Ricerche di Pedagogia e Didattica – Journal of Theories and Research in Education 9, 1 (2014). Special Issue. *Communicating certainty and uncertainty: Multidisciplinary perspectives on epistemicity in everyday life.* Edited by Andrzej Zuczkowski and Letizia Caronia

Audiovisual prosody of uncertainty: An overview

Charlotte Wollermann
University of Duisburg-Essen

Bernhard Schröder
University of Duisburg-Essen

Ulrich Schade Fraunhofer
Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE; University of Bonn

Abstract
The goal of this work is to give an overview of research conducted in the field of audiovisual prosody and uncertainty. We refer to previous studies investigating the relevance of prosodic cues for both the production and perception of uncertainty in natural conversation. Afterwards, we present our own experiments dealing with the role of prosodic cues of uncertainty and also of context for pragmatic focus interpretation. In a next step we discuss which role uncertainty plays for human-machine communication. In this context we summarise our own studies on the modelling and perception of uncertainty in speech from an articulatory speech synthesizer. We assume that speech synthesis systems with highly variable speech can help to improve human-machine communication.

Keywords: uncertainty, audiovisual prosody, context, focus interpretation, exhaustivity, speech synthesis

Introduction

Speakers and hearers use prosodic cues for signalling and detecting uncertainty in conversation. We discuss the role of audiovisual prosody of uncertainty in natural conversation and also in human-machine communication. In the context of natural conversation we present our own studies on the role of prosodic cues of uncertainty for pragmatic focus interpretation which serve as starting point for our further studies on audiovisual prosody and focus marking. With respect to human-machine communication we summarise our work on the modelling and perceiving of uncertainty in an articulatory speech synthesizer.

In this section we expose the theoretical background of our work and define the concepts of *uncertainty*, of *focus*, and of the *exhaustive interpretation of answers*. Finally, we are bringing these three concepts together and present our assumption.

Uncertainty

Suppose a communicative situation with two conversational partners, A and B. A is asking a question to B and B is not certain with respect to her answer. This uncertainty might be due to the fact that B only partially knows the answer. Uncertainty in general can be regarded as a complex phenomenon and different definitions are found in the literature. In some works uncertainty refers to a non-prototypical emotive state (Rozin, Cohen, 1993; Keltner, Shiota, 2003), in other works it refers to a cognitive state (Kuhltau, 1993) or it refers to both (Givens, 2009). However, according to Oh (2008: 8), the expression and perception of uncertainty is essential in conversation. He remarks the following:

> [...] people need to be able to express when they are uncertain of the information being put forth by their conversation partner. Expressing and perceiving uncertainty is thus an essential part of communication. (Oh, 2006: 8)

Regarding question-answering situations, the following questions arise: Which prosodic cues do speakers use when they express uncertainty in answers? And which prosodic cues are relevant for hearers for decoding uncertainty?

Focus

The term *focus* is used in several areas of semantics and pragmatics. It incorporates many, partly diverging intuitions. It is thus difficult to provide a single definition.

Various focus phenomena can be found in natural language and also different terminology is used in the literature (for an overview see e.g., Krifka, 2007; for an overview and a "co-ordinate system" see Fisseni, 2011). Szabolcsi (1980: 513) remarks the following:

> […] there exists a vast number of notions (topic-comment, theme-rheme, background-focus etc.), which seem to stem from some common intuitive basis but whose actual contents tend to vary almost from author to author although each appears to be useful in explaining some interesting aspects of syntax, semantics, and pragmatics.

In the literature on information structure (e.g., Büring, 1997; Kadmon, 2001; Baumann, 2006), pitch accent is assumed to correlate with new information in utterances, while given information is often deaccented. This is — at least — assumed for West Germanic languages like English, German or Dutch. From the cross-linguistic point of view there is evidence that focus is realised differently, e.g., by syntax or morphology (cf. Büring, 2009: 177).[1] However, in the following we will mainly address the issue of different focus phenomena in English and German.

Most focus theories agree that focus can be defined as "[…] the answer to the question being addressed […]" (Kadmon, 2001: 261). Question-answer focus, also referred to as *pragmatic focus*, usually applies to the constituent in the answer which corresponds to the interrogative pronoun in the question. In example (1b) taken from Szabolcsi (1980: 526) *John* corresponds to the interrogative pronoun *who* in (1a). If John is marked by stress like in (1b), the answer to the question is adequate. Contrary, this is not the case when stress is found on *Mary* as in (1c).

(1) a. Who kissed Mary?
(1) b. [John]$_F$ kissed Mary.
(1) c. John kissed [Mary]$_F$.

Semantic focus is a term that refers to special arguments of *focus operators* like *only*, *even* and *also* in English or *nur*, *auch* and *sogar* in German. With stress on *Bill*, (2b) (example taken from Krifka (1991: 18)) can be uttered as an answer to the question in (2a). But if *Sue* is stressed like in (2c) the answer does not fit the question.

(2) a. Whom did John introduce to Sue?
(2) b. John only introduced [Bill]$_F$ to Sue.
(2) c. John only introduced Bill to [Sue]$_F$.

In the case of *contrastive focus* the meaning of a sentence is determined by alternatives to the proposition of a sentence. These alternatives are identical with the proposition except for the focused constituent (cf. Selkirk, 2007: 126).[2] Consider example (3) given by Selkirk (2007: 126), (3) could be a possible answer to the question *Did you give an invitation to Caitlin?*. With stress on Sarah, the speaker signals that there is a contrast between the question and the answer, it was not Caitlin who got the invitation, but Sarah. Here, the alternative set is {I gave one to Sarah, I gave one to Caitlin, I gave one to Stella...}.

