Brain Injury, February 2011; 25(2): 206–220

informa
healthcare

ORIGINAL ARTICLE

A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury

BOAZ M. BEN-DAVID[1,2,3], PASCAL H. H. M. VAN LIESHOUT[1,2,3,4], & TALIA LESZCZ[1]

[1]*Department of Speech-Language Pathology, Oral Dynamics Lab, University of Toronto, Toronto, Ontario, Canada,* [2]*Research, Toronto Rehabilitation Institute, Toronto, Ontario, Canada,* [3]*Centre for Research on Biological Communication Systems, Department of Psychology, University of Toronto, Mississauga, Ontario, Canada, and* [4]*Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada*

(Received 1 June 2010; accepted 19 October 2010)

Abstract

Primary objective: The ability to identify emotions in spoken language is an essential component of communication and could be disrupted in persons with brain injury. Current tools to assess this function show important shortcomings. The aim is to present a set of validated and linguistically equated lexical sentences that can be used to separate the impact of lexical content and prosody on the processing of emotion in speech in persons with brain injury.

Methods and procedures: Using six-point Likert scales, a set of 125 sentences, carefully matched for linguistic variables, were rated by a group of young adults ($n = 48$) on their suitability to represent a particular emotion (anger, fear, happiness and sadness) in their lexical content.

Main outcomes and results: The findings identified a set of 50 sentences that were reliably associated with one particular emotion only or no emotion at all (neutral). Using less stringent criteria, 94 sentences were also found to be good representatives for these affective categories.

Conclusions: The findings generated a robust set of validated lexical stimuli necessary to reliably identify the specific contributions of verbal and prosodic information on difficulties in identifying emotions in speech with persons with brain injury.

Keywords: *Emotion, speech, sentences, rating, ABI*

Introduction

The identification of emotions in spoken language is part and parcel of spoken communication. Consider the following example, listening to the radio you hear the talk show host saying 'I am so happy'. The sentence conveys an emotion of happiness in its lexical meaning, but the speaker uses a tone of voice that conveys anger (e.g. with an inflection on the word 'so'). What emotion is this person trying to convey to the listeners? Is it happiness, anger or perhaps irony/sarcasm (i.e. a discrepancy between the expected emotion and the one that is presented)? This example portrays the complex interaction of the two dimensions that convey emotion in spoken language viz. the lexical content (text) and the prosody (tone of voice). To partake effectively in social dialogue, it is essential to identify, understand and respond appropriately to this interaction. Fortunately, in daily communications this is something most people achieve without any difficulty. Given its central role in communication, it is surprising to find only a limited number of studies

Correspondence: Boaz M. Ben-David, Oral Dynamics Lab, Department of Speech-Language Pathology, University of Toronto, 160-500 University Avenue, Toronto, Ontario, M5G 1V7, Canada. Tel: +1-416-946-8552. Fax: +1-416-978-1596. E-mail: boaz.ben.david@utoronto.ca

ISSN 0269–9052 print/ISSN 1362–301X online
DOI: 10.3109/02699052.2010.536197

examining this interplay between lexical and prosodic expression of emotions [1].

This gap in the literature may be explained by the complexity of mimicking this process in the lab, while controlling for the possible influence of other factors. In a typical spoken language study, participants are asked to identify the emotion expressed in spoken sentences or to judge whether two sentences convey the same emotion. In order to control for possible biases, it is imperative to have a set of validated lexical sentences that convey a distinct emotion in their verbal content (and thus are not mistaken with other possible emotions). This set must consist of sentences that are equated on relevant linguistic characteristics such as average word frequency and the number of syllables. These linguistic characteristics are known to influence cognitive processing in listeners [2] and may affect identification of prosodic cues in a spoken context. Word frequency has an early influence on lexical processing, with more frequent words being more easily accessed and activated in the mental lexicon [3]. This may leave more processing resources available to process and identify the prosody of a spoken sentence. Alternatively, the early access to the semantic content of frequent words may take precedence and overshadow the emotion expressed in the prosody. Similarly, the length of the sentences (as measured by the number of syllables) can conceivably have a contradictory effect on the perception of emotion in speech. On one hand, longer spoken sentences provide more prosodic cues; on the other hand, longer spoken sentences may put a heavier load on working memory leaving fewer resources available to analyse the prosodic cues (as evidenced in studies on children with specific language impairments [4]). For these reasons, it is clear that carefully controlling for linguistic characteristics of the test materials is essential in order to assess emotional processing correctly. This is of particular importance in studying emotional processing in populations where, due to brain injury or disease, cognitive resources may be more limited. One particular group of individuals, namely those who have a known difficulty in processing emotion in spoken language, will be presented next. In this paper it will be argued that in order to understand the nature of their problems, it is imperative to have access to a well-controlled set of verbal stimuli that researchers can use in a spoken context (adding different prosodic features that correspond to specific emotions). This can assist in disambiguating the influence of verbal and prosodic expressions of emotions in identification of emotions in speech.

Alexithymia

Perhaps the best illustration of the importance of correct identification of emotions in speech is in studies on individuals with Alexithymia—an impairment in the ability to experience, describe and identify emotions [5, 6]. The most extensively used tool of self-reported Alexithymia is the Toronto Alexithymia Scale (TAS-20 [7, 8]). Alexithymia has been associated with various pathologies including depression (both in the general population [9] and in persons with psychiatric disorders [10]), post-traumatic stress disorder [11] and eating disorders [12] (including both bulimia [13] and anorexia nervosa [14]). It has also been associated with a compound of borderline behaviours such as substance dependence [15], violent offences [16] and gambling [17]. Alexithymia, as evaluated by the TAS-20, has three main factors: difficulties in identifying emotions, describing emotions and externally-oriented thinking. This study is focused on one aspect of Alexithymia (that can be assessed behaviourally), namely the difficulty in identifying emotions in speech (EID). In order to assess this aspect (behaviourally) in the most reliable way, one needs to create a set of sentences for which the lexical content can uniquely be associated with specific emotions.

EID and acquired brain injury (ABI). A unique case study for the diverse effects related to EID is ABI (noting that most studies on EID were related specifically to traumatic brain injury (TBI), this paper is using the more general ABI terminology). ABI, a leading cause of disability, has wide-ranging effects on cognitive and behavioural functioning due to involvement of both the frontal and temporal lobes of the brain and diffuse injury to nerve axons. Persons with ABI suffer from reduced quality-of-life, increased anxiety and depression [18] and an inability to recognize emotional cues and respond appropriately during social interactions [19]. In several studies, EID was found to be a possible important cause of social isolation and behavioural dysfunction following an ABI, as communication is the medium for building social bonds. For example, Henry et al. [20] have shown that EID scores can be highly predictive of reported quality-of-life of persons with ABI, independent of other contributing factors, concluding that 'rehabilitation efforts should be focused on remediation of this specific deficit' (p. 70). Recently, there has been a growing interest in developing rehabilitation tools for EID in ABI [21], which so far are virtually non-existent [22] (an exception may be found in pitch feedback treatment of motor aprosodia [23]). This deficit in communication has not received enough attention in the literature and, as a result, its assessment is still very limited, as explained in the next paragraph.

