

# A QUALITATIVE ANALYSIS OF AFFECT SIGNS IN TELECOMMUNICATION DIALOGUES

Edmunds Vanags  
University of Latvia, Latvia

## ABSTRACT

There is growing evidence in the science of psychology that affective phenomena are not homogeneous and that their manifestations may vary across cultures and under the influence of contextual and demographic factors. Given that there is no clear universal expression of emotions and mood in human behavioral processes, it is necessary to continue to study the heterogeneity of the observed features in language and speech. This qualitative study analyzes the dialogues of 40 individuals in the field of telecommunications and, using a content analysis and phenomenological approach, describes lexical and non-lexical signs that could indicate features of affect. It can be observed that complete saturation has not been obtained within the framework of these data, which may indicate a wide variation of verbal and non-verbal affect features at both intra-individual and inter-individual levels and indicate different possible dialects of affect features. In addition, inter-rater reliability was determined and its results suggest that the determination of affective features may be subjective, contextual, in the absence of predefined reference criteria even in valence and activation dimensions of core affect.

**Keywords:** *affect features, lexical features, non-lexical features, paralinguistics.*

## Introduction

Increasingly, in the field of research on affect features, there is evidence that these traits are not clearly universal and can be observed in a wide variety in the research of facial expressions, lexical units of language and sound units of speech and recognizing these features is better within one's group, culture, and specific contexts (Barrett et al., 2019; Gendron et al., 2014; Paulmann & Uskul, 2014). Analyzing only the acoustic signals, there are no clear signs that indicate signs of affect in human speech (Scherer & Bänziger, 2004). In addition, meta-analyses point to a number of problems in this field of research, such as the vast majority of data being in English-speaking samples, based on the out of date theory of basic emotions, narrowing the search for traits to 6 or even less emotion categories

as well as data showing that there are no universal biological mechanisms in affect processes (Bağ, 2016; Clark-Polner et al., 2017). There are no consistent and constant acoustic signs associated with affective signs and there are not sufficiently reliable measurements to help measure them (Mauss & Robinson, 2009). In general, there are some universal features, for example, for some categories of emotions, but mostly there are culture-specific features that would allow to recognize the signs of affect in human speech and the categories of these traits, like emotions, can have rather fuzzed boundaries with large variation in coverage (Cowen & Keltner, 2017; Gendron et al., 2014; Paulmann & Uskul, 2014).

Research suggests that the features of emotional prosody – tone, rhythm, timbre, which is achieved through various modulations of the human voice, – allow another person to determine or recognize to a certain extent in different cultures both different categories of emotions (at least 12) and wider dimensions of affect – valence and activation (Cowen et al., 2019; Russell, 1994). Prosody has a different set of features in staged and ecologically true environments (Erickson et al., 2006; Jürgens et al., 2013) and the process of learning and expressing emotions is very closely linked to cultural factors (Matsumoto, 1989). Emotions are best recognized in one's own cultural environment (Pell et al., 2009) but also in the same national, ethnic, or regional group suggesting an in-group advantages (Elfenbein & Ambady, 2002).

The differences in the signs of affect in human speech are also related to significant features of the lexical properties of the language, for example, users of different languages in the same and different cultures may use different emotional labels (Mesquita & Walker, 2003) and languages may have different dictionaries of emotion and mood labels, words (Matsumoto & Assar, 1992). At the same time, it has been observed that speech sounds information may differ in different languages, such as intonation, rhythm with which different signs of affective state are expressed (Scherer et al., 2001). Researchers point out that, given the diversity of these signs of affect in a variety of factors, it is important to study all possible information channels and modalities, not limited to, for example, facial expressions or linguistic features (Elfenbein & Ambady, 2002; Katz & Assmann, 2019).

