

Beyond lexical meaning: The effect of emotional prosody on spoken word recognition

Seung Kyung Kim, and Meghan Sumner

Citation: *The Journal of the Acoustical Society of America* **142**, EL49 (2017); doi: 10.1121/1.4991328

View online: <https://doi.org/10.1121/1.4991328>

View Table of Contents: <https://asa.scitation.org/toc/jas/142/1>

Published by the [Acoustical Society of America](#)

ARTICLES YOU MAY BE INTERESTED IN

[Matched guise effects can be robust to speech style](#)

The Journal of the Acoustical Society of America **142**, EL18 (2017); <https://doi.org/10.1121/1.4990399>

[Effect of F0 contours on top-down repair of interrupted speech](#)

The Journal of the Acoustical Society of America **142**, EL7 (2017); <https://doi.org/10.1121/1.4990398>

[Effects of phonetically-cued talker variation on semantic encoding](#)

The Journal of the Acoustical Society of America **134**, EL485 (2013); <https://doi.org/10.1121/1.4826151>

[Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion](#)

The Journal of the Acoustical Society of America **93**, 1097 (1993); <https://doi.org/10.1121/1.405558>

[Comparing malleability of phonetic category between \[i\] and \[u\]](#)

The Journal of the Acoustical Society of America **142**, EL42 (2017); <https://doi.org/10.1121/1.4986422>

[Effects of emotional prosody on word recognition](#)

The Journal of the Acoustical Society of America **134**, 4232 (2013); <https://doi.org/10.1121/1.4831551>



Beyond lexical meaning: The effect of emotional prosody on spoken word recognition

Seung Kyung Kim^{a)} and Meghan Sumner

Department of Linguistics, Stanford University, Stanford, California 94305, USA
seung-kyung.kim@lpl-aix.fr, sumner@stanford.edu

Abstract: This study employs an auditory-visual associative priming paradigm to test whether non-emotional words uttered in emotional prosody (e.g., *pineapple* spoken in angry prosody or happy prosody) facilitate recognition of semantically emotional words (e.g., *mad*, *upset* or *smile*, *joy*). The results show an affective priming effect between emotional prosody and emotional words independent of lexical carriers of the prosody. Learned acoustic patterns in speech (e.g., emotional prosody) map directly to social concepts and representations, and this social information influences the spoken word recognition process.

© 2017 Acoustical Society of America

[MC]

Date Received: January 12, 2017 Date Accepted: June 19, 2017

1. Introduction

Spoken words convey much more meaning, or information, than simply the lexical meaning. If someone excitedly exclaims, “Pineapple!” this speech signal conveys lexical information, the spiky fruit, but it simultaneously conveys a wide range of additional information. For example, it communicates that this speaker is angry or happy, depending on the particular phonetic modification she made to express her emotional state. This additional dimension of information conveyed by phonetic variation in spoken language beyond the lexical meaning (which we refer to as *phonetically cued social information*, or simply as *social information*) is an integral, essential part of any speech signal.

How phonetically cued social information interacts, and is integrated, with linguistic information during the process of spoken word recognition is an open question. The general assumption has been that phonetically cued social information is either ignored or non-integral for spoken word recognition, but more recent approaches suggest this information is in fact integral to recognizing spoken words (e.g., Sumner *et al.*, 2014). The current study focuses on one type of phonetically cued social information, phonetically cued *emotional* information (i.e., emotional prosody) and investigates effects of emotional prosody on spoken word recognition. Specifically, we test whether information extracted from emotional prosody activates words related to that emotion independent of the words carrying the emotional prosody. From semantic priming studies, we know that recognition of a word is facilitated when preceded by a semantically related word (Swinney, 1979). If emotional information carried by emotional prosody can influence the network of the mental lexicon during word recognition just as lexical meaning does, we can expect that words uttered with emotional prosody (e.g., *pineapple* uttered with angry or happy prosody, henceforth *pineapple*_{Angry/Happy}), facilitate the recognition of non-lexically, but emotionally associated words (e.g., *mad* or *smile*).