(3) I gave one to [Sarah]$_F$, not to Caitlin.

Furthermore, *scalar implicatures* might be tied to focus. In theories on scalar implicature (e.g., Horn, 1972; Gazdar, 1979) it is assumed that sentences can be linked with expression scales, i.e. ordered sets of expressions (cf. van Rooij, Schulz 2004: 492). Rooth (1992: 82) gives the following example: After Mats took an exam in a self-paced calculus class, George asked him how it went.

(4) Well, I [passed]$_F$.

If George infers from (4) that Mats did not ace the examen, a *scalar implicature* exists. Thus, <fail,..., pass,..., ace> as hypothetical scale is assumed. George follows the Gricean (1975) *maxim of quantity*, he infers that Mats gives him no weaker information than available.

The described focus phenomena have in common that the scope of focus is *narrow*, i.e. one word in the sentence is the *focus exponent*. It might also be the case that the focus domain extends over a whole constituent or sentence. In this case, in which a *focus projection* occurs, the focus is *broad*. In example (5b), taken from Hanssen et al. (2008: 609), as answer to (5a), all information is new and the focus has scope over the whole sentence.

(5) a. What happened?
(5) b. [We went to London]$_F$.

The question how different focus types and different accent types relate to each other has been discussed in the literature, among others by Swerts and Krahmer (2001) and by Baumann et al. (2006).

Exhaustive interpretation of answers

In semantic-pragmatic theories (e.g., Groenendijk, Stokhof, 1984; Rooth, 1992) it is often assumed that focus is associated with a background question. If the later is interpreted as a *mention-all* question, the precondition for an *exhaustive* interpretation is given. In this context Groenendijk and Stockhof (1984: 276) remark the following:

> Sentences in isolation may carry focus on one or more of their constituents, and focus semantically results in an exhaustive interpretation of the focused constituent(s).

The following example of Szabolcsi (1980: 526) serves as illustration:

(6) a. Who kissed Mary?
(6) b. [John]$_F$ kissed Mary.

If the hearer concludes from (6b) that John is the only individual out of a number of persons in question who kissed Mary, (6b) is interpreted *exhaustively* with respect to the predicate in (6a). In the case of *non-exhaustive* interpretation there might be other individuals who kissed Mary as well.

An exhaustive interpretation depends on the knowledge about the situation in question which is ascribed to the speaker by the hearer. Suppose the following: "Quantity1-implicatures of a sentence *s* are, roughly speaking, sentences of the form ¬*s*' where *s*' is an alternative to *s* that is in some sense stronger than *s* itself" (van Rooij, Schulz, 2004: 494). The question arises if the hearer takes a *strong* or *weak* reading of the implicature being associated with the quantity-maxim of Grice (1975). The crucial factor is the following: Does the hearer assume that the speaker knows if *s*'? In the case of *epistemically strong* reading ¬*s*' is either indeed conversationally implicated or the hearer implicates that the speaker knows or believes ¬*s*'. Regarding an *epistemically weak* reading it is either the case that the hearer generates no implicature at all or that she does not know whether *s*' (cf. van Rooij, Schulz, 2004: 494f.).[3]

In semantic-pragmatic theories (e.g., Groenendijk, Stokhof, 1984; Rooth, 1992) it is often — at least implicitly — assumed that in the context of a question, the main accent in the answer is correlated with the focus induced by the question.

Thus, the accent should have an impact on exhaustive interpretation, especially in the context of a suitable question.

In order to empirically investigate such predictions, Fisseni (2011) conducted a series of interpretation experiments. In semantic-pragmatic theories (e.g., Groenendijk, Stokhof, 1984; Rooth, 1992) focus phenomena are usually discussed out of the blue, i.e. the discourse context is either not taken into account or only briefly sketched by considering the question. In contrast to that, Fisseni (2011) also considers the influence of the macro context in his studies. Results suggest that mere accent is not sufficient for affecting pragmatic focus interpretation, but the micro context and the macro context of the focus utterance are more important than suggested by semantic-pragmatic theories (e.g., Groenendijk, Stokhof, 1984; Rooth, 1992). Also, the expectations of the hearer and the sensitisation for focus phenomena are relevant for focus interpretation.

In the field of prosody and information structure, there is a lack of empirical evidence for the role of uncertainty in the interpretation of pragmatic focus. The analysis of Ward and Hirschberg (1985) showed for English that *fall-rise* intonation contributes to a context-independent meaning of utterance interpretation conveying speaker's uncertainty.[4] For German, it is less clear which intonation contour or prosodic cues are relevant for interpreting utterances in terms of uncertainty on the pragmatic level.

Assumption

We assume that if the speaker signals (un)certainty with respect to her answer, the hearer will use this prosodic information for decoding the utterance and will infer that the speaker is (un)certain with respect to her epistemic knowledge. In this case, the interpretation on the hearer's side should be biased towards a (non)exhaustive interpretation of the answer.

Uncertainty in natural conversation

In this section we present previous studies on the role of audiovisual prosody for the production and perception of uncertainty in natural conversation. Afterwards, we describe the empirical studies in which we tested our assumption.