Existing tools for EID assessment. The two most common tools for behavioural assessment of EID (as opposed to a self-report as provided by the TAS-20) are the FAB (Florida Affect Battery [24]) and the DANVA2 (Diagnostic Analysis of Nonverbal Accuracy version 2 [25]). However, one can maintain that these two tests are limited in their ability to reveal the full complexity of the interaction of emotions conveyed in both lexical content and prosody, specifically in patient populations. The DANVA2 only includes one lexical sentence 'I am going out of the room now, but I'll be back later' spoken in different prosodies. Since the largest differences between persons with brain injury and controls were found in the identification of irony, where the lexical content of the sentences contrasted with the emotion presented in the prosody [26, 27], the DANVA2 does not suffice. It is imperative to use sentences for which lexical content is systematically varied to reliably express different emotions. The FAB (in sub-tests 7–8a) uses 20 neutral sentences and (in sub-test 8b) 36 affective sentences. However, despite being more elaborate than the DANVA2 in this respect, it does have various shortcomings for studying EID. First, the authors of the FAB provide no information as to whether the lexical content of these sentences was equated for linguistic characteristics across the emotional categories (and it appears that this was not the case). Secondly, the affective sentences in this test are spoken only in two emotional prosodies, which are either matched or not with the emotion conveyed by the lexical content (limiting the scope of the test). However, most importantly, the authors did not validate these sentences by verifying that the emotion conveyed by their lexical content is distinct (i.e. cannot be interpreted as conveying other emotions) or at least they do not provide such information. Aside from these two tools, other studies found in the pertinent literature usually focus on the prosody and not on the complex interaction of prosody and lexical content. As a result, various study-specific lists of sentences have been developed, but the authors did not validate them for their emotional lexical content or equate them for their linguistic characteristics (see Pell et al. [28] for one such study).

A recent paper by Russ et al. [29] presents a first systematic attempt to present a list of sentences that convey distinct emotional information based on lexical content. Even though the Russ et al. [29] study has many merits, their list of sentences may not meet the criteria for tests of emotional processing with persons with ABI. First, the Russ et al. study used a relatively small ($n = 12$) group of undergraduate students who were 'motivated, highly proficient and closely supervised' (p. 938). The current study emphasizes the special nature of this group of raters, as it may introduce a (unknown) bias in the ratings. This bias may be avoided by recruiting a more diverse group of individuals, less 'motivated' and 'proficient', in other words a group more similar to people (patients and otherwise) for whom this stimulus material was designed. Secondly, their sentences were not equated for linguistic characteristics and contained a mix of questions and assertions. Questions impose specific variations in word order and, when spoken, in prosody (e.g. rising intonation at the end of a sentence). These variations may confuse listeners (especially persons with ABI) with respect to mixing linguistic and emotional signals (similar to findings on the impact of mixed signals of this kind in facial expression for Sign Language [30]). In addition, there is evidence that the processing of questions differs from other sentences in the involvement of selective parts of the brain [31], which may be of particular relevance in persons with ABI.

The current study

This paper proposes that the availability of a validated and linguistically equated set of sentences is instrumental to correctly assess EID in ABI and other patient populations. Mainly, in tests of identification of emotion in spoken language, it is important to separate the impact of the lexical content of the sentences from the impact of the emotional information conveyed by prosodic cues. To this end, a set of lexical sentences is presented in five affective categories: Anger, Fear, Happiness, Sadness and Neutral. These emotions were chosen as they represent the current convention in the literature (see Zupan et al. [1] for a review) and, in a spoken context, they are found to be easily recognized and distinguished in prosody [32, 33]. It is maintained that by using this set of sentences in a test of spoken language the researcher/clinician can have a clear evaluation of behavioural EID, controlling for specific emotions conveyed by the lexical content only.

In a pre-test, 500 different sentences were rated, from which a total of 125 sentences (25 in each affective category) were selected for a final test, that showed a relative distinctiveness in conveying a specific emotion in their lexical content. The next paragraphs portray in general terms how these sentences were further evaluated. More details can be found in the Method section.

Validity of the emotional (semantic) content. Raters were asked to rate all 125 sentences on four six-point Likert scales [34], corresponding to the four tested affective emotions (Anger, Fear, Sadness, Happiness). Participants were not asked to rate

sentences for a neutral emotion, since this type of rating was found to be inaccurate [35]. For each emotional category, sentences were chosen with the highest ratings on that category and lowest on all others. By choosing sentences whose averaged ratings in each scale had low standard deviations, it was verified that these ratings reflected a high level of agreement amongst raters. As a further method for assessing raters' agreement, responses were grouped on the six-point Likert scale in three sub-categories as high (6–5), medium (4–3) and low (2–1) emotional ratings. Based on the percentage of raters that provided high-score ratings for sentences (on their corresponding emotional scales), it was possible to ascertain that the lexical content of these sentences was taken to represent a particular emotion. In the neutral category, the authors targeted sentences that had low ratings on all four emotional scales, with low standard deviations indicating an agreement on those ratings across raters. It should be noted that the criteria used in the current study differ from Russ et al. [29], as they used a coefficient of variation (CoV) instead of standard deviations. Furthermore, for neutral sentences they looked for low overall agreement amongst raters (high variability). The authors of the current study do not concur with this choice of criteria. First, CoV is only recommended for data on a ratio scale [36] and is not appropriate for Likert scales. Secondly, low agreement among raters does not preclude ambiguity for a neutral sentence, whereas consistent low scores across raters, for all four emotions, provide a better litmus test for the 'neutrality' of sentences. Finally, this study verified that affective sentences (Anger, Fear, Happiness and Sadness) received an average rating that was significantly better than a set threshold on their corresponding emotional scale and an average rating that was significantly worse than a set threshold on their non-corresponding emotional scales. Similarly, it was verified that neutral sentences received ratings that were less than 2.0 on all emotional scales. By this, this study was able to validate that, with respect to lexical content, affective sentences represent their corresponding emotions and that the neutral sentences are indeed affectively neutral.

Linguistic characteristics. Sets of sentences were selected that were matched on *frequency of usage in the English language* and on *phonologic neighbourhood density*, which on their own or in combination affect the temporal course of cognitive processes [37]. With respect to the processing of emotional word stimuli, a voluminous literature on emotional Stroop can attest to the effects of linguistic characteristics on an emotion-related effect. The emotional Stroop paradigm tests the latency of responses when threat words are presented visually (for a review see Algom et al. [38]). For example, Larsen et al. [39] found in a meta-analysis a correlation between the linguistic characteristics of words and the emotional Stroop effect (but see [40]), where the emotional delay was related to lower frequency words and smaller orthographic neighbourhoods (note, these studies were visually-based and hence did not include phonologic neighbourhood density). A more recent study [41] also found that with proper controls for length, orthographic neighbourhood density and arousal, word frequency and emotional valence interact in contributing to the emotional Stroop effect.

Generally, words that are more frequent in the lexicon will make it easier for participants to identify them and their emotional content [42]. Neuroimaging studies have indicated that lexical word frequency and emotional valence interact (at least with respect to brain activation). For example, Scott et al. [43] found that high frequency negative words recruit additional cognitive resources as compared to other stimuli. It is clear then that if sentences in one emotional category (e.g. Fear) will have (on average) a higher lexical frequency than sentences in other categories (e.g. Anger, Happiness and Sadness), word frequency alone can be the cause of a difference in identification of emotions between these categories.