The processing of emotional prosody and emotional words has been extensively investigated by emotion researchers, especially in neuroscience (see Kotz and Paulmann, 2011, for a review). This body of literature provides valuable insights into emotion processing, especially into neural structures supporting emotion processing. Fewer studies examine emotional prosody with a clear focus on language processing, but they are gaining ground (e.g., Nygaard and Queen, 2008; Quené *et al.*, 2012; Schirmer *et al.*, 2002; Wurm *et al.*, 2001). Using four types of prosody (happy, disgusted, petrified, and neutral) and three types of emotional words (happy, disgusted, and petrified) in auditory lexical decision tasks, Wurm *et al.* (2001) tested whether recognizing emotional words is facilitated or inhibited when emotional words are spoken with a matching emotional prosody (e.g., *glad*_{Happy}) or with a mismatching emotional prosody (e.g., *glad*_{Disgusted}) compared to the emotional words spoken with neutral

^{a)} Author to whom correspondence should be addressed. Also at: Aix-Marseille Univ, CNRS, LPL, Aix-en-Provence, France.

prosody (e.g., *glad*_{Neutral}). When each participant heard only one of the four types of prosody throughout the experiment, emotional words spoken with the matching prosody were recognized faster than emotional words spoken with the neutral or mismatching prosodies. But when each participant experienced all four types of prosody randomly mixed throughout the experiment, no such congruence effect was found. Nygaard and Queen (2008) also obtained similar results using a naming task.¹

These studies provide initial evidence that emotional prosody can influence the spoken word recognition process, especially in the form of a congruence effect between prosodic/emotional information and lexical information. However, existing work has several limitations. First, congruence effects of emotional prosody on emotional words have been reliably demonstrated only in a single prosody condition but not in a multiple prosody condition. If we want to show that emotional prosody influences on-line spoken word processing, it is crucial to establish an effect in a mixed-prosody setting because a congruence effect in a single-prosody condition could easily arise from expectation-based strategic processing (Wurm *et al.*, 2001). When participants have a clear expectation about their experience, processing of stimuli that conforms to the expectation is facilitated. Thus, when listeners experience only one type of prosody throughout an experiment, it is likely they develop a prosodic expectation that can be manifested as a congruence effect.

A more general limitation of existing work is that almost all the previous studies examine the effect of emotional prosody carried by semantically emotional words. Solely examining emotional words makes it difficult to tease apart the effect of the lexical meaning from the effect of emotional prosody, and we cannot evaluate whether emotional prosody has a meaningful effect on spoken word recognition in general or whether it has an effect only for emotional words with which it is congruently paired. In other words, a congruence effect between emotional prosody and semantically emotional words has a limited domain. Congruence can be evoked when explaining data about semantically happy words (e.g., *exuberance*) spoken with an angry tone of voice, but congruence cannot be evoked the same way when explaining data about, for example, non-emotional words spoken with angry prosody (*refrigerator*). Responding to these limitations of existing work, in the current study we use semantically non-emotional lexical carriers with multiple types of prosody.

In sum, the purpose of the present investigation is to test whether emotional information conveyed by prosody, independent of a lexical carrier, can directly activate words associated with the corresponding emotion, just as lexical information is known to activate words semantically associated with it.

2. Experiment

We used an auditory-visual associative priming paradigm to test our hypothesis that non-emotional words uttered with emotional prosody (e.g., *pineapple*_{Angry/Happy}) facilitate recognition of corresponding emotional words (e.g., *mad*, *upset* or *smile*, *joy*).

2.1 Participants

Eighty-two native English speakers from the Stanford University community participated in the study for pay. No hearing- or vision-related issues were reported.

2.2 Auditory primes

For critical trials, 24 semantically non-emotional words (e.g., *pineapple*, *transmission*)² were recorded by a female speaker of American English with two types of emotional prosody (angry and happy) and with neutral prosody. Semantically non-emotional words can still have a positive or negative meaning, and this could potentially influence how emotional prosody is processed (e.g., Ben-David *et al.*, 2016). In a separate survey, we obtained valence ratings of the written prime words (60 raters per word), and the mean valence rating was 5.21 [0.96 standard deviation (SD)] on a 9-point scale, showing that the meaning of the words was not biased to negative or positive.³

Which acoustic features characterize a particular type of prosody is an important question, but to our purpose, it is more important to verify if the auditory stimuli are perceived as intended by listeners. For each of the 72 auditory stimuli, 25 naive listeners rated it on three 9-point scales: how emotional, how angry, and how happy a given spoken word sounded. The results are shown in Table 1.

