Journal of Memory and Language 88 (2016) 70–86

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Prosody conveys speaker's intentions: Acoustic cues for speech act perception

Nele Hellbernd *, Daniela Sammler

Otto Hahn Group "Neural Bases of Intonation in Speech", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

ARTICLE INFO

Article history:
Received 7 May 2015
revision received 5 January 2016

Keywords:
Prosody
Intention
Speech acts
Acoustics
Pragmatics

ABSTRACT

Action-theoretic views of language posit that the recognition of others' intentions is key to successful interpersonal communication. Yet, speakers do not always code their intentions literally, raising the question of which mechanisms enable interlocutors to exchange communicative intents. The present study investigated whether and how prosody—the vocal tone—contributes to the identification of "unspoken" intentions. Single (non-)words were spoken with six intonations representing different speech acts—as carriers of communicative intentions. This corpus was acoustically analyzed (Experiment 1), and behaviorally evaluated in two experiments (Experiments 2 and 3). The combined results show characteristic prosodic feature configurations for different intentions that were reliably recognized by listeners. Interestingly, identification of intentions was not contingent on context (single words), lexical information (non-words), and recognition of the speaker's emotion (valence and arousal). Overall, the data demonstrate that speakers' intentions are represented in the prosodic signal which can, thus, determine the success of interpersonal communication.

Introduction

During conversations, humans regularly decode not only *what* is said but also *why* (Bühler, 1934; Grice, 1957; Wittgenstein, 1953). Depending on the latter, we may understand the same statement "It's hard to be punctual in the morning" as empathic concern, criticism, or simply as a matter of facts. Pragmatic theory posits that it is particularly the *why*—the communicative intention of the speaker—that drives the recipient's behavior and is the motive of communication. Yet, how intentions are (de)coded in interpersonal communication is still not fully understood. Contemporary pragma-linguistic theories posit that listeners identify the speaker's goal via pragmatic inference (Wilson & Sperber, 2012), taking conversation context and "common ground" (Clark & Carlson, 1981; Levinson, 2013; Stalnaker, 2002; Tomasello, 2005; Wichmann, 2002) into account. Alternatively, other studies seek to identify extralinguistic cues that reveal a speaker's intention, such as facial expressions (Fridlund, 1994; Frith, 2009; Parkinson, 2005), properties of biological motion (Di Cesare, Di Dio, Marchi, & Rizzolatti, 2015), or gestures (Bucciarelli, Colle, & Bara, 2003; Enrici, Adenzato, Cappa, Bara, & Tettamanti, 2011). The present study will focus on speech prosody—the tone of the voice—and will weigh its potential to convey communicative intentions.

The question of how interlocutors decode the *why* of an utterance is grounded in *action-theories of language*. In the middle of the 20th century, scholars like Bühler (1934), Wittgenstein (1953), or Grice (1975) recognized that language is more than strings of symbols that are understood

* Corresponding author at: Otto Hahn Group "Neural Bases of Intonation in Speech", Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany. Fax: +49 341 9940 2204.

E-mail addresses: hellbernd@cbs.mpg.de (N. Hellbernd), sammler@cbs.mpg.de (D. Sammler).

http://dx.doi.org/10.1016/j.jml.2016.01.001
0749-596X/

by retrieving their conventional, *coded* meaning. In their view, language is an *intentional action* and gains meaning through its employment. Utterances become instruments to influence the behavior of the interlocutor. The meaning of an utterance must be found in its underlying intention. It was Grice (1957) who particularly promoted the central role of intentions in communication. He advocated the idea that intentions drive speakers' behaviors (e.g., utterances) whose sole function is to have an effect on the addressee by virtue of having their intention recognized (cf. Levinson, 2006). Notably, the intention of the speaker—the *speaker meaning* in Grice's terms—not necessarily surfaces in the overt lexical content of the utterance, as shown in the example on punctuality above, but needs to be interpreted by the listener.

This idea later became central to speech act theory by Austin (1962) and Searle (1969) who considered utterances as actions—or *speech acts*—with specific interpersonal goals such as promising, apologizing, or warning. Like Grice, they claimed that speakers convey information on at least two levels: (1) the *propositional content* carrying the lexical meaning of *what* is said, and (2) the *illocutionary force* representing the action and speaker's intention—the *why*. As mentioned above, it is this second level—what the speaker is attempting to accomplish with a remark—that is thought to predominantly drive the interlocutor's (conversational) reaction. Notably, illocutionary force is often expressed implicitly (i.e. without the performative verb) or even indirectly, hence requiring some sort of inference on the part of the listener (Austin, 1962; Bach, 1994).

