew [13], exhibiting a highly skewed distribution of 1–1–98%. The conclusion to be drawn from the aforementioned studies is that although morphology/phonotactics might play some role in what position of the word PRs might appear, other factors must play a role, too. Note that most of the aforementioned studies include analyses of what type of word is affected, including both open/closed words classes, phonological characteristics, and that this kind of analyses might provide further insight. Generally, although all types of segments (including unvoiced stops) can be subjected to prolongation, vowels, sonorants, and continuants are more prone to it within language-specific phonological constraints. For example, in German, vowel length is a distinctive feature and, assumedly for this reason, prolongation tends to be avoided on short vowels [12]. Prolongations have been described also in relation to filled pauses, being both durational and voiced speech management phenomena [10, 13]. However, they fundamentally differ in the degree of integration with the contextual speech units, prolongations belonging to word segments and filled pauses being independent elements. Moreover, filled pauses are found to have greater average duration than lengthenings (Swedish and Tok Pisin [7, 4]; German [12]; European Portuguese [14]; Italian [15, 16]). This suggests that they may be involved in different ways in speech planning.

In Italian, it has been shown that prolongations commonly occur on word-final vowels and, in the rare cases where the final segment is a consonant, it can be realized by producing a schwa sound [17, 15, 16], thus reproducing the canonical syllabic structure in Italian: CV [18]. Maybe due to this *filler-like* feature, in Italian studies, prolongations are perceived and

[1]This article is the result of the collaboration among the authors. However, for academic purposes, Loredana Schettino is responsible for sections 2 and 3, Robert Eklund for sections 1. Both the authors are responsible for section 4.

10.21437/DiSS.2023-17

classified as a particular type of *filled pause*. This work delves deeper into the study of the way voiced speech management phenomena, e.g., prolongations and filled pauses, appear in Italian by considering different speech styles and provides further evidence, this time from a Romance language, to the developing picture sketched by Eklund and colleagues. So, the study aims to foster the analysis of the way language-specific structures may affect the realisation of these phenomena and tests whether the "phonotactics matters" hypothesis holds true when observing Italian data.

2. Method

2.1. Data

The dataset considered for this work consists of a dialogic and a monologic section, both including approximately 40 minutes of semi-spontaneous speech by speakers of the Neapolitan variety of Italian (Table 1).

The dialogic section (DG) contains four task-oriented dialogues from the CLIPS corpus [19]. Two dialogues were elicited using the "Spot the Differences" method (*test delle differenze*, td), whereby the interlocutors are given similar pictures and asked to spot the differences without seeing each other and only relying on the verbal channel. The other two dialogues were obtained with the "Map-Task" elicitation method (mt). The interlocutors are given similar maps, one of which had a route marked on it, and the task of reproducing the route on the empty map. This settings provide mainly descriptive semi-spontaneous speech characterized by a low degree of discourse planning, and a high degree of collaboration in the interaction.

The monologic section (MG) includes audio-visual recordings of guided tours at San Martino Charterhouse (in Naples) led by three female expert guides (CHROME corpus [20]). It consists of informative semi-monologic, semi-spontaneous speech characterized by a high degree of discourse planning and an asymmetrical relationship between the speakers.

Table 1: *Dataset for the analysis*

Speech Type	File	Speaker	wrd(tkn)
DG, approx. 40'	DGtdA01	p1	1328
		p2	1551
	DGtdA02	p1	803
		p2	927
	DGmtA01	p1	540
		p2	615
	DGmtB01	p1	833
		p2	750
MG, approx. 40'	MGG01	G01	1491
	MGG02	G02	2788
	MGG03	G03	2138

2.2. Analysis

The phenomena that are considered objects of analysis in this work were annotated along with other speech management phenomena according to the system described in [21]. Prolongations (PRs) were identified as marked lengthening of segmental material [4, 22]; filled pauses (FPs) as non-verbal fillers realized as vocalization and/or nasalization. The identification of these phenomena did not rely on absolute measures but on perceptual judgments given their specific contexts of occurrence. To test the reliability of the system the Inter-Annotator Agreement was evaluated by measuring the values of Cohen's κ, i.e., 0.92 for dialogic data and 0.82 for monologic data, both standing for "high agreement" according to [23]. For both types of voiced speech management phenomena, the variation of the following variables was considered: frequency of occurrence (per word token); duration (ms) of lengthened segments; segmental composition (phonetic transcription). As for PR, the analysis also concerned: the lexical category of the word containing segmental lengthening (part-of-speech); the position of PRs within the word in which it occurs, namely initial, medial, or final position.