A series of linear regressions was conducted, and results confirm that our stimuli were perceived as intended. On a scale of how emotional a token sounded, both angry and happy stimuli scored much higher than neutral stimuli ($t_s > 24$, $p_s < 0.001$). On a sounds-angry scale, angry stimuli scored much higher than the other two prosodies ($t_s > 17$, $p_s < 0.001$). Finally, on a sounds-happy scale, happy stimuli scored much higher than the other two ($t_s > 21$, $p_s < 0.001$).

Table 1. Perceptual emotional ratings of critical auditory stimuli. Standard deviations are shown in parentheses. The scale ranges from 1 (not at all) to 9 (definitely).

	Sounds Emotional	Sounds Angry	Sounds Happy
Angry Prosody Stimuli	7.0 (0.7)	5.0 (0.9)	3.7 (0.9)
Happy Prosody Stimuli	6.6 (0.7)	1.7 (0.4)	6.9 (0.6)
Neutral Prosody Stimuli	3.6 (0.4)	2.3 (0.5)	3.9 (0.5)

2.3 Visual targets

To obtain a set of critical target words associated with a given emotion, we asked 100 people to provide five words that come to their mind for the words *angry* and *happy*. We selected the probe (the name of the emotion, *angry* and *happy*) and their eleven most frequently reported words for the target words. The 24 target words are provided in Table 2 along with their association strength to the corresponding emotion. For example, 37 out of 100 people provided the word *mad* as one of the five associated words for *angry*. The association strength ranges from 0.07 to 0.4, excluding the probe themselves.

Including 12 associated words in an associative priming study is very unusual. In fact, we are not aware of any study that uses such a broad set of associates. In a typical semantic priming study, the target word is simply the top semantic associate of the prime (e.g., table – chair), not a list of associates (e.g., table – chair, desk, leg, etc.). This is because semantic priming effects decrease significantly as the association strengths between prime and target decrease. In typical semantic priming studies the association strength between a prime and its top associate is often above 0.5 and at least 0.2 or 0.25. In the current study, however, including a broader set of associated words is necessary. It is methodologically infeasible to use only one associate for the two types of emotional prosody we investigate; we would have had only two critical trials.

Including a broad set of associates in the experiment requires examining the results in a different way than a typical priming study with a single top associate. Categorically comparing related trials to unrelated trials, which is the typical way of analyzing reaction time data in priming studies, is unlikely to yield a meaningful result because of the large range of association strengths, and we will need to analyze reaction times *within related trials* as a function of how strongly prime prosodies and target words are associated. In other words, we expect there will not be a main effect of preceding prosody *per se*, but rather an interaction between prosody and association strength.

2.4 Design

The experiment was a 2×2 within-participant design. Two types of target words, angry-related (e.g., *upset*, *mad*) and happy-related (*smile*, *joy*), were preceded by semantically unrelated, non-emotional words uttered either in emotionally related prosody (*pineapple*_{Angry/Happy}) or in unrelated neutral prosody (*pineapple*_{Neutral}). The 72 critical auditory primes (24 words \times 3 types of prosody) and the 24 critical visual target words (12 words \times 2 emotions) were pseudo-randomly paired and crossed in four experimental lists. Each list had 24 critical trials including all 24 emotional target words. In addition, each list had 72 filler trials (24 with real-word targets and 48 with non-word targets; 24 for each prosody type).

2.5 Experimental procedures

The experiment consisted of 96 trials and took approximately 10 min. Each trial started with a spoken prime word played through headphones while the computer screen remained blank. At 100 ms after the offset of the prime, a written target word was presented on the screen until the participant made a lexical decision. Participants

Table 2. Emotion-associated visual target words with association strengths.

Top 12 Angry Associated Words	angry (1.00), mad (0.37), upset (0.20), fight (0.16), furious (0.11), pissed (0.11), rage (0.11), violence (0.11), irritated (0.10), red (0.09), annoyed (0.08), hot (0.07)
Top 12 Happy Associated Words	happy (1.00), smile (0.40), joy (0.38), laugh (0.19), enjoy (0.18), cheerful (0.17), glad (0.15), fun (0.11), love (0.10), family (0.08), elated (0.07), friends (0.07)

were instructed to respond as quickly and as accurately as possible. The accuracy and the latency of each lexical decision were recorded.

