Proceedings of the 10th International Seminar on Speech Production (ISSP)

5 – 8 May 2014
Cologne, Germany

www.issp2014.uni-koeln.de

Edited by

Susanne Fuchs, Martine Grice, Anne Hermes, Leonardo Lancia, Doris Mücke

The role and prosodic characteristics of hesitation lengthening and filled pauses in speech planning in Maltese

Sarah Agius, Flavia Chetcuti, Alexandra Vella

University of Malta

agius.sarah@gmail.com, fchetcuti@hotmail.com, alexandra.vella@um.edu.mt

Abstract

This paper examines two phenomena which occur in speech, rather than during periods of silence: hesitation lengthening of lexical elements, and filled pauses involving non-lexical elements. An interesting characteristic of hesitation lengthening of lexical elements is that there do not seem to be any constraints on what element within the segmental and syllabic stream gets lengthened: although there is a tendency for the final (often unstressed) syllable to be lengthened, it seems possible for lengthening to affect any element within the segmental stream. The two phenomena share a number of features, the most prominent of which are marked lengthening and level F0. More significantly, both phenomena have a clear turn-holding effect: change of speaker rarely occurs when hesitation lengthening takes place. What is also interesting is that filled pauses come in two forms, one of which exhibits some sort of heightening of the usual characteristics for this sort of element. The discourse function of turn-holding which occurs in all cases of hesitation lengthening on lexical elements occurs also in the case of filled pauses but only when these also involve lengthening.

Keywords: hesitation lengthening, filled pauses, prosodic characteristics, turn-taking and turn-holding

1. Introduction

Spontaneous speech, unlike read speech, is marked by a variety of features, collectively often referred to as "normal disfluencies" (Shriberg 1994). Dialogue data is no different to spontaneous speech data of other sorts. Speakers in a dialogue need to keep the information flow going. They use a variety of strategies to keep the discourse moving forward and in doing so, may need to negotiate the floor. As part of this negotiation, it may sometimes be necessary for the current speaker to signal to her/his interlocutor that s/he is still thinking about what, and how best, to say what s/he has to say, and that therefore s/he is not yet ready to relinquish the floor.

These "talk-in-interaction" type requirements have some interesting effects on elements which have been observed to be found in dialogue data. Analysis of new spontaneous (but significantly non-Map Task) data from Maltese has brought to light a feature which appears to occur relatively frequently in these data. This involves lengthening which might not otherwise occur at a "natural" prosodic (or other type of) boundary and/or whose role is not one involving either simple demarcation of intonation phrases or straightforward hesitation. Such lengthening has been noted to be a distinct phenomenon which plays a role in planning similar to that noted to be involved in the use of both filled pauses (FPs) and, to a lesser extent, also of unfilled pauses in different languages (e.g. Cutler and Pearson 1986, Grosz and Hirschberg 1992, Swerts 1998, Clark and Fox Tree 2002, Campione and Véronis 2005), Maltese included (Vella et al., 2011 and 2014).

This study examines the role, as well as the phonetic, particularly prosodic, characteristics of hesitation lengthening, in two each of the Map and Conversation Task data available as part of the *MalToBI* corpus (Vella and Farrugia 2006) of spoken Maltese. It does so in the first instance by comparing hesitation lengthening to a phenomenon we believe to be closely related to it, that involving FPs. The study seeks to establish whether the functional and prosodic characteristics associated with these two – on the surface distinct – phenomena, are in fact one and the same thing.

The questions we examine are: What are the functional and prosodic characteristics of hesitation lengthening? And are the defining characteristics of hesitation lengthening the same or different to those of FPs?

2. Silence and speech

The stream of speech is broken up by silent intervals of different sorts. Silent intervals, or unfilled pauses, can perform a variety of functions in speech. Couper-Kuhlen (1986:75) provides a succinct summary of this variety of functions as resulting from "a performance-related origin – a pause for breath, a pause to search for a word or to plan". As hinted at earlier, the "pause to search for a word or to plan" is rendered that much trickier in conversation. If the interlocutor does not recognize the silence for what it is, there is the risk that s/he, mistakenly assuming that the floor is being relinquished, takes the floor. So what strategies are available to speakers engaged in conversation to minimize the chances of a breakdown in appropriate turn-taking?

