STROOP OF EMOTIONS IN SPOKEN LANGUAGE

Ben-David, B. M.[1,2,3], Joshi, D.[1,3] and Van Lieshout, P. H. H. M.[1,2,3,4,5]
[1]*Oral Dynamics Lab, Department of Speech-Language Pathology, University of Toronto, Canada*
[2]*Research, Toronto Rehabilitation Institute, Canada*
[3]*Department of Psychology, University of Toronto Mississauga, Canada*
[4]*Graduate Department of Rehabilitation Studies, University of Toronto, Canada*
[5]*Institute of Biomaterials and Biomedical Engineering, University of Toronto, Canada.*
E-mail: boaz.ben.david@utoronto.ca; Homepage: www.boazbendavid.org

Abstract

It has been argued that the emotional Stroop effect is a misnomer -- it is not a Stroop phenomenon but rather an emotional one. In this study we investigate a true Stroop of emotions. Does it take longer for listeners to correctly identify the emotion expressed in the prosody (tone of speech) of a spoken sentence when it conflicts with its lexical content, than when both dimensions are congruent? Would, in a similar fashion, a conflicting prosody interfere with the identification of the lexical content? Thirty-two participants listened to 88 spoken sentences and were asked to judge whether the speaker expressed Anger, Fear, Sadness or Happiness (forced-choice), in either the prosody or the lexical content. Significant Stroop effects were found for both prosodic and lexical judgments. We note that Stroop effects were impacted by the extent of the dimensional imbalance between baseline classification of prosody and lexical content.

Two of the hosts of this year's meeting, Drs. Algom and Chajut, have been advocating for the need to divorce the emotional Stroop effect (ESE) from its namesake, the Stroop effect (SE, e.g., Algom, Chajut & Lev, 2004; Chajut, Mama, Levy, & Algom, 2010). Indeed the two effects share a name and (to the most part) a common task -- naming the font color of a word while ignoring its content. However, the hallmark of the SE is the partition of the stimuli into congruent (e.g., the word RED printed in a red font) and incongruent (RED in green) items which defines the effect. In the ESE, participants are also asked to name the font color of words, but these words are divided into emotional and neutral semantic categories. The quality of congruity related to the SE does not apply to the ESE. The word CANCER in green is neither more nor less congruent than the word REPEAT in blue. The ESE is rather defined by the difference in performance due to the impact of the emotional content of words.

Based on the notion that the ESE is an emotional effect and not a Stroop effect, we embark on the task to define a "real" Stroop effect of emotions. That is, can we devise a task that gauges the impact of congruency using emotionally-laden stimuli? The pertinent literature provides only a few examples for this type of task. A notable example is provided by Kitayama and Ishii (2002; see also Ishii, Reyes, & Kitayama, 2003) who presented listeners with words that had positive (pleasant), neutral or negative (unpleasant) lexical meaning. These words were recorded using a prosody that was congruent with respect to the lexical meaning (i.e., LAUGH spoken in a pleasant prosody), incongruent (i.e., LAUGH in an unpleasant prosody) or neutral (i.e., LAUGH in a neutral prosody or TABLE in a pleasant prosody). Listeners were asked to either pay attention solely to the lexical content or to the prosody of the words. Word categorization (as a pleasant or an unpleasant stimulus) was facilitated (faster responses) for congruent stimuli, both in prosody and in lexical categorizations. The extent of this Stroop-like effect (for North American participants) was found to be larger for prosody classifications, indicating the significance of the lexical content to emotional judgments. However, the Kitayama & Ishii paradigm has several limitations. First, there are only two emotions presented along each dimension, both are extremes on a single continuum -- positive and negative. Yet, the semantic emotional spectrum includes emotions on various continua. Second, the paradigm makes use of single words rather than complete sentences. This choice limits the variability of emotions that can be presented in the lexical content. Moreover, words comprise of short duration speech segments whereas the majority of studies of perception of prosody used full sentences, as prosody classification is taken to be based on variation in acoustic cues that unfolds over time in the course of an utterance (see a meta-analysis by Juslin & Laukka, 2003).