Previous studies

Speakers and listeners use different cues in communication in order to signal and detect *uncertainty* in question-answering situations. The work of Smith and Clark (1993) serves as source of inspiration for many studies in this field. The authors investigated memory processes in question-answering situations. In order to test the hypothesis that speakers mark certainty differently from uncertainty, the *Feeling of Knowing* (FOK) paradigm according to Hart (1965) was used. With this method, it is possible to elicit metamemory judgements. Smith and Clark's experimental investigation brought to light that speakers express uncertainty for example on the lexical level by using phrases like "I guess", by means of prosodic cues like *rising intonation* and by *delay*.

In order to test how listeners perceive the FOK of a speaker, Brennan and Williams (1995) defined the *Feeling of Another's Knowing* (FOAK) paradigm. It was shown that the FOAK "[...] was affected by the intonation of answers, the form of nonanswers, the latency to response, and the presence of fillers" (Brennan, Williams, 1995: 396). The term *filler* is defined as interjections, e.g., "hmm", "um" and "uh" (cf. Brennan, Williams, 1995: 383).

Furthermore, *fillers* and *pauses* have been found as relevant cues with respect to self-repair in speech, especially to those self-repairs that do not contain lexical material (coined *c-repairs*) (Goldman-Eisler, 1967; Levelt, 1983). These repairs occur if the speaker recognises and corrects the slip of the tongue even before a speech signal is produced. A connectionist model of such a kind of repairs can be found in Schade and Eikmeyer (1991).

The studies mentioned focused on the influence of uncertainty on the acoustical signal, but the question remains open to what extent uncertainty affects the *audiovisual* modality. With respect to speech production, Swerts and Krahmer (2005) found that speakers use several cues for producing uncertain utterances. For the audio channel they reported that answers were characterized by *high intonation*, *delay*, and *fillers*. For the visual modality *eyebrow movements*, *gaze features*, *smiles* and *funny faces* were reported. In order to investigate the relevance of these cues for perception, audio-only, visual-only, and audiovisual stimuli had to be judged with respect to uncertainty. The experiment brought to light that subjects were able to distinguish certain from uncertain utterances for all three conditions, but the assignment was easier in the bimodal condition than in the unimodal conditions.

However, there is barely any empirical data and research with regard to the influence of uncertainty on *pragmatic focus interpretation* in natural speech. In the next subsection we will describe our interpretation studies investigating this topic.

Experimental studies on the role of uncertainty for focus interpretation

The starting point of our experimental studies was to test whether prosodic variation and also contextual variation influence the exhaustive interpretation of answers (Wollermann, Schröder, 2008a, see also Wollermann, 2012: 123ff.). In interpretation study I, the contextual variation referred to the micro context, i.e. the type of question which was preceding the focus utterance. In interpretation study II, (Wollermann, Schröder, 2008b; see also Wollermann, 2012: 135ff) the macro context was varied, i.e. the story with the embedded question-answering pair. Moreover, in the second study stronger cues of intended uncertainty were used. We will describe interpretation study I and II in some detail since there studies serve as source of inspiration for our further studies thereafter presented in this section.

Based on the results of interpretation study I and II we derived a model of focus interpretation (Wollermann et al., 2010; see also Wollermann, 2012: 161ff) which explains the role of prosody and of context for pragmatic focus interpretation. Since our experimental data did not provide evidence for clear influence of prosody on focus interpretation, we regarded it as necessary to conduct a production study to test which prosodic cues speaker use for focus marking.

Uncertainty, macro context and exhaustive interpretation of answers — interpretation study I

In a first study we experimentally investigated whether *intonation* as exclusive prosodic indicator of uncertainty affects exhaustive interpretation of answers and what role the *question-answering congruity* plays for exhaustivity (Wollermann, Schröder, 2008a; for a detailed description see Wollermann, 2012: 123ff.) .

For generating the audio stimuli, two speakers (one male and one female) were recorded. They were instructed to read six scripted dialogues. For three of six dialogues there is a question-answer pair in which the focus of the answer is one noun phrase (NP) (see (7b)). For the other three dialogues the focus constituent is a coordination of two noun phrases (see (8b)). The scenario is a fictitious student party in which different groups of students do different things. For every action, there is a question asking for the agent (see (7a)) and an answer providing the requested information. The subject of the answer denotes the respective group of students, which is also the focus of the answer (see (7b)). Focus accent was in general realised as L+H*, L*+H or H* in our data (for a description of the annotation scheme GToBI (German Tones and Break Indices) see e.g., Grice, Baumann, 2002).

(7) a. Wer hat um zehn Uhr getanzt? *Who danced at ten o'clock?*

(7) b. [Die Geografinnen]$_F$ haben um zehn Uhr getanzt. *[The geographers]*$_F$ *danced at ten o'clock.*

(8) a. Wer hat die Nachbarn durch lautes Lachen gestört? *Who disturbed the neighbours by laughing out loud?*

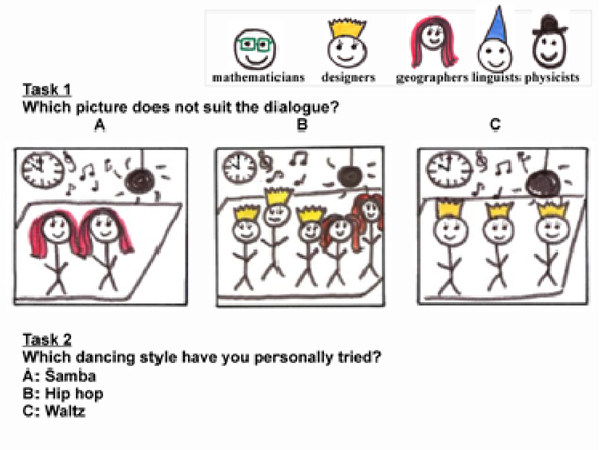
(8) b. [Die Mathematiker und Designerinnen]$_F$ haben die Nachbarn durch lautes Lachen gestört. *[The mathematicians and designers]*$_F$ *disturbed the neighbours by laughing out loud.*