To start with, an attempt is made to map out some of the distinctions we make in this study, in the light of the literature on this topic. Figure 1 below provides a schematisation.

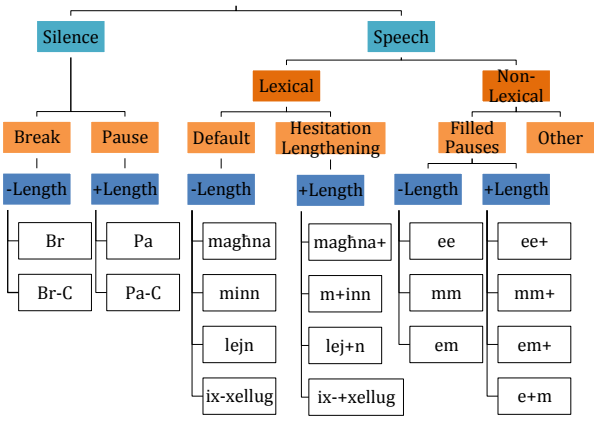

Figure 1: *Schematisation of distinctions*

One distinction which is often made is that between unfilled and filled pauses (Cruttenden 1997). However, such a distinction is far too simplistic. There are a number of ways in which speakers can slow things down, and these different

means can be used separately, or they can interact in interesting ways. The details of this interaction will be the subject of the analysis of data carried out and reported in Sections 4 and 5. Each distinction shown in the top three layers of the schematization is described in Figure 1 above.

Silence, when it occurs in speech often involves unfilled pauses. Pauses or breaks can occur for purely articulatory reasons such as when a speaker stops to breathe. They can also occur for junctural reasons involving a speaker organising text into phonological phrases of different types. One side-effect of unfilled pauses is that the stretch of speech immediately preceding such a pause is subject to the phenomenon often referred to as pre-boundary lengthening regardless of whether hesitation is involved. Unfilled pauses can however also arise in contexts of hesitation.

Speech can contain ***lexical*** as well as ***non-lexical*** elements. ***Lexical*** elements, although not normally associated with the phenomena of interest here (labelled as "default" in Figure 1) are interesting in the context of this study in their manifestation as elements involving "hesitation lengthening" (see also Figure 1).

Amongst the ***non-lexical*** elements of interest to us are the so-called FPs, examples of which include [e] in Scottish English and [n] in Russian (Cruttenden 1997:174) and at least [e], [m], [em] and [ʔ] in Maltese (Vella et al. 2011 and personal communication). Other ***non-lexical*** elements also occur in speech. These include backchannels such as *eħe* and *mhm* (orthographic *uh* and *um* for English), and vocalisations such as *ttt* (orthographic *tsk* in English).

3. Data and methodology

The data analyzed forms part of the *MalToBI* corpus consisting of 8 Map Tasks and 8 Conversation Tasks (see Vella and Farrugia 2006).

3.1 The Tasks

The Map Task involves quasi-naturalistic conversational data. The dialogue generated is oriented towards task completion, something which is not necessarily a prerequisite of spontaneous speech. The Maltese Map Task was designed with a view to collecting data involving the use of specific target items, items having different syllable structures and accent placement (final, penultimate or antepenultimate) but composed of all-sonorant material to better facilitate analysis in terms of F0.

The Conversation Task by contrast provided speakers with a specific scenario which they had to use as the basis for the conversation with their interlocutor. One of the speakers was asked to pretend that s/he knew someone who was seeking to fill a vacant post in the company s/he worked for. The task involved this speaker talking to the other speaker and trying to gather information which could be relayed on, regarding the suitability or otherwise of this person as a possible candidate for the job. Speakers changed roles when they felt they had no more to say.

