

Cross-modal validation of the Leipzig Affective Norms for German (LANG)

Philipp Kanske · Sonja A. Kotz

Published online: 7 February 2011
© Psychonomic Society, Inc. 2011

Abstract Affective stimuli are increasingly used in emotion research. Typically, stimuli are selected from databases providing affective norms. The validity of these norms is a critical factor with regard to the applicability of the stimuli for emotion research. We therefore probed the validity of the Leipzig Affective Norms for German (LANG) by correlating valence and arousal ratings across different sensory modalities. A sample of 120 words was selected from the LANG database, and auditory recordings of these words were obtained from two professional actors. The auditory stimuli were then rated again for valence and arousal. This cross-modal validation approach yielded very high correlations between auditory and visual ratings ($>.95$). These data confirm the strong validity of the Leipzig Affective Norms for German and encourage their use in emotion research.

Keywords Emotion · Language · Speech

Affective stimuli are increasingly used in emotion research. The most frequently used stimulus categories include emotional scenes and faces, but words have also been used in a variety of tasks including lexical decision (Eviatar & Zaidel, 1991; Kanske & Kotz, 2007; Kuchinke et al., 2005; Nakic et al., 2006; Schacht & Sommer, 2009; Scott, O'Donnell, Leuthold,

& Sereno, 2009), memory tasks (Kuchinke et al., 2006; Sim & Martinez, 2005), versions of the Stroop task (van Hooff, Dietz, Sharma, & Bowman, 2008), mental imagery (Osaka, Osaka, Morishita, Kondo, & Fukuyama, 2004), the attentional blink (Mathewson, Arnell, & Mansfield, 2008), attentional orienting (Stormark, Nordby, & Hugdahl, 1995), and emotionality judgments (Maddock, Garrett, & Buonocone, 2003). The biggest advantage of word stimuli is that they can be tightly controlled for physical attributes (size, complexity, color composition, luminance), frequency of occurrence in everyday life, or concreteness of the underlying concept. To make use of these advantages, normative data for word stimuli are needed.

Several databases offer affective norms for words in different languages, including English (Altarriba, Bauer, & Benvenuto, 1999; Stevenson, Mikels, & James, 2007), German (Lahl, Göriz, Pietrowsky, & Rosenberg, 2009; Vö et al., 2009; Vö, Jacobs, & Conrad, 2006), Spanish (Redondo, Fraga, Padrón, & Comesaña, 2007), and Finnish (Eilola & Havelka, 2010). Typically, these norms are obtained from rating studies in which participants evaluate the valence and arousal of the stimuli. However, the degree to which the norms are applicable in emotion research critically depends on their validity. Only if the ratings validly measure the emotional status of the stimuli will it be advisable to use normative data to select affective stimuli. The present study addresses this problem by measuring convergent validity of affective norms for words. *Convergent validity* is defined as the degree to which one variable correlates with another that it is theoretically predicted to correlate with. Therefore, we correlated valence and arousal ratings of word stimuli in different sensory modalities. We selected a subsample of words from the Leipzig Affective Norms for German (LANG) database (Kanske & Kotz, 2010) containing visual ratings of the written words. These

P. Kanske (✉) · S. A. Kotz
Neurocognition of Rhythm in Communication Group,
Max Planck Institute for Human Cognitive and Brain Sciences,
Leipzig, Germany
e-mail: philipp.kanske@zi-mannheim.de

P. Kanske
Department of Cognitive and Clinical Neuroscience,
Central Institute of Mental Health,
Square J5,
68159 Mannheim, Germany

words were spoken by professional actors to obtain an auditory version of each stimulus, which was then rated again in valence and arousal. Theoretically, the emotional status of the stimuli should be the same in different modalities, which we tested by correlating the visual and auditory rating scores. This is a critical test of convergent validity, as the two types of materials differ greatly; while visual stimuli are present at once, auditory stimuli evolve over time. In addition to the emotional word meaning, auditory stimuli also contain an emotional prosodic modulation. Furthermore, distinct neural pathways underlie the processing of visual and auditory affective signals (for a review, see Schirmer & Kotz, 2006). To our knowledge, ours is the first study that has tested the validity of affective norms cross-modally. We chose the valence and arousal dimensions because these have been shown to be the most unambiguous factors explaining variance in words (Hager & Hasselhorn, 1994). To reduce variance based on word class (Osterhout, 1997; Perani et al., 1999), we only included nouns.