For expressing *uncertainty* the speaker realising the focus utterance was instructed to produce a focus accent which was followed by a *rising* intonation (H-). The preceding question was either parallel to the answer or constituted a general question, e.g., "What happened?". It was hypothesised that *falling* intonation combined with a question *parallel* to the answer biased the interpretation towards *exhaustivity*, whereas *rising* intonation combined with *general* question biased the interpretation towards *non-exhaustivity*.

In order to test pragmatic focus interpretation and thus exhaustivity empirically, it is necessary to use an adequate method. Preference tests or other methods directly asking for perceptual judgements are not appropriate for testing focus interpretation since this aims at perception, but not at *interpretation* of the speech signal. Non-reactive methods of data collection, e.g., eye tracking, or measuring of reading times could also be used for measuring interpretation (cf. Fisseni, 2011: 69ff.).[5] Fisseni (2011: 71) remarks in this context that "[...] these measurements [...] are only meaningful if a sufficiently strong correspondence between cognitive activity and the measured data can be defined." However, in our approach we used pictures for testing focus interpretation. Hence, it should be avoided that the subjects' linguistic awareness is directed to the goal of the investigation.

In our study subjects had to choose between different pictures (see fig. 1, task 1). One picture represented the *exhaustive* reading (A), another one the *non-exhaustive* reading (B) and a third picture showed a scene functioning as distractor

(C). Also as a distractor, we asked questions about the subjects' personal opinion of an aspect of the dialogue and used three filler dialogues (task 2).



Task 1
Which picture does not suit the dialogue?

Task 2
Which dancing style have you personally tried?
A: Samba
B: Hip hop
C: Waltz

Figure 1. *Example for experimental task: measuring focus interpretation (task 1), distractor (task 2)*

Subjects were 71 students (24 males, 47 females) of the University of Bonn, all native speakers of German. They were tested in four groups, each time with a different kind of random order of the stimuli.

Results suggested that the *exhaustive* reading was considered as standard interpretation in that scenario. However, in four of six dialogues weak *prosodic* and *contextual* effects could be observed as expected.[6] Further, we concluded from the subjects' feedback that — against our intention — the purpose of the study was recognised easily by comparing the pictures showing *(non-)exhaustive* reading. This problem was addressed by an improved experimental design in interpretation study II which served as a follow-up study.[7]

Uncertainty, macro context and exhaustive interpretation of answers — interpretation study II

In interpretation study II, we investigated if the variation of the macro context of the focus utterance and also stronger prosodic indicators of intended uncertainty can bias the interpretation towards non-exhaustivity (Wollermann, Schröder, 2008b; for a detailed description see Wollermann, 2012: 135ff.).

Our audio stimuli consisted of the four dialogues with embedded question-answer pairs for which we could find prosodic and contextual effects in our interpretation study I. For conveying *(un)certainty*, we offered two different prosodic realisations of each answer:

I) For focus sentences with one noun phrase (NP), the NP and the sentence-final verb were both marked by *rising* intonation for expressing *uncertainty*, i.e. we had L+H* H- or L*+H H- as contour for the NP and H% as boundary tone. For focus sentences with two NPs, the contour was generally L*+H H- L*+H H- or L*+H L*+H H-. Further, the speech rate was significantly lower than in II). The intonation contour for example 7b is illustrated in figure 2.

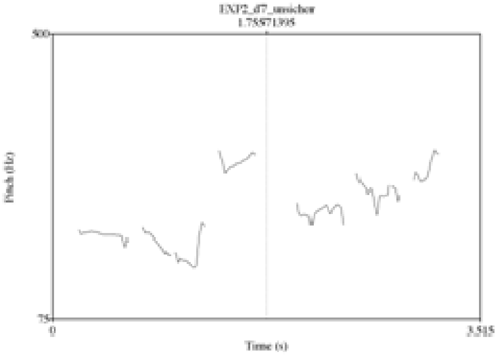


Figure 2. *Intonation contour intended to convey uncertainty for example 7b.*

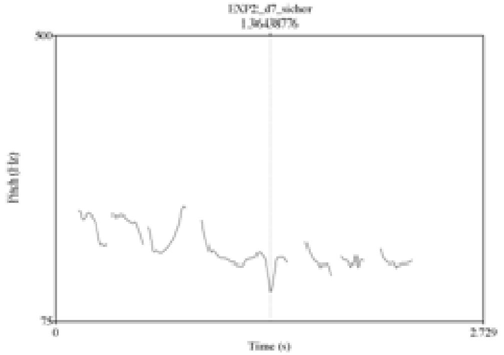


Figure 3. *Intonation contour intended to convey certainty for example 7b.*

II) Focus NP and sentence-final verb were realised by means of *falling* intonation for intended *certainty*. We have L% as boundary tone and no H- for realising the NP(s) after the accentuation. Figure 3 shows the intonation contour for example 7b.

Furthermore, we generated two kinds of linguistic contexts for each dialogue:

III) One student group usually carrying out the action under discussion was introduced at the beginning of the dialogue, e.g., the designers always dance (example 7). The following question was general and not congruent with the focus utterance. The group denotated by the focus of the answer did not overlap with the group introduced at the beginning of the dialogue.

