Emotional Speech Perception: A set of semantically validated German neutral and emotionally affective sentences

6 authors, including:

Sabrina Defren
Technische Universität Kaiserslautern
2 PUBLICATIONS **3** CITATIONS

SEE PROFILE

Patricia Wesseling
Technische Universität Kaiserslautern
2 PUBLICATIONS **2** CITATIONS

SEE PROFILE

Shanley Allen
Technische Universität Kaiserslautern
80 PUBLICATIONS **1,076** CITATIONS

SEE PROFILE

Boaz M Ben-David
Interdisciplinary Center Herzliya
50 PUBLICATIONS **577** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project InKraFT View project

Project ACQDIV: acquisition processes in maximally diverse languages View project

Emotional Speech Perception:

A set of semantically validated German neutral and emotionally affective sentences

Sabrina Defren[1], Patricia B. C. Wesseling[1], Shanley Allen[2], Vered Shakuf[3], Boaz Ben-David[3], Thomas Lachmann[1]

[1]Cognitive and Developmental Psychology, University of Kaiserslautern, Germany
[2]Psycholinguistics and Language Development, University of Kaiserslautern, Germany
[3]Communication, Aging and Neuropsychology Lab (CAN lab), School of Psychology, The Interdisciplinary Center (IDC), Herzliya, Israel

`sabrina.defren@sowi.uni-kl.de, castilho@sowi.uni-kl.de`

Abstract

In order to address the complex interplay of prosody and semantics, a set of sentences were generated, suitable for investigating emotional speech perception in German. Forty-seven German native speakers rated the emotional content of sentences on a 6-point Likert scale. From a set of 54 sentences, 10-11 each could reliably be associated with one of four distinct emotions. The remaining 11 were assessed as neutral (expressing no emotion). The unambiguous assignment of semantic (emotional) content enables the study of prosody as an independent factor. Moreover, the sentences were balanced regarding average word frequency, average phonological neighborhood density, and number of syllables per sentence. This linguistic balance enables an unbiased evaluation of the roles of semantic content and prosody in emotional speech.

Index Terms: emotional speech perception, German affective sentences, semantic validation, linguistic characteristics

1. Introduction

Recognition of emotions forms part of our everyday communication. Research on emotion perception often deals with the ability to interpret nonverbal cues, in particular from faces [1]. Affective speech is comparatively less investigated [2], [3], even though it greatly contributes to the perception of emotions, especially when facial cues are not available.

1.1. Prosody and semantics in emotional speech

Regarding emotional speech, two different factors are relevant: semantics, or what is said, and prosody, or how something is said. The latter includes, for instance, intonation, rhythm and stress. The relative contributions of semantics and prosody to emotion perception are still subject to debate. In [4] it is argued that these factors constitute two separate channels, yet engage in a complex interplay. Disentangling the contributions of prosodic and semantic cues, therefore, is not easily done. The authors asked native speakers of English to rate English sentences based on different emotions. Both semantics and prosody influenced the perception of emotional speech, even if one of them was irrelevant for the task.

Several tools might be offered to disentangle emotional speech processing (e.g. Florida Affect Battery [5];

Diagnostic Analysis of Nonverbal Accuracy scale [6]), but according to [3, p. 392], these tools lack the ability to "reveal the full complexity of the interaction of emotions conveyed in both lexical content and prosody".

This statement mirrors the current state in the literature; the entire extent of the relation is not fully understood. While several studies support prosodic primacy in the perception of spoken emotions [7], [8], others state that, at least in Western cultures, semantics dominates over prosody [9], [10]. Some studies suggest no processing advantage of combined prosodic-semantic cues or even a dominance of one factor [11], [12], while others show facilitation of emotional sentence processing if both channels are combined [13], [14]. Reduced ability to perceive the interaction of prosody and semantics amounts to an impaired ability to identify emotions. This inability is connected to depression [15] and leads to the reduction of both the quality and quantity of social relationships [2].

1.2. Assessment of emotional speech

Many studies of prosody used words that are emotionally neutral from a semantic perspective, or affective pseudo-utterances [13]. Since neither resembles natural speech, these methods are restricted in their applicability.

