

BLEKINGE TEKNISKA HÖGSKOLA
MANAGEMENTHÖGSKOLAN

**Talker-Contingent Simulation of
Voice in Silent Reading**

Michael Mania

Bachelor's thesis
Programme in Social Science for Communication and Learning
Spring semester 2010

Supervisor: Erik Lindström

Talker-Contingent Simulation of Voice in Silent Reading

Michael Mania

Blekinge Institute of Technology

Author Note

Michael Mania, School of Management, Blekinge Institute of Technology.

Correspondence concerning this article should be addressed to Michael Mania,
c/o Erik Lindström, School of Management, Blekinge Institute of Technology,
371 79 Karlskrona, Sweden.

E-mail: michael.mania@comhem.se

Abstract

An experiment was conducted to assess whether voices of familiar talkers are simulated while silently reading utterances from those talkers. In addition it was tested, whether talker-contingent simulation of voices in silent reading can occur after familiarization by reading utterances in context. After being familiarized with two talkers by watching video sequences alternatively by reading dialogues featuring those talkers, participants were asked to rate ambiguous ironic/non-ironic written utterances ascribed to these talkers. Participants rated the ambiguous utterances talker-contingent, i.e. utterances made by talkers who were ironic in study were rated as ironic even in test. The present study is discussed in terms of situated simulation and corroborates the view that talkers' voices are simulated while reading written utterances. In addition it is proposed that those simulations even can occur as simulations of earlier simulations.

Keywords: implicit prosody, situated simulation, talker-specificity, voice, silent reading

Talker-contingent Simulation of Voice in Silent Reading

Many forms of written communication such as e-mail, SMS-messages and chat exhibit characteristics of spoken language (Crystal, 2006, p. 29; Ko, 1996; Ling, 2005) which is particularly true for communicational situations that are defined by a high degree of communicative immediacy (Aeschbacher, 2008, pp. 48-49; Koch & Oesterreicher, 1985). The continuous concept of communicative immediacy proposed by Koch and Oesterreicher (1985) holds that the degree of distance and immediacy in communication is not tied to some dichotomous relation between written and spoken language in which writing would be seen as entailing great communicative distance and speech great communicative immediacy. Instead it claims that other factors such as the degree of intimacy, dialogicity and stylistic aspects account for distance or immediacy in verbal communication. For example, the final speech of prosecution in court, although performed orally, stands for a high degree of communicative distance while an SMS-correspondence between two lovers although delivered as written text meets a number of conditions for communicative immediacy. However, the possibilities of mediating paralinguistic features of speech (Crystal, 1975), including prosody and kinesic communication, through writing are obviously rather limited even though there are features of written text, like punctuation and emoticons, that give some prosodic and kinesic cues to the reader (Chafe, 1988; Crystal, 2006, p. 34 ff.; Derks, Fischer, & Bos, 2008; Steinhauer, 2003). This raises the question, whether the apparent “lack of voice” in written text necessarily entails a corresponding “lack of voice” in silent reading.

That is what would have to be assumed from a traditional, abstractionist point of view. Since abstractionist theories on language representation argue that words are encoded in an abstract and context-free way in memory, separated from episodic

information, these abstract lexical entries alone will be activated in visual word recognition (Tenpenny, 1995). However, abstractionist views are increasingly challenged by alternative accounts that deny the existence of different representations for words and episodes and instead claim the situated character of common episodic representations (Barsalou, 2003; Bowers, 2000; Garrod & Sanford, 2005, p. 237 ff.; Goldinger, 1996; Tenpenny, 1995; Walsh, Schütze, Möbius, & Schweitzer, 2007; Wu & Barsalou, 2009). Among those alternative, episodic accounts at least three approaches can be identified (Barsalou, 2003); connectionist models, depicting the lexicon as a dynamic network rather than a static list (e.g. McClelland & Rumelhart, 1985; Seidenberg & McClelland, 1989), exemplar models, suggesting, that every experience, such as the use of a word, is represented as one unique exemplar in memory (e.g. Goldinger, 1998; Walsh, Möbius, Wade, & Schütze, 2010) and finally models within the theory of grounded cognition (also termed "embodied cognition"; for a discussion see Barsalou, 2008), which posit, that multimodal states occurring during perception are stored in memory and that recognition, perception and imagery amongst others rely on re-enactments of those multimodal states, thus being simulations of perceptual events (Barsalou, 1999, 2008; Kent & Lamberts, 2008). As Barsalou (2003) points out, there are certain exemplar theories converging with theories of grounded cognition in several aspects.

These divergent views provide the setting for the extensive debate about the role of phonological code in visual word recognition during the last century. In particular the debate was concerned with the question whether phonological activation takes places at all while reading silently and provided it does whether it occurs pre- or postlexical (Dehaene, 2009, p. 26; R. Frost, 1998; Traxler & Gernsbacher, 2006, p. 325; Whitney,

1998, p. 197 ff.). According to Dehaene (2009, p. 26) this debate has resulted in the consentaneous view that phonological pathways to the lexicon exist alongside more direct pathways: “both [...] operate in parallel and reinforce each other”. On the account of silent reading there is strong evidence for phonological activation at least at some point of the process. A considerable number of studies point to early activation (Folk, 1999; Lee, Binder, Kim, Pollatsek, & Rayner, 1999; Perfetti & Bell, 1991; Pollatsek, Lesch, Morris, & Rayner, 1992) although there is some indication from a study of Simos et. al. (2002) for both early and late activation depending on the nature of the word to be read, i.e. early, prelexical activation for more “difficult” words, represented by pseudo-words as well as pseudo-homophones, and postlexical activation for less “difficult” words, represented by exception words.

Research on phonological activation provides a ground for considerations about “voice” in silent reading, but its concern is the level of word recognition, not prosody or idiosyncratic characteristics of voice. On the sentence-level it has been demonstrated in a number of studies, that prosody as a characteristic of speech is employed during silent reading too (Ashby & Clifton Jr., 2005; Ashby, 2006; Ashby & Martin, 2008; Luo & Zhou, 2010; Ren Gui-Qin & Yang Yufang, 2010; Steinhauer, 2003; Steinhauer & Friederici, 2001), a circumstance that is reflected in Fodor’s (2002, p. 2) implicit prosody hypothesis: “In silent reading a default prosodic contour is projected onto the stimulus and it may influence syntactic ambiguity resolution.” Fodor suggests (ibid.), that the occurrence of this projection might not be limited to syntactic ambiguity, a thought that has been elaborated further by Ashby (2006), Ashby and Clifton (2005), Steinhauer (2003) and Stolterfoht, Friederici, Alter and Steube (2007).

A growing number of online-studies corroborate the view, that the perceptual process of reading in many ways resembles that of speech perception (Aparicio, Gounot, Demont, & Metz-Lutz, 2007; Blau, van Atteveldt, Ekkebus, Goebel, & Blomert, 2009; S. J. Frost, Landi, & Mencl, 2009; Jobard, Vigneau, Mazoyer, & Tzourio-Mazoyer, 2007; Wheat, Cornelissen, & S. J. Frost, 2010), which is consistent with the research on phonological activation and implicit prosody presented above.

As opposed to abstractionist views the evidence for early phonological activation, implicit prosody and resembling perceptual processes for reading and speech perception suggests, that the “lack of voice” in the written text does not mean that there is a “lack of voice” in silent reading too. However, since written text does not actually contain idiosyncratic properties of the talker’s voice or even surface characteristics of speech in general, any conscious experience of voice or any unconscious cognitive process employing characteristics of speech in silent reading requires recollection of information from implicit or explicit memory and is to be regarded as auditory imagery or mental simulation.

Based on Damasio’s convergence-divergence-zone framework (K. Meyer & Damasio, 2009) Barsalou developed, within the aforementioned perspective of grounded cognition, his theory of situated simulation. According to this theory, modal perceptual states of experiences are stored in sensory-motor systems and can be re-enacted in order to simulate those earlier experiences. This encoding-retrieval process offers a framework for understanding a number of cognitive processes, such as perception, memory, mental imagery and language comprehension (Barsalou, 2008, 2009; Kent & Lamberts, 2008) and points to a common representational system for both perception and conception (Barsalou, 2003). While mental simulation often takes places

automatically and unconsciously (Barsalou, 2008), e.g. as part of implicit memory, mental imagery can be considered the conscious manifestation of modal mental simulation (Barsalou, 2008; Kent & Lamberts, 2008), e.g. as part of explicit memory. It must be emphasized that there is no claim of mental simulations being tantamount to actual experiences other than in rare cases (e.g. hallucinations), rather that “simulations are typically partial recreations of experience that can contain bias and error” (Barsalou, 2008, p. 620). Another but not necessarily incompatible way of conceptualizing mental and in this context especially auditory imagery can be theorized from Goldinger’s (1998) episodic theory, building on Hintzman’s idea of “echoes” (1986). According to this approach, which can be classified as exemplar theory, episodic traces of experiences are stored in memory and reactivated as “echoes” of those episodes, whereupon the term “echo” is especially appropriate for auditory experiences and their “echo”, the auditory image. What both the simulation and the “echo”-theory have in common and what is underpinned by evidence from online-studies is the assumption, that both perception and imagery make use of the same neural systems (Farah, Péronnet, Gonon, & Giard, 1988), as do speech perception and silent reading (*vide supra*). The crucial difference between situated simulation and Goldinger’s version of exemplar theory is the amodal and more static character of the latter (Barsalou, 2003).