Phonologic neighbourhood density was found to have a significant effect on spoken word recognition. Recall that the current study is designed to provide a set of sentences that can be used in future studies in a spoken format, as well as in a printed format. Thus, even though the sentences presented here were rated on lexical content (and presented visually, in a print form), it is important to control for this (phonological) factor as well. Phonological neighbourhood density is measured as the number of 'neighbours' (i.e. words differing from the target word by a single phoneme) a word has in the English lexicon. Namely, a word with a denser phonologic neighbourhood sounds very similar to other words in the mental lexicon. Several studies have shown that the identification of spoken words slows down as density increases [44–47]. Thus, the average phonologic neighbourhood density of a sentence can have a significant effect on the identification of emotion in speech. Recently, Goh et al. [45] showed that the effects of phonologic neighbourhood density and word frequency on spoken word recognition are not additive, but interact in a complex fashion. For the reasons described above, in the current study the average word frequency and phonologic neighbourhood density of words do not differ significantly between emotional categories. Thus, differences in identification of emotions between these

categories (in both spoken and print contexts) cannot be attributed to variation in those linguistic characteristics.

In addition to the linguistic characteristics of the *content words*, other linguistic features related to the *sentences as a unit* may also create a bias in the ability to identify emotions in spoken language. First, consider the length of an utterance. Longer sentences may be more complex and require extra resources to process. For example, in a study on speech comprehension, Lidestam et al. [49] found that shorter sentences may improve performance equally for negative and positive emotional sentences. This could suggest that sentence length (and complexity) may influence one's ability to identify emotion, in particular for patients with limited cognitive resources. Therefore, it was important to constrain the approximate length of the lexical sentences in this study to an average of seven syllables (range = 3–10 syllables) and to ensure that sets of sentences in the emotional categories were matched on the average number of syllables.

As mentioned earlier, mixing statements with questions may not be a good strategy, as it affects word order and prosody (when used in a spoken context). For these reasons, none of the sentences used in this study was phrased in the form of a question. Finally, the repetition of critical words in the sentences was kept to a minimum. This was done to prevent practice effects and word habituation. For example, when the content word '*better*' is seen and/or heard for the third time (as it appears in three sentences in the Happiness category of Russ et al.'s [29] list), the listener will adapt to it. This could either prime the emotion associated with the word or possibly dilute it. Furthermore, a variety of sentences can serve to trigger idiosyncratic performance profiles in EID, specifically in a pathological population that is characterized with Alexithymia (as in ABI).

Raters and validity. The ecological validity of the sentences was a priority in constructing the set of sentences. To this end, and in contrast to the study by Russ et al. [29], this study recruited a large group of (relatively) naïve raters (61, excluding 13 as outliers), that had not received any training in this type of task. This is important, as it is anticipated that these sentences will be used in assessments of the ability to identify emotions in a variety of populations, including and perhaps most importantly, persons with Alexithymia. There is a limitation in the age ranges one can accept for this type of task, as older individuals may respond quite differently from younger individuals in terms of emotional valence judgements. Recent studies show age-related differences in the encoding of emotional stimuli [50, 51], although the perception of prosody may be unaffected by age, as older adults depend more on amplitude envelope (suprasegmental) cues in speech recognition [52] and prosody was found to facilitate recognition to the same degree for younger and older adults [53]. To avoid potential biases and to keep the raters group relatively homogenous, only young adults were recruited. This study further excluded outlier raters by two measures of validity: inconsistency of responses and their infrequency. To maintain test validity this study controlled for both the order of sentence-presentation and the order of emotional scales (see Method section for a further discussion on the validity of raters and of the test).

Summary. Following the above depicted criteria, this study presents a set of 50 sentences, validated for their emotional (lexical) content in each of the following five categories: Anger, Sadness, Fear, Happiness and Neutral (10 sentences each). It also presents a set of 94 sentences in the five emotional categories that are based on less stringent criteria. It is suggested that these sets form, to date, the most optimized lexical stimuli to test the identification of emotion in both visually presented and spoken language, specifically with patient populations with acquired brain injury.

Method

Participants

Sixty-one young adults (ages ranging from 18–30 years old) participated in this study. They were either undergraduates at the University of Toronto, Mississauga, or individuals who had completed a bachelor's degree. Participants received either course credit or were paid $10/hour for their participation. All participants were native English speakers, as assessed by a self-report, and achieved a minimum Mill-Hill Vocabulary Test [54] score of 9/20, corresponding to normal vocabulary levels for native-English speakers. All participants indicated in a self-report that they did not suffer from any psychopathological state or took psycho-active medication. Data from 13 participants had to be discarded as these participants were flagged as outliers (based on the criteria mentioned above and in the following sections), leaving 48 raters for the current data set.

Stimuli

In a set of pre-tests, 88 undergraduate students rated emotional words found in the literature. Based on these ratings, a list of 500 sentences was constructed. Some of these sentences were constructed by altering sentences collected from various resources

available on-line or published in the literature (e.g. TIMIT [55] and SPIN [56]). In the next phase of the study, a group of 40 young adults (different from the ones mentioned above under *participants*) rated portions of these sentences (in a printed format) on their emotional content, until the authors were able to construct a list of 125 candidate sentences, 25 for each emotional category (Anger, Fear, Happiness, Sadness and Neutral). This selected list comprised the sentences used for the current study that either best described uniquely one intended emotion (for affective sentences) or none (for neutral sentences), based on these pre-test ratings.

Procedure

Each participant completed the rating study individually on a computer console using SurveyMonkey (https://www.surveymonkey.com). Participants were asked to rate each visually presented sentence four times on a 6-point Likert scale corresponding to the four emotions: Anger, Fear, Happiness or Sadness. The rating task was introduced as follows:

> How much do you agree with the following statement: 'this sentence conveyed a happy [fearful, sad, angry] emotion' from 1 'completely disagree' to 6 'completely agree'?

Note, the original scale used in the study was reversed, but, for ease of reporting the data, the ratings have been converted, so that a higher rating reflects a higher agreement. Responses were made by pressing the appropriate key on a keyboard. To preserve a high level of attention throughout the study and avoid fatigue, the list of 125 sentences was randomly divided into two sections, section A with 62 sentences and section B with 63 sentences. Each section was presented to a different group of 24 participants (after excluding outliers). Sentences were grouped into four experimental blocks. In each block, participants were asked to rate the list of sentences on a different emotional scale. Controlling for order effects is specifically important in the context of emotional ratings, as the emotion presented in one trial may serve as a context on which the next sentence will be rated. To counteract these biases, two methods were utilized. First, the order of sentence presentation within each block was randomized, to ensure that sentence order would vary between blocks and between participants. Secondly, the order of the four emotional blocks was counterbalanced across subjects using a Latin square procedure [57], making for four different block-order conditions. This counterbalancing procedure, when combined with the two sections (each including half of the stimuli), generated eight experimental groups (six participants per group). In general, each participant was asked to rate 248 (in section A) or 252 (in section B) critical items, with an additional 20 catch trials as outlined in the following section. A description of the eight experimental groups is presented in Table I. All together, the experimental sessions took no longer than 45 minutes for each subject.

Raters' response validity and outlier selection

To control for *inconsistency* of responses, five control-sentences, one in each category, were presented twice in each experimental block (in a randomized order), generating 20 catch trials (five sentences × four blocks). Participants who did not rate the repeated trials of the five control-sentences consistently (a rating difference of more than two

Table I. The order of the emotional rating blocks is denoted as group-orders 1, 2, 3 and 4 and the division of sentences as sections A and B.