There are various theoretical attempts to describe the variation and heterogeneity of these observed affect features, and one of them is the Dialect Theory, which suggests that, like languages, there are different dialects, and non-verbal dialects are also seen in groups of individuals in terms of emotions, moods (Elfenbein, 2017; Tomkins & McCarter, 1964). This theory is partially supported by data showing that there are the above-mentioned advantages of in-groups in recognizing other people's signs of affect in speech parameters (Elfenbein, 2017; Elfenbein & Ambady, 2002). A rich

field of research suggests that emotions are expressed not only in a wide variety, but also using repertoires of different techniques, such as communicating various vocal cries, bursts, short non-word utterances, laughs, moans, sighs, beeps and many other ways (Laukka et al., 2016; Sauter et al., 2010; Simon-Thomas et al., 2009). Thus, it can be thought that these broad manifestations of affect signs function in society and cultures functionally as a “grammar of social interaction” (Eibl-Eibesfeldt, 2007) and researchers cannot ignore specific cultural, social and contextual factors that encourage individuals to use sometimes similar but mostly different body, voice, facial and other component repertoires (Barrett et al., 2011; Keltner et al., 2019), to provide direct and indirect information on affective states.

For example, interesting observations suggest that individuals with lower socioeconomic status are more focused on the social context, but less on contextual information, recognizing the affective states of other individuals (Kraus et al., 2010). If we look at the process of recognizing states of affect, then instead researchers observe that recognition as such cannot be an objective process (because there are no objective, nomothetic features), so the process of recognition should rather be seen as a psychological construction process in which the signal receiver (speech sounds, words, face), gives his or her own interpretation, using his or her experience and knowledge of culture, context and other variable factors (Doyle & Lindquist, 2017).

Taking into account the arguments listed above and the fact that so far no description of the observed affect features has been performed in Latvian, it is necessary to study these features with the help of a qualitative study, so to speak, observing the phenomenon from a phenomenological point of view. On the one hand, this study analyzed human speech and specific units of speech – lexical, non-lexical, as well as other audible units (prosody units), which may indicate that specific content was analyzed. Thus, it can be assumed that such a study examines how emotions are conceptualized in speech and how it can be perceived by the person listening, analyzing the speech (researcher or another listener). Such an approach could be classified as social constructivist analysis (Averill, 1985). As is well known, content analysis is a method of research that aims to draw reliable and valid conclusions from text or other content about the context in which these items are used (Krippendorff, 2004). Thus, it can be assumed that both content analysis and phenomenological approach have been used in this study, studying the experience of one person and conceptualizing possible signs of affect speech. The contribution of phenomenological research is an opportunity to obtain the widest possible description of features (O’Leary, 2017). In the process of qualitative research, historical, social and situational factors that influence the researcher’s views

and work cannot and should not be avoided, thus gaining the opportunity to describe the phenomenon from as different perspectives as possible (Denzin & Lincoln, 2018). Phenomenological reflection, also called “intentional analysis,” allows to conclude that “human experience is practical, embodied, affective, social, linguistic, and temporal” (Wertz, 2011).

The main question of this qualitative research is: what signs of affect (emotions, moods) can be perceived in telecommunication dialogues? In addition, the question was raised about the inter-rater reliability of various annotators for one of the features of affect – valence and activation.

## **Methodology**

### **Participants**

In this study, telecommunication dialogues between 44 customers and 3 operators (1 man, 2 women) were analyzed. Recorded conversations have been analyzed, so demographics are unknown and cannot be ascertained. Purposive sampling was used (Welman & Kruger, 1999) to identify the primary participants: telecommunications customers. Sample were based on the purpose of the research (Babbie, 1995; Greig & Taylor, 1999; Schwandt, 1997), looking for those who “have had experiences relating to the phenomenon to be researched” (Kruger, 1988). At least 10 participants or research subjects as sufficient to reach saturation (Boyd, 2001; Creswell, 1998) for a phenomenological study. All annotations were performed by one evaluator, only a few were evaluated by five evaluators to determine inter-rater reliability.