An alternative method consists of using full sentences as utterances. To prevent their linguistic characteristics (e.g. word frequency [16], phonological neighborhood density [17], or sentence length [18]) from interfering with their emotional content, these characteristics must be balanced across emotional categories.

The need of equating word frequency is motivated by findings suggesting that frequent words are more easily identified, especially with regard to their emotional content [16]. Phonological neighbors are found to affect spoken word recognition. The identification of words decelerates as the phonological neighborhood density increases [17]. Balancing sentence length is important for controlling the impact of cognitive resources on the ability to identify emotions. Shorter sentences improve the identification of emotional sentences compared to longer ones [18].

In the present study, sentences were equated for all these features. The frequency and phonological neighborhood density of the content words in a sentence were averaged and

equated across sentences; sentence length was equated in terms of the number of syllables available.

1.3. Emotional categories

To clarify the interaction between prosody and semantics, emotional categories are needed that are comparable and well-distinguishable. Categories with angry, sad, fearful, happy, and neutral content represent the emotions most commonly investigated in the literature [2]. Of these, anger, sadness, and fear were recognized most accurately [19].

To complement these three negative emotions, happiness was added as a category. It comprises a broad range of intensities, from contentment to joy, which makes this emotion harder to recognize. It was shown [20], however, that happiness is better identified if it is the only positive response option provided. Neutral sentences were added as baseline condition.

Together, the categories cover a spectrum of social and physiological features of emotions. Happiness is related to "approach" (in the context of approach-withdrawal theory [21]) and left-sided anterior activation [22], whereas the negative emotions are connected to withdrawal and right-sided anterior activation available.

1.4. The present study

In studying the interaction of prosody and semantics, it is important to control for both independently. The present study was modeled after the T-RES (Test of Rating of Emotions in Speech) [3] and validated the sentences for their lexical emotional content in written format, irrespective of prosody.

The purpose of the present study was to provide a set of 50 sentences, ten per emotional category and neutral, that helps to assess emotional speech and to identify the complex interaction between prosodic and semantic cues. To implement the variety of potential studies that such a resource of sentences provides, they have to be balanced for several factors.

The sentences were validated for their emotional lexical content. In addition, emotional categories that are common and therefore well applicable in emotional speech research were used. Another novel aspect is the validation for the linguistic characteristics, namely word frequency, the number of syllables, and the phonological neighborhood density.

The resulting top sentences of each of the emotional categories anger, sadness, fear, happiness, and neutral enable future spoken language studies to evaluate the relative role of semantic cues in emotional speech perception.

2. Methods

2.1. Sentence selection

A selection of sentences was created based on emotional words and phrases found in scientific literature and public media. Emotionally affective sentences either included words and phrases that are related to one of the four selected emotions ("His words make me *smile*") or they described scenarios associated with one of these emotions ("Nobody is interested in my life").

The sentences were equated for their linguistic properties. The online data base CLEARPOND ([23], for details see [24]) served as a resource to gather the content words´ frequency of usage in the German lexicon as well as the phonological

neighborhood density. These linguistic characteristics were averaged across each sentence. Additionally, the sentences consist of 8 to 11 syllables.

Altogether, 400 sentences were created. These were randomly segmented into eight different lists so that each list contained 50 sentences. As a result, each emotional category was represented by ten sentences per list.

2.2. Pretests

The eight lists were separated into six different periods of time. In a series of pretests either one or two lists were tested. The lists were rated by 194 participants, between 21 to 32 participants per list. The sentences were rated on their emotional content using a 6-point Likert scale. For each sentence, the participants rated several emotions, the corresponding one and the three other emotions. The 65 sentences that described one single emotion (affective stimuli) or were clearly rated as neutral (neutral stimuli) were selected for the current study.

2.3. Participants

The main experiment was conducted with 47 participants (36 male; mean age = 22.7 years, ranging from 18 to 32 years). They received monetary compensation of 8 € for their participation in the experiment.