3. Results

3.1 Statistical analysis procedures

We used mixed-effects models for analyzing both accuracy (generalized linear models) and latency (linear models). All analyses were carried out using R's lme4 (Bates *et al.*, 2015) and lmerTest (Kuznetsova *et al.*, 2016) packages. Subjects and target words were specified as random factors. We report the results from models with a random effect structure justified by the data, which show the smallest Akaike Information Criterion (AIC) (Matuschek *et al.*, 2015; cf. Barr *et al.*, 2013). In all models, continuous variables were centered. Categorical variables were sum-coded unless a treatment coding was necessary for making relevant inferences. For latency analyses, target word frequency (log) and trial count were included as control factors. Also, we used median absolute deviation (MAD) around the median to determine outliers (Leys *et al.*, 2013). Data points that fell outside of 3 MAD of the median were discarded from the analyses.

3.2 Accuracy

Two participants whose lexical decision accuracy was below 3 MAD from the median were excluded. One additional subject was discarded as the data were not recorded due to a technical error. The analyses are thus based on data from 79 subjects. Overall, the mean accuracy rate was 96.8% (2.3% SD) with a range of 89.6%–100%. Within critical trials, the mean accuracy rate was 98.6% (2.4% SD) with a range of 91.7%–100%.

A generalized mixed-effects model was first fitted to the entire data including both critical trials and filler trials with factors of target word type (Angry, Happy, Real-Word Filler, or Non-Word Filler) and prime prosody (Emotional or Neutral). The lexical decision for non-word fillers were significantly less likely to be accurate than the average [$\beta = -0.53$, standard error (s.e.) = 0.24, $z = -2.2$, $p = 0.03$], and real-word fillers were marginally less likely to be accurate ($\beta = -0.46$, s.e. = 0.28, $z = -1.66$, $p = 0.097$). No other factors and their interactions emerged significantly influencing the accuracy rate. Another model was fitted only to critical trials with factors of target word type (Angry or Happy), prime prosody (Emotional or Neutral), and target word frequency. Target word frequency had a strong effect on accuracy rate. As lexical frequency increases, targets were more likely to be accurately responded to ($\beta = 1.37$, s.e. = 0.35, $z = 3.9$, $p < 0001$). Further, target word type had a significant effect. Angry related target words were more likely to be accurately responded to than the average ($\beta = 0.44$, s.e. = 0.21, $z = 2.1$, $p = 0.04$). No other factors and their interactions emerged significantly influencing the accuracy rate.

3.3 Latency

Latency analysis includes RTs to critical trials that are correctly responded to. All analyses were carried out with log RTs. Among 1896 critical trials (24 trials \times 79 participants), 26 trials incorrectly responded to (1.4%) were discarded. Then, RTs that fell outside of 3 MAD of the median were discarded (1.9%). The following analyses are based on the resulting 1835 RTs from 79 participants. The grand mean log RT was 6.28 (532 ms) with the standard deviation of 0.23.

The relationship between RTs and association strength is shown in Fig. 1. The possible association strength range is 0% to 100%. It translates in log space from $-\infty$ to 4.6. In Fig. 1, the range of x axis is set from 1.95 (7%) to 4.6 (100%). The solid line represents when the prime was spoken with the related emotional prosody, and the dotted line represents when the prime was spoken with the unrelated neutral prosody. For angry targets, as the target words get more strongly associated to the angry emotion, the solid line shows a clear decline in reaction times but no such decline with the dotted line. For happy targets, both the solid line and the dotted line show declines in reaction times as the target words get more strongly associated to the happy emotion.

