The functions of fillers, filled pauses and co-occurring gestures in Danish dyadic conversations

Costanza Navarretta

University of Copenhagen
costanza@hum.ku.dk

Abstract

Fillers, alone or accompanied by pauses and/or gestures, are quite frequent in all types of spoken communication. They have numerous and non-exclusive functions which are related to interaction management (feedback and turn management) or discourse planning. Fillers are part of the language and thus, to some extent, language dependent. This article presents an analysis of fillers, filled pauses and co-occurring gestures in a Danish multimodal corpus of first encounters. The aims of the study are to determine the most common fillers in the corpus, the gestures co-occurring with them, their functions, and possibly their most prototypical uses. The results of our study indicate that the most common fillers in the data are *øh*, *mm*, *øhm* which all are accompanied by one or more gestures in most of their occurrences. We also found that each filler type has a predominant or prototypical use. *Mm* often occurs alone as feedback marker and is accompanied by feedback gestures. *Øhm* has the longest duration and often precedes an utterance or a clausal phrase signaling discourse planning. Its co-speech gestures have also interaction management functions. Finally, *øh* often precedes a content word, has a shorter duration than *øhm* and signals lexical retrieval. Interestingly the prototypical uses of the vocal *øh* and the vocal-nasal *øhm* are the same as those of the English vocal *uh* and vocal-nasal *um*, respectively.

Index Terms: multimodal communication, gestures, filled pause

1. Introduction

Face-to-face communication is multimodal including at least auditory (speech) and visual (gestures) modalities. The modalities are not only temporally, but also semantically related at many levels. This papers is about a particular phenomenon of face-to-face communication, the so-called fillers, such as the English *uh* and *um*. Fillers are very frequent in spoken language and can occur alone or in conjunction with a speech pause (filled pauses). The language accounted for is Danish. The paper also addresses the gestures which co-occur with the Danish fillers and their functions. The gestures included in this study are head movements, facial expressions, body postures and hand movements.

Fillers have multiple, non-exclusive functions which are related to interaction management [1, 2, 3] and cognitive processes of discourse planning and word retrieval [4, 5]. Researchers have noticed that there is an inverse frequency relation between hand gestures and filled pauses [6, 7] and that many hold gestures co-occur with filled pauses [8, 9].

Fillers are an integral part of the language and have therefore language specific characteristics [10]. Clark and FoxTree [11] find that different English fillers are used in different contexts and therefore they suggest to consider them as words.

Because fillers and filled pauses are frequent in spoken language, it is important to exploit their use and functions as well as their relation to gestures in order to include them in spoken language models which reflect the type of conversation and the language. The present study wants to contribute to these models by determining a) which are the most common fillers in a Danish corpus of first encounters, b) whether fillers co-occur with gestures and with which functions, and c) whether the most frequent fillers in Danish have conventionalized uses as in English and what these uses are.

In section 2, we discuss relevant related studies, then in section 3 we shortly describe the data and the methodology used for studying Danish fillers and filled pauses. In section 4, we present the analysis of fillers and filled pauses in the Danish corpus and , in section 5 we discuss the data. Finally, in section 6, we conclude and suggest future work.

2. Related studies

The functions of fillers in spoken language have been related to both interaction management and discourse planning. The various functions are not mutually exclusive and they are often related. Interaction management comprises feedback, that is feedback giving, also known as backchanneling, feedback eliciting [1, 12], and turn exchange regulation [2, 3, 11]. Turn exchange regulation comprises inter alia turn keeping and turn giving signals. Turn exchange signals mark also discourse planning processes. For example, a speaker can move her head away from the interlocutor signaling at the same time that she is planning her discourse and wants to keep the turn or the speaker can signal with a filled pause and gestures that she wants to give the floor if she has difficulties in completing the discourse.