In the view of situated, modal, simulation being a part of perceptual top-down processing it could be assumed, that modal activations take place even in spite of absent stimuli within a particular modality. In an fMRI-study Baumann, Koenecke and Schmidt (2007) demonstrated transmodal information transfer between auditory and motor areas in professional pianists, who were either playing the piano silently or were listening to music. In another fMRI-study von Kriegstein and Giraud reported

transmodal information transfer for learned stimulus pairs, when only one of both stimuli was available (2006). Although Geiselman and Glenny (1977) and Pilotti, Gallo and Roediger (2000) did not explicitly refer to concepts of situated simulation, they reported priming effects from auditory imagery, i.e. without any actual perceptual sensory stimulus present at all, which further underpins the assumption of imagery as simulated perception.

As early phonological activation, implicit prosody and resembling perceptual processes for reading and speech perception as well as for perception and simulation all point to a simulated voice in silent reading, the question emerges, whether this simulated voice resembles idiosyncratic, i.e. talker-specific characteristics of speech. Written text does not specify those characteristics and therefore it is of great interest for the subject of written communication to study, on which grounds this simulation of voices operates. Prosody, even though simulated, plays a significant role in language comprehension as will be illustrated further down by means of the comprehension of non-literal meaning. As an author's intended meaning can be modified by the simulated voice in the process of reading a simulated voice must be regarded as a potential source of miscommunication.

With reference to the study at hand it should be pointed out, that although a distinction has been made between on the one hand mental simulation as being unconscious and automatic and on the other hand auditory imagery as consciously experienced, both are of equal interest. As will be shown in the results to follow, readers' own reports of experienced voice in reading point to considerable individual differences. It is suggested, that the phenomenon to be investigated here should not be

nailed down to either the implicit or the explicit domain. This corresponds well with the model of situated simulation:

Re-enactment is not necessarily conscious but may also be unconscious, probably being unconscious even more often than conscious. Unconscious re-enactments may occur frequently during perception, memory, conceptualization, comprehension and reasoning, along with conscious re-enactments. When re-enactments reach awareness, they can be viewed as constituting mental imagery, given that imagery is typically assumed to be conscious. (Barsalou, 2009, p. 1281)

Therefore, while reviewing the sparse research on the subject of talker-specificity of mental simulations presented below, no difference has been made between research on auditory imagery experiences and its implicit, automated equivalent. It appears that even studies deploying the notion of “imagery” do so with some latitude to whether the processes in question are consciously experienced or not (Alexander & Nygaard, 2008; Kurby, Magliano, & Rapp, 2009).

In an ERP-study using a match-mismatch paradigm in which messages did or did not match an expected speaker van Berkum, van Den Brink, Tesink, Kos and Hagoort (2008) demonstrated with regard to spoken language, that “the linguistic brain relates the message to the speaker immediately”. Nygaard, Sommer and Pisoni (1994) could show, also with regard to spoken language, that familiarity with talkers improved auditory word recognition. Both studies do support the view, that idiosyncratic voice characteristics are retained in memory and do facilitate comprehension and recognition of spoken words. Rosenblum, Miller and Sanchez (2007) conducted a study on the matter of cross-modal transfer between auditory and visual modalities in speech perception involving lip-reading and speech in noise, making evident, that talker-

specific information could be transferred from one modality to the other. Talker-specific priming effects in implicit memory tests both for hearing, reading and for imagined voice were found by Pilotti, Gallo, and Roediger (2000), pointing to a common cognitive ground for imagery and perception of speech, and corroborating the thesis, that talker-specific characteristics in fact play a part in implicit memory. Kurby, Magliano and Rapp (2009) adopted a match-mismatch-paradigm, in which participants read dialogs involving two characters. On an auditory recognition task during dialogue reading participants benefitted from previous familiarization with voices when the presented voices matched the characters presented in the written dialogues. Alexander and Nygaard (2008) hypothesized, that silent reading employing auditory imagery of a familiarized slow talker would consume more time, than reading a text featuring a familiarized faster talker, which proved to be true at least for more difficult text. Abramsson (2007) found voice-consistency after priming both for voices that had been learned by reading (without actually hearing a voice) and by listening to voice. The findings of the aforementioned three studies (Abramson, 2007; Alexander & Nygaard, 2008; Kurby et al., 2009) are all in support of the view, that talker-specific voice characteristics are encoded in memory and activated during reading. In an ERP-priming-study finally Regel, Coulson and Gunter (2010) investigated talker-specificity in silent reading by employing written ironic/non-ironic statements to familiarize participants with a speaker's communicative style, showing, that knowledge of a speaker's style affects the interpretation of ironic utterances early in the process.

The current study was designed to examine the hypothesis, that voices of familiar talkers are simulated while silently reading utterances from those talkers. In addition it was tested, whether talker-contingent simulation of voices in silent reading

can occur after familiarization by reading utterances in context. Similar to Regel et al. (2010) the paradigm of this experiment takes advantage of the ambiguity inherent to ironic utterances, as those utterances have a literal meaning as well as an intended meaning, of which the latter differs from the former¹ (Dews & Winner, 1999; Gibbs Jr., 2002; Pexman, 2008). There has been much debate (Creusere, 1999; Regel et al., 2010) as to how ironic statements can be understood as ironic despite of this discrepancy between what is said and what is implied. A number of different factors can be considered as possible cues for a listener's or reader's comprehension of irony. It is hereby suggested, that those possible cues can be broken down into four categories. The first category contains cues on the literal level, i.e. cues, which support the interpretation of a spoken or written utterance irrespective of context, speaker or prosody. This includes interjections, certain adjectives and adverbs (Utsumi, 2000), as well as direct cues (Kreuz, 1996, p. 28), like the negation of the utterance. The utterance "O dear, your presentation really made an impression on me – not!" contains both an interjection ("O dear!"), an adverb ("really") and a direct cue ("not!") in order to indicate, that the utterance is meant to be ironic. In the second category contextual cues are to be found, e.g. the counterfactual ("What a beautiful day!" when it's obviously raining) and attitude (e.g. a hostile conversation) (Creusere, 1999). The third category contains kinesic and prosodic cues (Bryant & Tree, 2005; Kreuz, 1996, p. 28), like facial expression and ironic intonation. Finally the fourth category contains other cues specific to the talker as a person, like the talker's character or profession (Regel et al., 2010). On the time course of irony comprehension the original standard pragmatic

¹ Although, as Gibbs (Gibbs Jr., 2002, p. 460) points out, it is arguable, if there is just one literal meaning. For the purposes of this paper it is sufficient, to state the inherent ambiguity of ironic statements.

model claimed, that literal meaning is accessed first, while irony is not understood until literal meaning is found to be inapplicable. This has been refuted by Glucksberg (2003), who argues, that both literal and ironic meaning are processed simultaneously.

In the current study participants were asked to rate written utterances that were ambiguous with respect to irony after participants had been familiarized with two talker's voices by watching video sequences with each talker responding to a third person (conditions A and B, see Table 1). One of the talkers made use of irony frequently while the other did not. When rating the written utterances, participants would depend on at least some of the abovementioned cues to determine, whether the reply should be interpreted as ironic or not. Due to the fact, that the dialogues were written and had been validated as ambiguous in a validation study (validation study 1), it can be said with certainty, that they in themselves did not contain cues sufficient to determine their ironic character. This holds especially for the first and the second categories of cues, i.e. cues on the literal level and contextual cues. If the ratings of the ambiguous replies would show to be biased by the fact, that the talker's name is indicated along with the written utterance, obviously participants would have relied on the third category, i.e. kinesic and prosodic cues and on the fourth category of cues, i.e. other talker-specific cues. Given phonological activation and implicit prosody, within the framework of situated simulation and according to the hypothesis advanced here, it should be expected, that participants interpret ambiguous written utterances as ironic in test, if the talker had been presented as being ironic frequently during study.