IV) Only one student group was salient during the dialogue. It was only mentioned in the answer which constituted the focus utterance. The preceding question was congruent with the answer (see 7a/8a).

For each stimulus, we generated four different variants as shown in table 1.

Table 1. *The four combinations of factors.*

Variant	Description
A	Exhaustive context (IV) + falling intonation (II)
B	Exhaustive context (IV) + rising intonation & low speech rate (I)
C	Non-exhaustive context (III) + falling intonation (II)
D	Non-exhaustive context (III) + rising intonation & low speech rate (I)

We assumed that variant A biases the interpretation towards *exhaustivity*, whereas D biases the interpretation towards *non-exhaustivity*. Furthermore, we hypothesised that in the case of B and C, exhaustification is less strong than for A, but stronger than for D.

Contrary to our interpretation study I we used only one picture, either intended to illustrate the *exhaustive* reading or the *non-exhaustive* reading, for testing focus interpretation.[8] Subjects had to rate on a Likert-scale how well the picture suited the dialogue.

152 students (122 female, 30 male) from the University of Bonn and the University of Duisburg-Essen participated in the experiment, all of them native speakers of German and nobody of them had taken part in the previous study. They were tested in eight groups, each time with a different kind of random order of the stimuli.

Results showed again a strong preference for the exhaustive reading, i.e. the picture illustrating the exhaustive interpretation was in general judged as better suiting the dialogue than the picture illustrating the non-exhaustive reading. However, when we compared the judgments for all cases for which the picture showing the non-exhaustive reading was presented, the following could be observed for three of four dialogues:

In general, a context with the alternatives mentioned combined with an uncertain way of speaking biased the interpretation towards non-exhaustivity (variant D). In contrast, effects of a context with no alternatives mentioned combined with a certain way of speaking on non-exhaustivity were weaker in a significant way (variant A). Moreover, the influence of a context with the alternatives mentioned combined with a certain way of speaking (variant C) on non-exhaustivity was generally stronger than a context without the alternatives mentioned combined with an uncertain way of speaking (variant B). Overall, the effect of the context on the exhaustive interpretation of answers was more evident than the prosodic influence. This result was not in line with our theoretical assumption.

A model of focus interpretation

Based on the results of interpretation study I and II we derived a model which explains the role of accentuation, prosodic indicators of uncertainty and context for pragmatic focus interpretation. Our model is described in detail in Wollermann et al. (2010) and in Wollermann (2012: 161ff.). It is assumed to be part of a complete language processing model and it is inspired by Levelt's language production model (Levelt, 1989). In our model, the prosodic information is processed bottom-up and the expectations of the hearer are processed top-down. The expectations are for instance affected by contextual information. Our model makes the following prediction: The influence of prosody increases when the hearer has less clear expectations.

Audiovisual prosody and pragmatic focus marking

Since our empirical data did not show evidence for the prosodic influence on exhaustivity to the extent we expected theoretically, we also investigated the audiovisual marking of pragmatic focus in a production study (Wollermann, 2009; for a detailed description see also Wollermann, 2012: 176ff.). The material from the interpretation study I and II was used and subjects were instructed to read scripted dialogues with focus utterances. The context was varied with respect to (un)certainty and (non)exhaustivity. Results showed a tendency that non-exhaustivity was marked by a *peak accent* accompanied by a *raising of eyebrows or head.* We interpreted this as a manifestation of the *metaphor of up and down* (Bolinger, 1986: 202ff.): if the pitch raises or falls, the body movement goes into the same direction.

Also, we regard the occurrence of the peak accent as evidence for the *frequency code* and for the *production code* according to Gussenhoven (2002: 48ff., 51ff.): on the informational level high fundamental frequency expresses uncertainty and continuity, whereas low fundamental frequency conveys certainty and finality.

In a follow up study, the audiovisual material from the production study was used to test the influence of audiovisual prosody on focus interpretation (Wollermann et al., 2011a; 2011b). Again, we used pictures for testing focus interpretation. Results show evidence for a contextual influence, whereas the audiovisual prosodic influence is less clear. This result is in line with the findings from our interpretation studies I and II.

Uncertainty in human-machine communication

In this section, we refer to previous studies on the audio(visual) modelling of uncertainty and also to studies dealing with the automatic detection of uncertainty. Also, our own studies on the acoustic modelling of uncertainty in an articulatory speech synthesizer are summarised.

Previous studies

Modelling uncertainty in speech synthesis can be useful to generate information systems with expressive abilities (cf. Marsi, van Rooden, 2007: 105). For the acoustical domain, it has been shown that the synthesis of *filled pauses* does not decrease the naturalness of speech from an unit selection synthesizer (Adell et al., 2010; Andersson et al., 2010).

For visual speech synthesis, Oh (2006) found that the variation of *facial expressions* and *head movements* affects the recognition of (un)certainty. Marsi and van Rooden (2007) observe that *head movement* alone as well as *head movement* combined with *eyebrow movement* influence the perception of uncertainty.

The automatic detection of uncertainty by dialogue systems is particularly useful for systems functioning as tutors. If such a system adapts to the student's uncertainty, the learning can be affected positively (Pon-Barry et al., 2006). For training these systems, corpora consisting of natural conversations between tutors and students are often used. By using prosodic cues covering *fundamental frequency*, *intensity*, *tempo* and *duration*, uncertain utterances have been detected with an accuracy of ca. 75% (Liscombe et al., 2005; Pon-Barry, Shieber, 2009).