As a requirement, all participants had to speak German as their native language. Due to exclusion and outlier criteria, 26 participants remained in the sample.

The study was conducted using the web survey tool Limesurvey (https://www.limesurvey.org/de/). They were instructed to assess each visually presented sentence on a 6-point Likert scale according to one emotion, resulting in four question blocks (as neutral was not assessed separately). The order of sentence blocks was randomized to control for order effects.

Note that participants were not given any situational context for a sentence to ensure that the emotional content was solely conveyed by the respective sentence. When asked how much they agreed that the sentence conveyed a happy [fearful, sad, angry] emotion, they chose an answer between 1 (completely disagree) and 6 (completely agree).

Altogether, participants rated 260 items. One sentence of each emotional category was presented twice during each block. This added 20 catch trials that were used as a criterion for the selection of outliers.

2.4. Data analysis

Three criteria for outlier selection were applied in succession. In order to ensure raters´ response validity and to control for inconsistency of responses, 20 catch trials were analyzed. One participant had to be removed from the data analysis because of this criterion. A second criterion was strikingly divergent ratings. Six participants were excluded because of this criterion. The last criterion to identify outliers was the amount of divergent ratings per participant. If the rating of a sentence deviated at least 3 points from the mean, this rating was excluded as an outlier. If a participant showed above 1 % of such outlying ratings, he/she was excluded from further analysis. This was the case for eight participants and 18 single sentence ratings (altogether .047 %).

ANOVAs were conducted in order to test whether there were significant differences between the emotional categories

regarding the linguistic characteristics. The independent factor was the Emotional category (Anger, Sadness, Happiness, Fear, and Neutral). The dependent variable was either Average word frequency, Phonological neighborhood density, or Number of Syllables.

Paired-sample *t*-tests were conducted to confirm that affective sentences were judged according to the corresponding emotion and significantly less on the non-corresponding ones. Altogether 43 of these tests were implemented, one for each affective sentence.

To examine whether single sentences received an average rating that differed from a set threshold, a series of one-sample *t*-tests was implemented. Forty-three of these were conducted to test whether ratings of affective sentences in non- corresponding cases were significantly below 2.6. Another 43 *t*-tests were conducted to examine whether corresponding ratings for affective sentences were significantly above 4.2. Eleven additional *t*-tests were conducted to verify whether average ratings of neutral sentences were below 2.0.

3. Results

The best sentences according to participants´ ratings in each of the five categories (anger, fear, happiness, sadness, and neutral) were chosen and matched for the linguistic criteria. This approach ensured that these sentences represent the lexical content according to their emotional category.

3.1. Ratings

Table 1 presents the average ratings on the four affective emotional scales for all five categories. Ratings were performed on a 6-point Likert scale with 1-3 indicating that the participants disagreed while 4-6 represents the participants´ agreement.

Table 1: *Average ratings for the four affective scales in emotional and neutral categories.*

| Category | *n* | Emotional scale | | | |
		A	F	H	S
A	11	4.8 (0.2)	1.9 (0.2)	1.1 (0.1)	2.2 (0.3)
F	11	1.7 (0.3)	5.1 (0.3)	1.1 (0.1)	1.8 (0.5)
H	11	1.0 (0.0)	1.2 (0.2)	5.5 (0.2)	1.1 (0.1)
S	10	1.9 (0.4)	2.4 (0.4)	1.2 (0.3)	5.2 (0.3)
N	11	1.1 (0.1)	1.1 (0.1)	1.4 (0.1)	1.1 (0.0)

Note. Shaded data present average ratings of affective sentences on their corresponding emotional scales. Numbers in parentheses show the standard deviation. *n* = number of sentences, A = Anger, F = Fear, H = Happiness, S = Sadness, N = Neutral

The average rating for corresponding emotions, i.e. the emotional category anger is rated on an anger-scale, fear on a fear-scale, and so on (see shaded data in Table 1), across all four categories was 5.2 (SD = .3). The ratings of single sentences ranged between 4.6 and 5.8. For non-corresponding ratings of affective sentences across categories, e.g. angry sentences rated on happy-scale, sad-scale, and fear-scale, the average was 1.5 (SD = .5). Single sentences ratings varied from 1.0 to 3.1. As the neutral category has no corresponding emotional scale (because these ratings have proven to be inaccurate, see [25]), it was examined separately. The average rating for the eleven neutral sentences was 1.2 (SD = 0.2) in the range of 1.0 to 1.7 for single sentences.