Rochester [4] finds that filled pauses are more frequent when speakers face an option or have to express something challenging, while Reynolds and Paivio [13] report that students used pauses and filled pauses much more frequently when they had to define abstract objects than when they described concrete objects. Filled pauses can also mark the process of lexical retrieval [5] and researchers have noticed that the frequency of filled pauses is inverse proportional to the frequency of gestures [6, 7]. Esposito et al. [8] find that hand gestures co-occurring with English filled pauses involving the fillers *uh*, *um* and *ah* are often augmented holds, that is holds in which a little movement of the hand is noticed. They interpret the function of these holds as parallel to that of the speech pauses with which they co-occur. The speaker signals with the filled pause that she is planning new spoken content and marks with gestural holds that she is planning new gestures.

Language specific studies of fillers have focused on their type and their position in the utterance. For example, English

researchers have focused on the uses of the two fillers *uh* and *um*. More specifically, Shriberg [14] 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 [11] propose to consider the English *uh* and *um* as words since speakers use them in a conventionalized way. The two researchers find that although the two fillers have many common uses, that is occur when speakers are looking for a word, planning the discourse, wanting to keep or give the floor, they have also a preferred or prototypical use. *Uhs* signal minor delays while *ums* signal major delays. Finally, Tottie [15] argues that *uh* and *um* can be used as discourse markers with a meaning similar to that of *well* and *you know*.

Swerts [16] analyzes the occurrences of filled pauses as markers of discourse boundaries in Dutch monologues, while De Leeuw [10] analyzes the realization of fillers in Dutch, English and German in order to determine their language specific characteristics. She finds that vocal-nasal fillers are predominant in English and German while vocal fillers are most common in Dutch. Vocalic-nasal fillers are only dominant in Dutch when they are surrounded by long pauses. English fillers are often preceded by a pause and followed by a lexical item, in De Leeuw's data, while in German and Dutch they are often surrounded by lexical items.

Possible effects of filled pauses on the listeners have also been investigated. For example, Fraundorf and Watson [17] prove that filled pauses have a positive effect on the listener's memory. Furthermore, different studies have determined that users perceive software agents to have more human-like behavior if they use fillers and therefore filled pauses have been included in the behavior of conversational software agents [18, 19, 20].

In a preceding study of pauses delimiting clause boundaries, and of the gestures which accompany them in the NOMCO corpus, we found that silent pauses and audible breath pauses are accompanied by head movements, facial expressions and body postures in 88% and 86% of their occurrences respectively, while filled pauses and pauses accompanied by other sounds are accompanied by the same gestures in only 77.5% and 70% of their occurrences respectively [21]. Furthermore, we found that the majority of clausal boundary pauses in the data were silent and breath pauses.

To our best knowledge, there are no previous general studies of fillers, filled pauses and the gestures which co-occur with them in Danish. However, it must be noted that the Danish fillers *hmm*, *øh* and *øhm* are included in a recent general language Danish lexicon *Den Danske Ordbog*[1]. In this lexicon, the three fillers are classified as interjections and are described as synonymous expressions of doubt. Furthermore, *øh* and *øhm* are analyzed as synonyms when used to fill in pauses while the speaker is thinking, and the filler *hmm* is defined as an interjection which expresses discontent, or a kind of disagreement or reservation with respect to the following word(s). In the following study of the functions of Danish fillers, filled pauses and co-occurring gestures, we will also investigate whether the lexicon definitions provided by the lexicon cover the uses of the fillers in the multimodal corpus of first encounters.

[1]Den Danske Ordbog is available on the internet at the address http://ordnet.dk/ddo/ordbog.

3. The data and method

The Danish NOMCO corpus consists of twelve multimodal annotated Danish first encounters which were collected and annotated under the Nordic NOMCO and the Danish VKK project. The NOMCO project's main aims were to create and analyze annotated comparable Nordic multimodal corpora, and first encounters were collected in more Nordic languages [22]. Furthermore, the conversations were annotated in all the corpora following a common theoretic framework [23], the so-called MUMIN annotation framework [24]. The Danish VKK project had the aim to analyze and model specific aspects of multimodal communication in Danish such as feedback and turn management [25, 26].

