Fillers, filled pauses and co-occurring gestures in Danish First Encounters

Costanza Navarretta

University of Copenhagen/ Njalsgade 140, 2300 Copenhagen S
Denmark
costanza@hum.ku.dk

1 Introduction

Filled pauses, which are pauses accompanied by so-called fillers, are very frequent in spoken language. Fillers have multiple non-exclusive functions which are both related to the management of communication (Allwood et al., 1992; Maclay and Osgood, 1959; Duncan and Fiske, 1977) and cognitive processes of planning the discourse and retrieving words (Rochester, 1973; Krauss et al., 2000). Researchers have found that there is an inverse frequency relation between hand gestures and filled pauses (Christenfeld et al., 1991; Rauscher et al., 1996) and that many hold gestures co-occur with filled pauses (Esposito et al., 2001; McNeill, 2014).

Fillers are an integral part of the language and have language specific characteristics (de Leeuw, 2007). Clark and FoxTree (2002) propose to consider fillers as words since they are used in different contexts.

In this study we want to determine a) which are the most common fillers in a Danish corpus of first encounters, b) whether fillers co-occur with gestures and their uses, c) whether the most frequent fillers in Danish have conventionalized uses as in English.

2 Related studies

The use of fillers has been related to both communicative management functions and to discourse planning functions. The first group of functions comprise feedback, (Allwood et al., 1992; Allwood, 2001), and to the regulation of turn exchange (Maclay and Osgood, 1959; Duncan and Fiske, 1977; Clark and Tree, 2002). The turn keeping function of fillers is partly related to their use as signals of speech planning processes. For example, Rochester (1973) finds that filled pauses are more frequent when speakers face an option or have to express something challenging, while

Reynolds and Paivio (1968) report that students used more pauses and filled pauses when they had to define abstract objects than when they described concrete objects. Filled pauses can also mark the process of lexical retrieval (Krauss et al., 2000) and the frequency of filled pauses and gestures have been found to be inverse proportional (Christenfeld et al., 1991; Rauscher et al., 1996), while Esposito et al. (2001) find that hand gestures co-occurring with filled pause (*uh*, *hum*, *ah* or *and*) are often augmented holds that is holds in which a movement of the hand is noticed.

Language specific studies of fillers have been focusing on their type and their position in utterances. Shriberg (1994) reports that vocal-nasal fillers are more frequent in the initial position of utterances in American English while vocal fillers occur most frequently when speakers have to find specific lexical items. Clark and FoxTree (2002) propose to consider the English *uh* and *um* as words since speakers use them in a conventionalized way. More specifically, they find that *uhs* signal minor delays while *ums* signal major delays. Similarly, Tottie (2014) argues that *uh* and *um* can be used as discourse markers with a meaning similar to that of *well* and *you know*.

De Leeuw (2007) analyzes the realization of fillers in Dutch, English and German and determines language specific characteristics. She finds that vocal-nasal fillers are predominant in English and German while vocal fillers are most common in Dutch. The effect of filled pauses on the listener's memory has also been proved (Fraundorf and Watson, 2011) and filled pauses have been included in the behavior of conversational software agents (Cassell et al., 1994; Traum and Rickel, 2002; Pfeifer and Bickmore, 2009).

3 The data

Our data are twelve multimodal annotated Danish first encounters which were collected and anno-

tated under the NOMCO and VKK projects (Paggio et al., 2010; Navarretta et al., 2012). The NOMCO project's main aims were to create and analyze annotated comparable Nordic multimodal corpora.

Six females and six males, aged 21-36, all native Danish speakers were involved in two encounters each. They talked freely while being audio and video recorded by three cameras. The annotations of the corpus include speech token transcriptions and shape and function descriptions of communicative co-speech gestures. These are connected to the speech tokens with which they were found to be semantically related. The gestures annotated are head movements, facial expressions and body postures. The gestural functions considered in this study are feedback, self-feedback and turn management features. More detailed description of the annotations are in (Paggio and Navarretta, 2011; Navarretta and Paggio, 2013).