In another condition of the current study participants were asked to rate ambiguous written utterances with respect to irony after that they had been familiarized with talkers by reading dialogues featuring those talkers along with context that

revealed if talkers employed irony in their replies (condition C, see Table 1). Again one of the talkers made use of irony frequently while the other talker did not. This condition was employed, to investigate, whether talker-contingent simulation of voices in silent reading can occur after familiarization by reading utterances in context. If participants would interpret ambiguous written utterances as ironic, when the talker had been presented as being ironic frequently during study, even though participants in fact never had heard the actual voice, within the framework of situated simulation it would demonstrate that simulation also is possible as re-enactment of an earlier simulation. An overview over the design of the main experiment is presented in Table 1.

Furthermore, all participants underwent an explicit memory test in which they were asked to describe the familiarized talkers, without depicting the talkers' physical appearance. This was to test, to what degree the participants' explicit memory of the talkers' characters was consistent with talkers' characters as they were presented during study. A poor learning result for the context-condition compared to the video-conditions in conjunction with a non-contingent result in the ambiguity task in the context-condition would raise questions about the validity of the results of this group, as a non-contingent result in the ambiguity task would point both to a lack of voice-contingency and to deficiencies in learning. Additionally participants were surveyed about their conscious experience of auditory imagery during silent reading.

Two validation studies were conducted to obtain valid materials for the main experiment described above.

Validation study 1

The objective of the first validation study was to obtain certain characteristics of sixty dialogues in order to make a selection of dialogues for the main experiment based

on those characteristics. Since the rating task of the main experiment required dialogues with ambiguous replies it was concluded, that dialogues exhibiting a high standard deviation when rated without instructional cues as well as an ample range between the mean of ratings when read with the instruction “without irony” and the mean of ratings when read with the instruction “with irony” should qualify for the rating task of the main experiment. Furthermore it was concluded, that dialogues exhibiting an ample range between the mean of ratings with the instruction “without irony” and the mean of ratings with the instruction “with irony” but not necessarily a high standard deviation when rated without instructional cues should be suitable to be used in the learning phase of the main experiment, more precisely for dialogues containing replies, that were supposed to be read with irony by one talker and without irony by the other talker. For the remaining dialogues in the main experiment it was only necessary to validate, that the replies could be understood without irony, as they were supposed to be uttered by the talkers without irony during the learning phase.

Method

Participants.

Sixteen college students from the Blekinge Institute of Technology, who were offered either a bonus lecture or a donation to charity in exchange for participation, and two acquaintances of the researcher participated in the study. Of the 18 participants all but 4 were females. Participants were between 26 and 47 ($M = 34.5$, $SD = 5.36$). Four participants had not learned Swedish during the five first years of their lives. Participants were randomly assigned to one of two groups (group I and group II).

Materials.

Sixty short dialogues were constructed, comprising one question or statement made by one person, labeled person “A” and a reply by another person, person “B”. The reply was supposed to be ambiguous, so that it could be read with irony or without irony. As illustrated in Table 2 there were three variants for each dialogue; a.), without reading instruction, b) with the instruction, to read the reply without irony and c) with the instruction to read the reply with irony.

The sixty dialogues were grouped into four batches (1, 2, 3, and 4) comprising fifteen dialogues each and allocated to two groups of participants. This was done with the objective to bisect the number of possible dialogues and their variants for each group in order to reduce the workload for each participant. An Internet-survey was constructed using Survey Gizmo, version 2, in which a seven-point rating scale (0 – 6), reaching from “very negative” to “very positive” was added to each dialogue in each variant. A first cycle was made up entirely of dialogues of the a-variant (without instruction), while a second cycle featured the b- and c-variant alternately (“without irony” and “with irony”). Since the study was designed as an Internet-experiment, participants had to log into Survey Gizmo on the Internet from a location of their choice using equipment of their choice in order to carry out the survey. An overview of the groups and cycles in this validation study is provided in Table 3.

Procedure.

Eighty college students were recommended by their professor to participate in the validation study for a bonus lecture; 15 of these students started and 13 completed the survey. One hundred eighty college students received an e-mail-invitation with the

offer, to participate for a donation to charity; one completed the survey². One hundred facebook-friends were invited by leaving an invitation on their message board; three completed the survey. One participant could not be traced to any of these invitation groups³. Each prospective participant with the exception of facebook-friends received a personalized e-mail-invitation with a link to the Internet-study and introductory instructions. Facebook-friends were provided with a non-personalized link to the study. By clicking on the link participants were directed to the Internet-survey, which opened with further information about the study and instructions. Thereafter participants were asked to provide personal details and then randomly assigned to one of two groups (group I or II). The following pages of the survey contained the first cycle of dialogues (variant a: without instruction). Participants were prompted, to read the dialogues silently and to rate the reply in each dialogue on a seven point rating scale from negative to positive. The instruction for the rating task advised the participants to rate the reply rather spontaneously and not to ponder too much about how to justify the rating. After rating 30 dialogues in each of the two groups in the first cycle, participants were presented with the second cycle containing 30 dialogues of the b-variant alternated with 30 dialogues of the c-variant (see Table 3), which they were asked to rate in the same way as in the first cycle. On a final page participants were thanked and dismissed.

Results and discussion.

Thirty dialogues were chosen for the rating task of the main experiment based on the assumption, that those dialogues should exhibit a high standard deviation in variant a. and a high value in range between variants b. and c. This was to validate, that

² This invitation was intended to recruit a few more participants to validation study 1 during the last days of data collection. Several of the students approached here participated in the main experiment eventually.

³ As an alternative technical solution all participants had access to a non-personalized link to the study.

the selected dialogues actually could be read with and without irony, that they were ambiguous and interpreted differently by different people and that the ratings reflected the variants in a proper way. As shown in Table 4 the 30 dialogues were grouped into two selections (Y and Z) that were supposed to be equivalent with regard to their central tendencies for the results obtained in this validation.

Another six dialogues were chosen to be used for two conditions in the learning phases of the main experiment; in one condition of the main experiment they would be read without irony and in the other condition with irony. The selection was made based on the assumption, that those dialogues should exhibit a high value in range between variant b. (instruction “without irony”) and variant c. (“with irony”). This selection of dialogues is labeled selection U. Two more selections were made, labeled S and T. They, too, were chosen to be used in two conditions of the learning phase, but would only be read without irony. Selections U, S and T are described in Table 5. A total of fifty-four dialogues were chosen for the main experiment, six dialogues were discarded.

Validation study 2

In one condition of the main experiment participants would not be familiarized with the talkers by watching video sequences but by reading the same dialogues as in the video conditions, with the talker’s reply given in context. This second validation study was conducted to validate that the context actually induces an ironic or non-ironic interpretation.

Method

Participants.

Seven college students from the Blekinge Institute of Technology, who were offered either a bonus lecture or a donation to charity in exchange for participation, and five acquaintances of the researcher and the researcher's wife respectively participated in the study. Of the twelve participants all but four were females. Participants were between 20 and 68 ($M = 37.5$, $SD = 12.64$). Two participants had not learned Swedish during the five first years of their lives. Participants were assigned to the study as a result of the assignment process described in connection with the main experiment (see below).

Materials.

The replies of the 24 dialogues labeled as selections U, S and T in the first validation study were supplemented with context information as exemplified in Table 6. This dialogue variant is referred to as variant d. The context information was designed to induce an ironic interpretation for six of the dialogues and a non-ironic interpretation for the remaining eighteen dialogues.

As part of the preparations for the main study an Internet-survey was constructed using Survey Gizmo, version 2, in which a seven-point rating scale (0 – 6), reaching from “very negative” to “very positive” was added to each dialogue. Since the study was designed as an Internet-experiment, participants had to log into Survey Gizmo on the Internet from a location of their choice using equipment of their choice to conduct the survey.

Procedure.

Participants had received an e-mail with a personal link to the Internet-study and introductory instructions. By clicking on the link in the e-mail the participants were directed to the Internet-survey, which opened with further information about the study and further instructions. Thereafter participants were asked to provide personal details. After this step the actual sampling process took place, which is described in more detail in connection with the main experiment. The following three pages of the survey contained eight dialogues each (variant d: with context). Participants were prompted to read the dialogues silently and to rate the reply in each dialogue on a seven point rating scale from negative to positive. The instruction for the rating task advised the participants to rate the reply rather spontaneously and not to ponder too much about how to justify the rating. On a final page participants were thanked and dismissed.

Results and discussion.

The objective of the second validation study was, to validate, that the dialogues used during the learning phase in the context condition (condition C) of the main experiment would be understood as intended, i.e. as non-ironical or ironical respectively. A paired-samples t-test was conducted to compare dialogues with a context suggesting a non-ironical interpretation and dialogues with a context suggesting an ironical interpretation. There was a significant difference in the scores for the non-ironical ($M = 5.03$, $SD = .68$) and the ironical ($M = 3.86$, $SD = .63$) conditions; $t(11) = 6.01$, $p = 0.00$, which validated, that the replies in context would be understood as intended.

Main Experiment

As has been explained in more detail earlier in this study the main experiment was conducted to examine, if participants would interpret ambiguous written utterances talker-contingent. The main experiment employed dialogues validated through the preceding validation studies.

