COMMENTARY

emotion review

Emotion Review
Vol. 13, No. 1 (January 2021) 51–56

ISSN 1754-0739
DOI: 10.1177/1754073920954288
https://journals.sagepub.com/home/emr

Comment: The Next Frontier: Prosody Research Gets Interpersonal

Marc D. Pell
School of Communication Sciences and Disorders, McGill University, Canada

Sonja A. Kotz
Department of Neuropsychology and Psychopharmacology, Maastricht University, The Netherlands
Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

Abstract

Neurocognitive models (e.g., Schirmer & Kotz, 2006) have helped to characterize how listeners incrementally derive meaning from vocal expressions of emotion in spoken language, what neural mechanisms are involved at different processing stages, and their relative time course. But how can these insights be applied to communicative situations in which prosody serves a predominantly interpersonal function? This comment examines recent data highlighting the dynamic interplay of prosody and language, when vocal attributes serve the sociopragmatic goals of the speaker or reveal interpersonal information that listeners use to construct a mental representation of what is being communicated. Our comment serves as a beacon to researchers interested in how the neurocognitive system "makes sense" of socioemotive aspects of prosody.

Keywords

emotion processing, pragmatics, social context, speech, vocal expression

Introduction

After decades of sporadic attention, empirical knowledge of how humans vocally communicate their emotions is progressing quickly, as showcased by work in this volume. The voice, it seems, has finally found its voice; coupled with psychoacoustic methods, new computational and neuroinvestigative approaches are helping to define the form and function(s) of emotional auditory signals and the neurocognitive apparatus that "senders" and "receivers" use to engage in emotional communication. The social and cultural context for expressing and interpreting emotional prosody is also under scrutiny (e.g., Laukka & Elfenbein, 2020), fostering new developments in theory. These converging efforts are promoting an enriched idea of vocal communication as a structured system (or multiple strategic communicative systems; Bryant, 2020) that operates "cheek-to-cheek" with language (Kotz & Paulmann, 2007), erasing historical views of emotional prosody as an unspecified melody that "gives colour" to what we say.

From a cognitive standpoint, a major gain that has been achieved in recent years is to establish a descriptive framework for testing how vocal emotion expressions are broken down, analyzed, and assigned meaning by listeners in running speech (Brück et al., 2011; Kotz & Paulmann, 2011; Schirmer & Kotz, 2006; Witteman et al., 2012). Among other factors, these models need to accommodate growing evidence that vocal expressions of basic emotion have unique "recognition points" over time (Pell & Kotz, 2011) as well as potentially distinct neural brain patterns (Giordano et al., 2018; Grandjean, 2020; Kotz et al., 2013). This temporal variability influences how the neurocognitive system responds to emotional attributes, and how they are used to understand a communicative situation. Spatial and temporal variability would tell us which brain areas respond *when* as vocal emotions are expressed. Neurophysiological data have been especially useful for developing hypotheses about the time course of effects due to emotional attributes in speech, highlighting three major stages at which vocal expressions are structured and assigned meaning during communication. As a springboard for advocating new directions in prosody research, we summarize principles and support for "three-stage models" of vocal expression analysis, taking the initial model of Schirmer and Kotz (2006) as our foundation.

Analysis of Vocally Expressed Emotions

As clinically focused research on emotional prosody transitioned to an information processing approach (Pell, 1998; Van Lancker & Sidtis, 1992), an attempt to formalize these data in a neurocognitive model was first proposed by Schirmer and Kotz (2006). A core feature of these (and more recent) models is that

Corresponding author: Marc D. Pell, School of Communication Sciences and Disorders, McGill University, 2001 Av. McGill College, 8th floor, Montréal, QC H3A 1G1, Canada.
Email: marc.pell@mcgill.ca

receivers "make sense" of vocal expressions during three consecutive structure-building phases.