In this study, we attempt to create a Stroop-like paradigm with emotional stimuli, building on and extending the Kitayama & Ishii paradigm. Instead of words that present two extreme ends of a single category, we selected validated lexical sentences that were reliably associated with one particular emotion of the following four: Anger, Sad, Fear and Happy. We also used neutral sentences that were found not to be associated with any of these emotions. All of the sentences were taken from a previous study by Ben-David, Van Lieshout, & Leszcz (2010). These sentences were recorded spoken in prosodies corresponding to the same set of emotions, creating three types of stimuli, viz. neutral, congruent and incongruent, with respect to the prosody and the lexical dimension. In one task, listeners were asked to pay attention only to the lexical content to classify the emotion presented by the sentence; in another task, they were asked to ignore the lexical content and classify the emotional content conveyed by the prosody. SE is measured as the latency advantage for classifications of a congruent block (e.g., Anger content with Anger prosody) over classifications of an incongruent block (e.g., Happy content with Anger prosody). Unlike the Kitayama paradigm, we presented stimuli in a blocked fashion, where congruent stimuli were presented together in a single congruent block, separate from incongruent stimuli (incongruent block) and neutral (baseline). This design is taken to decrease the dimensional uncertainty, as participants can learn about the relationship between the two dimensions in each block, which was found to inflate SE (for a discussion, see Melara & Algom, 2003).

An important feature of the SE relates to the impact of baseline dimensional imbalance (Melara & Algom; for an application in aging, see Ben-David & Schneider, 2009; and in brain injury, Ben-David, Nguyen & Van Lieshout, 2011). Namely, in a baseline condition, where the to-be-ignored dimension is held constant at neutral, the dimension that is more discriminable will be classified more swiftly than the other dimension, and will be more immune to interference from the other, slower dimension. In our study, dimensional imbalance was gauged in the baseline blocks by comparing the speed of emotion classification in the prosody dimension (with lexically-neutral sentences) with the speed of emotion classification in the lexical dimension (with sentences spoken in a neutral tone). This important part was missing in the Kitayama paradigm.

Two main hypotheses were constructed based on the above discussion. First, we hypothesized that the SE for classifications of the dominant dimension (the dimension that is classified faster in a baseline condition) will be smaller than the SE for the less salient dimension. In other words, the direction of the dimensional imbalance will match the direction of the Stroop asymmetry (faster baseline classifications will give rise to smaller SE, and vice versa). Second, given that we used spoken sentences rather than words, and a variety of four emotions rather than two, we might expect that prosody would have a more prominent role than was found by Kitayama & Ishii (2002). Finally, we wanted to examine the impact of the expressed emotion in the to-be-attended dimension on these factors.

Method

Participants

Thirty-two young adults (18-21 years old) participated in the study. They were undergraduates at the University of Toronto Mississauga, who received either course credit or $10/hour. All participants were native English speakers as assessed by a self-report and achieved a minimum score of 9/20 on the Mill Hill Vocabulary Test (Raven, 1965), corresponding to normal vocabulary levels for native-English speakers. All participants had pure-tone air-conduction thresholds within clinically normal limits for their age-range, from 0.25 to 3 kHz in both ears (≤ 20 dB HL).

Stimuli

We used a set of 88 linguistically equated sentences (Ben-David, Van Lieshout & Leszcz, 2010). These sentences were recorded by an actress (native English speaker) in various prosodies to generate 24 congruent sentences (6 sentences in each of the 4 emotions) and 24 incongruent sentences (2 for each incongruent combination of the four emotions in two dimensions). We also recorded a prosody-baseline condition, with 20 lexically-neutral sentences spoken in four prosodies (five in each emotion); for a lexical-baseline condition, we recorded 20 emotional sentences spoken in a neutral prosody (five in each emotion). In a pre-test, 24 participants (taken from the same population as the participants in this study) selected spoken sentences that present strong exemplars of the expressed prosodic emotion.

Procedure

Participants were tested separately in a sound-attenuated booth. Spoken sentences were presented on headphones and participants were asked to classify the emotion expressed as anger, sad, fear, or happiness by pressing the respective keys on a keyboard. Each participant took part in two different classification tasks: *lexical-classification*, where participants were asked to attend solely to the lexical content of the spoken words and classify their emotional content; and *prosody-classification*, where they were instructed to ignore the lexical content, focus only on the prosody and classify the prosodic emotion. In both tasks, there were three separate experimental blocks: baseline (where the to-be-ignored dimension was neutral), congruent and incongruent. Figure 1 portrays the combination of dimensions presented in the six experimental blocks.