Method

Participants.

Thirty-three college students from the Blekinge Institute of Technology, who were offered either a bonus lecture or a donation to charity in exchange for participation, thirty-five secondary school students and three acquaintances of the researcher participated in the study. The last two groups were not offered any reward for their participation. Fifty-five of the seventy-one participants were female, fifteen male. Participants were between 16 and 65 years of age ($M = 24.4$, $SD = 1.64$). Eight participants had not learned Swedish during the five first years of their lives. Participants were with few exceptions (see below) randomly assigned to one of three conditions (groups A, B and C).

Materials.

Based on selection T of the first validation study (see Tables 5 and 7) 12 video sequences were produced featuring the characters “Marie” and “Anna”, both played by experienced amateur actors. While Marie, who always opened the dialogue, could only be heard, not seen, Anna was captured on film. For each video sequence the actors played one of the dialogues of selection T. Character A of the T-selection was adapted as “Marie” and character B as “Anna”. The actor playing Anna was instructed to read her lines without irony.

Based on selection S of the first validation study (see Tables 5 and 7) six video sequences were produced featuring the characters “Marie” and “Karin”, both played by experienced amateur actors; Marie by the same actor as before. While Marie could only be heard, not seen, Karin, who replied to Marie in every dialogue, was captured on film. For each video sequence the actors played one of the dialogues of selection S. Character A of the T-selection was adapted as “Marie” and character B as “Karin”. The actor playing Karin was instructed to read her lines without irony.

Based on selection U of the first validation study (see Tables 5 and 7) six video sequences were produced featuring the characters “Marie” and “Karin”, both played by the same actors as before. While Karin was captured on film, Marie could only be heard, not seen; Marie opened every dialogue and Karin always replied. For each video sequence the actors played one of the dialogues of selection U. Character A of the T-selection was adapted as “Marie” and character B as “Karin”. The actor playing Karin was instructed to read her lines without irony. Thereafter the same six dialogues from selection U were picturized once more, but this time with the instruction to the actor playing Karin, to read the replies with irony.

Whenever one of the four persons being present during the shooting was unsecure about whether the replies of Anna or Karin would be understood in the intended way with respect to irony, the shooting was recapitulated.

In this way a total of 30 video sequences were produced, based on 24 dialogues; 12 dialogues featuring Marie and Anna and 12 dialogues featuring Marie and Karin. Six of the dialogues featuring Marie and Karin were produced in two versions; one in which Karin reads her replies without irony and one in which she reads the same replies with irony. Filming was conducted using a Nikon Coolpix L5 digital camera mounted on a

stand. All scenes were shot from the same angle, placing the actors in front of the same background. Both actors, Anna and Karin, were wearing the same, white T-shirt during the filming.

The video material was edited in Windows Movie Maker version 6.0. For each sequence a sign was added announcing the start of the video. Each sequence was edited to display the names of Anna and Karin for five seconds while they were speaking. Video sequences were made accessible using the Internet service “youtube”.

Character “A” in the written dialogues of the selections Y and Z of the first validation study was relabeled as “Marie”. Character “B” was relabeled as “Anna” with reference to selection Y and as “Karin” with reference to selection Z. In the same way the dialogues of the T, U and S selections were relabeled as “Anna” and “Karin” to be used in the learning phases of condition C.

An Internet-survey was constructed using Survey Gizmo, version 2. After the instructions and questions about the participants’ background a randomizer was inserted, to allocate the participants to either one of three conditions of the main experiment or to the second validation study, which has been described above. For the three conditions of the main experiment (groups A, B and C, see Table 1) the questionnaire was constructed in such a way, that learning phases consisting of embedded video sequences as described above (groups A and B) or dialogues with replies in context that had been validated in the second validation study (group C) alternated with rating phases containing rating tasks based on dialogues, that had been selected as the relabeled (Anna and Karin) selections Y and Z in the first validation study. The use of different selections in the construction of the questionnaire and the order of pages and phases are shown in Table 7. For each dialogue in the rating phases a

seven-graded rating scale was inserted ranging from “*very negative*” (0) to “*very positive*” (6). The order of pages 2, 4, 6 and 8, containing the rating tasks, was randomized as was the order of questions on each of those pages.

A number of statements were formulated to investigate the conscious experience of the participant’s inner voice in silent reading. The statements which are presented in detail in Table 8 ranged from “no experience at all” to “inner voices while reading characters’ voices in fiction” to “almost like reading aloud”.

Procedure.

Initially invitations applied for both the main experiment and the second validation study; first on a later stage in the process participants were assigned to either one as described below.

Four hundred college students at the Blekinge Institute of Technology were invited by the author to participate in the study for a donation to charity. Twenty-five of those students started the online-process and seventeen finished it. One hundred eleven college students from the same university college were invited by e-mail with a recommendation by their teacher of psychology to participate for a bonus lecture. Twenty-seven students started the online-process and nineteen finished it. One college student completed the survey but could not be located to one of the foregoing invitation groups. By leaving a message on their message board one hundred facebook friends were invited to participate. Nine started and six completed the survey. Fifty-five high school students from a school in Helsingborg, Sweden, were asked by their teacher of psychology to participate during class, although participation was voluntary. Thirty-nine students started the process and thirty-five finished it. Six prospects from the first validation study, which had not completed it though, were reinvited to participate in this

stage of the project; four started and three completed the process. Two acquaintances of the researcher were invited to participate either in condition C or the second validation study. Five acquaintances of the researcher and colleagues of the researcher's wife respectively were only invited to participate in the second validation study. All of the invited acquaintances and colleagues participated. In total 576 people were invited and 77 participated.

All invitations were sent by e-mail containing a direct and personal link to the study except for the Facebook-invitations, which only contained a non-personal link. By clicking on the link the participants were directed to the Internet-survey, which opened with further information about the study and with instructions on how to conduct it. A page with questions about the personal background of the participants ended with a question about whether the participant had been directed to the survey from facebook or not. All participants (with exceptions below) who were not linked to the survey from facebook were randomized to one of three conditions (A, B, C, see Table 1) or to the validation study. Because of the risk, that facebook friends would recognize at least one of the actors in the video sequences, their possible outcomes in the randomization process were limited to condition C and the second validation study. Furthermore, the probabilities to be selected for one of the four possible outcomes in the randomization process (conditions A, B, C and validations study 2) were adjusted subsequently with regard to the planned number of participants in each condition and the second validation study. Towards the end of the data collection invitees could only participate in the latter since there were enough participants in all conditions but not in the second validation study, which explains the five participants that were specified above as having received

an invitation to the second validation study only. The further procedures for the second validation study have been described above.

In the initial learning phase participants in conditions A and B were shown a number of video sequences while participants in condition C were presented with the written dialogues with replies in context. While Anna never replied with irony, Karin did so in half of her replies in conditions B and C. Participants were instructed, not to watch videos twice, but were otherwise in control of the time spent on each page. In the following rating phase participants were prompted to read the dialogues silently and to rate the reply in each dialogue on a seven point rating scale from negative to positive. The instruction for the rating task advised the participants to rate the reply rather spontaneously and not to ponder too much about how to justify their rating. Learning phase and rating phase alternated four times. After completing all learning- and rating phases participants were asked to describe Anna and Karin with five adjectives, without depicting their outward appearance.

Within two weeks form completion all participants who had completed the main experiment were invited by e-mail to describe their conscious experience of auditory imagery of voices during silent reading and 28 individuals commented on seven statements about experience of voice in reading by checking those statements they agreed to. The statements are presented in table 8.

Results.

For each individual the ratings for Anna and Karin were summed up separately. In the next step the difference between the ratings was calculated as Karin minus Anna, resulting in a negative value, when Karin had been rated less positive than Anna and in a positive value, when Karin had been rated more positive than Anna. An analysis of

variance⁴ showed a significant difference between the three conditions with regard to difference in rating of Anna and Karin, $F(2, 68) = 4.76$ $p = .01$. Post hoc Dunnett's T3 Tests indicate, that condition A (video, no irony) ($M = 3.96$, $SD = 8.52$) differs significantly ($p < .05$) from group B (video, irony) ($M = -6.43$, $SD = 16.06$) and from group C (context, irony) ($M = -2.88$, $SD = 8.94$). However group B did not differ from group C. The means of the groups are presented in Figure 1.

Additionally, the difference in rating of Anna and Karin was subjected to a two-way analysis of variance having the three levels of group (A, B, C) and furthermore two levels of Swedish as native language (yes, no). There was a main significant effect for group $F(2, 65) = 3.63$, $p < .05$, but none for Swedish as native language $F(1, 65) = .20$, $p > .05$ and no significant interaction effect $F(2, 65) = .55$, $p > .05$.