The annotation was conducted by integrating the use of the ELAN software [24], which facilitates multilevel linguistic annotations, and of Praat [25]. The statistical analysis was performed in R [26], using Generalized Linear and Linear Mixed Models in order to control for individual variability ('lme4' package [27]) by considering the Speaker as a random effect.

3. Results

3.1. General frequencies

In the selected dataset, 668 voiced phenomena were identified, namely 387 PRs and 281 FPs. As illustrated in Figure 1, on the amount of phenomena per type of speech (DG, N=372; MG, N=296), prolongations are more frequent than fillers in both dialogic (PRs, N=229; FPs, N=143) and monologic speech (PRs, N=158; FPs, N=138), but this difference is only statistically significant in dialogues (both td and mt: Est.= 0.55, SE = 0.17, $z = 3.26$, $p = 0.001$). No statistically relevant difference emerges between the frequency of voiced phenomena in dialogic and in monologic speech.

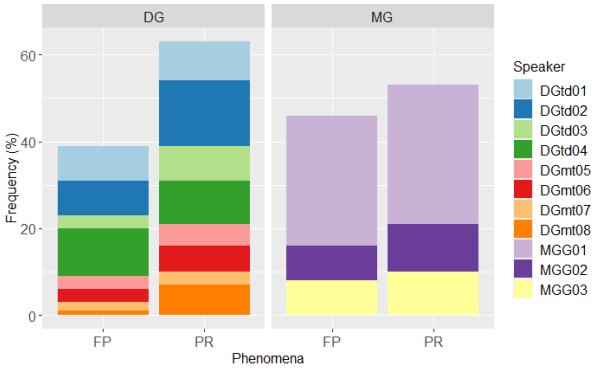

Figure 1: *Frequency of occurrence of FPs and PRs in dialogic and monologic speech.*

3.2. Duration

As reported in Figure 2 and in Table 2 FPs are significantly longer than PRs in both dialogic (Est.= 195, $SE = 24.6$, $t = 7.91$, $p < .0001$) and monologic speech (Est.= 153, SE = 26.2, $t = 5.82$, $p < .0001$). Moreover, both phenomena are slightly yet significantly longer in the dialogic than in the monologic speech data (Est.= 91.3, $SE = 39.9$, $t = 2.29$, $p = 0.041$).

Then, as shown in Figure 3, both voiced phenomena are significantly longer when consisting of a vowel sound (V) followed by a consonant (C), i.e., a nasal sound as described in the next section 3.3 (Est.= 405.04, $SE = 39.99$, $t = 10.13$, $p <$ 2e-16).

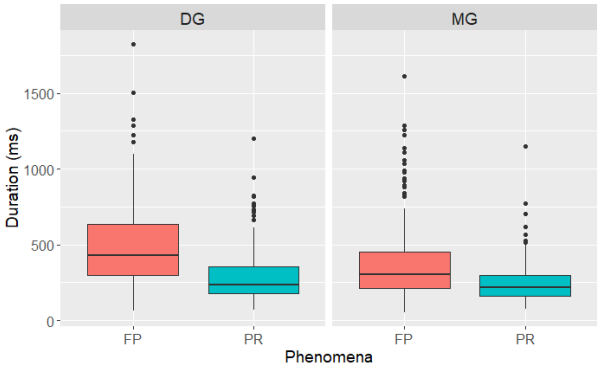

Figure 2: *Duration of FPs and PRs in dialogic and monologic speech.*

Table 2: *PRs and FPs duration per speech type*

Speech Type	Phenomena	M Dur (ms)	StDev
DG	FP	491	294
DG	PR	290	173
MG	FP	408	304
MG	PR	251	139