Experimental groups	Number of participants	Number of sentences	Number of trials	Order of emotional rating blocks			
				Anger	Sadness	Fear	Happiness
1A	6	62:	248	1st	2nd	3rd	4th
2A	6	12 Anger		4th	3rd	2nd	1st
3A	6	13 sadness		2nd	4th	1st	3rd
4A	6	13 fear		3rd	1st	4th	2nd
		12 happiness					
		12 neutral					
1B	6	63:	252	1st	2nd	3rd	4th
2B	6	13 Anger		4th	3rd	2nd	1st
3B	6	12 sadness		2nd	4th	1st	3rd
4B	6	12 fear		3rd	1st	4th	2nd
		13 happiness					
		13 neutral					
Sum	48	125					

points in one emotional scale, for at least one sentence) were flagged as outliers and removed from further analysis. Furthermore, participants who failed to rate more than 5% of all critical items (>5% missing trials) were also excluded. To control for *infrequent* responses, in each section five sentences were sampled, one from each emotional category: four affective sentences that were rated extremely high on their corresponding emotional scale during pre-tests and a neutral one that was rated low on all four scales (see Appendix C). Participants who did not agree that the sampled affective sentences represented their corresponding emotion (i.e. provided a rating lower than 4) or that the sampled neutral sentence did not represent any emotional scale (i.e. provided a rating higher than 3 for any given scale) were again excluded as outliers. In total, the ratings of 10 participants were excluded based on these criteria, the ratings of two others based on low scores on a vocabulary test and the ratings of another participant on the basis of a high rate of missing data (total: 13).

Linguistic characteristics

To ascertain the linguistic characteristics of the sentences, first, all content words were collected from each sentence (i.e. in the sentence 'I really love nature', the words 'really', 'love' and 'nature', but not the function word 'I'). Next, from the English lexicon project data-base [58], their frequency of usage in the English lexicon was gathered, as measured by the Hyperspace Analogue to Language frequency norms (HAL [59]), and the density of their phonological neighbourhood. Finally, for each sentence the linguistic characteristics of its content words were averaged. Furthermore, all of the sentences tested in the pre-tests and experimental tests had no more than 10 syllables and no less than three syllables, with an average length of seven syllables.

Data analysis

Missing responses on single cases were treated as missing data in the analysis (less than 0.3% of the responses). For each sentence on each rating scale, the average and the standard deviations of the responses of the 24 participants were calculated for each section. Responses that exceeded 2.5 SD from the mean were excluded from analysis (less than 1.0% of all responses).

Results

Top 50 sentences

The top 10 sentences in each of the five emotional-categories (Anger, Fear, Sadness, Happiness and Neutral) that in their lexical content best represented their designated emotions and are matched on linguistic characteristics were selected for the sentence list presented in Appendix A.

Ratings. Table II presents the average ratings on the four emotional scales for this list of the 'top 50' sentences. As a reminder, in the 6-point Likert scale used in this study a rating of 1–3 indicates that the rater disagrees that the sentences represented the emotional scale and a rating of 4–6 indicates that the rater agrees.

First, consider the shaded data cells that present the average ratings of affective sentences in each emotional category on their corresponding emotional scale (i.e. Anger-category rated on an Anger-scale, Fear on Fear, Happiness on Happiness and Sadness on Sadness). The average rating across the four categories was 5.45 (SD = 0.3), with ratings of single sentences ranging from 5.0–6.0. These robust high-scores reflect that the 40 affective sentences were judged by the raters as strong representations of their corresponding emotions. Next, consider the ratings of affective sentences in each emotional category on their non-corresponding emotions (e.g. sentences in the Anger-category rated on Happiness, Sadness and Fear scales). Across the 40 affective sentences, the average rating was 1.65 (SD = 0.3), with average ratings for non-corresponding emotions on single sentences ranging from 1.0–2.0. These low scores indicate that the 40 affective sentences were

Table II. Ratings were on a 6-point Likert scale, ranging from 1 (completely disagree that the sentence conveys a ___ emotion) to 6 (completely agree). Shaded data cells present average ratings of affective sentences on their corresponding emotional scales.

Emotional categories		*n*	Ratings on emotional scales			
			Anger	Fear	Happiness	Sadness
Affective	Anger	10	5.3 (0.3)	1.9 (0.2)	1.4 (0.2)	2.1 (0.4)
	Fear	10	2.2 (0.3)	5.5 (0.2)	1.5 (0.2)	1.8 (0.3)
	Happiness	10	1.1 (0.1)	1.2 (0.1)	5.6 (0.2)	1.2 (0.1)
	Sadness	10	2.0 (0.3)	2.1 (0.3)	1.3 (0.2)	5.4 (0.2)
	Neutral	10	1.3 (0.1)	1.3 (0.1)	2.2 (0.1)	1.3 (0.1)

Table III. High-score rating—the score 5 or 6 in a 6-point Likert scale corresponding to '(completely) agree that the sentence conveys a ___ emotion'. Shaded data cells present average ratings of affective sentences on their corresponding emotional scales.

Emotional categories		*n*	Ratings on emotional scales: Anger	Fear	Happiness	Sadness
Affective	Anger	10	87%	2%	0%	5%
	Fear	10	5%	92%	0%	2%
	Happiness	10	0%	0%	98%	0%
	Sadness	10	3%	2%	0%	86%
	Neutral	10	0%	0%	0%	0%

not confused by the raters with any of the other tested emotions. Finally, ratings for the neutral sentences were examined. The average rating for the 10 neutral sentences on all four emotional scales (by their nature, neutral sentences do not have a corresponding emotional scale) was 1.5 (SD = 0.03), with the range for single sentences (across emotions) varying from 1.5–1.6. These consistent low scores can attest to the fact that the 10 neutral sentences were not judged by the raters as reflecting any one of the four tested emotions.

To confirm that affective sentences were consistently judged better on their corresponding emotion than on their non-corresponding emotions, 40 separate paired sample *t*-tests (one for each affective sentence) were conducted, comparing the score on the corresponding emotion with the average score on all other emotional scales (e.g. for a sentence in the Fear category, the average score on the Sadness, Happiness and Anger scales was compared with the score on the Fear scale). All 40 tests were found to be significant ($t > 8.8$, $p < 0.001$, for all tests), indicating the validity of the ratings. Next, in a series of post-hoc one-sample *t*-tests (with a Bonferroni correction of the probabilities for two comparisons), it was verified that individual sentences received an average rating that was significantly different than a set threshold. First, it was confirmed that average ratings for each of the 40 affective sentences across all three non-corresponding emotions were significantly below 2.5, indicating that these sentences were judged as poor representations of the non-corresponding emotions ($t > 2.7$, $p < 0.05$, for all 40 tests). Next, it was found that ratings of the 40 affective sentences on their corresponding emotions were significantly above 4.5, indicating that they were considered an excellent representation of their corresponding emotions ($t > 2.8$, $p < 0.05$, for all 40 tests). Similarly, in a set of one-sample *t*-tests it was found that for all 10 neutral sentences the average ratings across the four emotional scales were significantly below 2.0 ($t > 2.7$, $p < 0.05$, for all 10 tests), confirming that these sentences were considered a poor representation of all the tested emotions.

