How Affective Properties of Voice Influence Memory and Social Perception

Author: Xuan Zhang

HOW AFFECTIVE PROPERTIES OF VOICE INFLUENCE MEMORY AND SOCIAL PERCEPTION

Xuan Zhang

A dissertation

submitted to the Faculty of

the department of psychology

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College
Morrissey College of Arts and Sciences
Graduate School

May 2016

HOW AFFECTIVE PROPERTIES OF VOICE INFLUENCE MEMORY AND SOCIAL PERCEPTION

Xuan Zhang

Advisor: Lisa Feldman Barrett, PhD

Human voice carries precious information about a person. From a brief vocalization to a spoken sentence, listeners rapidly form perceptual judgments of transient affective states such as happiness, as well as perceptual judgments of the more stable social traits such as trustworthiness. In social interactions, sometimes it is not just what we say – but how we say it – that matters. This dissertation sought to better understand how affective properties in voice influence memory and how they subserve social perception. To these ends, I investigated the effect of affective prosody on memory for speech by manipulating both prosody valence and semantic valence, I explored the fundamental dimensions of social perception from voice, and I discussed the relationship of those social dimensions to affective dimensions of voice.

In the first chapter, I examined how prosody valence influences memory for speech that varied in semantic valence. Participants listened to narratives spoken in neutral, positive, and negative prosody and recalled as much as they could of the narrative content. Importantly, the arousal level of the affective prosody was controlled across the different prosody valence conditions. Results showed that prosody valence influenced memory for speech content and the effect depended on the relationship between prosody valence and semantic valence. Specifically, congruence between prosody and semantic valence influenced memory. When people were listening to neutral content, affective prosody (either positive or negative) impaired memory. When

listening to positive or negative content, incongruent prosody led to better recall. The present research shows that it is not just what you say, but also how you say it that will influence what people remember of your message.

In the second chapter, I explored the fundamental dimensions of social perception from voices compared to faces, using a data-driven approach. Participants were encouraged to freely write down anything that came to mind about the voice they heard or the face they saw. Descriptors were classified into categories and the most frequently occurred social trait categories were selected. A separate group of participants rated the voices and faces on the selected social traits. Principal component analyses revealed that female voices were evaluated mostly on three dimensions: attractiveness, trustworthiness, and dominance; whereas male voices were evaluated mostly on two dimensions: social engagement and trustworthiness. For social evaluation of faces, a similar two-dimensional structure of social engagement and trustworthiness was found for both genders. The gender difference in social perception of voice is discussed with respect to gender stereotypes and the role voice pitch played in perceived attractiveness and dominance. This study indicates that both modality (voice vs. face) and gender impact the fundamental dimensions of social perception.

Overall, the findings of this dissertation indicate that the affective quality in our voice not only influence how our speech will be remembered but also relate to how we are being socially perceived by others. It would be wise to pay more attention to our tone of voice if we want to make our speech memorable and leave a good impression.

TABLE OF CONTENTS

CHAPTER 1

How much does your tone of voice matter?

Effects of prosody valence on memory for speech

Xuan Zhang[1,2], Jo-Anne Bachorowski[3], Michael Owren[4],

Spencer Lynn[2*], Lisa Feldman Barrett[2,5*]

[1]Department of Psychology, 300 McGuinn Hall, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA, USA 02467

[2]Affective Science Institute and Department of Psychology, 125 Nightingale Hall, 360 Northeastern University, 360 Huntington Avenue, Boston, MA, USA 02115

[3]Department of Psychology, Vanderbilt University, 2301 Vanderbilt Place, Nashville, TN, USA 37240

[4]Department of Psychology, Emory University, Clifton Road NE Suite 234, Atlanta, GA, USA 30322

[5]Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA 02129

*shared senior authorship

Correspondence to:

Xuan Zhang

E-mail: xuan.zhang@bc.edu

Abstract

We examined the effects of prosody valence on recall memory for speech that varied in semantic valence (neutral content in Sample 1, positive content in Sample 2, and negative content in Sample 3). In each study, participants listened to narratives spoken in neutral, positive, and negative prosody and recalled as much as they could of the narrative content, both immediately and after a 10-minute delay. As predicted, prosody valence influenced speech memory and the effect depended on the relationship between prosody valence and semantic valence. Specifically, congruence between prosodic and semantic valence influenced memory. When people were listening to neutral content, affective prosody (either positive or negative) impaired memory (Sample 1). When listening to positive or negative content, however, incongruent prosody led to better recall (Samples 2 and 3). The present research shows that it is not just what you say but how you say it that will influence what people remember of your message.

