"Your Tone Says It All": The processing and interpretation of affective language ☆

Jennifer M. Roche [a,*], Brett Peters [b], Rick Dale [c]

[a] School of Health Sciences, Kent State University, United States
[b] Clinical and Social Sciences in Psychology, University of Rochester, United States
[c] Cognitive and Information Sciences, University of California, Merced, United States

Abstract

Pragmatic interpretation of intent is essential for successful communication. The current studies evaluate the impact of affective prosody on the processing and interpretation of affectively spoken language. A production study provided further evidence of talker variability in the production of the emotionally-laden categories of Innuendo, Irritation, Compassion and Neutral, indicating a great deal of within and between talker variability, as well as talker systematicity within affect categories. Despite this talker variability, in a listening task, participants were asked to categorize the intent of the talkers statements (from the production study) to determine the relative accuracy of responding, while also tracking the perception of intent as it unfolded over time (i.e., via computer Wii-mote x, y coordinates). The results from the online measurement of the perception of intent indicated that even though our listeners were accurate in categorizing intent (~70% mean accuracy), the "dynamic signature" of their responses was laden with a great deal hesitation and indecision for some, but not all talkers. This suggests that during the perception of intent, the cognitive system is flexible enough to handle talker variability, but during perception, uncertainty will change the manner in which the intent is processed.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The interpretation of intent often goes beyond a single word, and its explicit meaning. Discourse often has embedded meanings that require attention to context and the appropriate decoding of paralinguistic information to facilitate a felicitous response. For example, a listener must not only pay attention to prosodic cues (linguistic and affective), but must also attend to speaker specific cues, in hopes to prompt the listener with a means to appropriately respond given the context and the speaker's intentions.

A great deal of research has focused primarily on talker variability, showing that talkers variably produce speaking rates, have different levels of spoken word intelligibility, have a range in voice quality, are not always systematic in their vowel production (Bachorowski and Owren, 1999; Mullennix et al., 1989; Mullennix and Pisoni, 1990; Pisoni, 1992), and, at the most basic level, have biological differences in the vocal tracts that provide very strong cues to the gender of the speaker (e.g., due to vocal tract length; Goldstein, 1980; Nordström, 1977). Listeners also have little difficulty in determining the race and even age of talkers (Ryalls et al., 1997). Though these cues may not necessarily

---

* Corresponding author. Address: Speech Pathology & Audiology, School of Health Sciences, Kent State University, Kent, OH, United States. Tel.: +1 330 672 0244.
 E-mail address: jroche3@kent.edu (J.M. Roche).
 URL: https://www.jennyroche.com (J.M. Roche).

contribute to the interpretation of intent, they are likely to guide the listener in deciding how to act upon that intent. In fact, some researchers suggest that the variability that exists between talkers is a prominent and necessary component of speech perception (e.g., Newman et al., 2001).

However, at the pragmatic level, it is possible that other factors, such as cultural experiences, may also strongly shape how interlocutors produce and perceive affect (e.g., Hawk et al., 2009; Ishii and Kitayama, 2002; Kitayama and Ishii, 2003; Kitayama et al., 2006). A speaker's ability (or inability) to produce affect has been shown to produce a negative effect on a listener's ability to perceive affect properly (Mullennix et al., 2002). Therefore, cues to speaker identity and ability to produce affect have important social ramifications with regards to how the listener might address and interact with the speaker.

Since communication is often layered with affective information that interacts with psycholinguistic processes, affective paralinguistic cues may further promote socially acceptable behavioral responses and the understanding of hidden meanings (Attardo et al., 2005; Nygaard and Lunders, 2002). Considering social cues and a speaker's ability to produce affective language may shed light on how we easily (and sometimes not so easily) are able to correctly identify speaker affect in novel situations, and with new people. Therefore, the presence of affective cues in speech should help guide social exchanges, imbued with varying emotions and intent, as it is integrated in a rich social context. Talker variability in spoken word production may help the listener better decode the message, but it is possible that talker variability as it relates to affect perception may make the interpretation of the message more difficult as research has shown that elicited emotions are often a blend of several disparate emotions (Scherer and Ceschi, 1997).

Difficulty may arise during affect speech perception because both the talker and listener must be aware of how the affective cues influence language. If the affective cues are not salient, or are somehow misinterpreted, then conflict may arise (e.g., both parties feeling negatively towards the other because they misunderstood something about the situation). Therefore, it is of particular importance for speakers to pay attention to context and use appropriate cues, in hopes that the listener will properly integrate the relevant cues during the interpretation of intent. Given the centrality of these cues in everyday interaction, it is important to understand the underlying mechanisms involved in processing them. In the current study we capitalize on the fact that speakers produce affect differently and assess how affective talker variability impacts the way listeners perceive the speaker's intent.

2. Background

Communication is often driven by behaviors related to the expression of affect cues. The tendency to respond affectively is important to decrease social distance, and maintain and develop social relationships. Additionally, responding affectively may promote the coordination of social activities, provide cues to others about how to respond in a socially appropriate manner, and may help promote the interpretation of another's behaviors that help regulate interpersonal interactions (Fischer and Manstead, 2008; Fridlund, 1994; Hawk et al., 2009; Keltner and Haidt, 1999; Scherer, 1980, 1988, 1994; van Kleef et al., 2004). The interpretation of affect is often multimodal (e.g., facial, gestural, postural, and vocal; Guerrero and Floyd, 2006). Other sensory modalities are highly interactive among one another (e.g., auditory and visual information during speaking). However, vocal expressions of affect may have general detectability advantages over the other modalities, because their expression has the ability to draw attention "omni-directionally and over long distances" (Hawk et al., 2009; pp. 294). In the current paper, we focus on this auditory channel during higher-level spoken language acts that require the interpretation of intent beyond the literal meaning of the words spoken.

Since "we don't always say what we mean, or mean what we say," (Galloway, 1974), we may rely on the vocal cues to disambiguate our intentions (e.g., Attardo et al., 2005; Nygaard and Lunders, 2002). A significant amount of work has been conducted to evaluate how we produce and perceive affective cues in speech, and motivates our studies here. During vocal production, affect has been evaluated based on a number of emotion/affect categories (e.g., ranging from basic emotions to more subtle pragmatic categories like sarcasm; e.g., Cheang and Pell, 2008; Rockwell, 2000; Scherer, 1986, 2003; Scherer and Banziger, 2004). The categories have been extensively evaluated for their relevant acoustic correlates across talkers during the production of single-word utterances (e.g., see Bachorowski, 1999; Banse and Scherer, 1996; Leionenen et al., 1997; Scherer, 2003; Scherer and Banziger, 2004), in addition to a number of studies evaluating nonsense sentential structures (e.g., Banse and Scherer, 1996; Scherer et al., 2010).

Studies of affective prosody are usually carefully controlled for lexical and semantic content. The evaluation of single words is practical, because single words carry the majority of the affective prosodic variation and it has also been shown that most of the affective information is carried in the vowel (Kaiser, 1962). However, the interpretation of affective speech in natural settings minimally, at best, involves the integration of lexical, semantic and prosodic content towards the interpretation of intent (e.g., for a review of natural vocal expression see Scherer, 2003). Here we consider that a single word in an utterance may carry a greater degree of affective prosody, but the surrounding words (with the interaction of their meaning) may contribute to and also have prosodic markers necessary to decode intent, especially when contextual cues may be less salient (e.g., on a cell-phone, which could require listeners to compare featural information held in the pre-categorical acoustic sensory store; Crowder and Morton, 1969; MacMillan et al., 1988).