The participants' descriptions of Anna and Karin, which were submitted by only 59 of the original 71 participants were coded in such a way, that descriptions indicating a learned difference between Anna and Karin in terms of "positive" and "negative", when such a difference was expected (groups B and C), were rewarded with one point, as were descriptions not indicating a learned difference, when no such difference was expected (group A). Furthermore, allocating a difference or non-difference to the right person was rewarded with one more point. Half points were awarded as soon as the description given by an individual was differentiated in terms of the positive or negative assessment of Anna and Karin or when the participants' comments indicated uncertainty about the identities and attributes of Anna and Karin. It could be argued, that this way of coding benefits group A (video, no irony). In cases where the participants of group A did characterize both Anna and Karin in a similar way, thus correctly not indicating a

⁴ Data collected using Likert-scales is treated as interval-data in this study.

difference, were no difference was expected, it cannot be concluded from looking at the describing adjectives alone, whether the participant actually has identified Anna and Karin correctly. However, participants in both the B- and the C-group did not in any case confound both characters without commenting that they were unsure about the identities of the both. Therefore, the identification of characters in group A is assumed to be correct, as long as there is no contradicting information given by the participant's comment. The obtained value is considered to be an indication of the learning results of the experiment.

An analysis of variance showed a significant difference between the three groups with regard to expected learning, $F(2, 56) = 10.24$, $p = .00$. Post hoc comparisons using Bonferroni adjustment indicate, that group B (video, irony) ($M = 1.57$, $SD = .46$) differs significantly ($p < .01$) from group A (video, no irony) ($M = .91$, $SD = .71$) and from group C (context, irony) ($M = .67$, $SD = .80$). However, group A did not differ from group C.

As to the questionnaire about conscious experience of inner voice in silent reading the multiple answers were recoded in such a way, that four main categories remained. The results that are presented in Table 9 showed individual differences in the conscious experience of inner voice in silent reading, although only one participant reported the he or she was not aware of any inner voice.

Discussion

The purpose of the present study was to examine, whether voices of familiar talkers are simulated while silently reading utterances from those talkers and if so, whether talker-contingent simulation of voices in silent reading can occur after familiarization by reading utterances in context.

In the first validation study sixty one-line dialogues between two talkers were validated with respect to the ambiguity for ironic interpretation of the replies of the second talker. On the basis of the results of the first validation study 24 of the 60 dialogues were selected for the learning sequences of the main experiment and thirty for the test sequences. In the second validation study the twenty-four dialogues that had been selected to be used for learning sequences as a result of the first validation study were altered by adding context-information prior to the second talker's reply. The context-information was supposed to indicate an ironic meaning for the replies of six of the dialogues and literal meaning for the remaining dialogues. That this was the case was validated and the dialogues were designated to be used in the learning sequences of the context-condition.

In the main experiment participants were familiarized with two talkers, Karin and Anna, in dialogue with a third person. One group was familiarized with Karin and Anna by watching video sequences in which either Anna or Karin were responding to a third person. For this group neither Anna nor Karin responded ironically in any of the sequences. Even a second group was familiarized with Anna and Karin by watching video sequences, although for this condition Karin responded with irony in half of the dialogues. A third group was familiarized with the two talker by reading written dialogues in which context information preceding Anna's or Karin's reply indicated literal or ironic meaning of the utterance. An ironic meaning was indicated for half of the replies given by Karin, but not for any reply given by Anna. Between learning phases participants in all three groups were asked to rate written replies by Anna and Karin on a positive to negative scale. The replies were ambiguous with respect to irony

and did not in themselves contain any cues to solve the ambiguity problem, other than the talkers' names.

As predicted, those groups, who had been familiarized with the talker's voices by watching video sequences featuring those talkers, interpreted talker's written utterances in test talker-contingent, i.e. ambiguous written utterances made by ironic talkers were interpreted as ironic and ambiguous utterances made by non-ironic talkers were interpreted as non-ironic. Even the group, who had been familiarized with the talkers' voices by reading dialogues with replies preceded by context obtained similar results.

Familiarization with a talker's voice resulted in talker-contingent interpretation of talker-specific ambiguous written replies in a dialogue. During familiarization participants were exposed to talker-specific communicative cues important for irony comprehension, such as communicative style and kinesic and prosodic cues. It is concluded, that the talker-contingent interpretation of the replies in the experiment relied on talker-specific cues given in study, as the talkers' names attributed to the replies were the only talker-specific cue given in test that participants could rely on for solving the ambiguity problem. However, from this it cannot be determined, to which extend different talker-specific cues contribute to the interpretation; i.e. if information about the speaker's character and communicative style encoded in memory, information about the speaker's voice characteristics or both are crucial for resolving ambiguity. Certainly even other talker-specific cues such as facial expressions, although not within the scope of this study might play a part in this cognitive process.

The quandary at hand is ultimately a question of the order of events. It could be argued that the literal meaning of an utterance is comprehended first and that

information about a possible nonliteral meaning is considered in a second stage of comprehension, which essentially complies with the standard pragmatic model presented earlier. It would eventually mean that no talker-specific prosody is required for either of those two steps, while information about the speaker's character and communicative style might be employed in the second stage. But the standard pragmatic model has been refuted (Glucksberg, 2003) and studies demonstrating early phonological activation (Folk, 1999; Lee et al., 1999; Perfetti & Bell, 1991; Pollatsek et al., 1992), early linking between message and talker (Van Berkum et al., 2008), reading-time effects for slow and fast talkers (Alexander & Nygaard, 2008) as well as early processing of a speaker's communicative style (Regel et al., 2010) all contradict such an order of events.

It has been demonstrated comprehensively in previous research, that prosody is projected onto the written text in silent reading (Ashby & Clifton Jr., 2005; Ashby, 2006; Ashby & Martin, 2008; Luo & Zhou, 2010; Ren Gui-Qin & Yang Yufang, 2010; Steinhauer, 2003; Steinhauer & Friederici, 2001). Since the participants in this study had never actually heard the talkers utter the exact wording presented in the written dialogues, all implicit talker-contingent prosody employed in silently reading that particular talker's replies must be considered to be a simulation of some kind. As had been pointed out above simulation within the framework of situated simulation, must not be regarded to be tantamount to actual experience, but as "partial recreations of experience that can contain bias and error" (Barsalou, 2008, p. 620). This is even supported by the reports of individual differences concerning experience of inner voice in silent reading given by the participants of this study.

Implicit prosody has to be regarded as simulation and as such it builds on information stored in memory including both prosodic cues and information about the speaker's character and communicative style. With regard to bias and error the simulation might even be influenced by information, which is not talker-specific; e.g. in the case of this study, a name attributed to a reply might also re-enact perceptual states attributed to another familiar person with just the same name. While silently reading the exact same text that the reader actually had heard as speech before, e.g. the transcript of a famous speech, it might be a matter of auditory imagery in the sense of vivid explicit memory recollection. While reading a novel, ambiguous text it is a matter of simulation, using whatever cues are at hand, including prosodic cues and information about the speaker's character and communicative style. Despite a possible bias from implicit memory deploying information, that is not talker-specific, the results of this study have demonstrated an apparent effect of the talker-specific cues on the resolution of the ambiguity problem. Thus at least within the frames of the experiment conducted here it can be concluded, that any possible bias from information not being talker-specific but rather reader-specific must be weaker, than the effect of the talker-specific cues.

Also familiarization with a talker's communicative style by reading dialogues with an ambiguous reply in context resulted in talker-contingent interpretation of talker-specific ambiguous written replies in a dialogue. As argued above in connection with those groups that were familiarized with the talkers by video, even in this case implicit prosody as a simulation must rely on information from the memory. For this group, referred to as group C, it is actually a question of two simulations; one while reading during study, simulating the voice of a talker, that had never been heard before, and one while reading during test. Within the framework of situated simulation the course of

events at hand can be termed “simulation of simulation”. On both reading occasions simulation builds on information encoded in memory, with the one difference, that the second simulation even has access to episodic traces of the first simulation. These findings are consistent with those of Abramson (2007) presented earlier in this study.

Because of the above discussed results for the context-condition the explicit memory test for expected learning does not have the same relevance as it would have had, had there not been an effect of familiarization during the rating phase. As this effect has been demonstrated, apparently learning was sufficient to bias the ratings during test. In the opposite case, with ratings not being biased by familiarization, one could have speculated that this was due to insufficient learning of talker-specific cues during study in the context-condition and a poor learning result in the explicit memory test would have corroborated this view, although it would not have been proof for insufficient learning. It may be objected that poor learning results as well could be interpreted as pointing to other factors affecting the ratings over and above those talker-specific cues mentioned earlier. This is based on the conclusion, that because of poor results in the learning test no talker-specific cues, which could affect rating, have been learned and therefore there must be other factors biasing the rating. This conclusion is precipitant, though. Firstly, the explicit memory test conducted here does obviously not rule out implicit memory. Secondly, although there is a difference in learning between the three groups with the video/non-irony-condition and the context condition achieving lower results than the video-irony-condition it cannot be concluded, that the lower scores are too low, to ensure sufficient learning. And thirdly and finally, as has been demonstrated, in all conditions the ambiguity task is solved in accordance with the talker-specific cues presented in study. The talker’s name as the only cue being inherent

to the written reply presented in the ambiguity task, clearly is the link to the learned talker-specific cues.