3.2 Material used, and adaptations to annotation

Two of the orthographically annotated files from each task were used in this study. All data had been annotated using the guidelines and conventions developed in the course of the projects *SPAN* and *ISMA post-SPAN* (Vella et al. 2010). The annotation of each file was checked and revised by a second transcriber. The analysis of the Conversation Tasks brought to light a feature which was relatively less present, and which had somehow not been noted when the annotation of the Map Tasks first took place. A specific annotation marker (+) was introduced within the orthographic tier. The *SPAN* guidelines were updated accordingly. The (+) is placed after the lengthened element (initially the lengthened element was assumed to be the syllable, but it was later observed to be possible for practically any segment within an element to be lengthened).

4. Hesitation lengthening

Our working definition of hesitation lengthening is the lengthening of segmental material in a lexical element for reasons other than those normally associated with a prosodic boundary of some sort. In what follows we look more closely at hesitation lengthening of lexical material which has undergone what transcribers intuitively marked as hesitation lengthening. Any differences in the occurrence of this phenomenon as a function of task type will not be considered in the analysis which follows.

4.1 Hesitation lengthening and preceding or following silence

The data analysed shows that hesitation lengthening can occur with silence preceding and/or following it, as well as with speech segments on both sides of it. Moreover, this phenomenon rarely occurs with silence on both sides of it.

There was silence on both sides of a lengthened element in only 16% of all the instances of lengthening in the data analysed. Lengthened elements mid-speech, i.e. with speech on both sides of the lengthened element, account for 27% of the instances analyed. Silence occurred on either the left or the right of a lengthened element in 57% of the instances analysed, although silence to the right of this phenomenon is more frequent, as compared to silence to the left.

4.2 Hesitation lengthening and change of speaker

Of the instances of hesitation lengthening in the data analysed, only 10% were followed by a change of speaker. In the overwhelming majority of cases, 90%, the speaker employing the lengthening continued speaking after the particular instance of lengthening. This clearly suggests that this type of lengthening serves as a cue to the listener that the speaker intends to hold her/his turn. The use of hesitation lengthening appears to be a clear signal to the interlocutor that the speaker needs to buy time for some difficult retrieval process or simply to process and formulate her/his thoughts.

4.3 Hesitation lengthening and F0

The data analysed shows that the lengthening phenomenon observed is linked to a level F0 which stays steady in a sort of "hold" for a while. The F0 before and after the hold seems to flow, in a way, as though there had been no interruption.

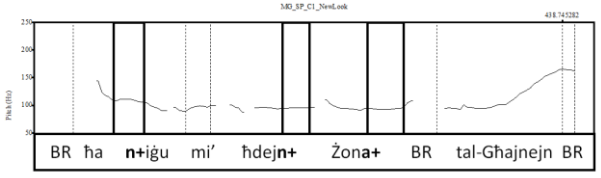

Figure 2: *Excerpt (436.5 to 438.9s) from MG_SP_C1*

Figure 2 above shows three instances of lengthening, one in which lengthening occurs at the beginning of a word and two in which lengthening occurs at the end of a word. The former

case, of lengthening on *n+iġu* 'we come', occurs in the middle of a stretch of speech with no surrounding silence and illustrates how F0 levels out on the segment which is being lengthened (in this case, the initial [n]). In the latter two instances, where lengthening occurs at the end of the word, one (*ħdejn+* 'near') occurs in a stretch of speech and one (*Żona+* 'zone/area') occurs preceding a break. What is interesting is that, in the case of *ħdejn+* 'near', where there is no surrounding silence on either side, F0 is level not only on the segment which is lengthened, but throughout the word, suggesting that here, it is this level F0 on its own that is serving as a turn-holding cue. This is different to what happens in the case of *Żona+* where the lengthening is followed by a break and where the pitch contour flattens out towards the end of the word and not throughout the word. Furthermore, the last of these instances in particular also illustrates how F0 is reset after each instance of lengthening, i.e. at the beginning of *Żona+*.