Mixed-effects models were applied separately to the Angry Target Condition and to the Happy Target Condition. For the Angry Target Condition, there was no main effect of prime prosody ($\beta = -0.008$, s.e. = 0.006, $t = -1.5$, $p = 0.1$), and the main effect of association strength was marginal ($\beta = -0.024$, s.e. = 0.013, $t = -1.8$, $p = 0.09$). Importantly, the interaction between prime prosody and association strength was significant ($\beta = -0.016$, s.e. = 0.008, $t = -2$, $p = 0.045$). Examining simple effects showed that

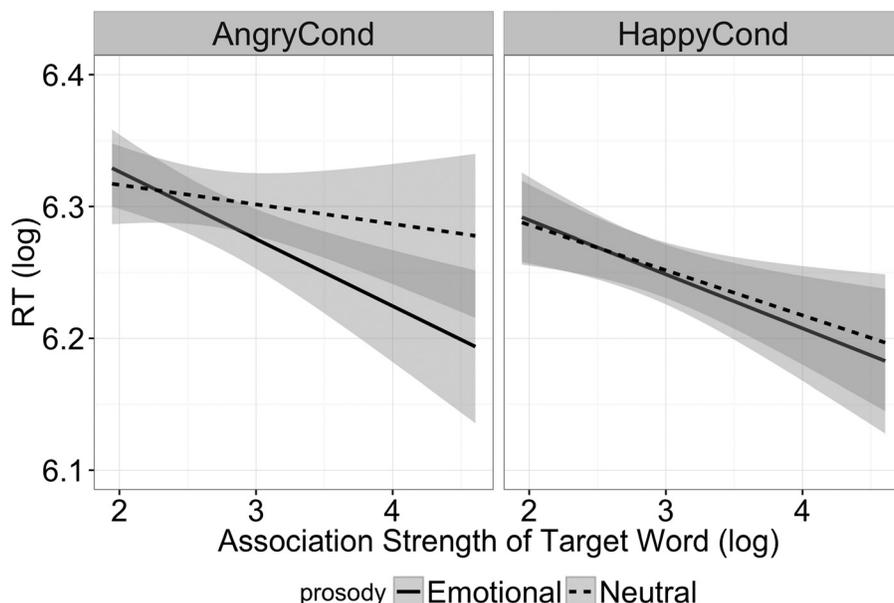


Fig. 1. Log RTs as a function of association strength of target words. The shaded area is 95% CI.

when prime prosody was emotionally related (i.e., angry prosody), association strength had a significant effect on log RTs ($\beta = -0.04$, s.e. = 0.015, $t = -2.6$, $p = 0.016$). That is, as the target word's association strength to the emotion increases, it was recognized faster. However, when prime prosody was unrelated (i.e., neutral prosody), association strength had no effect at all on the speed of target recognition ($\beta = -0.008$, s.e. = 0.015, $t = -0.5$, $p = 0.6$).

For the Happy Target Condition, there was no main effect of prime prosody ($\beta = -0.002$, s.e. = 0.006, $t = -0.04$, $p = 0.7$). The main effect of association strength was significant ($\beta = -0.04$, s.e. = 0.017, $t = -2.2$, $p = 0.045$). As the target word's association strength to the emotion increases, the target was recognized faster. This pattern was the same for both the related emotional (i.e., happy) prosody and for the unrelated (i.e., neutral) prosody; this can be seen in the lack of interaction between prime prosody and association strength ($\beta = 0.0001$, s.e. = 0.007, $t = -0.13$, $p = 0.9$).

4. Discussion and conclusion

In this study, we tested the hypothesis that phonetically cued emotional information, independent of the lexical carrier, facilitates recognition of *words* associated with the corresponding emotion. With angry prosody, we found supporting evidence of affective priming. When preceded by a non-emotional word uttered with angry prosody, RTs to an angry target word became faster as the target was more strongly associated to the emotion anger. In contrast, there was no such effect of association strength on RTs when the angry target words were preceded by words uttered with neutral prosody. Hearing *pineapple* uttered in an angry voice not only activates the word *upset*, but it activates the word sufficiently quickly to result in facilitation compared to the same target word preceded by *pineapple* uttered in a neutral prosody. With happy prosody, evidence for affective priming was also present, although less clear than the angry prosody. Identical to the angry target words, RTs to a happy target word became faster as the target was more strongly associated to the emotion happiness when preceded by a non-emotional word uttered with happy prosody. Unlike neutral prosody primes in the angry condition, neutral prosody primes in the happy condition showed the same effect of association strength as happy prosody primes. We will return to possible reasons why neutral prosody showed the same pattern as happy prosody, but for now we simply state that a significant effect of association strengths on RT with happy prosody is consistent with the interpretation that listening to happy prosody facilitates the recognition of happy associated words as a function of association strength.