The series of one-sample *t*-tests described above can also attest to the consensus among the raters. To further ascertain this conformity, Table III presents the average percentage of raters that provided a high-score of 5 or 6 on each emotional scale for sentences in the five emotional scales. A high percentage of high-score ratings indicates an agreement amongst raters that a sentence reflects a particular emotional scale, whereas when (almost) none of the ratings were 5 or 6, it indicates a consensus amongst raters that that sentence does not strongly reflect an emotional scale. Examining the shaded data cells that present ratings of affective sentences on their corresponding emotions, the average percentage across the 40 sentences was 91%, with a range of 70–100% for ratings of single sentences. This indicates a high consensus amongst the raters that the lexical content of these sentences is strongly associated with the corresponding emotions. The average percentage of high score ratings of the 40 affective sentences on their non-corresponding emotional scales was extremely low, 1.2%, with a range of 0–8% for averages of ratings of single sentences on their three non-corresponding emotional scales. Together, the data in Table III demonstrate a high consensus among raters that the affective sentences were not associated with emotions other than their own corresponding emotion. Finally, none of the 10 neutral sentences has received on any single emotional scale a score of 5–6 from even one of the 48 raters. Again, one can observe a high consensus among the raters that the 10 neutral sentences do not reflect any of the tested emotions.

The standard deviations of ratings of individual sentences were also used as a threshold criterion for the conformity of raters. The average standard deviation for emotional ratings of the top 50 sentences, across the four emotional scales, was 0.74. When selecting sentences for the top 50 list, sentences were excluded if they had an average standard deviation across the four emotional scales that was higher than 1.0. It is important to note that standard deviations were low for neutral sentences (0.69), as well as for the affective sentences (0.75), indicating low variability across raters in both classes.

Table IV. Data on word frequency in the English lexicon and word phonologic neighbourhood density are averaged across the content words in each sentence. The numbers of syllables are measured for the full sentence.

Emotional categories		*n*	Linguistic characteristics		
			Number of syllables	Frequency (HAL)	Phonologic neighbourhood
Affective	Anger	10	6.30 (1.9)	10.55 (1.8)	14.45 (12.4)
	Fear	10	7.00 (2.0)	10.06 (1.4)	14.48 (10.5)
	Happiness	10	7.00 (1.4)	10.14 (1.9)	9.03 (3.4)
	Sadness	10	6.50 (1.3)	10.06 (1.4)	11.36 (8.6)
	Neutral	10	6.70 (0.9)	10.03 (1.1)	13.64 (5.8)

Similar to Russ et al. [29], this study also found that Happy sentences showed in general more distinct ratings and more consistency among the raters compared to the other affective sentences. First, average rating of sentences from the Happiness category on the Happiness scale were significantly higher than the average rating of sentences in the other affective categories on their own corresponding scales (5.6 vs. 5.4, $t(38) = 2.0$, $p < 0.05$). Secondly, the average rating of sentences in the Happiness category on their non-corresponding scales was significantly lower than the average rating of sentences in the other affective categories on their non-corresponding emotional scales (1.2 vs. 1.8, $t(38) = 12.8$, $p < 0.001$). Taken together, this indicates that Happy sentences were rated as a more distinct representation of their corresponding emotion than sentences in other affective categories. Thirdly, it was found that ratings of Happy sentences were characterized by a higher conformity across raters than in the other affective categories, as indicated by the average standard deviations across the four emotional scales (0.40 vs. 0.87, $t(38) = 14.0$, $p < 0.001$). Finally, 98.3% of the ratings of Happy sentences on the Happiness scale fell in the high-score range (5–6). This contrasts with the percentage of the ratings of the other affective sentences that were high on their corresponding emotional scales (an average of 88.5%; $t(38) = 3.1$, $p < 0.005$).

Linguistic characteristics. The averages for the linguistic characteristics of the 50 sentences in each emotional category are presented in Table IV. To estimate whether there were any significant differences between the emotional categories, three separate ANOVAs were conducted, with the type of category (Anger, Sadness, Happiness, Fear and Neutral) as a between-participant factor. In the first ANOVA the dependent variable was the number of syllables, in the second, the average word frequency and, in the third, the density of the phonological neighbourhood. All three ANOVAs were non-significant ($F(4, 45) < 1$ and $p > 0.5$ for all three tests), therefore the five emotional categories (sets of sentences) were found to be balanced in their linguistic characteristics, with an overall average across all five emotional categories of 6.70 syllables (SD = 1.5) per sentence, an average frequency of content words of 10.17 (HAL scale, SD = 1.3) per sentence and an average phonological neighbourhood density of the content words of 12.59 (SD = 8.7) per sentence.

Top 94 sentences

Acknowledging that some experiments or tests may require more than 50 sentences, or more than 10 sentences in a specific emotional category, this study also included an analysis of the top 94 sentences from the study. These 94 sentences, 74 affective sentences in four emotional-categories and 20 neutral sentences, were found to reflect their corresponding emotion (affective sentences) or no emotion (neutral) and were balanced on their linguistic characteristics. As the top 94 sentences list includes the 50 sentences analysed earlier, Appendix B presents the list of *additional* sentences. These additional 44 sentences are based on *less stringent* criteria regarding their ratings than the ones used for the top 50 sentences; however, they still provide a set of validated sentences.

Ratings. Table V presents the average emotional ratings of the 94 sentences grouped into five emotional-categories, 16 in an Anger-category, 19 in a Fear-category, 20 in a Happiness-category, 19 in a Sadness-category and 20 in a Neutral-category. The average rating of the 74 affective sentences on their corresponding emotional scales (presented in the shaded cells) was 5.4 (SD = 0.3), with ratings of a single sentence ranging from 4.7–6.0. The average rating for affective sentences on their non-corresponding emotions was 1.7 (SD = 0.4), with a range of 1.0–2.5 for ratings of a single sentence (across three non-corresponding scales). Thus, the 74 emotional sentences were judged by the raters to reflect only their corresponding emotion. The average rating of the 20 neutral sentences on all four

Table V. Ratings were on a 6-point Likert scale, ranging from 1 (completely disagree that the sentence conveys a ___ emotion) to 6 (completely agree). Shaded data cells present average ratings of affective sentences on their corresponding emotional scales.

Emotional categories		*n*	Ratings on emotional scales			
			Anger	Fear	Happiness	Sadness
Affective	Anger	16	5.1 (0.3)	1.8 (0.2)	1.5 (0.2)	2.3 (0.4)
	Fear	19	2.3 (0.3)	5.4 (0.3)	1.5 (0.2)	2.1 (0.4)
	Happiness	20	1.2 (0.1)	1.3 (0.2)	5.5 (0.2)	1.2 (0.2)
	Sad	19	2.3 (0.5)	2.4 (0.5)	1.3 (0.2)	5.4 (0.3)
	Neutral	20	1.4 (0.1)	1.3 (0.1)	2.3 (0.1)	1.3 (0.1)

Table VI. High-score ratings—5 or 6 in a 6-point Likert scale corresponding to '(completely) agree that the sentence conveys a ___ emotion'. Shaded data cells present average ratings of affective sentences on their corresponding emotional scales.

Emotional categories		*n*	Ratings on emotional scales			
			Anger	Fear	Happiness	Sadness
Affective	Anger	16	81%	1%	0%	5%
	Fear	19	6%	89%	0%	4%
	Happiness	20	0%	0%	94%	0%
	Sad	19	4%	8%	0%	87%
	Neutral	20	0%	0%	1%	0%

emotional scales was 1.6 (SD = 0.05), with the range for single sentences from 1.5–1.7, indicating that the 20 neutral sentences were not judged by the raters as reflecting any of the four tested emotions. In a set of paired sample *t*-tests (one for each sentence) it was found that the score which raters provided for an affective sentence on its corresponding emotional scale was significantly different than the average rating on all of the non-corresponding emotional scales ($t > 10.2$, $p < 0.001$ for all tests). In post-hoc *t*-tests, Bonferroni corrected for two comparisons, it was verified that the average ratings of affective sentences across the three non-corresponding emotional scales were significantly below 3.0, a rating reflecting a poor representation of these emotions ($t > 2.5$, $p < 0.05$, for all tests). Similarly, average ratings of sentences on their corresponding emotions were significantly above 4.3, a rating reflecting a good representation ($t > 2.3$, $p < 0.05$, for all tests, Bonferroni corrected). The neutral sentences were all found to have average scores across the four emotional scales that were below 2.0 ($t > 2.7$, $p < 0.05$, for all tests), indicating very poor representation for any of the tested emotions.