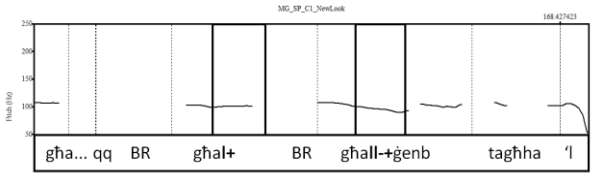

Figure 3: *Excerpt (166.5 to 168.4s) from MG_SP_C1*

Figure 3 illustrates two examples of lengthening of the lexical item *għal* 'towards', namely *għal+* 'towards' and *għall-ġenb* 'towards the side' (the apparent difference is purely orthographic), which both have silence in their surroundings. The first example has silence on both sides and interestingly F0 is level throughout. This parallels what we saw in Figure 2 above, where lengthening on a word, albeit one surrounded by speech rather than by silence, (in the above-mentioned case, *ħdejn+*), had level F0 throughout and not simply on the lengthened segment. The second instance of lengthening illustrated in Figure 3 also shows level F0 on the lengthened segment and a clear reset of the pitch by the speaker to continue speaking. These examples thus show that this lengthening does not disturb the flow of speech at all and seems to serve a turn-holding function.

4.4 Hesitation lengthening and location in the segmental stretch

The phenomenon in question has been found to occur at a variety of locations in the segmental stretch, both on "complete" stretches as well as on incomplete, abandoned ones of different types such as *għam+...* and *m+....* The lengthening of stretches of speech without disfluencies in them involves instances of lengthening at the beginning of a word, in the middle of a word, as well as at the end of the word. It is to be noted that "word" refers to some kind of phonological word. Examples of this include the following: *L+ejn[e]*, *Lej+n* and *lejn+* which one particular speaker uses in a specific conversation task analysed, together with other non-lengthened instances of the same lexical item, *lejn* 'towards'. This example illustrates perfectly the flexibility of this phenomenon, occurring as it does at all possible locations of a particular lexical item. Moreover, the phenomenon does not seem to be constrained by the phonological form of the segment in question. Thus, for example, in the case of word-initial lengthening, the analysis showed instances of lengthening at the beginning of a word where the word started with a sequence of a vowel, or a liquid, or a nasal, but also in stretches starting with obstruents or obstruent sequences. One example is lengthening in *S+qaq* 'alley'. Other examples are the lengthening at the beginning of *f+ilgħodu* 'morning' and *h+otel* 'hotel'.

In the case of lengthening in the middle of a word, once again, the phenomenon seems to occur irrespective of syllabic length or phonological constraints. Thus there were instances of lengthening in the middle of monosyllabic words, like the above-mentioned example of *lej+n* 'towards' and *xogħo+l* 'work'. Other examples included lengthening of the initial, penultimate and ultimate syllables in bisyllabic and polysyllabic words, such as *għall-+**ġenb*** 'towards the side', *mill+-**Baj**ja* 'from the bay', *il-+le**min*** 'the right-hand side', *l-Im**na+rj**a* – name of a specific Maltese feast and */**In**stitu+te/* 'institute'. Although the phenomenon seems to occur more frequently on non-stressed syllables, it also can occurs on stressed syllables as illustrated by the above-mentioned example of *l-Im**na+rj**a* – name of a specific Maltese feast. (The stressed syllable in the preceding examples is shown in **bold**.)

To sum up, word-final, post-stress location seems to be the preferred location for this phenomenon to occur, at least in the data analysed. It is clear however that hesitation lengthening in this position is not constrained by the phonetic nature of the lengthened element itself or by the nature of the segmental material in its immediate surroundings. A few examples of the many instances of lengthening noted in this context are the following: *lejn+* 'towards', ***ċir**ku+* 'circle', *nib**dew+*** 'we start', *il-**Ma**ra+* 'the wife' and *ċ-ċertifi**ka**ti+* 'the certificates'.

5. Filled pauses revisited

Vella et al. (2011) carried out an analysis of all the FPs in the Map Task data. Although no basis for a clear durational distinction, for example, for forms labelled as **m**, **mm** and **mmm** was found, it was nevertheless observed that there were cases where FPs themselves appeared to be "enhanced" by means of something which could well be hesitation lengthening. Comparison of the phenomenon of hesitation lengthening with the hesitation lengthening of FPs seems worth carrying out at this point.