To summarize, the present study probes the validity of affective word norms in German by correlating the ratings of word stimuli in vision and audition. This cross-modal approach should provide insights into the validity of affective rating studies and normative data.

Method

Participants

All procedures were in accordance with the ethical standards of the local committee on human experimentation and with the Helsinki Declaration of 1975, as revised in 1983.

A sample of 30 native German speakers was recruited from the University of Leipzig. There were 16 female participants; mean age was 23.2 years ($SD = 2.8$). All participants were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971), with a mean laterality quotient of 87.8 ($SD = 20.4$). All participants reported normal or corrected-to-normal vision, normal hearing, and no history of mental disorders or emotional problems according to the Depression Anxiety Stress Scales (DASS; Lovibond & Lovibond, 1995).

Materials and procedures

From the 1,000 German nouns in the LANG database, we selected a subset of 120 that were prototypical for the following categories: (1) negative and high-arousing, (2) neutral and low-arousing, and (3) positive and high-arousing. The descriptive statistics are displayed in Table 1. There were no significant differences in concreteness,

frequency of usage, number of letters, or number of syllables between the categories. Several auditory recordings of each word were made with the emotional expression corresponding to the word's emotional valence, using two professional actors who were native speakers of German. One of the speakers was female, the other male. Recordings were made with Algorec 2.1 (Algorithmix GmbH, Waldshut-Tiengen, Germany), and the sound files were further processed with PRAAT (Institute of Phonetics Sciences, University of Amsterdam). Two versions of each positive, negative, and neutral word from each speaker were chosen for the rating study, excluding any recordings with background noise or outlying in duration or intensity. In total, participants were presented with 480 different auditory stimuli. To control for differences in loudness, all stimuli were normalized in sound intensity to 75 dB SPL.

Participants completed one session during which they rated the words for valence (negative–neutral–positive) and for arousal (high arousing–low arousing). The order of the tasks was counterbalanced. Words were presented in a different randomization for each participant. Ratings were done on 9-point Likert scales. For valence and arousal ratings, Self Assessment Manikins (Bradley & Lang, 1994; Hodes, Cook, & Lang, 1985) were used. The assignment of the scale endpoints to the left and the right was counterbalanced across participants. During the rating, participants were seated in a comfortable chair in a sound-attenuating room and wore headphones (Sennheiser HD 202). Stimuli were presented with ERTS (Experimental Run Time System; Berisoft Cooperation, Frankfurt, Germany).

Results

Figure 1 shows the valence and arousal ratings of the 120 selected words in the visual and auditory modalities. (The visual ratings were taken from the LANG database.) Here, we observed quadratic relationships (visual: $r_{\text{quad}} = .88$, $p < .001$; auditory: $r_{\text{quad}} = .98$, $p < .001$), demonstrating the typical distribution of valence and arousal values. We then correlated visual and auditory ratings of the 120 words. Valence ratings correlated highly ($r = .98$, $p < .001$), as did the ratings in arousal ($r = .97$, $p < .001$). For the respective scatterplots, see Fig. 1.

To control for potential inflation of the correlation coefficients due to the selection of extreme groups (and the consequently enhanced variance), we used two approaches (Feldt, 1961). Firstly, assuming the same linear relation of written and spoken valence and arousal ratings (same regression coefficient, same residuals), but reduced variance (estimated through the variance in the entire list of 1,000 word ratings; i.e., for valence, $SD = 1.358$; for