The use of implicit prosody to solve ambiguity problems as a means for demonstrating talker-contingency in silent reading implicates one limitation to the generalizability of this study that cannot be overlooked. In this study implicit prosody is employed only to resolve ambiguity problems, and hence, strictly speaking, this study cannot make any claims that go beyond reading situations in which the reader is forced to choose between rival interpretations. To deal with this concern this study has to rely on the aforementioned research, which states, that implicit prosody might not be limited to syntactic ambiguity (Ashby, 2006; Ashby & Clifton Jr., 2005; Fodor, 2002; Steinhauer, 2003; Stolterfoht et al., 2007). Besides, one readily realizes the latency of non-literal meaning in each and every written utterance, which leads to the insight that ambiguity in silent reading may rather be the rule than the exception. (Just try to read almost any sentence in this text with a ridiculing intonation.)

Another issue of generalizability is the fact that orthography-phonology mapping differs between languages and writing systems (Perfetti & Liu, 2005) and therefore reading must be regarded as a cognitive process situated in cultural frameworks. Interestingly, there is growing evidence for the claim that the brain's reading network actually is formed depending on which language and writing system is acquired and in which way it is learned (Dehaene, 2009; Perfetti, Nelson, Liu, Fiez, & Tan, 2010). Thus, some caution is advisable with respect to the generalizability of the findings at hand beyond the group of readers represented by the sample of readers in this study, which simplified might be described as being composed of young, well-educated, skilled readers of the Swedish language. In the main experiment eight

participants had not learned Swedish during the first five years of their lives and since this study relies heavily on the ability to understand nuances of the Swedish language it might be suspected that this fact has biased the result of this study. As there was no effect of Swedish as native language in the two-way analysis of variance described in the result, this has been shown to be not true. With hindsight it becomes clear, however, that additional information about the participants' reading skills, such as reading impairments and cases of bilinguality, could have contributed further to the analysis of the results generated by this study. That individual differences have to be taken into consideration has also been underscored by the individual differences in conscious experience of inner voice in silent reading reported in this study.

Obviously, conducting an Internet-based experiment differs in some respects from laboratory-based experiments. Especially, the issue of control of extraneous variables needs to be addressed from a methodological point of view as the experimenter in the Internet-based experiment only has little control (i.e. by giving instructions and choosing materials and equipment) over the situation in which the participant chooses to run the experiment. In particular for between-subject designs noise produced by extraneous variables might state a serious threat to the validity of an experiment. This holds even for the experiment presented here. It is not farfetched to suspect that while participating in the experiment in a school or home environment participants could have been disturbed by voices of fellow students or family members during the learning and rating phases of the experiment, which might have resulted in poorer learning or in distorted ratings. To counteract the possible drawbacks of an Internet-mediated experiment two corresponding measures were taken. First, instructions were clear about finding both place and time to conduct the experiment

undisturbed and without interruptions. Second, participants were instructed to report any disturbing factors after the actual experiment. Of the 71 participants 23 made use of the possibility to comment on the course of the experiment, which can be seen as an indication for, that the participants in this study took responsibility for their participation. Only nine of the comments reported disturbing factors, such as technical problems with the video sequences, acoustical problems and in five of the cases disturbances by other persons or the television going in the background. Due to the marginal extent and the character of the disturbing factors it was decided to keep all participants' data records in the study. Of course, the approach of controlling the experiment by delegating part of the responsibility to the participant means to build research at least in part on trust. As Claire Hewson (2003, p. 293) puts it: "[...] a level of trust of participants on the part of the researcher will always be required, as is the case in traditional research." Since the start of Internet-mediated research a number of studies have demonstrated the validity of Internet-based research by successfully replicating laboratory-based studies (for a review see Birnbaum, 2004). Even though these studies can in no way be regarded as proof for the validity of the specific experiment at hand they are cited at this place, to support the trustworthiness of Internet-mediated research in general. In summarizing the research Gosling, Vazire, Srivastava and John also point to ecological validity as one of the advantages of Internet-mediated research as it is delivered to the participants' natural environments:

[...] more evidence is needed before we can be sure that the two methodologies [Internet- and laboratory-based] are consistent. Of course, if the two methods do yield inconsistent findings, it should not be concluded automatically that the

Internet method is the inaccurate one. Indeed, the real-world generalizability of findings from traditional methods often goes untested. (2004, p. 102)

By demonstrating, that talker-specific cues are employed during silent reading the present study corroborates the view advanced by Abramson (2007), Alexander and Nygaard (2008) and Kurby (2009) that talkers' voices are simulated while reading written messages, whereas this simulation must not be understood as tantamount to actual experience, but as a partial and deficient re-enactment of talker-specific information encoded in memory. In addition the results of this study indicate that simulations even may rely on memories of earlier simulations, which is in compliance with the findings of Abramson (2007). However the validity of these findings for talkers of other languages cannot be taken for granted. Furthermore it was shown, that talker-contingency in silent reading can be understood within the framework of the theory of situated simulation advanced by Barsalou (2003) and hence the results of this study give support to alternative views on language representation.

As these findings in some respects conform to research on transmodal information transfer (e.g. the abovementioned studies by Baumann, Koenecke and Schmidt (2007), Geiselman and Glenny (1977) von Kriegstein and Giraud (2006) and Pilotti, Gallo and Roediger (2000)) it should be investigated further, to what extent cognitive simulations could be regarded as a general principle underlying perceptual top-down processing, as has been proposed by Barsalou (1999, 2009). With specific regard to simulation in written communication future research should be extended to encompass even other modalities and communicational cues, e.g. facial expressions.

References

- Abramson, M. (2007). The written voice: implicit memory effects of voice characteristics following silent reading and auditory presentation. *Perceptual and Motor Skills, 105*(3 Pt 2), 1171-1186.
- Aeschbacher, C. (2008). *are you still awake...? : Personal Written Communication: From Early Modern English Letters to Electronic Communication of Today*. University of Zurich, Zürich. Retrieved from http://claudiaaeschbacher.ch/dissertation/Aeschbacher_Dissertation_full_text.pdf
- Alexander, J. D., & Nygaard, L. C. (2008). Reading voices and hearing text: Talker-specific auditory imagery in reading. *Journal of experimental psychology. Human perception and performance, 34*(2), 446.
- Aparicio, M., Gounot, D., Demont, E., & Metz-Lutz, M. (2007). Phonological processing in relation to reading: An fMRI study in deaf readers. *NeuroImage, 35*(3), 1303-1316.
- Ashby, J. (2006). Prosody in skilled silent reading: evidence from eye movements. *Journal of Research in Reading, 29*(3), 318-334.
- Ashby, J., & Clifton Jr., C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition, 96*(3), B89-B100.
- Ashby, J., & Martin, A. E. (2008). Prosodic phonological representations early in visual word recognition. *Journal of experimental psychology. Human perception and performance, 34*(1), 224.
- Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences, 22*(4), 577-609; discussion 610-660.

- Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes, 18*(5), 1. doi:10.1080/769813547
- Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology, 59*(1), 617-646.
- Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1521), 1281-1289. doi:10.1098/rstb.2008.0319
- Baumann, S., Koeneke, S., Schmidt, C. F., Meyer, M., Lutz, K., & Jancke, L. (2007). A network for audio-motor coordination in skilled pianists and non-musicians. *Brain Research, 1161*, 65-78. doi:10.1016/j.brainres.2007.05.045
- Birnbaum, M. H. (2004). Human Research and Data Collectinn via the Internet. *Annual review of psychology., 55*(1), 803-832.
- Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., & Blomert, L. (2009). Reduced Neural Integration of Letters and Speech Sounds Links Phonological and Reading Deficits in Adult Dyslexia. *Current Biology, 19*(6), 503-508. doi:10.1016/j.cub.2009.01.065
- Bowers, J. S. (2000). In defense of abstractionist theories of repetition priming and word identification. *Psychonomic Bulletin & Review, 7*(1), 83-99.
- Bryant, G. A., & Tree, J. E. F. (2005). Is there an Ironic Tone of Voice?. *Language & Speech, 48*(3), 257-278.
- Chafe, W. (1988). Punctuation and the Prosody of Written Language. *Written Communication, 5*(4), 395-426. doi:10.1177/0741088388005004001
- Creusere, M. A. (1999). Theories of Adults' Understanding and Use of Irony and Sarcasm: Applications to and Evidence from Research with Children.