To control for the possible effects of practice, participants were randomly assigned to one of four experimental groups (eight participants in each). For half of the participants, the first task was prosody-classification and for the other half, the first task was lexical-classification. For all groups, each task started with the baseline block. For half of the participants, the second block was a congruent block and for the other half, an incongruent block.

Analysis of Latencies

Latencies were measured from the offset of the spoken sentence to a key-press. Only accurate responses were analyzed, and responses that were slower than 1700 ms (M + 3 * SD) or conducted before the end of the sentence were removed (excluding 2.5% of the trials). The last line in Table 1 presents latencies for the baseline condition (left-most columns) and the SEs (right-most columns) averaged across the four expressed emotions.

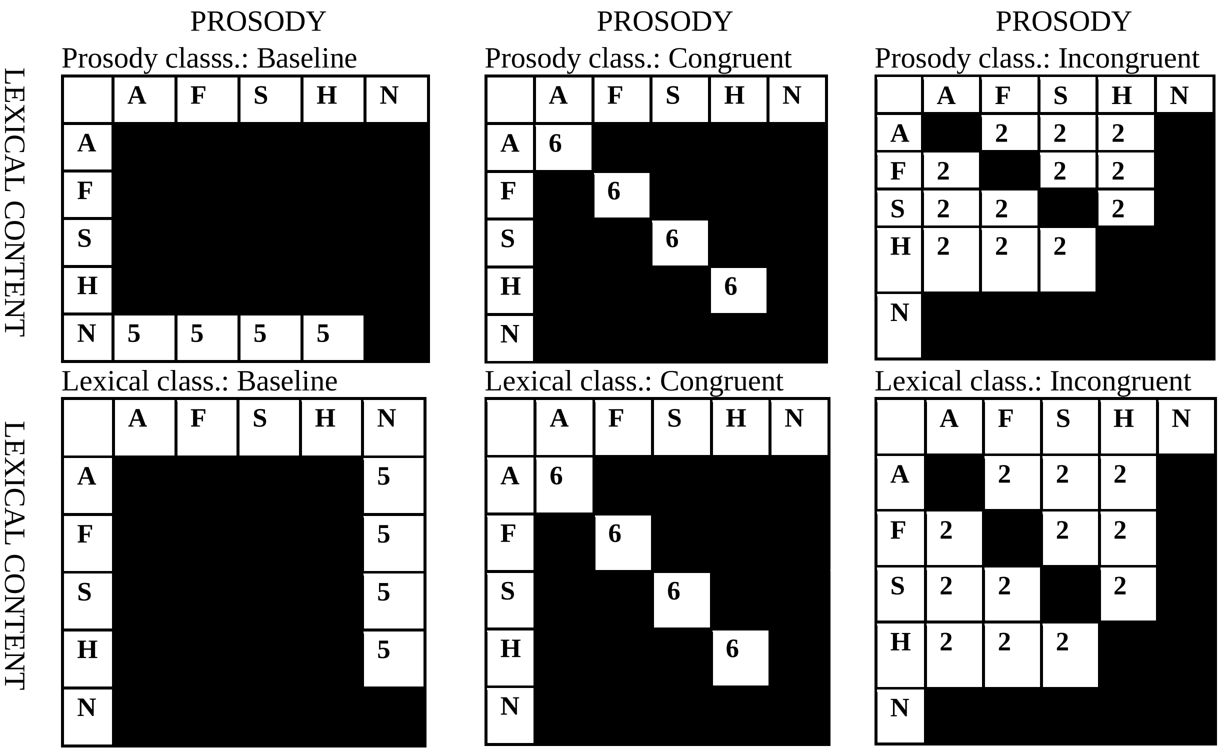

Figure 1. Distribution of trials in the six experimental blocks. In each panel, a row presents an emotional lexical content, and a column presents an emotional prosody. Emotions are: A – Anger, F – Fear, S – Sad, H – Happy and N – Neutral. The top three panels present trials in prosody-classification blocks, and the bottom three panels, present lexical classifications.

Table 1. Average response latencies for correct emotion classifications across 32 participants.