For the present study, we have identified all the fillers and filled pauses in the corpus and extracted the co-occurring gestures with a perl script. These and the speech tokens following the fillers and the filled pauses have been taken into account in the following analysis. Table 1 shows the Danish fillers, their occurrences in the first encounters, their multimodal occurrences and the percentage of multimodal occurrences. The most common

Filler	Nr	Multim	%
øh	411	308	75
mm	106	89	84
øhm	91	70	77
årh	9	8	89
åh	9	9	100
hm/ehm	8	4	50
Total	634	488	77

Table 1: Filler types and co-occurring gestures

fillers in the Danish encounters are the vocal *øh*, the nasal *mm* and the vocal-nasal *øhm*. In over two-thirds of the cases, they co-occur with gestures. In the following we focus on the uses of these three fillers.

Most of the occurrences (76%) of the vocal *øh* are connected to speech planning in these data. In the large majority of the cases, *øh* occurs inside utterances as signal that the speaker is searching a word (it precedes an adjective, noun or verb). It also occurs before self repairs. In 42% of its

occurrences, *øh* co-occurs with gestures having a turn management function. This confirms preceding studies that indicate that speakers signal with their body that they want to keep the turn while searching for a word or planning their discourse, i.a. (Kendon, 2004), or that they wish to give the turn if they have difficulties in completing the discourse (Clark and Tree, 2002). *Øh* is only related to feedback gestures in 15% of its occurrences, and it co-occurs with self-feedback gestures (especially smiles) in 30 % of the cases.

The vocal-nasal *øhm* precedes in the majority of cases (78% of the occurrences) utterances and phrases, but it can also occur in the middle of a phrase. *Øhm* co-occurs with turn-management gestures in 42% of its uses, and it is most frequently related to feedback gestures (62% of the occurrences), and to a lesser extent (30% of the occurrences) to self-feedback.

Finally, the filler *mm* co-occurs often (66% of its occurrences) with feedback gestures, hereunder especially head movements, and more rarely to self-feedback and turn management gestures (11% and 16% of the occurrences respectively). In few cases, 6% of the occurrences, *mm* occurs inside a phrase as signal of lexical retrieval. A first analysis of the data indicates that holds in gestures often occur when fillers are related to lexical retrieval and discourse planning. There are no holds when fillers are related to feedback giving and self-feedback.

4 Discussion

Even though all fillers in the Danish data occur as signals in communication management and/or discourse planning contexts as it was the case for fillers in other languages (de Leeuw, 2007; Clark and Tree, 2002), each filler has some more prototypical uses. More specifically, the nasal filler *mm* and the vocal-nasal filler *øhm* are mostly used for backchannelling, while the vocal filler *øh* accompanied by a pause is more often connected to speech planning and turn management. As discourse marker *øhm* often precedes sentence boundaries, while *øh* often occurs inside phrases. This indicates that the Danish *øh* and *øhm* are used similarly to the corresponding English *uh* and *um* (Clark and Tree, 2002; Tottie, 2014) and that they and the filler *mm* also have conventionalized uses. With respect to gestures co-occurring with fillers and filled pauses, our data show, not surprisingly,

that their form depends on their function. The multimodal aspect of fillers and filled pauses will be adressed in more details in the final version of the paper.

References

Jens Allwood, Joakim Nivre, and Elisabeth Ahlsén. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9:1–26.

Jens Allwood. 2001. Dialog Coding - Function and Grammar: Gteborg Coding Schemas. *Gothenburg Papers in Theoretical Linguistics, University of Gteborg, Dept of Linguistics*, 85:1–67.

Justine Cassell, Catherine Pelachaud, Norman Badler, Mark Steedman, Brett Achorn, Tripp Becket, Brett Douville, Scott Prevost, and Matthew Stone. 1994. Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 413–420. ACM.

Nicholas Christenfeld, Stanley Schachter, and Frances Bilous. 1991. Filled pauses and gestures: It's not coincidence. *Journal of Psycholinguistic Research*, 20(1):1–10.