The score on the corresponding emotions was compared to the average score of all other emotions in a series of 43 paired sample *t*-tests. One test was done for each affective sentence to confirm that affective sentences were consistently better rated on their corresponding emotion than on their non-corresponding emotions. All 43 *t*-tests were found to be significant ($t(\geq 24) > 9.0$, $p < .001$, for all tests).

In a second step, a series of post hoc one-sample *t*-tests verified that individual sentences received an average rating that was significantly different to a set threshold. For the 43 affective sentences, this threshold was set to 2.6 for the average of all their non-corresponding emotions. The *t*-tests confirmed that the ratings of all participants and for all 43 sentences were significantly below the set threshold ($t(25) \geq 2.3$, $p < .05$, for all tests). The rating score on the corresponding emotions had to be significantly above the threshold of 4.2 which was true for all sentences ($t(\geq 23) \geq 2.1$, $p < .05$, for all tests). Eleven additional one-sample *t*-tests were conducted for the neutral sentences. The average ratings of all four emotional scales had to stay below the threshold of 2.0. All ratings for neutral sentences were even significantly below a threshold of 1.4 ($t(25) \geq 2.1$, $p < .05$).

The sentences that conveyed a specific emotional content or no emotional content according to the scores on participants´ ratings are shown in Table 2.

Table 2: *A selection of German sentences for four emotional categories and neutral.*

Emotional category	Sentences
Anger	Du gehst mir total auf die Nerven. *You are a pain in the neck.* Provoziere mich bloß nicht noch weiter. *Do not provoke me any further.* Kümmere dich um deinen eigenen Mist. *Take care of your own crap.*
Fear	Ich verlier die Gewalt über mein Auto. *I´m losing control over my car.* Das Feuer hat die Gasleitung fast erreicht. *The fire has almost reached the gas line.* Es kann jeden Moment explodieren. *It can explode any moment.*
Happiness	Meine Mannschaft hat gestern gewonnen. *My team won yesterday.* Ich fühle mich heute großartig. *I feel great today.* Seine Worte bringen mich zum Lächeln. *His words make me smile.*

Sadness	Mein Hund musste eingeschläfert werden.
	My dog had to be euthanized.
	Du fehlst mir an jedem einzelnen Tag.
	I miss you every single day.
	Niemand interessiert sich für mein Leben.
	Nobody is interested in my life.
Neutral	Der Kunde kauft eine graue Hose.
	The customer buys a pair of gray trousers.
	Der Teller steht auf dem runden Tisch.
	The plate is on the round table.
	Meine Tasche liegt im Zimmer.
	My bag is in the room.

For the anger, fear, happiness, and neutral categories it was possible to gather eleven (for sadness ten) sentences. The additional sentences were added, as it is our aim to record these sentences for future research on prosody.

3.2. Linguistic criteria

The results of the CLEARPOND analysis for the linguistic characteristics of 54 sentences revealed an overall average of 9.6 syllables (SD = 1.1) per sentence, an average frequency of 610.0 (SD = 507.8) for content words and an average phonological neighborhood density of 9.6 (SD = 4.2). The ANOVAs did not reveal any significant differences between the five categories regarding Number of syllables ($H(4)$ = 5.53, p = .237), Average word frequency per million ($F(4, 49)$ = 2.14, p = .090), or Average phonological neighborhood density ($F(4, 49)$ = 1.65, p = .178). Slight deviations in word frequency resulted from well-known but rather infrequent words used for sentences in the neutral category, such as "*Hai*"/"shark".

4. Discussion

This study provides a set of 54 emotionally affective or neutral sentences. These were validated according to their main linguistic characteristics across four emotional (anger, fear, happiness, sadness) and one neutral category.