Vocal expression analysis begins with a basic level of sensory encoding, which gives structure to the incoming acoustic stream in the form of an auditory gestalt. These operations define the physical properties and spectral composition of acoustic events, allowing human voices to be differentiated from music and other sound categories (Meyer et al., 2007; Rigoulot et al., 2015). These preattentive processes are quickly followed by a stage of "salience detection," which serves to attentionally highlight motivationally significant vocal features of the unfolding stimulus, such as emotionality (i.e., emotional vs. neutral utterances; Liu et al., 2012; Paulmann & Kotz, 2008), high arousal (Paulmann et al., 2013; Sauter & Eimer, 2009), and other relevant characteristics. Due to their endogenous properties and/or strategic importance to the receiver, *salient* vocal attributes are potentiated to a greater extent by the neurocognitive system at this stage. In addition to stimulus properties, registering the salience of vocal expressions pivots on the expectations, goals, and individual characteristics of the receiver (Garrido-Vásquez et al., 2018; Paulmann et al., 2013).

It can be said that early processing stages "prepare" vocal expressions for detailed cognitive appraisal by categorizing and tagging the potential relevance of the input; this yields a rough semantic encoding of vocal expressions that marks the type (nonlinguistic, speech-embedded) and discrete emotional qualities of the emerging stimulus (Pell et al., 2015). The time course of these operations resembles other forms of sensory processing (faces, written language) occurring within the first 100–300 milliseconds postonset of the vocal expression (Charest et al., 2009; Kotz & Paulmann, 2007). Measures of event-related potentials (ERPs) have traced these operations to amplitude differences in the N100 (basic level) and P200 (salience detection) components (see Paulmann et al., 2013). As we suggest in what follows, changes in these early components may be useful to study how prosodic contrasts of a more interpersonal nature influence "feedforward" processes that characterize the salience of the voice in a variety of communicative situations (see Grandjean, 2020).

A third, protracted period of cognitive analysis allows receivers to refine a mental representation of vocally expressed meanings and to specify their individual and contextual relevance to what is being communicated. Cognitive analysis of vocal expressions can serve multiple purposes and engage an array of (associative and inferential) processes, which dynamically evolve as voice information itself unfolds in time. These cognitive operations begin around 300 ms postonset of the vocal expression and continue in a sustained manner as the speech signal accumulates (Kotz & Paulmann, 2011). For example, listeners may recalibrate their initial representation of vocal meanings or activate graded details about their significance in reference to stored knowledge of vocal behaviour, social norms, and individual experience (e.g., previous interactions with the speaker). Data show that certain emotional expressions (e.g., anger; Pell et al., 2015) exert differential demands on these processes and require sustained monitoring, reflected in an increased late positive potential (or late positive complex [LPC]) in the 300–900 ms time window following stimulus onset. These data highlight a variable but critical time period in which vocally expressed meanings become "recognized," consciously accessible, and during which dynamic voice information can lead to an updating of mental representations of what is being communicated.

While certain cognitive operations hone a mental representation of the speaker's vocally expressed meaning, other processes register the interplay of salient vocal characteristics with meanings activated by the linguistic message (e.g., words or phrases which could bear on an emotional interpretation of the utterance). Vocal and lexical (word) meanings build up simultaneously during speech perception, but linguistic contrasts tend to be realized over shorter time scales (Poeppel, 2003). This means that as listeners construct a word-by-word representation of what is being communicated, initial representations of what the speaker is conveying through their vocal expression may have added cognitive effects at any number of potential "attachment" points in the utterance which promote comparative analysis of the two information sources (Hoeks & Brouwer, 2014). As such, researchers can examine language-related ERP components that index word retrieval and the contextual integration of sentence meaning (e.g., N400, P600) to infer how vocal expressions of emotion incrementally promote an understanding of the communicative situation in concert with what the speaker is saying. For instance, when emotional characteristics of prosody conflict with word valence or the semantic content of the message, N400-like deflections are observed in the 300–500 ms time window following the critical emotion-related word (indexing voice meaning agreement; e.g., Schirmer & Kotz, 2005; for fMRI evidence, see Kotz et al., 2015; Wittfoth et al., 2010). At slightly later time points, listeners may engage in processes to update a mental representation of what is being communicated based on the prosodic context (P600-like effects) and draw pragmatic inferences about the speaker, the speech act, or other dimensions of the communicative situation based on a combinatorial analysis of available cues (Brouwer et al., 2017). It should be noted that for many utterances, an analysis of vocal expressions often precedes an analysis of contingent word meanings (imagine someone saying, "I didn't make the team" expressed in a happy tone). The sequencing of information flow is therefore an important factor to keep in mind as we study the interplay of prosody and language, which is especially important as we look at how vocal expressions are used as a sociopragmatic device.