### **Data collection and analysis**

Within the framework of a qualitative study, a detailed annotation of 40 dialogues (actual telephone conversations in call centers) was completed, creating 10 synchronous layers or levels. Annotations were created by Annotation Pro software (Klessa & Wydział Neofilologii (Uniwersytet im. Adama Mickiewicza ; Poznań), 2016). In these layers for each speaker were described and annotated linguistic, paralinguistic units that could have an affective meaning. Layers were created in several iterative listening sessions. After each hearing, it was decided whether to create a new layer that contains categorically different features. Gradually, the number of layers increased to 10, of which two layers (one for the client, one for the operator) was created to obtain accurate speech transcriptions, preserving the mistakes made by the speakers. Linguistic units were annotated in next two layers (one for the client, one for the operator) using a circular affect model (Barrett & Russell, 1999). Next two layers for emotion words (one for the client, one for the operator) using a Latvian thesaurus, and two layers

for paralinguistic units (speech changes, vocal bursts, laughter, breathing changes, etc.). Finally, two layers were created (one for the customer, one for the operator) in which specific linguistic units were identified that could directly, semantically indicate signs of affect. Rich descriptions of phenomena and their settings (Bentz & Shapiro, 1998; Kensit, 2000) were made. The dialogues were listened to several times and in each layer of the annotation it was written what emotions could be heard in the speech, what non-verbal features could be heard in the speech and how each segment should be assessed using circular affect dimensions – valence and activation. In addition, it was noted which words or semantic units could directly and unambiguously indicate the affective state (for example, if the client said he was desperate). There were no pre-defined categories or criteria in this process. The conceptualization of the features of the affect was done according to author's experience, feelings, with the assumption that "it could be" which corresponds more closely to the phenomenological approach. Deep subjectivity cannot be avoided by describing the observed phenomenon in this way. Such annotation of speech segment emotion labeling, affect dimension labeling and synchronous paralinguistic features could allow to find multimodal feature profiles. Besides were made observational notes, theoretical notes, methodological notes, analytical memos (Bailey, 2007). In iterative process of annotation memoing were done (Miles & Huberman, 1994) – field notes recording what the researcher hears, sees, experiences and thinks in the course of collecting and reflecting on the process. To answer the additional question of the study, inter-rater reliability determination was performed for one annotation layer – valence and activation dimensions. To determine reliability, several dialogues were annotated by 5 annotators. The annotators were told about the annotation process, but no examples of annotation were shown (blind approach). The annotators listened to specific dialogues and performed valence and activation assessments for each segment using the Annotation Pro environment, which were used in the analysis. The results are described in more detail in the subsection.

## Results

A total of 378,28 minutes of dialogue ( $M = 9,46$   $SD = 3,38$ ) were annotated and analyzed, describing 17,638 segments. Looking at what is described in all layers, there are linguistic, paralinguistic (voice parameters), as well as other audible signs that could indicate the affective states of the clients – emotions and mood. For example, customer dissatisfaction may be manifested by a variety of verbal, nonverbal traits (see Table 1), and traits vary not only between individuals but also within a single individual's observed repertoire.

## Emotions

Particular emotions were annotated from all segments using different emotion labels. The names of the emotions or the words denoting them were fixed based on the annotator's experience, feelings, current understanding, without pre-defined criteria. Several emotions were repeatedly marked many times, such as "dissatisfaction" "anxiety" "hope" "surprise" "guilt" "despair" "doubt" but most rarely such as "disappointment" "regret" "slight confusion" "unpleasant surprise". This may be due to the assessor's inability to find the exact word of the emotion. It can be observed that there are mostly emotions characterizing negative situations, which could be related to the content of specific conversations. In most conversations, customers are dissatisfied with some aspect of the service and want a solution.