Based on the results, sentences were found to be distinctive in conveying their corresponding emotion or no emotion at all. Note for the interpretation of the results that the *t*-tests were not corrected for inflated Type I errors for the sentences were tested independently of each other.

These sentences bear some crucial advantages over those used in traditional studies investigating the perception of emotional speech. First, a rating task was applied instead of classical forced-choice paradigms. This method allowed to observe possible competitor emotions when looking at the ratings of the sentences. In a forced-choice task, slight differences between ratings in corresponding or non-corresponding categories could have remained undetected, whereas the rating paradigm enables us to assess the relative weights of ratings to each dimension.

Secondly, questions and exclamations were excluded as items for this study. Not only could the varying structure of such sentences influence findings with regard to linguistic criteria, but they were also found to bias results of testing for identifying emotional content when they were mixed with other types of sentences [26], [27].

Furthermore, the selected sentences were carefully matched for basic linguistic criteria to ensure that differences in ratings arouse from distinct lexical emotional content and not from biases due to varying linguistic factors.

Future steps include the recording of these sentences in corresponding and non-corresponding prosodies and validation for their respective prosodic content.

Such a database of validated sentences allows assessing the ability to perceive and identify emotions in speech. It contributes a meaningful part to the discussion of how prosody and semantics are integrated in emotional spoken language. With clear sentences regarding emotional content and linguistic criteria it is possible to test for the predominance of one factor, prosody or semantics. It has proven difficult to perceive one without the other without mutual influence [4]. Isolation of one factor is challenging as they are naturally intertwined and linked in spoken language. The present study allows an independent variation of prosody and semantics in planned studies with factorial design.

As a similar set of sentences exists for English [26], it also offers a tool for an intercultural comparison of emotional speech perception with participants from different linguistic background.

5. Acknowledgements

6. References

[1] D. Radice-Neumann, B. Zupran, D. R. Babbage, and B. Willer, "Overview of impaired affect recognition in persons with traumatic brain injury," *Brain Injury*, vol. 21, no. 8, pp. 807-816, 2007.

[2] B. Zupan, D. Neumann, D. R. Babbage, and B. Willer, "The importance of vocal affect to bimodal processing of emotion: Implications for individuals with traumatic brain injury," *Journal of Communication Disorders,* vol. 42, no. 1, pp. 1-17, 2009.

[3] B. M. Ben-David, N. Multani, N. A.-M. Durham, F. Rudzicz, and P. H. van Lieshout, "T-RES: Test of rating of emotions in speech: Interaction of affective cues expressed in lexical content and prosody of spoken sentences," *Proceedings of the 27th annual meeting of the international society for psychophysics,* pp. 391-396, 2011.

[4] B. M. Ben-David, N. Multani, V. Shakuf, F. Rudzicz, and P. H. van Lieshout, "Prosody and semantics are separate but not separable channels in the perception of emotional speech: Test for rating of emotions in speech," *Journal of Speech, Language and Hearing Research*, vol. 59, no. 1, p. 72, 2016.

[5] L. Blonder, D. Bowers, and K. Heilman, "Florida affect battery," *Center for Neuropsychological Studies, Department of Neurology*, 1998. [Online]. Available: https://books.google.de/books?id=-Fq6PgAACAAJ.

[6] S. Nowicki and M. P. Duke, "Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale," *Journal of Nonverbal Behavior,* vol. 18, no. 1, pp. 9-35, 1994.

[7] H. Jacob, C. Brück, C. Plewnia, and D. Wildgruber, "Cerebral processing of prosodic emotional signals: Evaluation of a network model using rTMS," *PloS ONE,* vol. 9, no. 8, e105509, 2014.

[8] J. M. Roche, B. Petters, and R. Dale, "Your tone says it all: The processing and interpretation of affective language," *Speech Communication*, vol. 66, pp. 47-64, 2014.

[9] K. Ishii, J. A. Reyes, and S. Kitayama, "Spontaneous attention to word content versus emotional tone: Differences among three cultures," *Psychological Science*, vol. 14, pp. 39-46, 2003.