To validate the conformity across raters, Table VI presents the percentage of raters that provided a score of 5 or 6 on an emotional scale. Eighty-eight per cent of ratings of affective sentences on their corresponding emotions (shaded cells) were high, with a range of 65–100% for ratings of single sentences. Only 2% of the scores given to these 74 sentences on their non-corresponding emotional scales were 5–6 (with a range of 0–11% for single sentences). Less than 1% of the ratings given to neutral sentences were high on any of the four emotional scales (with a range of 0–2%, for single sentences). Therefore, ratings indeed reflected a high consensus amongst the raters. The standard deviations of ratings of individual sentences were further used as a criterion for the conformity of raters. The average standard deviation for the top 94 sentences, across the four emotional scales, was 0.78. This average was low for both affective (0.80) and neutral sentences (0.72). The criterion for exclusion from the top 94 sentences list was an average standard deviation above 1.1.

The advantage of the Happy sentences over sentences in the other affective categories was preserved in the list of 94 sentences. Average ratings of Happy sentences on the Happiness scale were significantly higher than the average ratings of other affective sentences on their corresponding scales (5.54 vs. 5.34, $t(72) = 2.7$, $p < 0.01$) and ratings of Happy sentences on their non-corresponding emotional categories were significantly poorer than the equivalent ratings of the other affective sentences (1.2 vs. 1.9, $t(38) = 14.0$, $p < 0.001$). Ratings of Happy sentences were characterized by a higher

Table VII. Data on word frequency in the English lexicon and word phonologic neighbourhood density are averaged across the content words in each sentence. The numbers of syllables are measured for the full sentence.

Emotional categories		*n*	Linguistic characteristics		
			Number of syllables	Frequency (HAL)	Phonologic neighbourhood
Affective	Anger	16	6.13 (1.7)	10.81 (1.8)	12.95 (10.3)
	Fear	19	6.53 (1.8)	9.71 (1.6)	10.74 (8.9)
	Happiness	20	6.60 (1.4)	10.43 (1.6)	7.83 (4.3)
	Sad	19	6.58 (1.3)	10.14 (1.2)	11.00 (7.3)
	Neutral	20	6.85 (1.2)	9.81 (1.4)	11.28 (6.5)

conformity across raters, as indicated by the average standard deviations across the four emotional scales (0.50 vs. 0.91, $t(72) = 13.1$, $p < 0.001$) and by the percentage of high-score ratings given to sentences on their non-corresponding emotional scales (0.1% vs. 3.3%, $t(72) = 3.2$, $p < 0.005$) and the percentage of high-score ratings given to sentences on their corresponding scales (94.0% vs. 85.9 %, $t(72) = 4.9$, $p < 0.001$). Taken together, these analyses on a larger set again show that sentences in the Happy category were characterized with stronger ratings and a higher conformity among raters.

Linguistic characteristics. Table VII presents the averages for the linguistic characteristics of the 94 sentences in each emotional-category. Three separate ANOVAs were conducted, with the type of category (Anger, Sadness, Happiness, Fear and Neutral) as a between-participant independent factor and as the dependent variables the number of syllables, the average frequency of words in the English lexicon and the average density of the phonological neighbourhood. All three ANOVAs were non-significant ($F(4, 89) < 1.6$ and $p > 0.1$ for all three tests), confirming that the five emotional categories were not significantly different in their linguistic characteristics, with an overall average across all five emotional categories of 6.55 syllables per sentence (SD = 1.8), an average frequency of content words of 10.16 (HAL scale, SD = 1.6) per sentence and an average phonological neighbourhood density of content words of 10.67 (SD = 7.6) per sentence.

Discussion

The current study was set up to generate a list of validated lexical sentences that can convey a specific emotional content (or no emotional content) in a reliable manner while being matched for linguistic features, such as number of syllables, word frequency and neighbourhood density. This list was designed to allow researchers and clinicians to assess the ability of individuals to identify emotions in both written and spoken language, in particular individuals who have a known deficit in this faculty (e.g. persons with ABI). Ratings, performed by a group of 48 representative and naïve young adults, clearly identified sentences that on the basis of their lexical content are well-chosen prototypes for each of the five selected emotional categories (Anger, Sadness, Happiness, Fear and Neutral). To accommodate different experimental and clinical needs, data were provided on a smaller set of 50 sentences that showed the strongest consensus (and highest ratings) and a larger set of 94 sentences that were still excellent representatives for each emotional category, but to a slightly lesser degree.

As discussed in the introduction to this paper, this is not the first attempt to generate a list of emotional content sentences and in particular the recent paper by Russ et al. [29] is an example of such previous efforts. However, it is believed the current set is an important improvement on the sentences presented by these authors for various reasons. First, across the different emotional categories, sentences in the present lists are carefully matched for basic linguistic characteristics. Secondly, this study did not use questions which could bias the findings in testing for identifying emotional content when mixed with other types of sentences [30]. Thirdly, there are also some important differences with the current study in terms of methodology, in particular with respect to the number and type of raters (this study recruited a large number of naïve raters) and the way to establish reliability for the ratings (variance vs. coefficient of variance), as detailed in the outset of this study. Similar to the Russ et al. [29] study, the current study found a strong consensus among the raters showing high (5–6 on a 6-point Likert scale) and consistent agreement ratings for affective sentences belonging to their corresponding emotional category, while at the same time these sentences were consistently rated as low (1–2) for alternative emotional categories. For neutral sentences, raters were also in agreement that they did not belong to any particular emotional category. Given the lack of clear gender differences in distinguishing between emotional categories of verbal

content [29], this study did not further pursue gender-based analyses of the data.

This study also replicated the Russ et al. [29] finding that Happy sentences in general show higher and more consistent ratings compared to other emotions. This confirms that overall an emotion of Happiness seems somewhat easier to identify in a print context, perhaps in the way emotional verbal content may induce an emotional state in the reader, which in turn may influence attention differently for Happy and non-Happy emotions [60, 61]. Another possible explanation is that Happiness was the only positive emotion tested (both in this study and in Russ et al. [29]), whereas Anger, Fear and Sadness are all negative emotions. A Happy sentence may be unique in its emotional ratings only in the context of the other emotions investigated here.

The main weakness of the current study is that raters were only recruited from a specific cohort of young adult English-speaking residents of southern Ontario, Canada. Although the group was larger and more representative than the one used in a previous study [29], it remains to be determined in future studies if the sentences are equally applicable to other groups, especially where there are cultural and language background differences. However, based on the strong consistency of the emotional ratings and the fact that the sentences were rated as extremely (smaller set, top 50 sentences) or very (larger set, top 94 sentences) prototypical for a given emotional category, there is no reason to expect any major differences with other speakers from a North-American background.

Conclusion

The database presented in this study provides a meaningful set of validated lexical stimuli that can be used for assessing the ability to identify emotions in spoken language in people with and without disorders in this faculty. This set, when used in a spoken context, can further assist to separately identify the impact of prosody and lexical content on identification of emotions in speech. Since difficulty in identifying emotions in speech has significant impacts on the rehabilitation of persons with ABI [20–22], it is believed this database can be useful as a tool to improve the reliability of the assessment and the rehabilitation of communication skills of persons with ABI.