Some emotions were described in a number of words, suggesting confusing situations when it is not really clear which is decisive or whether the emotions are blended, such as "helpless anger". Some emotions were accompanied by different adjectives that indicate their different intensity, such as "deeply unhappy," "a little anxious," "a little confused," "formally kind."

Some emotions were also presented differently, conflicting in terms of valence, such as "sad but constructive" "sad but determined" "unpleasant surprise".

With the exception of repetitive emotions and those that could be synonymous in meaning, there are 125 emotion labels that denoted different segments in telecommunication dialogues. Most of them are related to negative affect states. The following sections will look at lexical and non-lexical signs that may be related to the observed emotions.

## Lexical features

485 lexical items (phoneme, word or phrase) were identified in all conversation segments, which could indicate the affective state or its changes. In all 40 dialogues, only 2 clients use direct, unambiguous words that indicate their affective state, such as: "I'm afraid," "I doubt." The other lexical items used do not semantically have a direct meaning that would indicate the client's affective state. Several words, phrases such as "problem" "mismatch" "difficulty" "crazy" "somehow stupid" "idiots" "beyond nowhere" could have negative meanings, but they can also imply for the affect of the individual. There are also words and phrases that are semantically positive words, such as "all the best" "you helped me a lot" "great" "funny" but with the help of intonation, prosody, their meaning is probably communicated as negative. It is often said in dialogues that the client cannot do something or does not know how to do it, it can indicate his affective state. Many lexical items have different meanings, depending on the

context of the conversation. For example, “it” can be used repeatedly, in one place with a more negative meaning, in another with a more positive meaning. The same word can be repeated in one phrase, such as “so, so, so”. Similarly, the words “either” “and” “vo” “or” “it” “this” “no” “no” “if” “yes” and their quantity, combination and syntactic techniques may be of affective importance (Tab. 1).

**Table 1.** Observed various affective features in linguistic or lexical, paralinguistic (and mixed) items related to customer’s dissatisfactions or complaints (in Latvian)

Features		
oiai.. ar nopūtu	nē nu	Tfuu-u-u-ūū
pfff... (ar nopūtu)	nu jā	hum, hum mmm
elpas ievilkšana un aizturēšana	nūūo	elpas ievilkšana ar āā
skaļa elpas ievilkšana caur degunu	tāām, tamm	mņū, mņe
klakšķis ar m	ja	Mmmhu..
mņņēē	njā	hu, hu, huuu
ēēūūū	un un un un un	Mhmēū
Eee, ar zemu, lēnu frekvenci, kā rūcot rīklē	ee	mēles klakšķis ar ieelpu
ummm	vai vai	Elpas ievilkšana ar tsss (ar nozīmi, ka ir sāpīgi, bet neko darīt)
Mmmmm (rūcoši, neizpratnē)	vienkārši (seko darbības vārds)	hmm
Nopūta ar sākuma burtu B	tā tā tā	nnn.. nnn..
mhm	jāā-a	gaisa ievilkšana caur nāsīm, nošņaukāšanās
aamm	vārda pirmais burts tiek ilgi vilkts (piemēram, “nnn-neibraucu..”)	Mmm-e-m.
Āāhh.. ar nopūtu	jā, nē nu	skaļi ievēl elpu un aiztur
bum, bum, bum	tātad-e (ē)	gaisa ievilkšana caur degunu
āāā	nu tad!	nopūta ar balsīgu izskaņu (ehme)
oua!	vot	mēles klakšķis
ēūufff (nopūta)	vot tā man	nepacietīga elpas ievilkšana ar nelielu klakšķi
mhm, mhm (dubultais)	tā tā tā man	oi ar nopūtu un šņākoņu beigās

The analysis of lexical units suggests that each individual has expressions, phrases that, after repeated use, give the impression that they carry an affective meaning. Some units are for multiple people, such as “well, well,” “no, no, no,” but not for most.