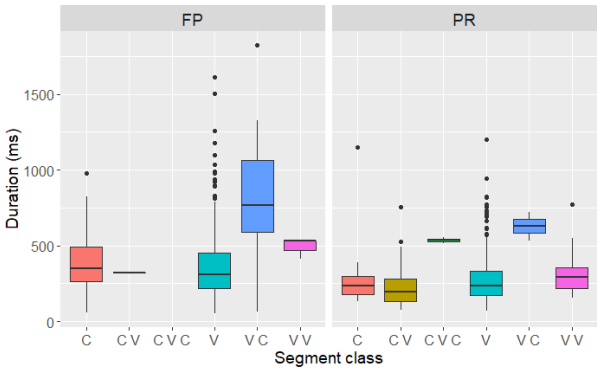

Figure 3: *Duration of different realizations of FP and PR in dialogic and monologic speech.*

3.3. Segmental Content

As for the distribution of the surface realizations of PRs (Figure 4), most instances are realized by lengthening vowel sounds (76%), the remainder includes the lengthening of word-final consonants followed by a schwa sound (about 16%), diphthongized and lengthened word-final vowels with a schwa sound (about 4%), the lengthening of continuants consonants ([l] and [s]) or nasals ([n], 3%) and rare cases of word-final vowels plus nasal sound lengthening (1%). As for the realization of FPs (Figure 5), the picture is less varied than the one obtained for PRs but quite rich though. In this case, most instances are realized as central schwa sounds [ə] (52%), followed by nasalized vocalizations [m] (19%) and [əm] (8%). The other cases consist of open-mid or closed-mid front vowels, [ɛ], [e], less frequently, by their nasalized counterparts, [ɛm], [em], and very few open-central realisations, [a], [am].

The comparison of the realisations of PRs and FPs grouped by their sound class, i.e., vowel (V) or consonats (C), in Figure 6, shows that both voiced phenomena are mostly realised by vocalic sounds. The differences between these two types of phenomena concern the vowel quality (as previously described) and the structure of less frequent instances.

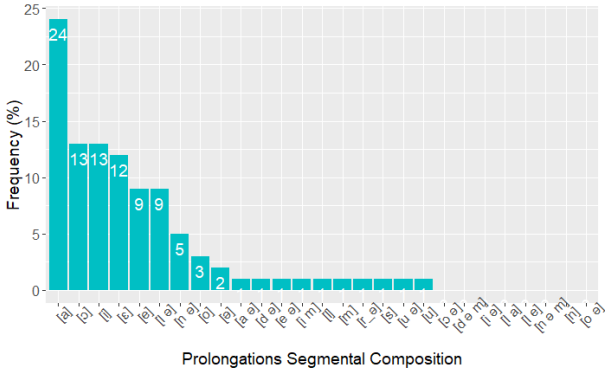

Figure 4: *Segmental content of PRs.*

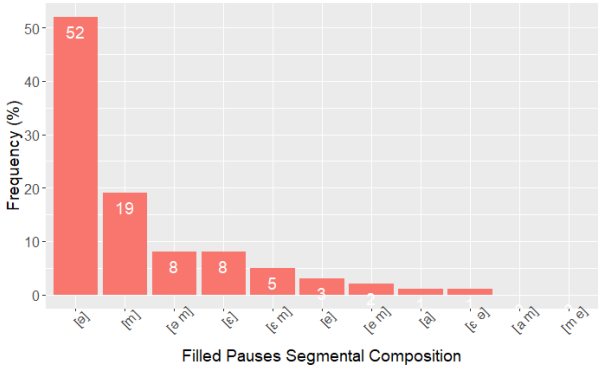

Figure 5: *Segmental content of FPs.*

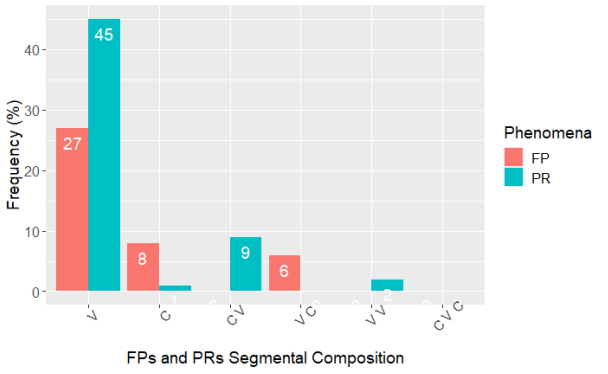

Figure 6: *Phone class of FPs and PRs.*

The segmental material most frequently prolonged consists of vowels, the fewer cases of lengthened consonants are mostly followed by a schwa. As for filled pauses, non-nasalised vocalisations are the most frequent realisations, followed by nasalised vocalisations and vowel+nasal realisations. No relevant difference emerged between the realisations of PRs and FPs in dialogic and monologic speech.