Vella et al. (2011) have shown that FPs, by their very nature, vary in their duration. Therefore, in order to account for the observed phenomenon of lengthening while keeping in mind this inherent variability in duration, a decision was taken to mark FPs, which seemed to be proportionately longer than the "default" equivalent filled pause, using the same annotation marker (+) of lexical elements used in the Speaker tiers.

Analysis of these FPs showed that a distinction between lengthened and non-lengthened FPs is actually made by speakers, with lengthened FPs having an average duration of 0.50 seconds and non-lengthened ones having an average of 0.22 seconds in the data analysed. Furthermore, not only can one make this distinction for the FPs analysed in the data, but analysis showed that a small majority of them (67 out of a total of 121, or 55% of all FPs analysed) are actually lengthened.

Figure 4 below illustrates the difference in duration of lengthened (the first *ee*) as compared to non-lengthened (the second *ee*) FPs.

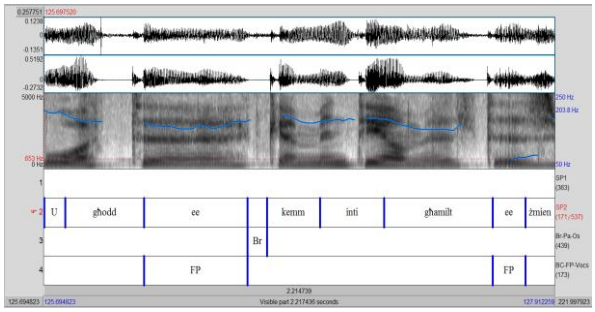

Figure 4: *Excerpt (166.5 to 168.4s) from MG_SP_C1*

5.1 FPs and preceding or following silence

Both lengthened and non-lengthened FPs seem to occur with silence in their surroundings (Vella et al. 2011). However, it seems that lengthened FPs tend to occur with silence on both sides, not just preceding or following them. This might be due to the nature of the lengthening in itself, which seems to favour contexts of following silence (cf. Section 4.1 above). Out of the lengthened FPs, only 7% are found in a stretch of speech with no silence in their vicinity. However, in the case of non-lengthened FPs, a much higher proportion, 22%, are found with no break or pause in their vicinity. This strengthens the argument that non-lengthened FPs, unlike their lengthened counterparts, do not serve the function of turn-holding.

5.2 FPs and discourse function

FPs serve a turn-taking as well as a turn-holding function. This is illustrated in Figure 5 below, which includes two instances of non-lengthened FPs. As we can see from this Figure, Speaker 2 uses the first FP, *ee* (which is 0.20 seconds long) to take the turn from Speaker 1. It is clear here that the speaker, although taking the turn, is still planning her utterance, and the use of the FP actually shows this. The same speaker then uses the same filled pause *ee*, a second time, with a shorter duration (0.10 seconds) to show that she is still planning her utterance.

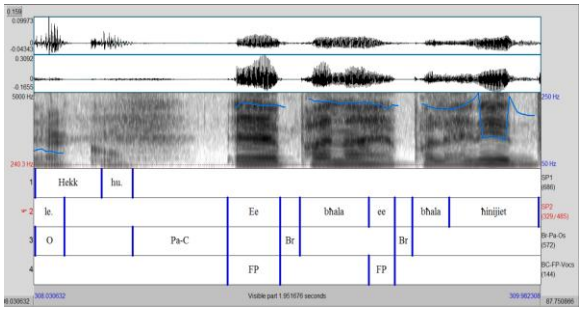

Figure 5: *Excerpt (308.03 to 309.98s) from AS_IV_C4*

However, in this case this second filled pause is not serving to take the turn and is not lengthened. What is happening here is that instead of lengthening the FP (or any other element), the speaker repeats the lexical word *bħala* 'like', to serve as a turn-holder.