### **Non-lexical features**

More than 70 different non-lexical features were observed in the annotations, which could indicate the current affective state or its fluctuations in the speaker’s speech. Non-lexical signs include: breathing (inhalation, exhalation, sighs, duration, length, severity, rapidity and accompanying sounds such as snoring, moaning, muttering), articulation (blurred, stressed, syllable, stretched, fainting), speech volume (quieter, louder, sudden changes in volume), intonation (low, high, sudden changes, ascending, descending, urgent, falling, revival, sarcastic, poisonous), rhythm (steady, jerky, with sudden changes), laughter (nervous, short, with exhalations, inhalations, formal, played, natural), coughing, pausing, changes in speech rate, swallowing saliva, tongue clapping, perseverances. It can be subjectively observed in the speeches that these features can be related to different affective states, both negative and positive, and their use may differ not only between different individuals, but also in the repertoire of one individual. For example, a change in the volume of one speaker’s speech in one passage of the dialogue is associated with joy because a solution has been found to his problem, but in another passage with dissatisfaction that the operator does not know the required answer. For another client, inhaling air through the nose may indicate irritation, numbness, while nasal breathing by another speaker may indicate difficulty breathing. Only the tongue clicks used by the speaker serve as a confirmation and on average may indicate a positive affect, the tapping of another speaker’s tongue may indicate overt dissatisfaction.

In general, it can be observed that speakers use different repertoires of features and their use strongly depends on the dynamics of conversation, contextual factors, as well as the methods of other modalities used, which suggests that each speaker has his own specific dialect of affective expressions.

### **Annotator memos**

In the process of annotation it becomes clear that the evaluation of the same lexical, non-lexical units using the circular affect model and emotion words is not the same. The assessment of the core affect is broader, more ambiguous based on personal feeling, while emotions are attributed based on the pragmatic aspects of speech. In addition, possible words of emotion were named based on subjective experience rather than pre-determined



criteria. During the conceptualization, there was no evidence that such an approach would lead to a wider set of features. This probably suggests that the two rating systems are not parallel, but possibly involve differently affective processes. Observations suggest that the current mood of the annotator itself can certainly influence the ratings that can be deduced from several iterations by listening to the same passage. Repeated listening also makes it possible to observe that when you hear the same passage of speech for the first time and for the third time, it is understood slightly differently. This may be due to all the knowledge of the context of the conversation, which is not available for the first time. During the conversations, it is heard that both speakers use and rely on some specific contextual information (more often previous conversations), but this is not known and applicable in this study.

### Inter-rater reliability

Considering that the annotation process was performed by one person, inter-rater reliability determination was performed. The results suggest that there is a relatively higher consistency in customer speech ratings than in operator speech ratings (see Table 2 and 3), which could be related to speech duration (customers speak more, longer on average). In addition, it could be assumed that the determination of valence and activation is subjective in the absence of any initial reference criteria or predefined characteristics. With all this, however, it must be said that inter-rater reliability is in line with what has been studied in other languages. In general, the reliability indicators (see tables 2 and 3 in Appendix) correspond to the findings of the meta-analyzes, which show that the reliability indicators for annotating emotions via phone do not exceed 0.47 in total (Siegert et al., 2014).

**Table 2.** Intraclass correlations (ICC) of 5 annotator's core affect features in customer's speech

Customer's speech	Single measures	CI 95%	Average measures		CI 95%	
<i>Consistency</i>						
Valence	.26***	.14	.40	.64***	.46	.77
Arousal	.24***	.13	.38	.61***	.42	.75
<i>Absolute agreement</i>						
Valence	.25***	.14	.39	.63***	.44	.76
Arousal	.18***	.07	.31	.52***	.28	.69