Six females and six males, aged 21-36 and native Danish speakers, were engaged in two encounters each, one with a female and one with a male. The participants talked freely about themselves, their studies and work while being audio and video recorded. Two microphones and three cameras were used and the encounters took place in a studio at the University of Copenhagen. Two snapshots from the data showing the three camera views are in Figure 1 and Figure 2.

Each encounter lasts between four and seven minutes, and the corpus has a duration of one hour. The annotations of the corpus comprise speech token transcription and shape and function descriptions of communicative co-speech gestures. In the speech transcriptions pauses are annotated as tokens and are annotated as a plus sign +. Furthermore, filled pauses, breath and other audible sounds accompanying pauses are also annotated.

The annotations of gestures are connected to speech tokens produced by either participant if the annotators found them to be semantically related. The gestures annotated are head movements, facial expressions and body postures [25]. For this study, we have added shape annotations of hand gestures co-occurring with fillers. The gestural functions considered in this study are feedback, self-feedback and turn management.

Table 1 shows the shape features of the gestures which are relevant to the present research while the function features of the gestures are in Table 2. The features describing the shape of gestures are coarse grained and only the most general shape features are used in this study. It must also be noted that information about gestural phases is not available.

Table 1: Shape features

Attribute	**Value**
HeadMovement	Nod, Jerk, HeadForward, HeadBackward, Tilt, SideTurn, Shake, Waggle, HeadOther, None
General face	Smile, Laugh, Scowl, FaceOther, None
BodyDirection	BodyForward, BodyBackward, BodyUp, BodyDown, BodySide, BodyTurn, BodyDireOther, None
Handedness	SingleHand, BothHands

The first two function features in Table 2 are related to feedback. The values of the attribute *FeedbackBasic* are assigned if feedback expresses Contact, Perception and Understanding (CPU) and if feedback only shows Contact or Contact and Perception but no Understanding (FeedbackOther) [1]. A positive feedback attribute is accompanied with the values of the *FeedbackDirection* attribute indicating whether feedback is given or

Figure 1: Two frontal snapshots from the corpus

Table 2: Function features

Attribute	Value
FeedbackBasic	CPU, FeedbackOther, None
FeedbackDirection	FeedbackGive, FeedbackElicit, None
TurnManagement	TurnTake,TurnHold, TurnAccept, TurnElicit, None

elicited.

The third function attribute, *TurnManagement* describes turn related behavior. The following four turn management values are relevant to the present study: a) TurnTake is assigned if the speaker signals that she wants to take a turn that wasn't offered; b) TurnHold: the speaker signals that she wishes to keep the turn; c) TurnAccept: the speaker signals that she is accepting a turn that is being offered; d) TurnElicit: the speaker signals that she is offering the turn to the interlocutor [26].

Inter-coder agreement were run on the data and resulted in Cohen's kappa scores in between 0.6 and 0.9 depending on the attributes. The transcriptions and annotations were made by one annotator, corrected by a second annotator and, in case of disagreement between the two main annotators, a third expert annotator took the final decision. We have used the final version of the data in this study. A more detailed description of the annotation procedure is in [25].

For the present study, we have identified all the fillers and filled pauses in the NOMCO corpus and we have extracted the co-occurring gestures with a perl script. Co-occurring gesture are defined as those gestures which temporally overlap with fillers or filled pauses. No limitation to the extension of the overlap were given. we have then extracted the duration of the fillers, and manually analyzed the context in which they occur, that is the speech tokens which precede and follow the fillers as well as the gestures which co-occur with them.