[10] S. Paulmann and S. A. Kotz, "An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context," *Brain and Language*, vol. 105, pp. 59-69, 2008.

[11] R. Schwartz and M. D. Pell, "Emotional speech processing at the intersection of prosody and semantics," *PLoS ONE*, vol. 7, no. 10, e47279, 2012.

[12] M. D. Pell, A. Jaywant, L. Monetta, and S. A. Kotz, "Emotional speech processing: Disentangeling the effects of prosody and semantic cues," *Cognition and Emotion*, vol. 25, no. 5, pp. 834- 853, 2010.

[13] S. A. Kotz and S. Paulmann, "When emotional prosody and semantics dance cheek to cheek: ERP evidence," *Brain Research*, vol. 1151, pp. 107-118, 2007.

[14] C. S. Min and A. Schirmer, "Perceiving verbal and vocal emotions in a second language," *Cognition and Emotion*, vol. 25, no. 8, pp. 1376-1392, 2011.

[15] J. S. Carton, E. A. Kessler, and C. L. Pape, "Nonverbal decoding skills and relationship well-being in adults," *Journal of Nonverbal Behavior*, vol. 23, no. 1, pp. 91-100, 1999.

[16] A. J. Bruce, M. J. Harman, and M. T. Turner, "Reactions to heard words: Varying emotional content and frequency of exposure," *Journal of General Psychology*, vol. 134, no. 1, pp. 67-82, 2007.

[17] P. A. Luce and D. B. Pisoni, "Recognizing spoken words: The neighborhood activation model," *Ear and Hearing*, vol. 19, no. 1, pp. 1-36, 1998.

[18] B. Lidestam, B. Lyxell, and G. Andersson, "Speech-reading: Cognitive predictors and displayed emotion," *Scandinavian Audiology*, vol. 28, no. 4, pp. 211-217, 1999.

[19] M. D. Pell, S. Paulmann, C. Dara, A. Alasseri, and S. A. Kotz, "Factors in the recognition of vocally expressed emotions: A comparison of four languages," *Journal of Phonetics*, vol. 37, no. 4, pp. 417-435, 2009.

[20] J. Bachorowski and M. J. Owren, "Sounds of emotion: Production and perception of affect-related vocal acoustics," *Annals of the New York Academy of Sciences*, vol. 1000, no. 1, 244-265, 2003.

[21] F. C. Murphy, I. Nimmo-Smith, and A. D. Lawrence, "Functional neuroanatomy of emotions: a meta-analysis," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 3, no. 3, 207-233, 2003.

[22] R. J. Davidson, P. Ekman, C. D. Saron, J. A. Senulis, and W. V. Friesen, "Approach-withdrawal and cerebral asymmetry: emotional expression and brain physiology. I.," *Journal of Personality and Social Psychology*, vol. 58, no. 2, pp. 330-341, 1990.

[23] V. Marian, J. Bartolotti, S. Chabal, and A. Shook, "CLEARPOND," *Northwestern University Bilingualism and Psycholinguistics Research Group*, 2016. [Online]. Available: http://clearpond.northwestern.edu/. [Accessed: Dec. 12, 2017].

[24] V. Marian, J. Bartolotti, S. Chabal, and A. Shook, "CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities," *PLoS ONE*, vol. 7, no. 8, e43230, 2012.

[25] S. McDonald and S. Flanagan, "Social perception deficits after traumatic brain injury: Interaction between emotion recognition, mentalizing ability, and social communication," *Neuropsychology*, vol. 18, no. 3, pp. 572-579, 2004.

[26] B. M. Ben-David, P. H. van Lieshout, and T. Leszcz, "A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury," *Brain Injury*, vol. 25, no. 2, pp. 206-220, 2011.

[27] C. De Vos, E. van der Kooij, and O. Crasborn, "Mixed signals: Combining linguistic and affective functions of eyebrows in questions in sign language of the Netherlands," *Language and Speech*, vol. 52, no. 2/3, pp. 315-339, 2009.