Table 1 Descriptive statistics of the 120 words selected

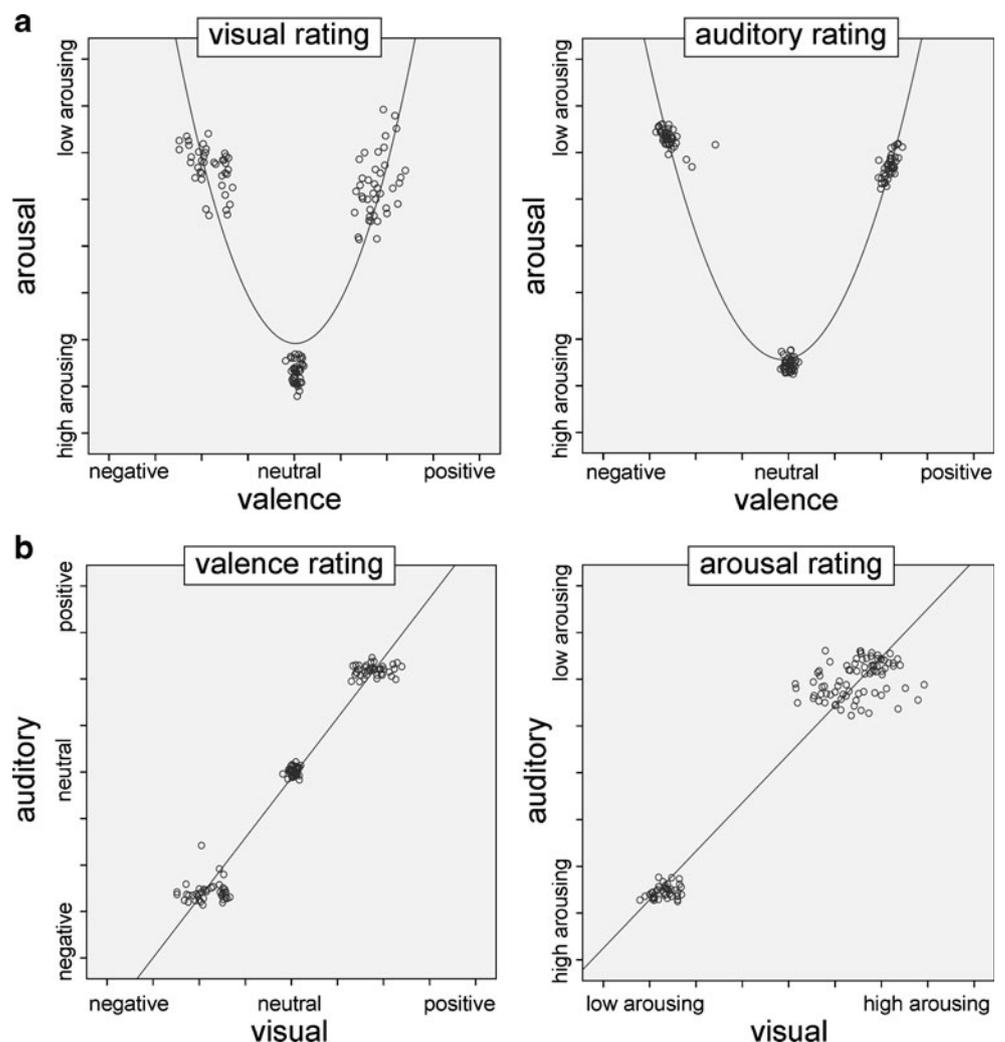
| | Negative | Neutral | Positive |
|------------------------|----------------------|------------------------|----------------------------|
| Example word | Tod [<i>death</i>] | Phase [<i>phase</i>] | Triumph [<i>triumph</i>] |
| Rated visual valence | 2.4 (0.4) | 5.0 (0.1) | 7.5 (0.3) |
| Rated visual arousal | 6.8 (0.5) | 2.3 (0.3) | 6.7 (0.6) |
| Rated auditory valence | 2.3 (0.2) | 5.0 (0.1) | 7.3 (0.2) |
| Rated auditory arousal | 7.4 (0.3) | 2.4 (0.2) | 6.8 (0.4) |
| Rated concreteness | 6.3 (1.5) | 6.0 (1.3) | 6.3 (1.8) |
| Word frequency | 11.0 (1.9) | 11.3 (1.9) | 10.7 (2.2) |
| Number of letters | 5.5 (1.1) | 5.9 (1.1) | 5.7 (1.3) |
| Number of syllables | 1.7 (0.5) | 1.8 (0.4) | 1.7 (0.4) |
| Duration | 0.6 (0.1) | 0.6 (0.1) | 0.8 (0.2) |
| Mean pitch | 280.8 (58.4) | 159.7 (22.7) | 262.1 (52.2) |

The means, standard deviations, and ranges for valence, arousal, concreteness, frequency of usage, number of letters, and number of syllables are given

arousal, $SD = 1.597$), we found only slightly reduced correlation coefficients (valence, $r = .979$; arousal, $r = .949$). Secondly, the most conservative approach assumes random variation of the intermediate data points. We therefore repeated the analysis including pseudorandomly

created data. These were created to be intermediate in visual valence and arousal ratings (e.g., to fall between the negative and neutral stimuli). We then assigned auditory ratings that randomly varied between the minimum and maximum valence or arousal values obtained in the auditory rating.