These details inform what mental processes listeners seem to undergo as they interpret the meaning of vocal expressions of emotion in speech. But how can these insights be applied to situations in which prosody serves a predominantly interpersonal function in social interactions? For example, how is prosody used when it conveys the speaker's level of certainty or marks their stance in relation to what they're saying? Or when the speaker's prosody marks their (im)polite attitude towards the receiver or another person? Long-standing views have considered facial expressions of emotion as social tools that serve

the various needs and goals of the sender (Crivelli & Fridlund, 2018; Fridlund, 1994; Russell et al., 2003). Well-articulated theories of the affective pragmatic functions served by spoken utterances have also been put forth (Caffi & Janney, 1994; Scarantino, 2017). Along these lines, we believe that examining vocal expressions beyond their biological functions as emotional signals promises to inform cognitive processing (e.g., three-step) models of vocal expression analysis, and ultimately to fill the gaps in our knowledge of how the human brain uses voice information to interpret communicative situations. Emotive communication involves the conventionalized expression of feelings, attitudes, and relational orientations towards topics, actions, and other people. In what follows, we consider three examples of how vocal expressions are treated when prosody is "emotive" rather than emotional, driven by the sociopragmatic goals rather than the biological needs of the speaker.

Confidence, Persuasion, and Trust

One fertile area to explore considers how prosody is used to communicate mental states of confidence, certainty, and commitment to messages transmitted in spoken language. Vocal speech cues provide graded information about a speaker's confidence level as they make a statement ("We will finish it during the pandemic!"), allowing listeners to estimate whether the message is exact, truthful, or will be performed. But how are vocal confidence expressions analyzed for meaning and used to form a holistic impression of what is being communicated?

These questions were recently evaluated in a series of ERP studies (Jiang et al., 2020; Jiang & Pell, 2015, 2016). Data show that, like emotional expressions, vocal expressions of confidence are rapidly assigned meaning from the acoustic onset of speech and refined with increased exposure to the input (Jiang & Pell, 2015, 2016). Vocally expressed confidence is robustly detected at the stage of salience detection, differentiating the P200 response to confident versus doubtful voices. Similar to when vocal emotion expressions are analyzed, the directionality of P200 effects depends on the listener's task focus (Paulmann et al., 2013); for example, while highly confident vocal expressions are more salient in certain contexts (Jiang & Pell, 2015), vocal cues marking the speaker's hesitation are more salient (increased P200 amplitude) when listeners must decide whether or not to trust the speaker (Jiang et al., 2020; see also Caballero & Pell, 2020). Following initial semantic differentiation of the stimulus (confident vs. doubtful), fine-tuning of a mental representation of the speaker's vocal confidence level appears to build up in the 300–700 ms time window, ensuring that a finer gradient of meaning is achieved ("she seems just slightly uncommitted to this idea"). Vocal confidence expressions can also promote a variety of late inferential processes about the communicative situation (e.g., that a confident speaker's goal is actually to manipulate or persuade the listener; Jiang & Pell, 2015; Jiang et al., 2017).