Acknowledgements

The authors wish to thank the following students for their contribution: Nicole Durham, Eshika Mehreen, Namita Multani and Linh (Le Truc) Nguyen.

***Declaration of interest*:** This research was undertaken, in part, thanks to funding from the Canada Research Chairs program (303712CRC) awarded to the second author, and a Research Opportunity Program grant from the Faculty of Arts & Science at the University of Toronto Mississauga awarded to the first and second author. B. M. Ben-David was partially supported by a grant from the Ontario Neurotrauma Foundation (2008-ABI-PDF-659).

References

1. Zupan B, Neumann D, Babbage DR, Willer B. The importance of vocal affect to bimodal processing of emotion: Implications for individuals with traumatic brain injury. Journal of Communication Disorders 2009;42:1–17.
2. Kuchinke L, Võ MLH, Hofmann M, Jacobs AM. Pupillary responses during lexical decisions vary with word frequency but not emotional valence. International Journal of Psychophysiology 2007;65:132–140.
3. Strijkers K, Costa A, Thierry G. Tracking lexical access in speech production: Electrophysiological correlates of word frequency and cognate effects. Cerebral Cortex 2010;20: 912–928.
4. Marton K, Schwartz R. Working memory capacity and language processes in children with specific language impairment. Journal of Speech, Language, and Hearing Research 2003;46:1138–1153.
5. Nemiah JC, Freyberger H, Sifneos PE. Alexithymia: A view of the psychosomatic process. In: Hill OW, editor. Modern trends in psychosomatic medicine. Vol. 2, London: Butterworth; 1976. pp 26–34.
6. Larsen JK, Brand N, Bermond B, Hijman R. Cognitive and emotional characteristics of alexithymia. A review of neurobiological studies. Journal of Psychosomatic Research 2003;54:533–541.
7. Bagby RM, Taylor GJ, Parker JD. The Twenty-item Toronto Alexithymia Scale–II. Convergent, discriminant, and concurrent validity. Journal of Psychosomatic Research 1994;38: 33–40.
8. Parker J, Taylor G, Bagby R. The 20-Item Toronto Alexithymia Scale: III. Reliability and factorial validity in a community population. Journal of Psychosomatic Research 2003;55:269–275.
9. Honkalampi K, Hintikka J, Tanskanen A, Lehtonen J, Viinamäki H. Depression is strongly associated with alexithymia in the general population. Journal of Psychosomatic Research 2003;48:99–104.
10. Wise T, Mann L, Hill B. Alexithymia and depressed mood in the psychiatric patient. Psychotherapy and Psychosomatics 1990;54:26–31.
11. Cloitre M, Scarvalone P, Difede J. Posttraumatic stress disorder, self- and interpersonal dysfunction among sexually retraumatized women. Journal of Trauma and Stress 1997;10:437–452.
12. Cochrane C, Brewerton T, Wilson D, Hodges E. Alexithymia in the eating disorders. International Journal of Eating Disorders 1993;14:219–222.
13. De Groot J, Rodin G, Olmsted M. Alexithymia, depression, and treatment outcome in bulimia nervosa. Comparative Psychiatry 1995;36:53–60.

14. Bourke M, Taylor G, Parker J, Bagby R. Alexithymia in women with anorexia nervosa. A preliminary investigation. British Journal of Psychiatry 1992;161:240–243.
15. Haviland MG, Hendryx MS, Shaw DG, Henry JP. Alexithymia in women and men hospitalized for psychoactive substance dependence. Comparative Psychiatry 1994;35: 124–128.
16. Louth S, Hare R, Linden W. Psychopathy and alexithymia in female offenders. Canadian Journal of Behavioural Science 1998;30:91–98.
17. Mitrovic D, Brown J. Poker mania and problem gambling: A study of distorted cognitions, motivation and alexithymia. Journal of Gambling Studies 2009;25:489–502.
18. Dijkers MP. Quality of life after traumatic brain injury. A review of research approaches and findings. Archives of Physical Medicine and Rehabilitation 2004;85:S21–S35.
19. Milders M, Fuchs S, Crawford JR. Neuropsychological impairments and changes in emotional and social behaviour following severe traumatic brain injury. Journal of Clinical and Experimental Neuropsychology 2003;25:157–172.
20. Henry JD, Phillips LH, Crawford JR, Theodorou G, Summers F. Cognitive and psychosocial correlates of alexithymia following traumatic brain injury. Neuropsychologia 2006;44:62–72.
21. Dimoska A, McDonald S, Pell M, Tate R, James C. Recognizing vocal expressions of emotion in patients with social skills deficits following traumatic brain injury. Journal of the International Neuropsychological Society 2010; in press.
22. Bornhofen C, McDonald S. Emotion perception deficits following traumatic brain injury: A review of the evidence and rationale for intervention. Journal of the International Neuropsychological Society 2008;14:511–525.
23. Stringer A. Treatment of motor aprosodia with pitch biofeedback and expression modelling. Brain Injury 1996;10:583–590.
24. Bowers D, Blonder LX, Heilman KM. Introduction to Florida affect battery revised. University of Florida, Center for Neuropsychological Studies, Cognitive Neuroscience Laboratory; 1999. Available online at: http://www.neurology.ufl.edu/forms/fab_manual.pdf, accessed 15 April 2010.
25. Nowicki Jr S, Duke M. Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. Journal of Nonverbal Behavior 1994;18:9–35.
26. Tonks J, Williams W, Frampton I, Yates P, Slater A. Reading emotions after child brain injury: A comparison between children with brain injury and non-injured controls. Brain Injury 2007;21:731–739.
27. Simpson A, Schmitter-Edgecombe M. Intactness of inhibitory attention mechanisms following severe closed head injury. Neuropsychology 2000;14:310–319.
28. Pell MD, Paulmann S, Dara C, Alasseri A, Kotz SA. Factors in the recognition of vocally expressed emotions: A comparison of four languages. Journal of Phonetics 2009;37:417–435.
29. Russ J, Gur R, Bilker W. Validation of affective and neutral sentence content for prosodic testing. Behavioral Research Methods 2008;40:935–939.
30. De Vos C, Van Der Kooij E, Crasborn O. Mixed signals: Combining linguistic and affective functions of eyebrows in questions in sign language of the Netherlands. Language and Speech 2009;52:2–3.
31. Doherty C, West W, Dilley L, Shattuck-Hufnagel S, Caplan D. Question/statement judgments: An fMRI study of intonation processing. Human Brain Mapping 2004;23: 85–98.
32. Scherer K. Vocal communication of emotion: A review of research paradigms. Speech Communication 2003;40: 227–256.
33. Laukka P. Categorical perception of emotion in vocal expression. Annals of the New-York Academy of Science 2003;1000:283–287.
34. Likert R. A method for measuring the sales influence of a radio program. Journal of Applied Psychology 1936;20:175–182.
35. McDonald S, Flanagan S. Social perception deficits after traumatic brain injury: Interaction between emotion recognition, mentalizing ability, and social communication. Neuropsychology 2004;18:572–579.
36. Norris N. Some efficient measures of relative dispersion. The Annals of Mathematical Statistics 1938;9:214–220.
37. Adelman J, Brown G. Phonographic neighbors, not orthographic neighbors, determine word naming latencies. Psychonomic Bulletin and Review 2007;14:455–459.
38. Algom D, Chajut E, Lev S. A rational look at the emotional stroop phenomenon: A generic slowdown, not a Stroop effect. Journal of Experimental Psychology: General 2004;133:323–338.
39. Larsen RJ, Mercer KA, Balota DA. Lexical characteristics of words used in emotional Stroop experiments. Emotion 2006;6:62–72.
40. Estes Z, Adelman JS. Automatic vigilance for negative words in lexical decision and naming: Comment on Larsen, Mercer, and Balota (2006). Emotion 2008;8:441–444, discussion 445.
41. Kahan TA, Hely CD. The role of valence and frequency in the emotional Stroop task. Psychonomic Bulletin and Review 2008;15:956–960.
42. Bruce A, Harman M, Turner M. Reactions to heard words: Varying emotional content and frequency of exposure. Journal of General Psychology 2007;134:67–82.
43. Scott GG, O'Donnell PJ, Leuthold H, Sereno SC. Early emotion word processing: Evidence from event-related potentials. Biological Psychology 2009;80:95–104.
44. Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. Ear and Hearing 1998;19: 1–36.
45. Vitevitch MS, Luce P. Probabilistic phonotactics and neighborhood activation in spoken word recognition. Journal of Memory and Language 1999;40:374–408.
46. Goldinger S, Luce P, Pisoni D. Priming lexical neighbors of spoken words: Effects of competition and inhibition. Journal of Memory and Language 1989;28:501–518.
47. Ziegler J, Muneaux M, Grainger J. Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. Journal of Memory and Language 2003;48:779–793.
48. Goh W, Lidia S, Yap M, Hui Tan S. Distributional analyses in auditory lexical decision: Neighborhood density and word-frequency effects. Psychonomic Bulletin and Review 2009;16:882–887.
49. Lidestam B, Lyxell B, Andersson G. Speech-reading: Cognitive predictors and displayed emotion. Scandinavian Audiology 1999;28:211–217.
50. Addis D, Leclerc C, Muscatell K, Kensinger E. There are age-related changes in neural connectivity during the encoding of positive, but not negative, information. Cortex 2010;46:425–433.
51. Fischer H, Nyberg L, Bäckman L. Age-related differences in brain regions supporting successful encoding of emotional faces. Cortex 2010;46:490–497.
52. Ben-David BM, Chambers C, Daneman M, Pichora-Fuller MK, Reingold E, Schneider BA. Effects of aging and noise on real-time spoken word recognition: Evidence from eye movements. Journal of Speech Lang and Hearing Research 2010; in press.
53. Wingfield A, Lindfield KC, Goodglass H. Effects of age and hearing sensitivity on the use of prosodic information in