Fig. 1 (a) Mean valence and arousal ratings for each word in the visual and in the auditory modality. (b) Mean valence and arousal ratings across the visual and auditory modalities, including regression slopes



To complement this, we also used the opposite strategy (intermediate in auditory ratings and randomly varying in visual ratings). These analyses represent a lower boundary for the correlation, as it is unlikely that empirical values would have been completely random. Since there were 30 words in each group (negative, neutral, positive), we also used 30 words for each intermediate group (i.e., 30 neutral–negative mean arousing and 30 neutral–positive mean arousing data points). The results showed reduced, but still highly significant, correlation coefficients between the visual and auditory ratings, ranging from $r = .78 - .82$ (all $ps < .001$).

Discussion

The present study aimed to probe the validity of normative data regarding valence and arousal ratings of emotional and neutral word stimuli. The correlations between the ratings of visual and auditory words were very high for both affective dimensions, confirming that affective norms of the words are valid measures. These data justify the use of word stimuli selected on the basis of affective norms for experimental studies of emotion.

Valence and arousal ratings displayed the typical quadratic function that has also been observed in other normative studies with words and pictorial stimuli (Eilola & Havelka, 2010), indicating that highly negative and positive words are more likely to be rated as high arousing, while neutral words are more likely to be rated as low in arousal. This underlines the comparability of the present results with previously established norms. We used this dimensional approach because it seems well suited to the characterization of word stimuli, for which it is difficult to always assign a primary emotional category (e.g., to words like “bomb, vacation, pizza”). However, we acknowledge attempts to use discrete emotional categories (Stevenson et al., 2007), and future studies could validate the affective norms in dimensional and categorical ratings.

The major result provided by the present approach is the strong correlation between the visual and auditory word ratings. This result indicates the high validity of the present normative data. Using this cross-modal validation strategy is an unorthodox approach. It does not compare different ways of measuring valence and arousal, but manipulates the stimulus characteristics themselves. These include the change in the sensory modality and the additional prosodic manipulation, which also conveys emotional information. These alterations should, however, not change the affective character of the stimuli, which we confirmed in the high correlations between the ratings.

One limitation of the present study is the use of extreme groups (i.e., words that are high or low in arousal and either

very negative, neutral, or positive). This may inflate the observed correlation coefficients (Feldt, 1961). To address this issue, we used two approaches. Making the plausible assumption that the regression slope is the same with data that are not affected by an extreme group variance increase, we found only slightly reduced correlation coefficients. We also repeated the analyses including artificial, randomly varying data between the extreme groups. This is a conservative approach, because it inserts data with absolutely no correlation of the visual and auditory ratings, which is unlikely to be the case if ratings were acquired on intermediate data points. Nevertheless, the correlations remained highly significant, indicating that the relation between visual and auditory ratings was stable, despite the use of extreme groups.

A critical question is whether the reported results would generalize to other normative data and different samples. We believe the latter to be true, as we previously showed that correlations between the affective ratings of different samples and of the same sample at different time points are very high (Kanske & Kotz, 2010). However, it remains an open question whether the observed validity of the present norms is specific to the current database or indicates that the word norms are generally highly valid. This seems very likely, as the instructions and methodology used in the different rating studies are very similar. Nevertheless, it is an empirical question and should be addressed in future studies.

The present results demonstrate the validity of the Leipzig Affective Norms for German (LANG), which provides a comprehensive database of German nouns with a wide range of emotional status ratings. Subsamples of these words have already been successfully used to study the neural basis of emotional processing with functional magnetic resonance imaging and electroencephalography (Kanske & Kotz, *in press a, b*). Therefore, these norms can support experimental studies on emotion by helping researchers to select highly controlled word samples.

Acknowledgments This work was supported by the German Research Council (DFG; Graduate Program [1182]: Function of Attention in Cognition, University of Leipzig, Germany and Grant DFG-FOR-499 to S.A.K. at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany).

References

- Altarriba, J. B., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers*, *31*, 578–602.
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy & Experimental Psychiatry*, *25*, 49–59.