3.4. Prolongation – positioning and word part-of-speech

In the Italian speech data considered, PRs almost exclusively occur in word-final position according to the following distribution: 3–1–96%. As for the lengthened word part of speech, functional words (206) are slightly more frequently lengthened than content words (181) but the difference is not statistically significant (Est.= 0.21, *SE* = 0.18, z = 1.15, p = 0.25). As illustrated in Figure 7, most lengthened words belong to the class of prepositions, followed by verbs, articles, nouns and conjunctions.

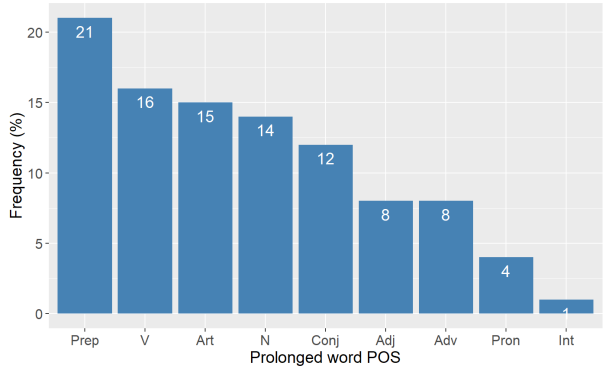

Figure 7: *Prolonged word part of speech.*

4. Discussion and Conclusions

This study follows the framework set by the strand of corpus studies aimed at investigating and unveiling the role of prolongations as linguistic elements by comparing their use in different languages, ranging from Germanic languages (Swedish, American English [6], German [12]) to Tok Pisin [7], Chinese Mandarin [9], Hungarian [10] and Hebrew [13] and provides evidence from a Romance language, namely Italian.

Caution is advised in drawing strong conclusions when comparing findings from analyses based on different amounts and types of speech data. However, the findings reported in this work, along with those presented in previous studies, contribute, to a certain extent, to add pieces to the general picture.

The findings concerning prolongations distribution show that in the considered Italian speech data, unlike in other languages, content words (i.e., verbs and nouns) are prolonged almost as much as functional word (i.e., prepositions, articles and conjunctions). This does not necessarily counter the idea that speakers generally produce prolongations before or on items with high cognitive load, e.g., either the preposition or article before a semantically heavy item [7], since verbs and nouns may as well precede semantically relevant elements such as a specific verb argument, or a noun's adjective (in Italian, qualifying adjectives mostly follow the noun they describe). Within word segments, prolongations most likely occur on word-final segments, i.e., 3–1–96%, resembling the distribution observed in Hebrew [13], Mandarin Chinese [9] and Japanese [8] and quite differently from those in Tok Pisin, German and from the somewhat more even distributions that characterize some Germanic languages, American English and Swedish [6], and Hungarian [10], see Table 3.

At the segmental level, much like what is observed in other languages, the segmental material preferentially lengthened is vocalic and the fewer cases of consonant prolongations consist

Table 3: *Summary of PR appearance. W-i, the first segment is prolonged compared to normal duration, perceptually normalized for speech rate; W-f, ditto final segment; W-m, not word-initial or word-final.*

Language	**W-i%***	**W-m%**	**W-f%**	**Source**
Swedish	31/24**	18/17**	59/58**	[6, p. 2632]
Swedish	28	20	52	[7, p. 6]
				[4, p. 246]
Am. English	32	22	50	[6, p. 2632]
Tok Pisin	18	0	82	[7, p. 7]
				[4, p. 249]
Japanese	10	5	85	[8, p. 88]
Mandarin	4	1	85	[9, p. 2182]
Hungarian	18	19	63	[10, p. 30]
				[11, p. 43]
German	7	15	78	[12, p. 14]
Hebrew	1	1	98	[13, p. 50]
Italian	3	1	96	this study

*Most figures are slightly rounded off. **Two corpora.

of sonorants or continuants sounds. However, in Italian data, we observe that the cases of word-final consonant prolongation are mostly realised also by the insertion of a central vocalic sound, thus reconstructing the canonical CV syllable structure in Italian [28, 18]. Moreover, it has been attested that the mid-central vocalic variant is a characterizing trait of the dialectal substrate of the Neapolitan variety of Italian [29].