So far, empirical work suggests that three-step models can be usefully applied to situations in which the listener forms a mental representation of speaker (un)certainty or (un)commitment in the context of particular speech acts, such as statements of fact, opinion, or intentions. This research creates a starting point for broader work that examines how vocal expressions of confidence contribute to social cognitive processes related to competence, persuasion, and trust (e.g., Caballero & Pell, 2020; McAleer et al., 2014; van Zant & Berger, 2019), and which explore their neural underpinnings (e.g., Hellbrand & Sammler, 2018). Examining other communicative situations in which vocally expressed confidence is used as a pragmatic device—for example, in persuasive communication and marketing, to convince people of untruths in the political arena, or using cues of uncertainty solely to convey politeness, etc.—is an especially promising research area to explore further.

Attitudes and Stance

Another emotive function of prosody is to mark the speaker's stance towards particular aspects of the communicative situation (e.g., the message being conveyed, the receiver, or other people). Often, these cues point the receiver to intended nonliteral meanings through the "atypical" use of prosody while speaking (Rigoulot et al., 2020; Wilson & Wharton, 2006). For example, vocal expressions are instrumental in marking the valence of the speaker's attitude toward the receiver when expressing verbal irony ("You're such a wonderful cook!"), dictating whether or not a comment is meant to be taken as sincere and interpreted literally (Rigoulot et al., 2014). When compared to how vocal emotions are processed, how is a mental representation of vocal stance formed and combined with other emerging details of what the speaker is communicating in different social situations?

Recent work exemplifies how the brain registers vocal stance when speakers make interpersonal requests (e.g., "Lend me a nickel") expressed with polite (positive) or impolite (negative) attitudes towards the listener (Vergis et al., 2020). Differences in vocally expressed stance modulated P200 amplitudes in anterior brain regions from the acoustic onset of the request; this demonstrates that attention is quickly diverted to the potential significance of stance-related vocal cues, in a similar way that vocal expressions of emotion and other emotive contrasts are initially structured (Zougkou et al., 2017). Once again, it seems that increased salience may be assigned to either positive or negative vocal cues, depending on the evaluative goals of the listener and how emotive details are likely to contribute to an understanding of what is being communicated (see Mauchand et al., 2020). Also, like emotional expressions (Pell et al., 2015), there is evidence that vocal expressions of stance are cognitively monitored, and their meanings presumably updated in an ongoing manner over time (Rigoulot et al., 2020; Vergis et al., 2020). In Vergis et al.'s (2020) study of interpersonal requests, ERPs time-locked to a late-occurring target word ("Lend me a nickel") displayed significant interactive effects of the polite/impolite voice on both target word retrieval (N400) and on the contextual integration of emotive information with utterance meaning (indexed by P600-like effects occurring 500–800 ms postonset of the critical word; see also Rigoulot et al., 2014). While such data are still limited, they clearly exemplify the rapid uptake of

interpersonal cues from prosody and reveal situations in which stance-related information biases what is understood by the linguistic formulation of the message. Much more research could be undertaken in this area to illuminate how vocal cues encoding stance inform communicative meaning and what mechanisms are involved.

Social Identity Features

A third way that prosody research can get more interpersonal is by progressively defining and operationalizing social identity features of the voice in the communicative situation under scrutiny. From the aforementioned, it seems clear that vocal expressions dynamically reveal information about the mental state and interpersonal stance of the speaker, which guide interpretative processes in social interaction. However, vocal attributes simultaneously betray indexical features of the sender (age, sex, social group, etc.), which may enter into pragmatic interpretations of spoken language (van Berkum et al., 2008). Social characteristics of the receiver are also a determinant of how (and if) prosodic cues are incorporated into a representation of what is being communicated and which brain mechanisms are engaged (Jiang et al., 2018). Thus, understanding the impact of social identity features and social categorization processes on how communicative situations are interpreted represents another challenge for prosody researchers.