4. Analysis

There are 18,556 speech tokens (words, fillers and pauses) in the Danish first encounters, while there are 3,117 head movements, 1,448 facial expressions, 982 body postures and 566 hand gestures. The fillers in the Danish corpus are *øh*, *øhm*, *mm*, *årh*, *åh*, *hm/ehm*. Their frequency is in Table 3. Thus the most common filler is *øh*, *øhm* and *mm*.

Table 3: Filler types and their frequency

Filler	Occurrences
øh	375
mm/hmm	109
øhm	84
årh	9
åh	9
ehm	1
Total	587

Table 4 shows the fillers, their occurrences, their multimodal occurrences and the percentage of multimodal occurrences of fillers.

Table 4: Filler types and co-occurring gestures

Filler	Occurrences	Multimodal	%
øh	411	308	75
mm	113	92	81
øhm	91	70	77
årh	9	8	89
åh	9	9	100
ehm	1	1	100
Total	634	488	77

The number of the occurrences of the fillers in table 4 is higher than that in table 4 because when gestures are added to the speech tokens, some speech tokens are doubled. This is for example the case if two head movements co-occur with the same filler as indicated in figure 3.

Slightly over two-thirds of the occurrences of the fillers co-occur with gestures, and the number is the same as that of gestures co-occurring with filled paused [21]. In the rest of the study, we focus on the three most common fillers, that is the

Figure 2: A total view snapshot from the corpus

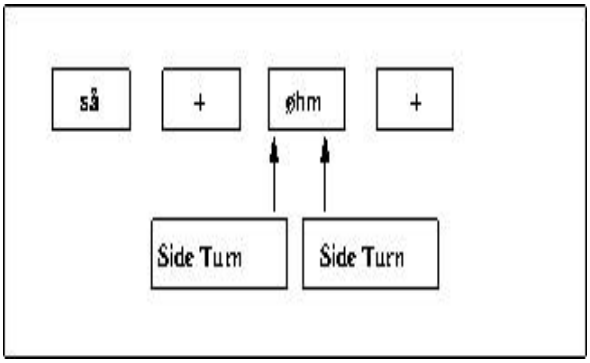

Figure 3: Two head movements co-occur with a filler

vocal *øh*, the nasal *mm* and the vocal-nasal *øhm*, on theirs uses and the types of gestures which co-occur with them.

Table 5 shows the percentage of *øhs*, *mms* and *øhms* which co-occur with head movements, facial expressions, body postures and hand gestures. In the table, it is not accounted for the fact that more gesture types can co-occur with the same filler occurrence.

Table 5: Filler types and co-occurring gestures

Filler	Occurrences	Head	Face	Body	Hand
øh	411	50%	18%	26%	11%
mm	113	68%	28%	19%	2%
øhm	91	55%	38%	23%	8%

The three most frequent fillers co-occur in most cases with head movements which are the most common body behavior. The filler *mm* co-occurs most frequently with head movements (half of its co-occurrences) while in 1/3 of the cases it co-occurs with facial expressions. The filler *øh* co-occurs frequently with head movements and body postures (50% and 26% respectively) and more seldom with facial expressions and hand gestures (18% and 11% of the occurrences respectively), while *øhm* co-occurs frequently with head movements, facial expressions and body postures (55%, 38% and 23% of the occurrences) and, less frequently that is in 8% of the occurrences with hand gestures.

In Table 6 are the mean and standard deviation of use of the three fillers. Table 6 shows that standard deviation for the use of

Table 6: Mean and standard deviation of fillers' use

Filler	Mean	Stand.dev.
mm	9.17	12.02
øh	31.64	27.84
øhm	7.75	3.49
all	48.42	25.2

fillers is quite large especially for the two fillers *øh* and *mm*, and some participants used more fillers than others. For analyzing the spoken contexts in which the fillers occur in the first encounters we have distinguished the following context categories:

- The filler occurs inside a phrase preceding a content word (adjective, adverb, noun, or verb). In this context the filler is often accompanied by a pause and signal lexical retrieval, e.g. A: *helt* + *øh* + *ubehøvlet* (completely + uh + boorish)
- The filler precedes a phrase or a feedback word such as *okay*, *ja* (yes) or *no*, e.g. + + *øhm jeg har også musiklinjen fra seminariet* (++ um, I have also music from the teacher school).In this context the filler is always accom-

panied by a longer pause and signals that the speaker is planning the discourse.