The current results help us understand how listeners understand spoken words in two ways. First, phonetically cued emotional information is automatically encoded to an emotion concept, independent of the lexical carrier, during spoken word processing. Second, the extracted emotional information directly activates associated words, just as the lexical meaning of a word spreads its semantic activation to semantically

associated words. Our result supports an approach in spoken word recognition where both social encoding (e.g., recognizing the emotional state of the speaker) and linguistic encoding (e.g., recognizing what word was said) occur simultaneously and are essential components of spoken word recognition (e.g., Sumner *et al.*, 2014).

One might argue that the current result can be explained via episodic linguistic encoding alone, and social encoding is extraneous to the spoken word recognition process. Given an acoustically detailed lexicon, words uttered with angry prosody are expected to activate stored lexical episodes that are acoustically similar to them. This means words that are experienced frequently enough with angry prosody can be activated by an incoming spoken word like *pineapple*_{Angry}. A problem with this account, however, is that most lexical items (perhaps with the exception of a small set of cuss words) are unlikely to be experienced frequently enough with a particular emotional prosody to result in robust activation that can lead to priming in a lexical decision task. In other words, stored lexical episodes of the word *upset* are not acoustically similar to spoken words like *pineapple*_{Angry}, and there is no other associative link between *pineapple* and *upset* that can be used to activate the word *upset*.

Let us now go back to the finding that in the happy target condition, neutral prosody primes were as sensitive to association strength as happy prosody primes. This suggests that neutral prosody as well as happy prosody activated happy associated words as a function of association strength. Why did neutral prosody activate happy words? One possibility is that neutral prosody was actually perceived as happy prosody, not as neutral prosody. This is unlikely, however, given that neutral prosody prime words and happy prosody prime words were clearly different in the perceptual analysis.

Another possibility is that neutral prosody stimuli sounded pleasant enough because the speaker for the current study happened to have a pleasant voice. When producing neutral prosody, the speaker is not actively conveying happiness (which is the case when she produces happy prosody), but her neutral prosody could still sound pleasant. This could be why her angry prosody was rated only around 5 (out of 9) on the angry scale, while her happy prosody was rated around 7 on the happy scale. Her baseline voice was perhaps too pleasant for her angry prosody to score higher than 5. If our speculation is correct that the speaker's baseline vocal quality is behind the pattern of neutral prosody behaving like happy prosody, neutral prosody produced by a speaker with different baseline voice qualities should exhibit different patterns. A different possibility is that the distance between neutral and happy emotion is simply closer than the distance between neutral and other emotion regardless of a vocal quality. If this is the case, neutral prosody will behave like happy prosody with any voice. Future research will need to address the relationship between emotional prosody and vocal quality.

Yet another possibility is that the result of the happy target condition has nothing to do with prime prosody. It might be that words are inherently recognized faster as they are more strongly associated with happiness. Research on processing of emotional stimuli often report that positive valence stimuli are processed faster than negative valence stimuli (e.g., Algom *et al.*, 2004; Estes and Adelman, 2008; but see Baumeister *et al.*, 2001, for the opposite pattern). This inherent processing advantage for positive items might be responsible for the results of happy condition. Perhaps words with a stronger association strength toward happiness have an inherent processing advantage over words with a weaker association strength toward happiness. This is certainly a possible explanation, but we are not aware of any study showing processing differences among positive valence items.

In conclusion, the current study showed that, when preceded by semantically non-emotional words uttered with a particular emotional prosody, listeners recognized target words faster as association to the corresponding emotion increases. This was observed with both angry prosody and happy prosody. However, as neutral prosody exhibited the same pattern as happy prosody, how different types of emotional prosody and different voices engage in affective priming remains as an open question for future studies.

The affective priming effect illuminated here with angry prosody parallels the standard semantic priming effect. The only difference is that semantic priming stems from the lexical semantic meaning while affective priming stems from the non-lexical emotional meaning conveyed by prosody. Spoken words carry multiple layers of meaning beyond lexical meaning, and this additional meaning is critical to the spoken word recognition process and speech processing in general.

Acknowledgments

This research has been supported by a Mellon Foundation Dissertation Fellowship, the Diversity Dissertation Research Opportunity Grant and the Graduate Research Opportunity Award from Stanford University granted to the first author. Part of this work has been also supported by Excellence Initiative of Aix-Marseille University – A*MIDEX, a French “Investissements d’Avenir” program.