spoken word recognition. Journal of Speech Language and Hearing Research 2000;43:915–925.
54. Raven JC, Court JH, Raven J. Coloured progressive matrices. Oxford, U.K.: Oxford Psychological Press; 1995.
55. Garofolo JS. Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database. Gaithersburg, MD: National Institute of Standards and Technology (NIST); 1988.
56. Bilger R, Nuetzel J, Rabinowitz W, Rzeczkowski C. Standardization of a test of speech perception in noise. Journal of Speech and Hearing Research 1984;27:32–48.
57. Grant D. The latin square principle in the design and analysis of psychological experiments. Psychological Bulletin 1948; 45:427–442.
58. Balota D, Yap M, Cortese M, Hutchison K, Kessler B, Loftis B, Neely JH, Nelson DL, Simpson GB, Treiman R. The English lexicon project. Behavioral Research Methods 2007;39:445–459.
59. Lund K, Burgess C. Producing high-dimensional semantic spaces from lexical co-occurrence. Behavior Research Methods, Instruments, and Computers 1996;28:203–208.
60. Srinivasan N, Srivastava P, Lohani M, Baijal S. Focused and distributed attention. Progress Brain Research 2009;176: 87–100.
61. Rowe G, Hirsh J, Anderson A. Positive affect increases the breadth of attentional selection. Proceedings of the National Academy of Sciences of the USA 2007;104: 383–388.

Appendix A: A list of the top 50 sentences, 10 sentences in each emotional category

Emotional category	Sentences	
Anger	Do not push your luck.	I'm sick of you being late.
	Do not waste my time.	Quiet, this is a library.
	Get out of my room.	Stop what you're doing and listen to me.
	Go to hell.	This is infuriating.
	I am very angry.	You over-charged me for that.
Fear	He has a knife.	It's about to explode.
	I can hear footsteps in the night.	Look out there's a car coming.
	I can't see the bear but I can hear it.	Run for your life.
	I hear a sharp scream from behind.	The fire is spreading to the gas pipe.
	I smell gas leaking from the stove.	Watch out for that tiger.
Happiness	Congratulations, you're hired.	I won an award.
	Good job, the crowd loved you.	It's a beautiful day outside.
	Great, you got first place.	This food tastes very good.
	I feel wonderful today.	This is my favourite part.
	I got promoted in my job.	This is the happiest day of my life.
Sad	Gray clouds make me feel gloomy.	My pet died today.
	I am so lonely.	The weather is depressing.
	I'm going to a funeral.	This is a sad moment.
	I've been crying all day.	This scene makes him feel blue.
	My best friend is moving away.	This song makes me cry.
Neutral	A bag is in the room.	I see a rug on the floor.
	Containers have a blue lid.	Lots of bins are in the room.
	Digital clocks are common.	Our body is made of water.
	Four drawers are in the cabinet.	Red pipes are metallic.
	He stands on the deck.	This is a garbage can.

Appendix B: A list of the 44 sentences (in five emotional categories) that complement the list in Appendix A to make for the Top 94 sentences found in the study

Emotional category	Sentences	
Anger	I wasn't talking to you.	You disgust me.
	Stop wasting my time.	You need to grow up.
	This is not your concern.	You think you know everything.
Fear	I am choking.	The cobra is on the loose.
	I'm so scared.	This place is creeping me out.
	Someone is following me.	Watch out, he's got a gun
	Something is creeping up my leg.	You're starting to scare me.
	That man terrifies me.	

(*continued*)

Appedix B. Continued.

Emotional category	Sentences	
Happiness	I really love nature.	I'm marrying the one I love.
	His words make me smile.	Thanks for the present.
	I am going on vacation.	The clouds are pretty today.
	I am graduating today.	This food tastes very good.
	I love you so much.	Your kids are so cute.
Sad	I am so very sorry.	She is filled with despair.
	I have no friends.	She lost her whole family.
	I think we should see other people.	She said she wants a divorce.
	My son is miserable.	The orphans never saw their father.
	No one sat beside me at lunch.	
Neutral	Her book is under her bed.	Some tablecloths are in the basket.
	Her camera is in the bag.	The earth is round.
	His glasses are on the table.	There are magnets on the fridge.
	My spoon is on the table.	This table is brown.
	One towel is folded.	Your music sheets are on the stand.

Appendix C: A list of the 10 sentences (two in each emotional category) that were presented twice to participants, five in version A and five in version B of the survey

Emotional category	Version	Sentences
Anger	A	Do not push your luck.
	B	I'm sick of you being late.
Fear	A	It's about to explode.
	B	I can hear footsteps in the night.
Happiness	A	Good job, the crowd loved you.
	B	I feel wonderful today.
Sad	A	Gray clouds make me feel gloomy.
	B	This scene makes him feel blue.
Neutral	A	He stands on the deck.
	B	One towel is folded.