- The filler occurs in self-repairs and false starts. Also in these cases the filler is often accompanied by pauses, e.g. *læst + øh + hvor jeg læser*(studied + uh + where I study...)
- The filler (often a filled pause) occurs at the end of a turn, that is before the interlocutor takes the floor, E.g A: *det var sådan lidt +øh + (breath)* (it was such a little + uh + (breath), B: *(smack) + okay det junne du + det kan du godt lide* ((smack) + okay you did + you like it).
- The filler occurs alone as feedback marker or co-occurs with laughter, e.g. A: *er det på dansk?* (is it at Danish?) B: *mm* (yes).

In Table 7, we show the average duration in milliseconds of the three most frequently occurring fillers and of eventual pauses surrounding them. The percentage of occurrences of each filler in each spoken context is also given.

72% of the occurrences of the vocal filler *øh* are connected to speech planning in these data. More specifically in 50% of its occurrences, *øh* precedes a content word signaling lexical retrieval, while in 32% of the occurrences the filler precedes a phrase or a feedback word. *Øh* also occurs in self repairs (7% of the occurrences) and it precedes turn ends (9% of the cases), while only in 2% of its occurrences *øh* is used alone as feedback marker.

The functions of the gestures co-occurring with *øh* corresponds not surprisingly to the function of the filler indicated by the spoken context. In fact, the filler often co-occurs with gestures having a turn keeping function and, in fewer cases, a turn giving/yielding function. This finding confirms preceding studies that indicate that speakers signal with their body behavior that they want to keep the turn while searching for a word or while planning an utterance, i.a. [5, 27], or that speakers wish to give the floor if they have difficulties in completing the discourse [11]. *Øh* is only related to feedback gestures in 15% of its occurrences, and it co-occurs with self-feedback gestures, predominantly facial expressions, in 30 % of the cases.

The vocal-nasal *øhm* often precedes phrases or feedback words (70% of its occurrences). The majority of the phrases preceded by *øhm* are clauses and the filler often follows the conjunctions *og* (and) and *så* (so, therefore). *Øhm* occurs in the middle of a phrase in only 10% of the occurrences and in self repairs in 11% of the cases. Finally, it occurs at the end of a turn in 8% of its occurrences. The gestures that co-occur with *øhm*s are often feedback head movements (62% of the cases) and/or turn-management gestures (42% of the cases). In 30% of the occurrences gestures co-occurring with the filler *øhm* are facial expressions having a self-feedback gesture (own communication management).

The nasal filler *mm* occurs in most cases alone (65% of the occurrences) as feedback signal. It precedes a phrase or another feedback word in 35% of the occurrences. Similarly to *øhm*, in these cases it often follows the conjunctions *og* (and) or *så* (so, therefore) and precedes a clausal phrase.

Not surprisingly, the gestures which co-occur with *mm* are often related to feedback giving (backchanneling) (71% of the cases) and they are often nods. More rarely self-feedback and turn management gestures co-occur with *mm* (11% and 16% of the cases respectively).

A first analysis of half of the data indicates that holds in gestures often occur when fillers are related to lexical retrieval and discourse planning. There are no gestural holds when fillers are related to feedback giving and self-feedback.