Experimental studies on the expression of uncertainty in articulatory speech synthesis

The goal of our studies (Wollermann, Lasarcyk, 2007; Lasarcyk, Wollermann, 2010) was to model different degrees of uncertainty in speech synthesis and to test their impact on perception. We used the articulatory speech synthesizer of Birkholz (2005) by varying *intonation* (falling vs. rising), *pause* (absent vs. present) and the *hesitation particle* "hm" (absent vs. present). The *falling* intonation was intended to convey certainty, whereas the *rising* intonation was expected to signal uncertainty.

In Wollermann and Lasarcyk (2007), we modelled four different levels of (un)certainty: (1) *falling intonation* (intended certainty), (2) *rising intonation* (intended uncertainty), (3) *rising intonation* combined with *pause* (intended uncertainty), (4) *hesitation* combined with *rising intonation* and *pause* (intended uncertainty). Afterwards we carried out a perception study in order to test whether listeners are able to distinguish different levels of intended (un)certainty. Results showed that *rising intonation* as prosodic indicator of uncertainty had a stronger influence on the perception of uncertainty than *falling intonation* as prosodic indicator of certainty. This observation was in line with our expectation. Contrary to our expectation, the combination of *rising intonation* and *pause* did not contribute to a stronger degree of perceived uncertainty than *rising intonation* alone. The combination of *rising intonation*, *pause* and *hesitation* leaded to stronger degree of perceived uncertainty than (1) *falling* intonation alone as a prosodic indicator of certainty and (2) a *rising intonation* alone or a *rising intonation* combined with a *pause* as prosodic indicators of uncertainty.

In a follow-up study (Lasarcyk, Wollermann, 2010), we used all eight possible combinations of the three cues *intonation*, *pause* and *hesitation*. The scenario was also modified. In accordance with our previous findings, the combination of *rising intonation*, *pause* and *hesitation* as prosodic indicators of uncertainty had a stronger effect on the perception of uncertainty than *falling intonation* alone as prosodic indicator of certainty. Our data show only weak evidence for the effect of the *pause*. Furthermore, *rising intonation* and *hesitation* have a similar impact on the perception of uncertainty.

Conclusion

Several studies have provided evidence that uncertainty is communicated by using prosodic cues in natural conversation. On the one hand speakers convey uncertainty of answers prosodically in the acoustical and in the visual channel, on the other hand hearers use these cues for the *perception* of uncertainty. However, with respect to utterance *interpretation*, our studies suggest only a weak effect of prosodic indicators of uncertainty on pragmatic focus interpretation. In contrast to that, contextual cues have a stronger impact on the pragmatic focus interpretation than expected.

In the context of human-machine interaction, the communication of uncertainty has also been investigated. In speech synthesis, different degrees of uncertainty can be expressed and perceived by using prosodic cues. The results of our study suggest that the relative contribution of acoustic cues for the perception of uncertainty in speech from an articulatory synthesizer varies.

More studies are necessary in order to investigate which role uncertainty plays for generating speech synthesis systems with highly variable speech and in which scenarios a benefit for human-machine communication would occur.

Notes

[1] An overview of language specific focus marking can for instance be found in Gussenhoven (2004) and Büring (2009).

[2] In another case of *contrastive focus* a constituent which was focused in a previous utterance can be revised in the actual utterance. Here, the focus has the function of *correction*.

[3] For an overview of a more fine-grained differentiation of epistemically strong vs. weak reading and terminological variation according to Gazdar (1979) and Horn (1972) see van Rooij and Schulz (2004: 494f.). However, in our work we are investigating the *epistemically strong* reading.

[4] *Fall-rise* intonation is defined as follows: Firstly, the pitch peak is reached late in the accented syllable and a relatively abrupt drop in pitch must appear in the two following syllables. Secondly, a sentence-final rise in pitch is at hand (cf. Ward, Hirschberg, 1985: 748).

[5] For instance Weber et al. (2006) tested whether prosodic cues can affect the interpretation of grammatical functions by using eye-tracking.

[6] In one of the four dialogues no significant effects occurred. However, since this dialogue served as introduction for the whole scenario, we also used it for interpretation study II.

[7] The results of interpretation study I and II are described in detail by using diagrams in Wollermann (2012: 129ff., 139ff.).

[8] Hence, a direct comparison between the pictures should be excluded.

References

Adell, Jordi, Bonafonte, Antonio and David Escudero-Mancebo. 2010. "Modelling Filled Pauses. Prosody to Synthesise Disfluent Speech". In *Proceedings of Speech Prosody 2010*, 100624:1-4, Chicago, IL.

Andersson, Sebastian, Georgila, Kallirroi, Traum, David, and Matthew Aylett. 2010. "Prediction and Realisation of Conversational Characteristics by Utilising Spontaneous Speech for Unit Selection". In *Proceedings of Speech Prosody 2010*, 100116:1-4, Chicago, IL.

Baumann, Stefan. 2006. "The Intonation of Givenness - Evidence from German." In *Linguistische Arbeiten*, 508, Tübingen: Niemeyer.

Baumann, Stefan, Martine Grice, and Susann Steindamm. 2006. "Prosodic Marking of Focus Domains - Categorical or Gradient? " In *Proceedings of Speech Prosody*, 301-304, Dresden, Germany.

Birkholz, Peter. 2005. *3D-Artikulatorische Sprachsynthese.* Dissertation, Rostock: Universität Rostock.