For example, when listeners perceive speakers as outgroup members based on having a regional or foreign accent, it appears that the nature and time course of processes which extract social meaning from their vocal expression are systematically altered (Jiang et al., 2020). In this study, in which participants rated how much they believed the speaker based on their vocally expressed confidence, P200 effects, which mark the relative salience of doubtful versus confident expressions from utterance onset, were only observed for in-group speakers who shared the same accent as the listener. Rapid social categorization of the speaker as an outgroup member reduced and delayed operations which extract confidence-related information from vocal expressions; downstream, this seemed to increase overall demands on contextual integration with language and to form a social impression of outgroup speaker believability (see Jiang et al., 2020, for details). These data imply that social identity information can influence vocal expression analysis prior to the stage of salience detection, with long-lasting effects on how communicative representations are built. Other analyses emphasize that individual characteristics of the listener (attitudes about particular outgroups, anxiety levels, tendency to trust, etc.) must also be taken into account. Given these results, new work which expands our knowledge of how social characteristics of speakers and listeners affect vocal expression analysis would be highly worthwhile.

Going Forward

The types of meanings that can be derived from vocal expressions extend beyond the basic emotions and are much more varied and socially rich. In addition to the excellent work that is being conducted on emotion, this comment encourages reflection on how three-stage models of vocal expression analysis can be applied when prosody acts not so much as a biological signal but as a cue to intended meanings of the communicative situation—a complex window into the speaker's feelings, attitudes, and relational disposition.

Our selective review suggests that some of the processing mechanisms that act on vocal expressions of emotion, especially those that contribute to significance detection, are similar to those that decipher vocally expressed confidence and the interpersonal stance of a speaker. However, exploring the interpersonal functions of prosody obliges researchers to look more carefully at how and when meaning is revealed through the interplay of vocal cues and linguistic features of an utterance (e.g., linguistic choices, activation of message-level representations), which engage different processing routines operating over distinct temporal domains. Increasingly, deriving meaning from socially complex speech signals in communication should be explored in natural (Verga & Kotz, 2019) and interactive communicative settings (Kotz & Schwartze, 2016). This would require not only the integration of social features coded in an utterance but also the timing of interactive speech in communication by considering dedicated temporal processing brain networks (Kotz & Schwartze, 2010; Kotz et al., 2018). A collective effort in this new direction will lead to valuable insights into human brain system(s) of vocal communication, giving an even more prominent "voice" to prosody-based research.

As part of this project, emotion researchers should also look more closely at situations in which a speaker strategically simulates vocal expressions of emotion to achieve distinct social or interpersonal effects (Scarantino, 2017). Imagine the person who uses their voice to feign intense anger when encountering an obviously insignificant obstacle, with the goal of communicating humour; or the speaker who vocally expresses fear with the purpose of mocking an anxious listener. The extent to which representations of the speaker's emotional state are used in these and other social contexts to "make sense" of the communicative situation, and the types of pragmatic inferences they generate, will add to our growing knowledge base of how vocal emotion expressions are analyzed in real-life encounters.

ORCID iD

Marc D. Pell https://orcid.org/0000-0002-9947-1552

References

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, *41*(S6), 1318–1352.

Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, *8*(4), 383–403. https://doi.org/10.1016/j.plrev.2011.10.002

Bryant, G. A. (2020). The evolution of human vocal emotion. *Emotion Review*, https://doi.org/10.1177/1754073920930791.

Caballero, J., & Pell, M. D. (2020). Implicit effects of speaker accents and vocally expressed confidence on decisions to trust. *Decision* http://dx.doi.org/10.1037/dec0000140.

Caffi, C., & Janney, R. (1994). Toward a pragmatics of emotive communication. *Journal of Pragmatics*, *22*(3–4), 325–373.