These pieces of evidence corroborate the assumption that, to a certain extent, the distribution of prolongations may be subjected to the phonotactic constraints of languages. Indeed, this could contribute to explaining the distinction between distributions in Italian, where less complex syllable structures are allowed, i.e., C^3VC [30], as compared to some Germanic languages such as Swedish, which allows quite complex syllable structures, i.e., C^3VC^9 [4]. Furthermore, results highlight that for the segmental realisation of prolongations, speakers tend to rely on articulatory models related to their phonological inventory, and the mid-central vocalic realisations may not just result from speakers' articulatory economy but be rather connected to the inventory of dialectal sounds that emerge in the local variety of Italian [17].

As for the duration, in line with the literature, in Italian data, prolongations are generally shorter than filled pauses, which supports the idea that these two types of phenomena may be used for different purposes and be involved in different speech management processes. However, both prolongations and filled pauses are longer when finishing with a nasal sound, which suggests that the insertion of nasals represents a common strategy for gaining more time when speech production processes require it. This finding, together with the observation of no relevant difference between the frequency of nasalized phenomena in the pre-scripted monologic informative speech and in the informal dialogic speech, suggests that nasal realisations in Italian may not sound as "less salient" and pleasant as found by [31] in a perception study in German.

Future work will be devoted to further investigating the uses of prolongations and filled pauses in different languages by also considering their functional role given their context of occurrence [21]. Also, it would be interesting to integrate the production perspective with perception evaluations.

5. Acknowledgements

Work funded by the Italian National Project PRIN Cultural Heritage Resources Orienting Multimodal Experiences (CHROME)(B52F15000450001).

6. References

[1] W. J. Levelt, *Speaking: From intention to articulation*. Cambridge and London: MIT press, 1993.

[2] J. Allwood, J. Nivre, and E. Ahlsén, "Speech management—on the non-written life of speech," *Nordic Journal of Linguistics*, vol. 13, pp. 3–48, 1990.

[3] M. Voghera, *Dal parlato alla grammatica*. Roma: Carocci, 2017.

[4] R. Eklund, "Disfluency in Swedish Human–Human and Human–Machine travel booking dialogues," Ph.D. dissertation, Linköping University Electronic Press, 2004.

[5] J. Ginzburg, R. Fernández, and D. Schlangen, "Disfluencies as intra-utterance dialogue moves," *Semantics and Pragmatics*, vol. 7, pp. 9–1, 2014.

[6] R. Eklund and E. E. Shriberg, "Crosslinguistic disfluency modelling: A comparative analysis of Swedish and American English Human–Human and Human–Hachine dialogues," in *Proceedings of ICSLP 98, The 5th International Conference on Spoken Language Processing*, vol. 6. 30 November – 4 December, 1998, Sydney, Australia, 1998, pp. 2627–2630.

[7] R. Eklund, "Prolongations: A dark horse in the disfluency stable," in *Proceedings of DiSS 2001, ISCA Tutorial and Research Workshop on Disfluency in Spontaneous Speech*. 29–31 August 2001, University of Edinburgh, UK, 2001, pp. 5–8.

[8] Y. Den, "Some strategies in prolonging speech segments in spontaneous Japanese," in *Proceedings of DiSS 2003, ISCA Tutorial and Research Workshop on Disfluency in Spontaneous Speech*. 5–8 September 2003, Göteborg, Sweden, 2003, pp. 87–90.

[9] T.-L. Lee, Y.-F. He, Y.-J. Huang, S.-C. Tseng, and R. Eklund, "Prolongation in spontaneous Mandarin," in *Proceedings of ICSLP 2004*, vol. III. 4–8 October 2004, Jeju Island, Korea, 2004, pp. 2181–2184.

[10] M. Gósy and R. Eklund, "Segment prolongation in Hungarian," in *Proceedings of DiSS 2017, The 8th Workshop on Disfluency in Spontaneous Speech*, R. Rose and R. Eklund, Eds. 18–19 August 2017, KTH Royal Institute of Technology, Stockholm, Sweden, 2017, pp. 29–32.

[11] ——, "Language-specific patterns of segment prolongation in hungarian," *The Phonetician. Journal of the International Society of Phonetic Sciences (ISPhS)*, vol. 115, pp. 36–52, 2018.