Developmental Review, 19(2), 213-262.

Crystal, D. (1975). Paralinguistics. In J. Benthall & T. Polhemus (Eds.), *The body as a medium of expression : essays based on a course of lectures given at the Institute of contemporary arts, London* (pp. 162-174). London: Allen Lane.

Crystal, D. (2006). *Language and the Internet*. Cambridge: Cambridge University Press.

Dehaene, S. (2009). *Reading in the brain : the science and evolution of a human invention*. New York: Viking.

Derks, D., Fischer, A., & Bos, A. (2008). The role of emotion in computer-mediated communication: A review. *Computers in Human Behavior*, 24(3), 766-785.

Dews, S., & Winner, E. (1999). Obligatory processing of literal and nonliteral meanings in verbal irony. *Journal of Pragmatics*, 31(12), 1579-1599. doi:10.1016/S0378-2166(99)00005-3

Farah, M. J., Péronnet, F., Gonon, M. A., & Giard, M. H. (1988). Electrophysiological evidence for a shared representational medium for visual images and visual percepts. *Journal of experimental psychology. General*, 117(3), 248.

Fodor, J. D. (2002). Psycholinguistics Cannot Escape Prosody. In *Proceeding of the first conference on Speech Prosody*. Presented at the First conference on Speech Prosody, Aix-en-Provence.

Folk, J. R. (1999). Phonological codes are used to access the lexicon during silent reading. *Journal of experimental psychology. Learning, memory, and cognition*, 25(4), 892.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological bulletin*, 123(1), 71.

Frost, S. J., Landi, N., & Mencl, W. E. (2009). Phonological awareness predicts

- activation patterns for print and speech. *Annals of Dyslexia*, 59(1), 78-98.
- Garrod, S., & Sanford, A. J. (2005). Language in action. In *Cognitive psychology* (pp. 231-264). Oxford: Oxford University Press.
- Geiselman, R. E., & Glenny, J. (1977). Effects of imagining speakers' voices on the retention of words presented visually. *Memory & Cognition*, 5(5), 499-504.
- Gibbs Jr., R. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*, 34(4), 457-486.
- Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, 7(2), 92-97.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of experimental psychology. Learning, memory, and cognition*, 22(5), 1166.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review*, 105(2), 251.
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. *The American psychologist*, 59(2), 93.
- Hewson, C. (2003). Conducting research on the internet. *Psychologist*, 16(6), 290-294.
- Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological review*, 93(4), 411.
- Jobard, G., Vigneau, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2007). Impact of modality and linguistic complexity during reading and listening tasks. *NeuroImage*, 34(2), 784-800.
- Kent, C., & Lamberts, K. (2008). The encoding-retrieval relationship: retrieval as

- mental simulation. *Trends in Cognitive Sciences*, 12(3), 92-98.
doi:10.1016/j.tics.2007.12.004
- Ko, K. (1996). Structural Characteristics of Computer-Mediated Language: A Comparative Analysis of InterChange Discourse. *Electronic Journal of Communication/La Revue Electronique de Communication*, 6(3).
- Koch, P., & Oesterreicher, W. (1985). Sprache der Nähe ? Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In *Romanistisches Jahrbuch (1985)*, Romanistisches Jahrbuch (Vol. 36, pp. 15-43). Berlin, New York: Walter de Gruyter. Retrieved from <http://dx.doi.org/10.1515/9783110244922.15>
- Kreuz, R. J. (1996). The Use of Verbal Irony: Cues and Constraints. In *Metaphor : implications and applications* (pp. 23-38). Mahwah, N.J.: Erlbaum.
- von Kriegstein, K., & Giraud, A. (2006). Implicit Multisensory Associations Influence Voice Recognition. *PLoS Biol*, 4(10), e326. doi:10.1371/journal.pbio.0040326
- Kurby, C., Magliano, J., & Rapp, D. (2009). Those voices in your head: Activation of auditory images during reading. *Cognition*, 112(3), 457-461.
- Lee, Y., Binder, K. S., Kim, J., Pollatsek, A., & Rayner, K. (1999). Activation of phonological codes during eye fixations in reading. *Journal of experimental psychology. Human perception and performance*, 25(4), 948.
- Ling, R. (2005). The Sociolinguistics of SMS: An Analysis of SMS Use by a Random Sample of Norwegians. In *Mobile Communications* (pp. 335-349). Retrieved from http://dx.doi.org.miman.bib.bth.se/10.1007/1-84628-248-9_22
- Luo, Y., & Zhou, X. (2010). ERP evidence for the online processing of rhythmic pattern during Chinese sentence reading. *NeuroImage*, 49(3), 2836-2850.

- McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of experimental psychology. General*, *114*(2), 159.
- Meyer, K., & Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends in Neurosciences*, *32*(7), 376-382.
doi:10.1016/j.tins.2009.04.002
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech Perception as a Talker-contingent Process. *Psychological Science*, *5*(1), 42-47.
- Perfetti, C. A., & Bell, L. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and priming. *Journal of Memory and Language*, *30*(4), 473-485.
- Perfetti, C. A., & Liu, Y. (2005). Orthography to Phonology and Meaning: Comparisons Across and within Writing Systems. *Reading and Writing*, *18*(3), 193-210.
doi:10.1007/s11145-004-2344-y
- Perfetti, C. A., Nelson, J., Liu, Y., Fiez, J., & Tan, L. (2010). The Neural Bases of Reading - Universals and Writing System Variations. In P. L. Cornelissen, P. Hansen, M. L. Kringelback, & K. Pugh (Eds.), *The Neural Basis of Reading* (pp. 147-172). Oxford University Press US.
- Pexman, P. M. (2008). It's Fascinating Research: The Cognition of Verbal Irony. *Current Directions in Psychological Science*, *17*(4), 286-290.
doi:10.1111/j.1467-8721.2008.00591.x
- Pilotti, M., Gallo, D. A., & III, H. L. R. (2000). Effects of hearing words, imaging hearing words, and reading on auditory implicit and explicit... *Memory & Cognition*, *28*(8), 1406-1419.

- Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of experimental psychology. Human perception and performance*, *18*(1), 148.
- Regel, S., Coulson, S., & Gunter, T. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain Research*, *1311*, 121-135.
- Ren Gui-Qin, & Yang Yufang. (2010). Syntactic boundaries and comma placement during silent reading of Chinese text: evidence from eye movements. *Journal of Research in Reading*, *33*(2), 168-177.
- Rosenblum, L. D., Miller, R. M., & Sanchez, K. (2007). Lip-Read Me Now, Hear Me Better Later: Cross-Modal Transfer of Talker-Familiarity Effects. *Psychological Science*, *18*(5), 392-396. doi:10.1111/j.1467-9280.2007.01911.x
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological review*, *96*(4), 523.
- Simos, P. G., Breier, J. I., Fletcher, J. M., Foorman, B. R., & Castillo, E. M. (2002). Brain Mechanisms for Reading Words and Pseudowords: an Integrated Approach. *Cerebral Cortex*, *12*(3), 297-305.
- Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language*, *86*(1), 142-164.
- Steinhauer, K., & Friederici, A. D. (2001). Prosodic Boundaries, Comma Rules, and Brain Responses: The Closure Positive Shift in ERPs as a Universal Marker for Prosodic Phrasing in Listeners and Readers. *Journal of Psycholinguistic Research*, *30*(3), 267-295.

- Stolterfoht, B., Friederici, A. D., Alter, K., & Steube, A. (2007). Processing focus structure and implicit prosody during reading: Differential ERP effects. *Cognition*, *104*(3), 565-590. doi:10.1016/j.cognition.2006.08.001
- Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review*, *2*(3), 339-363.
- Traxler, M. J., & Gernsbacher, M. A. (2006). *Handbook of psycholinguistics*. Amsterdam ;: Elsevier. Retrieved from <http://www.scribd.com/doc/26814104/Handbook-of-Psycholinguistics#fullscreen:on>
- Utsumi, A. (2000). Verbal irony as implicit display of ironic environment: Distinguishing ironic utterances from nonirony. *Journal of Pragmatics*, *32*(12), 1777-1806. doi:10.1016/S0378-2166(99)00116-2
- Walsh, M., Möbius, B., Wade, T., & Schütze, H. (2010). Multilevel Exemplar Theory. *Cognitive Science*, *34*(4), 537-582. doi:10.1111/j.1551-6709.2010.01099.x
- Walsh, M., Schütze, H., Möbius, B., & Schweitzer, A. (2007). An Exemplar-Theoretic Account of Syllable Frequency Effects. In *ICPhS XVI*. Presented at the ICPhS XVI, Saarbrücken. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.72.6381>
- Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The Neural Integration of Speaker and Message. *Journal of Cognitive Neuroscience*, *20*(4), 580-591. doi:10.1162/jocn.2008.20054
- Wheat, K. L., Cornelissen, P. L., & Frost, S. J. (2010). During Visual Word Recognition, Phonology Is Accessed within 100 ms and May Be Mediated by a Speech Production Code: Evidence from Magnetoencephalography. *Journal of*

Neuroscience, 30(15), 5229-5234.