\*\*\*  $p < .001$

**Table 3.** Intraclass correlations (ICC) of 5 annotator's core affect features in operator's speech

Operator's speech	Single measures	CI 95%		Average measures	CI 95%	
<i>Consistency</i>						
Valence	.15**	.02	.32	.46**	.11	.70
Arousal	.19**	.06	.37	.54**	.23	.74
<i>Absolute agreement</i>						
Valence	.12**	.02	.27	.40**	.08	.65
Arousal	.13**	.03	.28	.42**	.12	.66

\*\*  $p < .01$  \*\*\*  $p < .001$

Another aspect that suggests inter-rater relatively low scores is that evaluators each use their own experience, their current context, and a range of other socio-demographic factors that determine what the other person's speech valence and activation may seem at that time. This is in line with the fact that there are no universal, well-known preconditions for determining affective signs, such as other people's emotions.

It is important to note that inter-rater reliability was performed for only one of all layers – core affect valence and activation dimensions. Perhaps the results would be different when comparing the ratings of other strata between different people.

## Discussion

The first qualitative analysis of the telecommunications dialogue data in Latvian allows to conclude that the observed affective features vary widely and suggest a large heterogeneity of the repertoires of affective features. Unified system of features has not been observed, at least in particular telephone conversations. The multimodal features may differ in age, gender groups, as well as vary depending on certain contextual circumstances of the conversation. In this study, the assumptions about the features of affect are based only on the subjective ideas of one annotator, which on the one hand provides an opportunity to study phenomenologically, on the other hand allows to look at a set of elements from a content analysis perspective. Inter-rater reliability analysis leads to the conclusion that there are more consistent assessments of customer speech segments (in opposite of the operator's speech assessment) that could be related to the duration of content units. At the same time, however, it should be emphasized that inter-rater reliability is performed on only one of the annotation layers – the valence and activation dimensions of the affect. The low consistency

can carry positive, negative or neutral meanings. It can be concluded that there are individual profiles of affective features or affective dialects that may indicate certain idiographic traits, but there are also trait profiles that may be characteristic of certain demographic, cultural, contextual or other aspects. Low levels of inter-rater reliability suggest that features of affect are conceptualized situationally, contextually, and based on personal experience.

## Acknowledgement

This research has been supported by the European Regional Development Fund within the joint project of SIA TILDE and University of Latvia “Multilingual Artificial Intelligence Based Human Computer Interaction” No. 1.1.1.1/18/A/148.

## References

- Averill, J. R. (1985). The Social Construction of Emotion: With Special Reference to Love. In K. J. Gergen & K. E. Davis (Eds.), *The Social Construction of the Person* (pp. 89–109). Springer New York. [https://doi.org/10.1007/978-1-4612-5076-0\\_5](https://doi.org/10.1007/978-1-4612-5076-0_5)
- Bailey, C. (2007). *A Guide to Qualitative Field Research* (2<sup>nd</sup> ed.). <https://doi.org/10.4135/9781412983204>
- Bak, H. (2016). The State of Emotional Prosody Research—A Meta-Analysis. In H. Bak, *Emotional Prosody Processing for Non-Native English Speakers* (pp. 79–115). Springer International Publishing. [https://doi.org/10.1007/978-3-319-44042-2\\_5](https://doi.org/10.1007/978-3-319-44042-2_5)
- Barrett, L. F. (2006). Are Emotions Natural Kinds? *Perspectives on Psychological Science*, 1(1), 28–58. <https://doi.org/10.1111/j.1745-6916.2006.00003.x>
- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, 20(1), 1–68. <https://doi.org/10.1177/1529100619832930>
- Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in Emotion Perception: *Current Directions in Psychological Science*. <https://doi.org/10.1177/0963721411422522>
- Barrett, L. F., & Russell, J. A. (1999). The Structure of Current Affect: Controversies and Emerging Consensus. *CURRENT DIRECTIONS IN PSYCHOLOGICAL SCIENCE*, 8(1), 5.
- Barsalou, L. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, 18(5–6), 513–562. <https://doi.org/10.1080/01690960344000026>
- Clark-Polner, E., Johnson, T. D., & Barrett, L. F. (2017). Multivoxel Pattern Analysis Does Not Provide Evidence to Support the Existence of Basic Emotions. *Cerebral Cortex*, 27(3), 1944–1948. <https://doi.org/10.1093/cercor/bhw028>
- Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, 114(38), E7900–E7909. <https://doi.org/10.1073/pnas.1702247114>