[12] S. Betz, R. Eklund, and P. Wagner, "Prolongation in German," in *Proceedings of DiSS 2017, The 8th Workshop on Disfluency in Spontaneous Speech*, R. Rose and R. Eklund, Eds. 18–19 August 2017, KTH Royal Institute of Technology, Stockholm, Sweden, 2017, pp. 13–16.

[13] V. Silber-Varod, M. Gósy, and R. Eklund, "Segment prolongation in Hebrew," in *Proceedings of DiSS 2019, The 9th Workshop on Disfluency in Spontaneous Speech*, R. Rose and R. Eklund, Eds. 12–13 September 2019, ELTE Eötvös Loràand University, Budapest, Hungary, 2019, pp. 47–50.

[14] H. Moniz, A. I. Mata, and M. C. Viana, "On filled-pauses and prolongations in European Portuguese," in *Proceedings of Interspeech 2007*. 27–31 August 2007, Antwerp, Belgium, 2007, pp. 2645–2648.

[15] J. Di Napoli, "Filled pauses and prolongations in Roman Italian task-oriented dialogue," in *Proceedings of the Laughter and Other Non-Verbal Vocalisations Workshop*. 5 October 2020, Bielefeld, Germany, 2020, pp. 24–27.

[16] L. Schettino, S. Betz, F. Cutugno, and P. Wagner, "Hesitations and individual variability in Italian tourist guides' speech," in *Speaker Individuality in Phonetics and Speech Sciences: Speech Technology and Forensic Applications, STUDI AISV 8*. Milano: Officinaventuno, 2021, vol. 8, pp. 243–262.

[17] A. Giannini, "Hesitation phenomena in spontaneous Italian," in *Proceedings of ICPhS 2003, The 15th International Congress of Phonetic Sciences*, M. J. Solé, D. Recasens, and J. Romero, Eds., 2003, pp. 2653–2656.

[18] G. Berruto and M. S. Cerruti, *La linguistica. Un corso introduttivo*. Torino: UTET De Agostini, 2011.

[19] R. Savy and F. Cutugno, "Diatopic, diamesic and diaphasic variations in spoken Italian," in *Proceedings of CL2009, The 5th Corpus Linguistics Conference*, M. Mahlberg, V. González-Díaz, and C. Smith, Eds. 20–23 July 2009, Liverpool, UK, 2009, pp. 20–23.

[20] A. Origlia, R. Savy, I. Poggi, F. Cutugno, I. Alfano, F. D'Errico, L. Vincze, and V. Cataldo, "An audiovisual corpus of guided tours in cultural sites: Data collection protocols in the CHROME project," in *Proceedings of the 2018 AVI-CH Workshop on Advanced Visual Interfaces for Cultural Heritage*, vol. 2091, 2018, pp. 1–4.

[21] L. Schettino, "The role of disfluencies in Italian discourse. Modelling and speech synthesis applications." Ph.D. dissertation, Università degli Studi di Salerno, 2022.

[22] S. Betz, "Hesitations in spoken dialogue systems," Ph.D. dissertation, Universität Bielefeld, 2020.

[23] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, pp. 159–174, 1977.

[24] H. Sloetjes and P. Wittenburg, "Annotation by category-elan and iso dcr," in *Proceedings of LREC 2008, The 6th international Conference on Language Resources and Evaluation*, 2008, pp. 816–820.

[25] P. Boersma and D. Weenink, *Praat: doing phonetics by computer [Computer program]*, 2021. [Online]. Available: http://www.praat. org/

[26] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, 2021. [Online]. Available: https://www.R-project.org/

[27] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[28] M. Nespor, *Fonologia*. Bologna: Il Mulino, 1993.

[29] A. Ledgeway, "The dialects of southern Italy," in *The Oxford guide to the Romance languages*, A. Ledgeway and M. Maiden, Eds. Oxford: Oxford University Press, 2016, vol. 1, pp. 245–269.

[30] P. Maturi, *I suoni delle lingue, i suoni dell'italiano: nuova introduzione alla fonetica*. Bologna: il Mulino, 2014.

[31] O. Niebuhr and K. Fischer, "Do not hesitate! – Unless you do it shortly or nasally: How the phonetics of filled pauses determine their subjective frequency and perceived speaker performance," in *Proceedings of Interspeech 2019*. 15–19 September 2019, Graz, Austria, 2019, pp. 544–548.