Hebert H. Clark and Jean E. Fox Tree. 2002. Using uh and um in spontaneous speaking. *Cognition*, 84:73–11.

Esther de Leeuw. 2007. Hesitation markers in english, german, and dutch. *Journal of Germanic Linguistics*, 19:85–114, 6.

S. Duncan and D.W. Fiske. 1977. *Face-to-face interaction*. Erlbaum, Hillsdale, NJ.

Anna Esposito, Karl Erik McCullough, and Frank Quek. 2001. Disfluencies in gesture: gestural correlates to filled and unfilled speech pauses. In *Proceedings of IEEE International Workshop on Cues in Communication*, Hawai.

S.H. Fraundorf and D.G. Watson. 2011. The disfluent discourse: Effects of filled pauses on recall. *Journal of memory and language*, 65(2):161–175.

Adam Kendon. 2004. *Gesture - Visible Action as Utterance* . Cambridge University Press.

R.M Krauss, Y. Chen, and R. F. Gottesman. 2000. Lexical gestures and lexical access: a process model. In D. McNeill, editor, *Language and gesture*, pages 261–283. Cambridge University Press.

Howard Maclay and Charles E. Osgood. 1959. Hesitation phenomena in spontaneous English speech. *Word*, 15:19–44.

David McNeill. 2014. *The Conceptual Basis of Language*. Routledge Library.

C. Navarretta and P. Paggio. 2013. Classifying Multimodal Turn Management in Danish Dyadic First Encounters. In *Proceedings of the 19th Nordic Conference of Computational Linguistics (Nodalida 2013)*, pages 133–146, Oslo, Norway, May. NEALT.

Costanza Navarretta, E. Ahlsén, J. Allwood, K. Jokinen, and Patrizia Paggio. 2012. Feedback in nordic first-encounters: a comparative study. In *Proceedings of LREC 2012*, pages 2494–2499, Istanbul Turkey, May.

Patrizia Paggio and Costanza Navarretta. 2011. Head movements, facial expressions and feedback in danish first encounters interactions: A culture-specific analysis. In Constantine Stephanidis, editor, *Universal Access in Human-Computer Interaction- Users Diversity. 6th International Conference. UAHCI 2011, Held as Part of HCI International 2011*, number 6766 in LNCS, pages 583–690, Orlando Florida. Springer Verlag.

Patrizia Paggio, E. Ahlsén, J. Allwood, K. Jokinen, and Costanza Navarretta. 2010. The NOMCO multimodal Nordic resource - goals and characteristics. In *Proceedings of LREC 2010*, pages 2968–2973, Malta, May 17-23.

LauraM. Pfeifer and Timothy Bickmore. 2009. Should agents speak like, um, humans? the use of conversational fillers by virtual agents. In Z. Ruttkay, M. Kipp, A. Nijholt, and H.H. Vilhjálmsson, editors, *Intelligent Virtual Agents*, volume 5773 of *Lecture Notes in Computer Science*, pages 460–466. Springer Berlin Heidelberg.

F.H. Rauscher, R.M. Krauss, and Y. Chen. 1996. Gesture, speech and lexical access: The role of lexical movements in speech production. *Psychological Science*, 7:226–231.

Allan Reynolds and Allan Paivio. 1968. Cognitive and emotional determinants of speech. *Canadian Journal of Psychology*, 22:164–175.

Sherry R. Rochester. 1973. The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research*, 2:51–81.

Elisabeth Shriberg. 1994. *Preliminaries to a theory of speech disfluencies*. Ph.D. thesis, University of California, Berkeley.

Gunnel Tottie. 2014. Uh and um in British and American English: Are they words? Evidence from co-occurrence with pauses. In Nathalie Dion, André Lapierre, and Rena Torres Cacoullos, editors, *Linguistic Variation: Confronting Fact and Theory*, pages 38–54. New York: Routledge.

David Traum and Jeff Rickel. 2002. Embodied agents for multi-party dialogue in immersive virtual worlds. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 2*, AAMAS '02, pages 766–773, New York, NY, USA. ACM.