Whitney, P. (1998). *The psychology of language*. Boston: Houghton Mifflin.

Wu, L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination:

Evidence from property generation. *Acta Psychologica*, 132(2), 173-189.

doi:10.1016/j.actpsy.2009.02.002

Table 1

Simplified overview over the design of the main experiment

	Condition A	Condition B	Condition C
Learning phase	Video	Video	Written, context
	Marie talking to Anna, <u>Anna</u> <u>never</u> responds with irony	Marie talking to Anna, <u>Anna</u> <u>never</u> responds with irony	Marie talking to Anna, <u>Anna</u> <u>never</u> responds with irony
	to Karin, <u>Karin</u> <u>never</u> responds with irony in half of the dialogues	to Karin, <u>Karin</u> responds <u>with</u> irony in half of the dialogues	to Karin, <u>Karin</u> responds <u>with</u> irony in half of the dialogues
	▼	▼	▼
Rating phase		Written dialogues	
		Marie in dialogue with Anna, Anna responding Marie in dialogue with Karin, Karin responding <i>No other cues for irony comprehension provided but name</i>	
	▼	▼	▼
Expected outcome	Anna rated positive Karin rated <u>positive</u>	Anna rated positive Karin rated <u>negative</u>	Anna rated positive Karin rated <u>negative</u>

Table 2

Example of a dialogue used in the first validation study in three variants (translated from Swedish)

Variant	Dialogue
a. without instruction	A: Could you call the travel agency today? B: Well, I don't have anything else to do
b. with instruction "without irony"	A: Could you call the travel agency today? B (without irony): Well, I don't have anything else to do
c. with instruction "with irony"	A: Could you call the travel agency today? B (with irony): Well, I don't have anything else to do

Table 3

Arrangement of dialogue batches in the first validation study with regard to variants, groups and cycles

Variants			
	First cycle	Second cycle	
	a. without instruction	b. instruction “without irony”	c. instruction “with irony”
Group I	batches 1 and 2	batches 1 and 3	batches 2 and 4
Group II	batches 3 and 4	batches 2 and 4	batches 1 and 3

Table 4

Means of the mean ratings of the Y- and Z-selections of dialogues with 15 dialogues in each as well as means of standard deviations for those dialogues as obtained in validation study 1

without instruction						
Selection	Means		Standard Deviations			
	M	SD	M	SD		
Y	3.44	0.68	1.71	0.42		
Z	3.61	0.63	1.77	0.77		

with instruction						
Selection	without irony		with irony		range	
	Means		Means		Means	
	M	SD	M	SD	M	SD
Y	4.54	0.51	1.12	0.47	3.42	0.50
Z	4.59	0.77	1.18	0.44	3.41	0.86

Table 5

Means of the mean ratings of U-, S- and T-selections of dialogues as well as means of standard deviations for those dialogues as obtained in validation study 1

		without instruction					
		Means		Standard Deviations			
Selection	N of dialogues	M	SD	M	SD		
U	6	4.83	0.87	1.01	0.35		
S	6	3.63	1.04	1.46	0.43		
T	12	2.86	1.44	1.32	0.43		

		with instruction					
		without irony		with irony		range	
		Means		Means		Means	
Selection	N of dialogues	M	SD	M	SD	M	SD
U	6	5.11	0.58	0.94	0.29	4.17	0.81
S	6	3.96	0.69	1.44	0.17	2.52	0.73
T	12	3.84	0.75	1.46	0.54	2.41	1.03

Table 6

Example of a dialogue in two variants used in the second validation study (translated from Swedish)

	Variant	Dialogue
a.	without instruction	A: Could you call the travel agency today? B: Well, I don't have anything else to do
d.	with context	A: Could you call the travel agency today? B (having a very busy day): Well, I don't have anything else to do

Table 7

Flow of learning- and ratings sections for each condition in the main experiment as well as selections of dialogues used in the sections

Section	Phase	# d ^a	Reply	Select. ^b	Condition		
					A	B	C
1	Learning	3	Anna	T	video, no	video, no	context, no
		3	Karin	U ^c ,S	video, no	video, yes	context yes
2	Rating	3	Anna	Y	Rating	Rating	Rating
		4	Karin	Z			
3	Learning	3	Anna	T	video, no	video, no	context, no
		3	Karin	U ^c ,S	video, no	video, yes	context, yes
4	Rating	4	Anna	Y	Rating	Rating	Rating
		4	Karin	Z			
5	Learning	3	Anna	T	video, no	video, no	context, no
		3	Karin	U ^c ,S	video, no	video, yes	context, yes
6	Rating	4	Anna	Y	Rating	Rating	Rating
		3	Karin	Z			
7	Learning	3	Anna	T	video, no	video, no	context, no
		3	Karin	U ^c ,S	video, no	video, yes	context, yes
8	Rating	4	Anna	Y	Rating	Rating	Rating
		4	Karin	Z			

^a number of dialogues, ^b selections as result of the first validation study, ^c read with irony in the B-condition

Table 8

Statements about the conscious experience of inner voice (translation from Swedish)

While reading silently

- I read without an inner voice.
 - I lend an inner voice to the word in such a way that silent reading almost is like reading aloud
 - I do read with some kind of an inner voice. But it is not as clear and articulated as when I read aloud.
 - I sometimes read with an inner voice depending on the kind of text I read.
 - I vocalize certain words/sentences in my head, when words/sentences are a little more difficult.
 - I read different characters in fiction with different voices
 - I am not aware of hearing an inner voice.
-

Table 9

Conscious experience of inner voice

Almost like reading aloud	8
Not as articulated	15
Characters in fiction	4
Not aware	1

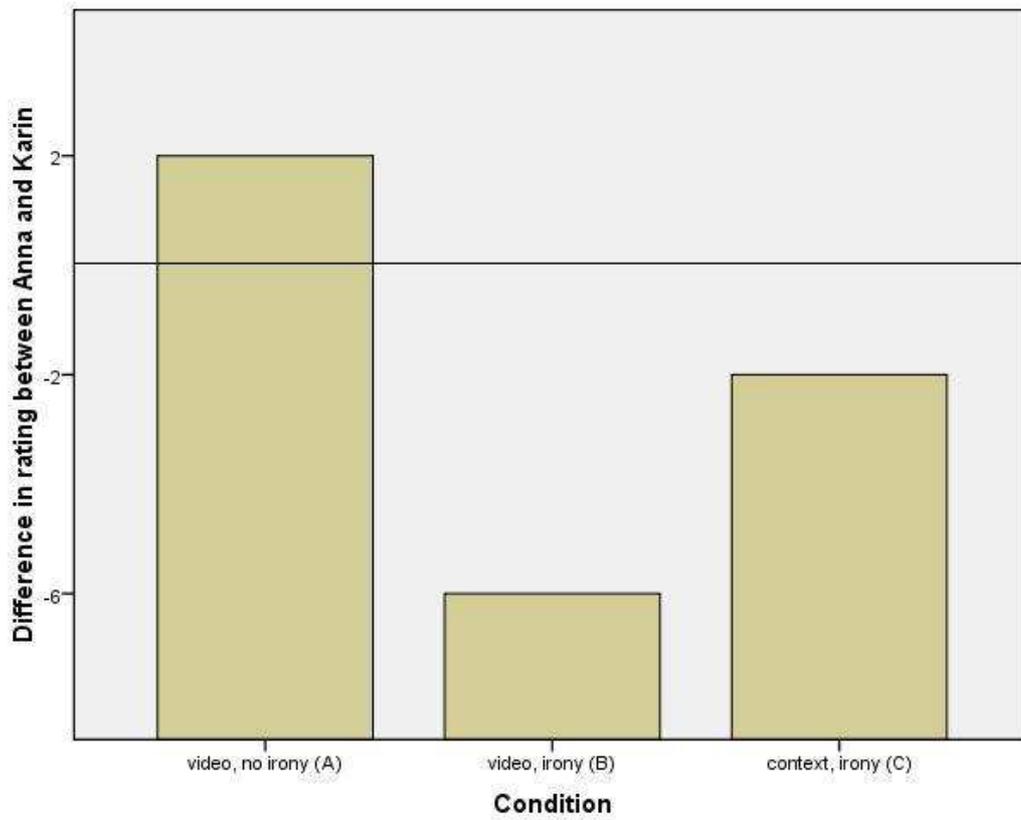


Figure 1. Differences in ratings between Anna and Karin. A negative value indicates, that participants rated Karin more negative than Anna; a positive value indicates the opposite relationship