Bolinger, Dwight. 1958. "A theory of pitch accent in English." In *Word* 14: 109-149.

Brennan, Susan E., and Maurice Williams. 1995. "The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers." In *Journal of Memory and Language* 34: 383-398.

Büring, Daniel. 2009. "Towards a Typology of Focus Realization". In *Information Structure*, edited by Malte Zimmermann, and Caroline Fery, 177-205, Oxford: Oxford University Press.

Fisseni, Bernhard. 2011. *Focus: Interpretation? Empirical Investigations on Focus Interpretation.* Duisburg: Universitätsverlag Rhein-Ruhr.

Gazdar, Gerald. 1979. *Pragmatics: Implicature, Presupposition and Logical Form*. New York: Academic Press.

Givens, David B.. 2009. *The Nonverbal Dictionary of Gestures, Signs, and Body Language Cues*. http://center-for-nonverbal-studies.org/uncert.htm.

Goldman-Eisler, Frieda. 1967. "Sequential temporal patterns and cognitive processes in speech." In *Language and speech* 10: 122-132.

Grice, H. Paul. 1975. "Logic and Conversation." In *Syntax and Semantics*, Vol. 3: Speech Acts, edited by Peter Cole, and John L. Morgan, 41-58, New York: Academic Press.

Grice, Martine and Stefan Baumann. 2002. "Deutsche Intonation und GToBI." In *Linguistische Berichte* 191: 267-298.

Groenendijk, Jeroen, and Martin Stokhof. 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*. Dissertation, Amsterdam: Amsterdam University of Amsterdam.

Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Gussenhoven, Carlos. 2002. "Intonation and interpretation: Phonetics and Phonology". In *Proceedings of Speech Prosody*, 47-57, Aix-en-Provence, France.

Hanssen, Judith, Jörg Peters, and Carlos Gussenhoven. 2008. "Prosodic Effects of Focus in Dutch Declaratives." In *Proceedings of Speech Prosody 2008*, 609-612, Campinas, Brazil.

Hart, J.T.. 1965. "Memory and the feeling-of-knowing experience." In *Journal of Educational Psychology* 56: 208-216.

Horn, Laurence R.. 1972. *On the Semantic Properties of the Logical Operators in English*. Mimeo: Indiana University Linguistic Club.

Kadmon, Nirit. 2001. *Formal Pragmatics. Semantics, Pragmatics, Presupposition, and Focus*. Oxford: Blackwell.

Keltner, Dacher, and Michelle N. Shiota. 2003. "New Displays and New Emotions: A Commentary on Rozin and Cohen (2003)." In *Emotion* 3(1): 86-91.

Krifka, Manfred. 1991. "A Compositional Semantics for Multiple Focus Constructions." In *Informationsstruktur und Grammatik*, Sonderheft 4 der Linguistischen Berichte, edited by Joachim Jacobs, 17-53.

Krifka, Manfred. 2007. "Basic Notions of Information Structure." In *Working Papers of the SFB632*, Interdisciplinary Studies on Information Structure (ISIS) 6, edited by Caroline Fery, Gisbert Fanselow, and Manfred Krifka, 13-56, Potsdam: Universitätsverlag Potsdam.

Kuhlthau, Carol C.. 1993. *Seeking Meaning: A Process Approach to Library and Information Services*. Norwood, NJ: Ablex.

Lasarcyk, Eva, and Charlotte Wollermann. 2010. "Do prosodic cues influence uncertainty perception in articulatory speech synthesis? " In *Proceedings of the 7th ISCA Workshop on Speech Synthesis*, 230-235, Kyoto, Japan.

Liscombe, Jackson, Julia Hirschberg, and Jennifer J. Venditti. 2005. "Detecting certainness in spoken tutorial dialogues." In *Proceedings of Interspeech 2005*, 1837-1840, Lisboa, Portugal.

Levelt, Willem J.M.. 1983. "Monitoring and self-repair in speech." In *Cognition* 14: 41-104.

Levelt, Willem J.M.. 1989. *Speaking: From Intention to Articulation.* Cambridge: MIT Press.

Marsi, Erwin, and Ferdie van Rooden. 2007. "Expressing Uncertainty with a Talking Head in a Multimodal Question-Answering System." In *Proceedings of the Workshop on Multimodal Output Generation*, 105-116, Aberdeen, UK.

Oh, Insuk. 2006. *Modeling Believable Human-Computer Interaction with an Embodied Conversational Agent: Face-to-Face Communication of Uncertainty.* Dissertation, New Jersey: Rutgers The State University of New Jersey.

Pon-Barry, Heather, Karl Schultz, Elizabeth O. Bratt, Brady Clark, and Stanley Peters. 2006. "Responding to Student Uncertainty in Spoken Tutorial Dialogue Systems." In *International Journal of Artificial Intelligence in Education* 16(2): 171-194.

Pon-Barry, Heather, and Stuart Shieber. 2009. "The Importance of Subutterance Prosody in Predicting Level of Certainty." In *Proceedings of the Companion Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT/NAACL) 2009*, 105-108, Boulder, CO.

Rooth, Mats. 1992. "A theory of focus interpretation." In *Natural Language Semantics* 1: 75-116.

Rozin, Paul, and Adam B. Cohen. 2003. "High Frequency of Facial Expressions Corresponding to Confusion, Concentration, and Worry in an Analysis of Naturally Occurring Facial Expressions of Americans." In *Emotion* 3(1): 68-75.