The analysis shows that in the majority of cases, non-lengthened FPs are found in the vicinity of lengthened FPs or lengthened lexical items, or else are used together with other turn-holding devices such as repetition. Figure 5 illustrates an example where this is the case. It therefore seems that FPs on their own do not serve a turn-holding function.

6. Conclusion

The data we report on reveal that hesitation lengthening and (both types of) FPs are different in some respects. First, the former tends to affect the final, rather than the stressed syllable, suggesting that this type of lengthening is an edge phenomenon of some sort; this is clearly not the case for FP because of their very structure (monosyllabic elements). Secondly, hesitation lengthening can, but needn't be followed, and is not usually preceded, by pause; FPs, by contrast, usually occur with a pause to their left or to their right or on both sides (Vella et al. 2011).

Nevertheless, hesitation lengthening and (lengthened) FPs are similar in that they enable the speaker to pause without relinquishing the floor, and both are possible at a wide range of locations, not always ones at which one would expect a boundary to be placed. All of these phenomena are characterised by level F0. In the case of FPs this is always present. In the case of hesitation lengthening on lexical items, this is always present on the lengthened element and could also stretch out over the surrounding elements.

Whereas lengthened FPs can be treated as having the same function as hesitation lengthening, non-lengthened FPs might serve a different function, and do not necessarily serve to hold the turn in the same way as hesitation lengthening does. This distinction between long FPs and short FPs however needs to be further analysed in order for any conclusions to be reached.

Given that it has been noted that FPs, but more especially, hesitation lengthening, occur more frequently in the Conversation Task-type data, it would be interesting to investigate this issue further.

7. Acknowledgements

We would like to thank the University of Malta for funding *SPAN* and *ISMA post-SPAN*.

8. References

Campione, E., and J. Véronis (2005). "Pauses and hesitations in French spontaneous speech". In: *Proceedings of DiSS 2005, Disfluency in Spontaneous Speech Workshop*, pp. 43–46.

Clark, H., and J. E. Fox Tree (2002). "Using uh and um in spontaneous speech". In: *Cognition* 84, pp. 73–111.

Couper-Kuhlen, Elizabeth (1986). *An introduction to English prosody*. London: Edward Arnold, 239 p.

Cruttenden, Alan (1997). *Intonation*. 2nd edition. Cambridge: Cambridge University Press, 220 p.

Cutler, A., and M. Pearson (1986). "On the analysis of prosodic turn-taking cues". In: Johns-Lewis, Catherine (Ed.), *Intonation in discourse* (pp. 139–155). London: Croom Helm, 302 p.

Grosz, B., and J. Hirschberg (1992). "Some intentional characteristics of discourse structure". In: *Proceedings of International Conference on Spoken Language Processing*, pp. 429–432.

Shriberg, Elizabeth Ellen (1994). *Preliminaries to a theory of speech disfluencies*. PhD thesis. University of California, Berkeley, CA.

Swerts, M. (1998). "Filled pauses as markers of discourse structure". In: *Journal of Pragmatics* 30, pp. 485–496.

Vella, A., and P.-J. Farrugia (2006). "MalToBI – building an annotated corpus of Spoken Maltese". In: *Proceedings of Speech Prosody 2006*.

Vella, A., F. Chetcuti, S. Grech, and M. Spagnol (2010). "Integrating annotated spoken Maltese data into corpora of written Maltese." In: *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pp. 83-90.

Vella, A., M. Spagnol, S. Grech, and F. Chetcuti (2011). "A preliminary investigation of filled pauses in Maltese". In: Caruana, Sandro, Ray Fabri and Thomas Stolz (Eds.), *Variation and change: the dynamics of Maltese in space, time and society* (pp. 255–275). Bremen: Akademie Verlag, 326 p.

Vella, A., M. Spagnol, S. Grech, and F. Chetcuti (2014). "The role of silence in spoken Maltese: durational and distributional characteristics of breaks and pauses". In: Caruana, Sandro, Ray Fabri and Thomas Stolz (Eds.), *Perspectives on Maltese inguistics* (pp. 89–113). Bremen: Akademie Verlag, 354 p.