5. Discussion

The main function of the filler *mm* is that of signaling feedback, alone or in connection with other feedback words. More seldom *mm* marks the start of a phrase. As feedback marker *mm* is often accompanied by feedback head movements, especially nods. The fact that this filler is mostly used as a feedback marker is also reflected by its average duration, which is shorter than that of the other two fillers. The use of the filler as feedback mark is in line with its analysis in the Danish dictionary *Den Danske Ordbog*, but in these data the filler is only connected to disagreement in four cases. In the majority of occurrences it simply signals feedback giving. This might be due to the type of interaction. Furthermore, different meanings of e.g. feedback words can only be identified when the multimodal context is available (audio and video), and the effect of intonation and gestures on the interpretation of the semantics of feedback words in Danish data has been proved previously [28].

The vocal-nasal filler *øhm* frequently occurs with pauses and precedes clauses or even utterances and therefore signals discourse planning processes. *Øhm* also occurs at the end of speech turns signaling that it has been interpreted by the interlocutor as a turn giving signal. *Øhm* also occurs with a certain frequency inside a phrase marking lexical retrieval (10% of the occurrences) or in self repairs (11% of occurrences), while it only seldom occurs alone as feedback marker (1% of the occurrences). *Øhm* lasts longer than the other two fillers (0.48 milliseconds in average) and this finding is in line with what has been noted about English data: English filled pauses that mark larger syntactic units have a longer duration that filled pauses that signal lexical retrieval inter alia [11, 15]. It is not strange that the gestures which co-occur with *øhm* have feedback and turn management functions since feedback and turn management signals often occur at utterance or clausal boundaries.

The most common filler *øh*, which has an average duration between that of the two other fillers (0.4 milliseconds) most often precedes a content word, signaling lexical retrieval and, less frequently, precedes a phrase in these data. Other uses of *øh* mark self repairs or turn end. Only seldom *øh* is used as feedback marker (2% of its occurrences). The most common function of the gestures co-occurring with *øh* is that of self-feedback. This is not surprising since speakers often produce self-feedback gestures in self repairs or at the end of their spoken contributions.

Even though all fillers in the Danish data occur as signals in interaction management and/or discourse planning contexts as it was the case for fillers in other languages [10, 11], each filler has some more common or prototypical uses, as it also is the case for English fillers [11]. Our data indicates that the use of fillers variates from one participant to the other, and some participants use more frequently one or two fillers. The standard deviation was particularly high for the two fillers *øh* and *mm*, while it was lower for the filler *øhm*. The reason for this variation should be investigated in the future.

Thus, the analysis of the first encounters confirms overall the uses of the three fillers described in the Danish lexicon, but it also indicates that even though *øh* and *øhm* can occur in the same contexts and can be used as synonyms, they have different preferred/prototypical uses. Moreover, the three fillers have even more uses than those described in the Danish lexicon. Interestingly, our data indicate that the prototypical uses of the

Table 7: Fillers, duration and spoken contexts

Filler	Duration	Inside phrase	Before phrase	Self repair	End turn	Alone
øh	0.4	50%	32%	7%	9%	2%
øhm	0.48	10%	70%	11%	8%	1%
mm	0.3	0	35%	0	0	65%

two Danish fillers *øh* and *øhm* are the same as those of the English fillers *uh* and *um* [11, 15] that is *øh* and *uh* often precede a content word and signal lexical retrieval while *øhm* and *um* precede a clausal phrase or an utterance signaling planning of a larger discourse part and having a function similar to discourse markers [15]. This is also in line with studies on silent pauses. For example Tøndering [29] finds that silent pauses preceding subordinated phrases are shorter than those between independent phrases in a Danish spoken corpus, the DANPASS corpus.

The fact that gesture holds were only found when the gestures co-occurred with filled pauses confirms the study by Esposito et al. [8] which found that augmented hand gestural holds co-occurred with speech pauses.

6. Conclusion and future work

In the paper, we presented a study of fillers, filled pauses and co-occurring gestures in the Danish NOMCO corpus of first encounters. In these data, the majority of the fillers are accompanied by gestures and the gestures reinforce the filler's function.

The Danish fillers have the same functions as fillers in other languages, that is they have functions related to feedback and/or turn management, or they signal discourse planning processes, hereunder lexical retrieval. The various functions are not mutually exclusive.