Charest, I., Pernet, C. R., Rousselet, G. A., Quiñones, I., Latinus, M., Fillion-Bilodeau, S., Chartrand, J. P., & Belin, P. (2009). Electrophysiological evidence for an early processing of human voice. *BMC Neuroscience*, *10*, Article 127. https://doi.org/10.1186/1471-2202-10-12

Crivelli, C., & Fridlund, A. J. (2018). Facial displays are tools for social influence. *Trends in Cognitive Sciences*, *22*(5), 388–399.

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. Academic Press.

Garrido-Vásquez, P., Pell, M. D., Paulmann, S., & Kotz, S. A. (2018). Dynamic facial expressions prime the processing of emotional prosody. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00244

Giordano, B. L., Whiting, C., Kriegeskorte, N., Kotz, S. A., Belin, P., & Gross, J. (2018). *From categories to dimensions: Spatio-temporal dynamics of cerebral representations of emotion in voice*. bioRxiv. https://doi.org/10.1101/265843

Grandjean, D. (2020). Brain networks of emotional prosody processing. *Emotion Review*, https://doi.org/10.1177/1754073919898522.

Hellbrand, N., & Sammler, D. (2018). Neural bases of social communicative intentions in speech. *Social Cognitive and Affective Neuroscience*, *13*(6), 604–615. https://doi.org/10.1093/scan/nsy034

Hoeks, J., & Brouwer, H. (2014). Electrophysiological research on conversation and discourse processing. In T. Holtgraves (Ed.), *Oxford University handbook of language and social cognition* (pp. 365–386). Oxford University Press.

Jiang, X., Gossack-Keenan, K., & Pell, M. D. (2020). To believe or not to believe: How voice and accent information in speech alter listener impressions of trust. *Quarterly Journal of Experimental Psychology*, *73*(1), 55–79.

Jiang, X., & Pell, M. D. (2015). On how the brain decodes vocal cues about speaker confidence. *Cortex*, *66*, 9–34.

Jiang, X., & Pell, M. D. (2016). Neural responses towards a speaker's feeling of (un)knowing. *Neuropsychologia*, *81*, 79–93.

Jiang, X., Sanford, R., & Pell, M. D. (2017). Neural systems for evaluating speaker (un)believability. *Human Brain Mapping*, *38*, 3732–3749.

Jiang, X., Sanford, R., & Pell, M. D. (2018). Neural architecture underlying person perception from in-group and out-group voices. *NeuroImage*, *181*, 582–597.

Kotz, S. A., Dengler, R., & Wittfoth, M. (2015). Valence-specific conflict moderation in the dorso-medial PFC and the caudate head in emotional speech. *Social Cognitive and Affective Neuroscience*, *10*(2), 165–171. https://doi.org/10.1093/scan/nsu021

Kotz, S. A., Kalberlah, C., Bahlmann, J., Friederici, A. D., & Haynes, J. D. (2013). Predicting vocal emotion expressions from the human brain. *Human Brain Mapping*, *34*(8), 1971–1981.

Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, *1151*, 107–118.

Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistic Compass*, *5*(3), 108–125.

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The evolution of ryhthm processing. *Trends in Cognitive Science*, *10*(12), 896–910. https://doi.org/10.1016/j.tics.2018.08.002

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Science*, *14*(9), 392–399. https://doi.org/10.1016/j.tics.2010.06.005

Kotz, S. A., & Schwartze, M. (2016). Motor timing and sequencing in speech production: A general-purpose framework. In S. Small & G. Hickok (Eds.), *Neurobiology of language* (pp. 717–724). Elsevier Press.

Laukka, P., & Elfenbein, H. A. (2020). Cross-cultural emotion recognition and in-group advantage in vocal expression: A meta-analysis. *Emotion Review*, https://doi.org/10.1177/1754073919897295.

Liu, T., Pinheiro, A. P., Deng, G., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *Neuroreport*, *23*, 108–112.