References and links

- ¹Nygaard and Queen (2008) reported congruence effects *both* when prosody was a between-subject factor and when it was a within-subject factor. However, a more careful analysis of their data reveals that a congruence effect is present when prosody was a between-subject factor but *not* when it was a within-subject factor (see Kim, 2015, Chap. 2, for the reanalysis of their data).
- ²The 24 non-emotional words were *academy, adequate, after, bounce, collide, compact, compose, conference, copy, galaxy, garbage, listen, ministry, multiply, pending, pineapple, planet, question, recall, salary, scale, specialist, stem, and transmission*.
- ³We obtained the valence ratings using the same method as the valence norms published by Warriner *et al.* (2013). If we use their ratings (excluding one word that does not have an entry in their list), we obtain essentially the same result with a mean valence of 5.27 (1.05 SD).
- Algom, D., Chajut, E., and Lev, S. (2004). “A rational look at the emotional stroop phenomenon: A generic slowdown, not a stroop effect,” *J. Exp. Psychol.* **133**(3), 323–338.
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). “Random effects structure for confirmatory hypothesis testing: Keep it maximal,” *J. Mem. Lang.* **68**(3), 255–278.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). “Fitting linear mixed-effects models using lme4,” *J. Stat. Softw.* **67**(1), 1–48.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D. (2001). “Bad is stronger than good,” *Rev. Gen. Psychol.* **5**(4), 323–370.
- Ben-David, B., Multani, N., Shakuf, V., Rudzicz, F., and Van Lieshout, P. (2016). “Prosody and semantics are separate but not separable channels in the perception of emotional speech: Test of Rating of Emotions in Speech (T-RES),” *J. Speech Language Hear. Res.* **59**, 72–89.
- Estes, Z., and Adelman, J. S. (2008). “Automatic vigilance for negative words in lexical decision and naming: Comment on Larsen, Mercer, and Balota (2006),” *Emotion* **8**(4), 441–444.
- Kim, S. K. (2015). “Speech, variation, and meaning: The effect of emotional prosody on word recognition,” Doctoral dissertation, Stanford University, Stanford, CA.
- Kotz, S., and Paulmann, S. (2011). “Emotion, language, and the brain,” *Lang. Ling. Compass* **5**(3), 108–125.
- Kuznetsova, A., Brockhoff, P., and Christensen, R. (2016). “lmerTest: Tests in Linear mixed effects models, R package (version 2.0-30),” <https://CRAN.R-project.org/package=lmerTest> (Last viewed June 28, 2017).
- Lays, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). “Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median,” *J. Exp. Soc. Psychol.* **49**(4), 764–766.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2015). “Balancing type I error and power in linear mixed models,” arXiv:1511.01864.
- Nygaard, L. C., and Queen, J. S. (2008). “Communicating emotion: Linking affective prosody and word meaning,” *J. Exp. Psychol. Hum. Percept. Perf.* **34**(4), 1017–1030.
- Quené, H., Semin, G. R., and Foroni, F. (2012). “Audible smiles and frowns affect speech comprehension,” *Speech Commun.* **54**(7), 917–922.
- Schirmer, A., Kotz, S. A., and Friederici, A. D. (2002). “Sex differentiates the role of emotional prosody during word processing,” *Cognit. Brain Res.* **14**(2), 228–233.
- Sumner, M., Kim, S. K., King, E., and McGowan, K. B. (2014). “The socially weighted encoding of spoken words: A dual-route approach to speech perception,” *Front. Psychol.* **4**, 1015.
- Swinney, D. A. (1979). “Lexical access during sentence comprehension: (Re)consideration of context effects,” *J. Verb. Learn. Verb. Behav.* **18**(6), 645–659.
- Warriner, A. B., Kuperman, V., and Brysbaert, M. (2013). “Norms of valence, arousal, and dominance for 13,915 English lemmas,” *Behav. Res. Methods* **45**, 1191–1207.
- Wurm, L. H., Vakoch, D. A., Strasser, M. R., Calin-Jageman, R., and Ross, S. E. (2001). “Speech perception and vocal expression of emotion,” *Cognit. Emotion* **15**(6), 831–852.