Schade, Ulrich, and Eikmeyer, Hans-Jürgen. 1991. ""wahrscheinlich sind meine Beispiele soo sprunghaft und und und eh ehm zu zu telegraph" - Konnektionistische Modellierung von "covert repairs"". In *GWAI-91 1. Fachtagung für Künstliche Intelligenz*, edited by Thomas Christaller, 264-272, Berlin: Springer Verlag.

Selkirk, Elisabeth. 2007. "Contrastive focus, givenness and the unmarked status of "discourse-new"". In *Working Papers of the SFB632*, Interdisciplinary Studies on Information Structure (ISIS) 6, edited by Caroline Fery, Gisbert Fanselow, and Manfred Krifka, 125-145, Potsdam: Universitätsverlag Potsdam.

Smith, Vicki L., and Herbert H. Clark. 1993. "On the course of answering questions." In *Journal of Memory and Language* 32: 25-38.

Swerts, Marc, and Emiel Krahmer. 2001. "On the alleged existence of contrastive accents." In *Speech Communication* 34: 391-405.

Swerts, Marc and Emiel Krahmer. 2005. "Audiovisual prosody and feeling of knowing." In *Journal of Memory and Language*, 53(1): 81-94.

Szabolcsi, Anna. 1980. "The semantics of topic-focus articulation." In *Formal methods in the study of language*, 2, edited by Jeroen Groenendijk, Theo Janssen, and Martin Stockhof, 513-540, Amsterdam: Mathematical Centre Tracts.

van Rooij, Robert, and Katrin Schulz. 2004. "Exhaustive interpretation of complex sentences." In *Journal of Logic, Language and Information* 13: 491-519.

Ward, Gregory, and Julia Hirschberg. 1985. "Implicating Uncertainty: The Pragmatics of Fall-Rise Intonation". In *Language* 61(4): 747-776.

Weber, Andrea, Martine Grice, and Matthew Crocker. 2006. "The role of prosody in the interpretation of structural ambiguities: a study of anticipatory eye movements." In *Cognition* 99(2): 63-72.

Wollermann, Charlotte, and Eva Lasarcyk. 2007. "Modeling and Perceiving of (Un)Certainty in Articulatory Speech Synthesis." In *Proceedings of the 6th ISCA Workshop on Speech Synthesis*, 40-45, Bonn, Germany.

Wollermann, Charlotte, and Bernhard Schröder. 2008a. "Does Uncertainty Effect the Case of Exhaustive Interpretation? " In *Proceedings of the ISCA Tutorial and Research Workshop on Experimental Linguistics*, 233-236, Athens, Greece.

Wollermann, Charlotte, and Bernhard Schröder. 2008b. "Certainty, Context and Exhaustivity of Answers." *Paper presented at the Workshop on Speech and Face to Face Communication*. Grenoble, France.

Wollermann, Charlotte, and Bernhard Schröder. 2009. "Effects of Exhaustivity and Uncertainty on Audiovisual Focus Production." In *International Conference on Auditory-Visual Speech Processing (AVSP) 2009*, 145-150, Norwich, UK.

Wollermann, Charlotte, Ulrich Schade, Bernhard Fisseni, and Bernhard Schröder. 2010. "Accentuation, Uncertainty and Exhaustivity: towards a model of pragmatic focus interpretation." In *Proceedings of the 5th International Conference on Speech prosody 2010*, 100063:1-4, Chicago, Illinois.

Wollermann, Charlotte, Ulrich Schade, and Bernhard Schröder. 2011a. "Variation of Accent Type and of Context - Influences on Pragmatic Focus Interpretation." In *Proceedings of Interspeech 2011*, 1077-1080, Florence, Italy.

Wollermann, Charlotte, Bernhard Schröder, and Ulrich Schade. 2011b. "Pragmatic focus interpretation: interplay between context and audiovisual prosody? " In *Proceedings of the Workshop on Gesture and Speech in Interaction*, Bielefeld, Germany.

Wollermann, Charlotte. 2012. *Prosodie, nonverbale Signale, Unsicherheit und Kontext - Studien zur pragmatischen Fokusinterpretation.* Dissertation, Duisburg-Essen: Universität Duisburg-Essen.

Charlotte Wollermann is research assistant at the Institute of German Linguistics at the University of Duisburg-Essen. She holds a master degree in computational linguistics and phonetics from the University of Bonn (2004). During her studies, she spent time abroad at the Dublin City University. In her PhD thesis (2012), she investigated the role of prosody and of context for pragmatic focus interpretation. Her research interests include experimental pragmatics, audiovisual prosody, and emotion and human machine-interaction.
Contact: charlotte.wollermann@uni-due.de

Bernhard Schröder is professor for German Linguistics at the University of Duisburg-Essen. He studied computational linguistics, phonetics, general linguistics and philosophy at the Universities of Bonn, Cologne and Helsinki. He received his master degree in 1991 and his PhD in 1997 and finished his habilitation in 2004. His main research areas are formal and experimental pragmatics, the epistemology of linguistics, models of language change, and the linguistics of mathematical proof texts. Contact: bernhard.schroeder@uni-due.de

Ulrich Schade is senior research scientist with Fraunhofer FKIE (since 2002), responsible for the development of NLP applications, and is associate professor to the Department of English, American and Celtic Studies, Bonn University. His received his PhD (1990) and habilitation (1996) from Bielefeld University with studies on modelling the cognitive process of language production.
Contact: ulrich.schade@fkie.fraunhofer.de.