Interestingly, the notion of implicitness and indirectness conflicts with Grice's *cooperative principle* (1975), which describes principles for effective communication in conversation in four maxims. Following his *maxim of manner*, speakers ought to shape their utterances in ways that support the purpose of the conversation. Hence, speakers should produce unambiguous cues that make their intentions comprehensible to listeners. The fact that this seems often not to be the case but listeners still efficiently recognize the speaker's intent has fueled research on the cognitive and neural bases of comprehending communicative intentions. A great deal of work has focused on implicit speech acts, i.e. utterances that express the speaker's intention and illocutionary force without inclusion of the performative verb (e.g. "I will be there." expressing a promise without including the verb "promise"). These studies demonstrated the psychological reality of speech acts (Holtgraves, 2005), their automatic (Holtgraves, 2008a; Liu, 2011) and early recognition during conversation turns (Egorova, Pulvermüller, & Shtyrov, 2014; Egorova, Shtyrov, & Pulvermüller, 2013; Gisladottir, Chwilla, & Levinson, 2015), and their importance for conversation memory (Holtgraves, 2008b). However, despite their importance for understanding human communication, these studies remain incomplete in one particular way: They often rely on written linguistic material and, thus, miss out on extralinguistic cues that are usually available during natural spoken conversations. These cues comprise signals expressed via additional communicative channels like eyes, face, body, or voice and may render the speaker's intention less implicit and indirect than typically thought. The present study will focus on vocal acoustic cues, i.e., prosody, as one non-verbal channel in interpersonal conversation that may play an important role for speakers and listeners to express and recognize communicative intentions.

The term prosody refers to variations in pitch, loudness, timing, or voice quality over the course of an utterance (Warren, 1999) that can modify the communicative content of a message, both linguistically and paralinguistically (Bolinger, 1986). Linguistically, prosody has direct effects on the information structure of an utterance. It conveys, for example, semantic relationships (Cutler, Dahan, & van Donselaar, 1997; Wagner & Watson, 2010), disambiguates the syntactic constituent structure (Carlson, Frazier, & Clifton, 2009), and marks declarative vs. interrogative sentence mode (Sammler, Grosbras, Anwander, Bestelmeyer, & Belin, 2015; Schneider, Lintfert, Dogil, & Möbius, 2006; Srinivasan & Massaro, 2003). Paralinguistically, the "manner of saying" conveys additional information that goes beyond the linguistic content. Whether or not this includes intentions is a matter of debate (Bolinger, 1986) and will be topic of the present research.

Until now, most studies on paralinguistic prosody either focused on the speaker's emotion (Banse & Scherer, 1996; Bänziger & Scherer, 2005; Frick, 1985; Simon-Thomas, Keltner, Sauter, Sinicropi-Yao, & Abramson, 2009) or, more recently, on their attitude, for example, the politeness, confidence, or sincerity of the speaker (Jiang & Pell, 2015; Monetta, Cheang, & Pell, 2008; Rigoulot, Fish, & Pell, 2014) and often sought to determine links between the acoustics of the prosodic signal and the listeners' comprehension of the paralinguistic message. Although opinions diverge on whether prosody as such can convey meaning, i.e. without contextual information (see below) (Cutler, 1976; Wichmann, 2000, 2002), studies revealed distinct acoustic properties for the prosodic expression of different emotions (Banse & Scherer, 1996; Szameitat, Alter, Szameitat, Darwin, et al., 2009; Szameitat, Alter, Szameitat, Wildgruber, et al., 2009) and attitudes (Blanc & Dominey, 2003; Morlec, Bailly, & Aubergé, 2001; Uldall, 1960). Similarly, on the perception side, researchers showed that participants were able to identify the speaker's attitude (Morlec et al., 2001; Uldall, 1960) and emotion by prosodic differences alone, in verbal (Banse & Scherer, 1996; Morlec et al., 2001) and non-verbal utterances (Monetta et al., 2008; Sauter, Eisner, Calder, & Scott, 2010), in laughter (Szameitat, Alter, Szameitat, Darwin, et al., 2009; Szameitat, Alter, Szameitat, Wildgruber, et al., 2009), and to some extent even cross-culturally (Sauter, Eisner, Ekman, & Scott, 2010).

Compared to this active field of research, only little is known about the perceptual reality, relevance and effectiveness of prosodic cues in conveying *intentions*. We consider communicative intentions as the goals of interpersonal actions (e.g., language) that are meant to be recognized by the interlocutor and to influence her (conversational) reactions. This differentiates communicative intentions from basic emotions that do not necessarily need another person to be displayed, and attitudes that are not necessarily meant to purposefully influence conversation partners (Wichmann, 2000). Certainly, both