The analysis of the Danish data shows that each filler type has a predominant use, even though the most common fillers are often used synonymously, that is they can also occur in the same contexts. This finding confirms for Danish what has been also found to be the case for fillers in other German languages and especially in English.

More specifically, we found that the nasal filler *mm* is often used as a feedback giving marker and it is nearly always accompanied by feedback head movements. Moreover, in this corpus, it mostly indicates positive feedback and co-occurs with nods. This is not surprising since in first encounters participants are kind and try to give the interlocutor a positive impression [30].

The vocal *øh* often signals lexical retrieval, but it is also used in other contexts. The gestures which accompany this filler have mainly turn management functions or signal self-feedback. *Øh* has the same prototypical use as the English vocal filler *uh*.

The vocal-nasal filler *øhm* often precedes a clausal phrase and marks discourse planning. It is often accompanied by gestures having an interaction management function. The prototypical function of *øhm*is the same as that of the English nasal-vocal filler *um*. As in English, filled pauses occurring at the boundaries of larger discourse units have longer duration than filled pauses preceding lexical entries.

In the future, we will analyze fillers in more types of spoken data, including monologues and dialogs involving more than two participants as well as in conversations between interlocutors who know each other in advance. Individual differences in the use of fillers and filled pauses should also be investigated and the uses and occurrences of fillers and filled pauses in the Danish first encounters should be compared with the uses of fillers in the other Nordic first encounters corpora.

7. Acknowledgments

I would like to thank my colleague Patrizia Paggio, the Danish NOMCO and VKK annotators Anette Luff Studsgård, Sara Andersen, and Bjørn Wessel-Tolvig. Special thanks also go to the NOMCO project's Nordic partners, Elisabeth Ahlsén, Jens Allwood and Kristiina Jokinen.

8. References

[1] J. Allwood, J. Nivre, and E. Ahlsén, "On the semantics and pragmatics of linguistic feedback," *Journal of Semantics*, vol. 9, pp. 1–26, 1992.

[2] H. Maclay and C. E. Osgood, "Hesitation phenomena in spontaneous English speech," *Word*, vol. 15, pp. 19–44, 1959.

[3] S. Duncan and D. Fiske, *Face-to-face interaction*. Hillsdale, NJ: Erlbaum, 1977.

[4] S. R. Rochester, "The significance of pauses in spontaneous speech," *Journal of Psycholinguistic Research*, vol. 2, pp. 51–81, 1973.

[5] R. Krauss, Y. Chen, and R. F. Gottesman, "Lexical gestures and lexical access: a process model," in *Language and gesture*, D. McNeill, Ed. Cambridge University Press, 2000, pp. 261–283.

[6] N. Christenfeld, S. Schachter, and F. Bilous, "Filled pauses and gestures: It's not coincidence," *Journal of Psycholinguistic Research*, vol. 20, no. 1, pp. 1–10, 1991.

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

[8] A. Esposito, K. E. McCullough, and F. Quek, "Disfluencies in gesture: gestural correlates to filled and unfilled speech pauses," in *Proceedings of IEEE International Workshop on Cues in Communication*, Hawai, 2001.

[9] D. McNeill, *The Conceptual Basis of Language*. Routledge Library Editions: Linguistics, 2014.

[10] E. de Leeuw, "Hesitation Markers in English, German, and Dutch," *Journal of Germanic Linguistics*, vol. 19, pp. 85–114, 6 2007.

[11] H. H. Clark and J. E. F. Tree, "Using uh and um in spontaneous speaking," *Cognition*, vol. 84, pp. 73–11, 2002.

[12] J. Allwood, "Dialog Coding - Function and Grammar: Gteborg Coding Schemas," *Gothenburg Papers in Theoretical Linguistics, University of Gteborg, Dept of Linguistics*, vol. 85, pp. 1–67, 2001.