- Cowen, A. S., Laukka, P., Elfenbein, H. A., Liu, R., & Keltner, D. (2019). The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nature Human Behaviour*, 3(4), 369–382. <https://doi.org/10.1038/s41562-019-0533-6>
- Denzin, N. K., & Lincoln, Y. S. (2018). *The SAGE Handbook of Qualitative Research*. 1688.
- Diller, J. V. (2011). *Cultural diversity: A primer for the human services* (4<sup>th</sup> ed). Brooks/Cole Cengage Learning.
- Doyle, C. M., & Lindquist, K. A. (2017). Language and emotion: Hypotheses on the constructed nature of emotion perception. In *The science of facial expression* (pp. 415–432). Oxford University Press.
- Eibl-Eibesfeldt, I. (2007). *Human ethology*. Aldine Transaction.
- Elfenbein, H. A. (2017). *Emotional Dialects in the Language of Emotion* (Vol. 1). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190613501.003.0025>
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128(2), 203–235. <https://doi.org/10.1037/0033-2909.128.2.203>
- Erickson, D., Yoshida, K., Menezes, C., Fujino, A., Mochida, T., & Shibuya, Y. (2006). Exploratory Study of Some Acoustic and Articulatory Characteristics of Sad Speech. *Phonetica*, 63(1), 1–25. <https://doi.org/10.1159/000091404>
- Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). Cultural Relativity in Perceiving Emotion From Vocalizations. *Psychological Science*, 25(4), 911–920. <https://doi.org/10.1177/0956797613517239>
- Jürgens, R., Drolet, M., Pirow, R., Scheiner, E., & Fischer, J. (2013). Encoding Conditions Affect Recognition of Vocally Expressed Emotions Across Cultures. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00111>
- Katz, W. F., & Assmann, P. F. (Eds.). (2019). *The Routledge handbook of phonetics*. Routledge.
- Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional Expression: Advances in Basic Emotion Theory. *Journal of Nonverbal Behavior*, 43(2), 133–160. <https://doi.org/10.1007/s10919-019-00293-3>
- Klessa, K. & Wydział Neofilologii (Uniwersytet im. Adama Mickiewicza ; Poznań). (2016). *Annotation Pro: Enhancing analyses of linguistic and paralinguistic features in speech*. Wydział Neofilologii UAM.
- Kraus, M. W., Côté, S., & Keltner, D. (2010). Social Class, Contextualism, and Empathic Accuracy. *Psychological Science*, 21(11), 1716–1723. <https://doi.org/10.1177/0956797610387613>
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2<sup>nd</sup> ed). Sage.
- Laukka, P., Elfenbein, H. A., Thingujam, N. S., Rockstuhl, T., Iraki, F. K., Chui, W., & Althoff, J. (2016). The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. *Journal of Personality and Social Psychology*, 111(5), 686–705. <https://doi.org/10.1037/pspi0000066>
- Matsumoto, D. (1989). Cultural Influences on the Perception of Emotion. *Journal of Cross-Cultural Psychology*, 20(1), 92–105. <https://doi.org/10.1177/0022022189201006>
- Matsumoto, D., & Assar, M. (1992). The effects of language on judgments of universal facial expressions of emotion. *Journal of Nonverbal Behavior*, 16(2), 85–99. <https://doi.org/10.1007/BF00990324>

- Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition & Emotion*, 23(2), 209–237. <https://doi.org/10.1080/02699930802204677>
- Mesquita, B., & Walker, R. (2003). Cultural differences in emotions: A context for interpreting emotional experiences. *Behaviour Research and Therapy*, 41(7), 777–793. [https://doi.org/10.1016/S0005-7967\(02\)00189-4](https://doi.org/10.1016/S0005-7967(02)00189-4)
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*, 2<sup>nd</sup> ed. (pp. xiv, 338). Sage Publications, Inc.
- O’Leary, Z. (2017). *The Essential Guide to Doing Your Research Project*. 751.
- Paulmann, S., & Uskul, A. K. (2014). Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. *Cognition and Emotion*, 28(2), 230–244. <https://doi.org/10.1080/02699931.2013.812033>
- Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing Emotions in a Foreign Language. *Journal of Nonverbal Behavior*, 33(2), 107–120. <https://doi.org/10.1007/s10919-008-0065-7>
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115(1), 102–141. <https://doi.org/10.1037/0033-2909.115.1.102>
- Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual Cues in Nonverbal Vocal Expressions of Emotion. *Quarterly Journal of Experimental Psychology*, 63(11), 2251–2272. <https://doi.org/10.1080/17470211003721642>
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology*, 32(1), 76–92. <https://doi.org/10.1177/0022022101032001009>
- Scherer, K. R., & Bänziger, T. (2004). *Emotional Expression in Prosody: A Review and an Agenda for Future Research*. 8.
- Siegert, I., Böck, R., & Wendemuth, A. (2014). Inter-rater reliability for emotion annotation in human–computer interaction: Comparison and methodological improvements. *Journal on Multimodal User Interfaces*, 8(1), 17–28. <https://doi.org/10.1007/s12193-013-0129-9>
- Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion*, 9(6), 838–846. <https://doi.org/10.1037/a0017810>
- Tomkins, S. S., & McCarter, R. (1964). What and where are the primary affects? Some evidence for a theory. *Perceptual and Motor Skills*, 18(1), 119–158. <https://doi.org/10.2466/pms.1964.18.1.119>
- Wertz, F. J. (Ed.). (2011). *Five ways of doing qualitative analysis: Phenomenological psychology, grounded theory, discourse analysis, narrative research, and intuitive inquiry*. Guilford Press.

- Paulmann, S., & Uskul, A. K. (2014). Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. *Cognition and Emotion*, 28(2), 230–244. <https://doi.org/10.1080/02699931.2013.812033>
- Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing Emotions in a Foreign Language. *Journal of Nonverbal Behavior*, 33(2), 107–120. <https://doi.org/10.1007/s10919-008-0065-7>
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115(1), 102–141. <https://doi.org/10.1037/0033-2909.115.1.102>
- Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual Cues in Nonverbal Vocal Expressions of Emotion. *Quarterly Journal of Experimental Psychology*, 63(11), 2251–2272. <https://doi.org/10.1080/17470211003721642>
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology*, 32(1), 76–92. <https://doi.org/10.1177/0022022101032001009>
- Scherer, K. R., & Bänziger, T. (2004). *Emotional Expression in Prosody: A Review and an Agenda for Future Research*. 8.
- Siegert, I., Böck, R., & Wendemuth, A. (2014). Inter-rater reliability for emotion annotation in human–computer interaction: Comparison and methodological improvements. *Journal on Multimodal User Interfaces*, 8(1), 17–28. <https://doi.org/10.1007/s12193-013-0129-9>
- Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion*, 9(6), 838–846. <https://doi.org/10.1037/a0017810>
- Tomkins, S. S., & McCarter, R. (1964). What and where are the primary affects? Some evidence for a theory. *Perceptual and Motor Skills*, 18(1), 119–158. <https://doi.org/10.2466/pms.1964.18.1.119>
- Wertz, F. J. (Ed.). (2011). *Five ways of doing qualitative analysis: Phenomenological psychology, grounded theory, discourse analysis, narrative research, and intuitive inquiry*. Guilford Press.