	Baseline		Dimensional	Stroop Effects		Stroop
	Lexical	Prosody	Imbalance (lex – pros)	Lexical	Prosody	Asymmetry (lex-pros)
Anger	703 (39)	421 (36)	282 (50)**	17 (25)	-41 (25)	57 (31)#
Sad	570 (38)	485 (45)	86 (52)*	303 (35)**	83 (22)**	220 (40)**
Fear	536 (36)	593 (45)	-57 (49)	2 (32)	116 (35)**	-114 (45)*
Happy	375 (26)	443 (35)	-68 (25)*	78 (24)**	100 (27)**	-23 (35)
Overall	*540 (30)*	*476 (36)*	*64 (35)**	*99(19)***	*50 (14)***	*49 (20)**

** $p < .01$, * $p < .05$, # $p = .07$

Results

Accuracy of responses in general was high, 96.8% (*SE* = .3%) and higher for congruent trials (98.7%) compared to incongruent ones (94.2%, $p < .001$), but there was no significant difference between the accuracy for prosody and lexical classifications and the two factors did not interact. Furthermore, we note that accuracy for prosody-classification of Fear was lower than the average for the other emotions (88% vs. 99%), indicating either the difficulty in identifying the fear prosody in our sample of recorded sentences, or an aversion from classifying fearful stimuli (= the ESE). Accuracy will not be further discussed.

Table 1 presents the latency data for correct emotional classification. First, we examine the overall average of baseline blocks. Baseline classifications of prosody were on average faster than baseline classifications of the lexical content, resulting in a significant dimensional imbalance of 64 ms. Next, latencies in the congruent and incongruent blocks were submitted to a repeated-measure ANOVA with congruency (congruent vs. incongruent) and task (prosody vs. lexical classification) as within-participants factors (designation to experimental groups was used as a covariate). We found a significant interaction between the task and congruency factors, indicating a larger SE for lexical classification. Examining the data closer, we found significant SEs for both lexical and prosody classifications, yet the SE for lexical classifications was twice as large, resulting in a Stroop asymmetry of 49 ms. That is, overall averages indicate a dominance of the prosodic dimension.

The four top lines in Table 1 present the latency data separately for each conveyed emotion. Baseline classifications were submitted to a repeated measure ANOVA with task (prosody vs. lexical classification) and conveyed emotion (Anger, Sad, Fear or Happy) as within-participants factors. We found a significant interaction of emotion and task, indicating that two emotions, Anger and Sad, presented an advantage for prosody, whereas baseline classification of the two other emotions, Fear and Happy, showed the *reverse* pattern. Next, we investigated responses on the congruent and incongruent blocks for each emotion separately. For Anger and Sad there was an interaction of congruency and task, indicating larger SEs for the lexical dimension. In contrast, for Fear and Happy the interactions were *reversed*, with larger SEs for the prosody dimension (at least nominally).

Discussion

This study presents a Stroop paradigm comprised of stimuli laden with emotional content: Spoken sentences, where four emotions varied along two dimensions, prosody and lexical content. The findings indicate significant SEs for both prosody- and lexical-classification. These SEs may reflect the interactive nature of both dimensions in the process of understanding emotions in speech. Second, unlike previous studies, the prosody was found to be dominant in emotion classification, when latencies were averaged across all 4 emotions. This larger role of prosody may be related to the length of the speech segments: sentences (in our study) vs. single words. This may also relate to the increased difficulty of the task, as we presented four emotions rather than two. Alternatively, it could be specific to our stimulus-set.

However, the general findings need to be qualified as shown in Figure 2. Our data reaffirms the “seesaw relationship” (Algom et al., 2004) of dimensions in the SE. A dimensional imbalance in the direction of one dimension (indicating that it was classified faster in baseline) was matched by a Stroop asymmetry in the same direction (indicating smaller SEs in the classifications of this dimension). That is, the dimension that was more swiftly classified was harder to ignore. This trend was found both in in an overall analysis and in a separate analysis of each expressed emotion. Finally, Figure 2 shows that not all emotions are the same when it comes to speech perception. These results resonate with data collected in our lab and presented in a different paper in this volume (“T-RES: Test of rating of emotions in speech”). Specifically, the prosodic prominence in classifications of Anger and Sadness may relate to the findings reported in a meta-analysis by Juslin and Laukka (2003) indicating that these two emotions are better communicated in prosody (and music) than Fear and Happiness.

Acknowledgments

Portions of this study are based on an undergraduate independent research project completed at the University of Toronto Mississauga by D. Joshi, supervised by the other authors. B. M. Ben-David was partially supported by a grant from the Ontario Neurotrauma Foundation (2008-ABI-PDF-659). This research was undertaken, in part, thanks to funding from the

Canada Research Chairs program (303712CRC) awarded to P. H. H. M. van Lieshout. We thank Kareem Lezama, Jeannyne Salvatierra and Aisha Yusuf Ibrahim for their assistance.