[13] A. Reynolds and A. Paivio, "Cognitive and emotional determinants of speech," *Canadian Journal of Psychology*, vol. 22, pp. 164–175, 1968.

[14] E. Shriberg, "Preliminaries to a theory of speech disfluencies," Ph.D. dissertation, University of California, Berkeley, 1994.

[15] G. Tottie, "Uh and um in British and American English: Are they words? Evidence from co-occurrence with pauses," in *Linguistic Variation: Confronting Fact and Theory*, N. Dion, A. Lapierre, and R. T. Cacoullos, Eds. New York: Routledge, 2014, pp. 38–54.

[16] M. Swerts, "Filled pauses as markers of discourse structure," *Journal of Pragmatics*, vol. 30, pp. 485–496, 1998.

[17] S. Fraundorf and D. Watson, "The disfluent discourse: Effects of filled pauses on recall," *Journal of memory and language*, vol. 65, no. 2, pp. 161–175, 2011.

[18] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone, "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*. ACM, 1994, pp. 413–420.

[19] D. Traum and J. Rickel, "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*, ser. AAMAS '02. New York, NY, USA: ACM, 2002, pp. 766–773.

[20] L. Pfeifer and T. Bickmore, "Should Agents Speak Like, um, Humans? The Use of Conversational Fillers by Virtual Agents," in *Intelligent Virtual Agents*, ser. Lecture Notes in Computer Science, Z. Ruttkay, M. Kipp, A. Nijholt, and H. Vilhjálmsson, Eds. Springer Berlin Heidelberg, 2009, vol. 5773, pp. 460–466.

[21] C. Navarretta, "Pauses delimiting semantic boundaries," in *Proceedings of the 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2015)*, IEEE, Ed., Giör, Hungary, October 2015, pp. 533–538.

[22] P. Paggio, E. Ahlsén, J. Allwood, K. Jokinen, and C. Navarretta, "The NOMCO multimodal Nordic resource - goals and characteristics," in *Proceedings of LREC 2010*, Malta, May 17-23 2010, pp. 2968–2973.

[23] C. Navarretta, E. Ahlsén, J. Allwood, K. Jokinen, and P. Paggio, "Feedback in Nordic First-Encounters: a Comparative Study," in *Proceedings of LREC 2012*, Istanbul Turkey, May 2012, pp. 2494–2499.

[24] J. Allwood, L. Cerrato, K. Jokinen, C. Navarretta, and P. Paggio, "The MUMIN Coding Scheme for the Annotation of Feedback, Turn Management and Sequencing," *Multimodal Corpora for Modelling Human Multimodal Behaviour. Special Issue of the International Journal of Language Resources and Evaluation*, vol. 41, no. 3–4, pp. 273–287, 2007.

[25] P. Paggio and C. Navarretta, "Head Movements, Facial Expressions and Feedback in Danish First Encounters Interactions: A Culture-Specific Analysis," in *Universal Access in Human-Computer Interaction- Users Diversity. 6th International Conference. UAHCI 2011, Held as Part of HCI International 2011*, ser. LNCS, C. Stephanidis, Ed., no. 6766. Orlando Florida: Springer Verlag, 2011, pp. 583–690.

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

[27] A. Kendon, *Gesture - Visible Action as Utterance* . New York: Cambridge University Press, 2004.

[28] C. Navarretta and P. Paggio, "Classification of feedback expressions in multimodal data," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*, Upssala, Sweden, Juli 2010, pp. 318–324.

[29] J. Tøndering, "Prosodiske fraser og syntaktisk struktur i spontan tale," *NyS - Nydanske Sprogstudier*, vol. 39, pp. 166–198, 2010.

[30] C. Navarretta, "Annotating and analyzing emotions in a corpus of first encounters," in *Proceedings of the 3rd IEEE International Conference on Cognitive Infocommunications*, IEEE, Ed., Kosice, Slovakia, December 2012, pp. 433–438.