Mauchand, M., Caballero, J., Jiang, X., & Pell, M. D. (2020) Submitted. Immediate online use of prosody reveals the ironic intentions of a speaker: Neurophysiological evidence.

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say "hello"? Personality impressions from brief novel voices. *PLoS ONE*, *9*(3), Article E90779. https://doi.org/10.1371/journal.pone.0090779

Meyer, M., Elmer, S., Baumann, S., & Jancke, L. (2007). Short-term plasticity in the auditory system: Differential neural responses to perception and imagery of speech and music. *Restorative Neurology and Neuroscience*, *25*(3–4), 411–431.

Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, *4*, Article 345. https://doi.org/10.3389/fpsyg.2013.00345

Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *Neuroreport*, *19*, 209–213.

Pell, M. D. (1998). Recognition of prosody following unilateral brain lesion: Influence of functional and structural attributes of prosodic contours. *Neuropsychologia*, *36*(8), 701–715.

Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE*, *6*(11), Article e27256. https://doi.org/10.1371/journal.pone.0027256

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, *111*, 14–25.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time." *Speech Communication*, *41*, 245–255.

Rigoulot, S., Fish, K., & Pell, M. D. (2014). Neural correlates of inferring speaker sincerity from white lies: An event-related potential source localization study. *Brain Research*, *1565*, 48–62.

Rigoulot, S., Pell, M. D., & Armony, J. L. (2015). Time course of the influence of musical expertise on the processing of vocal and musical sounds. *Neuroscience*, *290*, 175–184.

Rigoulot, S., Vergis, N., Jiang, X., & Pell, M. D. (2020). Neurophysiological correlates of sexually evocative speech. *Biological Psychology*, *154*, Article 107909. https://doi.org/10.1016/j.biopsycho.2020.107909

Russell, J. A., Bachorowski, J.-A., & Fernandez-Dols, J.-M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology*, *54*, 329–349.

Sauter, D. A., & Eimer, M. (2009). Rapid detection of emotion from human vocalizations. *Journal of Cognitive Neuroscience*, *22*, 474–481.

Scarantino, A. (2017). How to do things with emotional expressions: The theory of affective pragmatics. *Psychological Inquiry*, *28*(2–3), 165–185.

Schirmer, A., & Kotz, S. A. (2005). ERP evidence for a sex-specific Stroop effect in emotional speech. *Journal of Cognitive Neuroscience*, *15*, 1135–1148.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Science*, *10*(1), 24–30.

Van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591.

Van Lancker, D., & Sidtis, J. J. (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research*, *35*(5), 963–970.

Van Zant, A. B., & Berger, J. (2019). How the voice persuades. *Journal of Personality and Social Psychology*, *118*(40), 661–682.

Verga, L., & Kotz, S. A. (2019). Putting language back into ecological communication contexts. *Language, Cognition, and Neuroscience*, *34*(4), 536–544.

Vergis, N., Jiang, X., & Pell, M. D. (2020). Neural responses to interpersonal requests: Effects of imposition and vocally expressed stance. *Brain Research*, *1740*, Article 146855. https://doi.org/10.1016/j.brainres.2020.146855

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics*, *38*(10), 1559–1579.

Witteman, J., van Heuven, V. J., & Schiller, N. O. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia*, *50*(12), 2752–2763. https://doi.org/10.1016/j.neuropsychologia.2012.07.026

Wittfoth, M., Schröder, C., Schardt, D. M., Dengler, R., Heinze, H. J., & Kotz, S. A. (2010). On emotional conflict: Interference resolution of happy and angry prosody reveals valence-specific effects. *Cerebral Cortex*, *20*(2), 383–392. https://doi.org/10.1093/cercor/bhp106

Zougkou, K., Weinstein, N., & Paulmann, S. (2017). ERP correlates of motivating voices: Quality of motivation and time-course matters. *Social Cognitive and Affective Neuroscience*, *12*(10), 1687–1700.