References

Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional Stroop phenomenon: a generic slowdown, not a Stroop effect. *Journal of Experimental Psychology: General*, *133*, 323-338.

Ben-David, B. M., van Lieshout, P. H. H. M.., & Leszcz, T. (2010). A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. *Brain Injury, 25*, 206-220.

Ben-David, B. M., Nguyen, L. L., & van Lieshout, P.H.H.M. (2001). Stroop effects in persons with traumatic brain injury: Selective attention, speed of processing, or color naming? A meta-analysis. *Journal of the International Neuropsychological Society. 17,* 354-363.

Ben-David, B. M., & Schneider, B. A. (2009). A sensory origin for color-word Stroop effects in aging: A meta-analysis. *Neuropsychology, Development and Cognition*. *16,* 505-534.

Chajut, E., Mama, Y., Levi, L., & Algom, D. (2010). Avoiding the approach trap: A response bias theory of the emotional Stroop effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1567-1572.

Ishii, K., Reyes, J. A., & Kitayama, S. (2003). Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychological Science, 14,* 39-46.

Juslin, P.N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770–814.

Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition and Emotion, 16,* 29-59.

Melara, R. D., & Algom, D. (2003). Driven by Information: A Tectonic theory of Stroop effects. *Psychological Review*, *110,* 422-471.

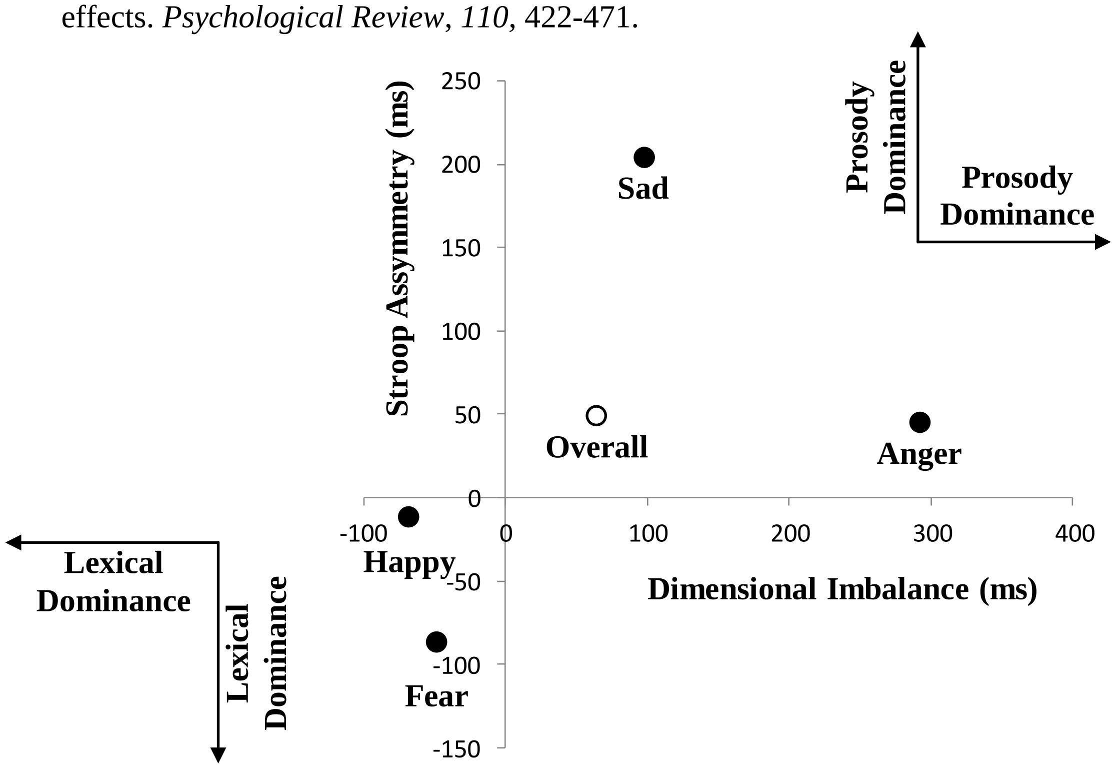

Figure 2. Stroop Asymmetry -- the difference between SE for lexical and prosody classifications -- as a function of dimensional imbalance -- the difference between baseline-latencies for lexical and prosody classifications.