- Eilola, T. M., & Havelka, J. (2010). Affective norms for 210 British English and Finnish nouns. *Behavior Research Methods*, *42*, 134–140.
- Eviatar, Z., & Zaidel, E. (1991). The effects of word length and emotionality on hemispheric contribution to lexical decision. *Neuropsychologia*, *29*, 415–428.
- Feldt, L. S. (1961). The use of extreme groups to test for the presence of a relationship. *Psychometrika*, *26*, 307–316.
- Hager, W., & Hasselhorn, M. (1994). *Handbuch deutschsprachiger Wortnormen*. Göttingen: Hogrefe.
- Hodes, R. L., Cook, E. W., & Lang, P. J. (1985). Individual differences in autonomic response: Conditioned association or conditioned fear? *Psychophysiology*, *22*, 545–560.
- Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research*, *1148*, 138–148.
- Kanske, P., & Kotz, S. A. (2010). Leipzig affective norms for German: A reliability study. *Behavior Research Methods*, *42*, 987–991.
- Kanske, P., & Kotz, S. A. (in press a). Emotion speeds up conflict resolution: A new role for the ventral anterior cingulate cortex? *Cerebral Cortex*. doi:10.1093/cercor/bhq157
- Kanske, P., & Kotz, S. A. (in press b). Emotion triggers executive attention: Anterior cingulate cortex and amygdala responses to emotional words in a conflict task. *Human Brain Mapping*. doi:10.1002/hbm.21012
- Kuchinke, L., Jacobs, A. M., Grubich, C., Vö, M. L., Conrad, M., & Herrmann, M. (2005). Incidental effects of emotional valence in single word processing: An fMRI study. *Neuroimage*, *28*, 1022–1032.
- Kuchinke, L., Jacobs, A. M., Vö, M., Conrad, M., Grubich, C., & Herrmann, M. (2006). Modulation of prefrontal cortex activation by emotional words in recognition memory. *NeuroReport*, *17*, 1037–1041.
- Lahl, O., Göritz, A. S., Pietrowsky, R., & Rosenberg, J. (2009). Using the World-Wide Web to obtain large-scale word norms: 190, 212 ratings on a set of 2, 654 German nouns. *Behavior Research Methods*, *41*, 13–19.
- Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress scales*. Sydney: Psychology Foundation.
- Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2003). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Human Brain Mapping*, *18*, 30–41.
- Mathewson, K. J., Arnell, K. M., & Mansfield, C. A. (2008). Capturing and holding attention: The impact of emotional words in rapid serial visual presentation. *Memory & Cognition*, *36*, 182–200.
- Nakic, M., Smith, B. W., Busis, S., Vythilingam, M., Blair, R., & James, R. (2006). The impact of affect and frequency on lexical decision: The role of the amygdala and inferior frontal cortex. *Neuroimage*, *31*, 1752–1761.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–98.
- Osaka, N., Osaka, M., Morishita, M., Kondo, H., & Fukuyama, H. (2004). A word expressing affective pain activates the anterior cingulate cortex in the human brain: An fMRI study. *Behavioural Brain Research*, *153*, 123–127.
- Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, *59*, 494–522.
- Perani, D., Cappa, S. F., Schnur, T., Tettamanti, M., Collina, S., Rosa, M. M., et al. (1999). The neural correlates of verb and noun processing: A PET study. *Brain*, *122*, 2337–2344.
- Redondo, J., Fraga, I., Padrón, I., & Comesaña, M. (2007). The Spanish adaptation of ANEW (affective norms for English words). *Behavior Research Methods*, *39*, 600–605.
- Schacht, A., & Sommer, W. (2009). Time course and task dependence of emotion effects in word processing. *Cognitive, Affective & Behavioral Neuroscience*, *9*, 28–43.
- Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*, 24–30.
- Scott, G. G., O'Donnell, P. J., Leuthold, H., & Sereno, S. C. (2009). Early emotion word processing: Evidence from event-related potentials. *Biological Psychology*, *80*, 95–104.
- Sim, T.-C., & Martinez, C. (2005). Emotion words are remembered better in the left ear. *Laterality*, *10*, 149–159.
- Stevenson, R. A., Mikels, J. A., & James, T. W. (2007). Characterization of the affective norms for English words by discrete emotional categories. *Behavior Research Methods*, *39*, 1020–1024.
- Stormark, K. M., Nordby, H., & Hugdahl, K. (1995). Attentional shifts to emotionally charged cues: Behavioral and ERP data. *Cognition and Emotion*, *9*, 507–523.
- van Hooff, J. C., Dietz, K. C., Sharma, D., & Bowman, H. (2008). Neural correlates of intrusion of emotion words in a modified Stroop task. *International Journal of Psychophysiology*, *67*, 23–34.
- Vö, M. L., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*, 534–538.
- Vö, M. L., Jacobs, A. M., & Conrad, M. (2006). Cross-validating the Berlin Affective Word List. *Behavior Research Methods*, *38*, 606–609.