Key words: affective prosody, memory, acoustic parameters, mediation

Introduction

Have you wondered if you said the same thing in a different tone of voice, would it be remembered differently? A small child may remember a story more vividly if it were told in a melodic tone of voice rather than a monotone. Yet a college student may find a course harder to follow if the professor's tone of voice were so dramatic that it distracted them away from the course content. Considered to be 'the music of speech', *prosody* refers to the melodic and rhythmic aspects of speech (Wennerstrom, 2001). Independent of the semantic content of speech (i.e., what is said), two forms of prosody are typically distinguished. *Linguistic prosody* provides cues regarding syntax and pragmatics (Beach, 1991). *Affective prosody* provides cues regarding the perceived affective state of the speaker (Fairbanks & Provonost, 1938). Affective prosody attains its quality through different patterns of acoustic parameters, such as fundamental frequency (F_0, perceived as pitch), intensity (perceived as loudness), duration (perceived as rhythm), and spectral characteristics (indicating voice quality; see reviews by Bachorowski & Owren, 2010; Banse & Scherer, 1996; Laukka, Juslin, & Bresin, 2005; Scherer, 2003). For example, when speech is spoken in a louder, faster, with a more variable pitch and a smooth voice quality, it is often experienced as pleasant; whereas speech that is spoken in a softer, slower, and with a coarse voice quality is often experienced as unpleasant (Busso & Rahman, 2012; Goudbeek & Scherer, 2010). While many prior studies of affective prosody have focused on the perception of speakers' affective state and correlated acoustical patterns, the functional consequences of affective prosody have not been much investigated. So far, only a handful of studies explored the

effect of affective prosody on memory with mixed findings (Chappuis & Grandjean, 2014; Kitayama, 1996; Schirmer, Chen, Ching, Tan, & Hong, 2013; Schirmer, 2010).

On the *word* level, studies exploring the effects of affective prosody on memory found inconsistent results (Chappuis & Grandjean, 2014; Schirmer et al., 2013; Schirmer, 2010). For instance, with old/new *recognition* memory test, one study demonstrated that words spoken in neutral prosody were better recognized than words spoken in portrayed sad prosody (Schirmer et al., 2013), while another study found no difference in the recognition memory of words spoken in neutral prosody and portrayed happy or sad prosody (Schirmer, 2010). Another study explored the effects of affective prosody on *recall* memory for *words* using a single-word presentation paradigm (Chappuis & Grandjean, 2014). Although the overall results showed a better recall of the affectively spoken words than neutrally spoken words, when individually examining the effect of each prosody condition, the pattern was not clear: compared to neutral prosody, the posed happy and angry prosody resulted in better recall, but posed fearful prosody did not.

On the *sentence* level, the effect of affective prosody on *incidental* memory was found to be dependent upon the cognitive load (the total amount of mental effort being use working memory) during encoding (Kitayama, 1996). Using a surprise free recall test, the study found that affective prosody improved incidental verbal memory when the load of the memory span task was heavy (leaving little available attention to allocate to the spoken sentence), in which case the author inferred that affective prosody captured more attention and resulted in better memory for the sentence in comparison to the negligible effects found for neutral prosody. But affective prosody impaired incidental verbal memory when the memory load was light, in which case, despite more available

attention, attention was divided between prosody and content under affective prosody condition but devoted entirely to the content under neutral prosody condition. However, it is unknown how affective prosody influences *intentional* memory for sentence-length speech.

Several caveats limit interpretation of the above studies. First, none of these studies separated the influence of valence and arousal. As two important properties of affective experience in the perceivers, *valence* varies from positive/pleasant to negative/unpleasant; *arousal* refers to a sense of energy or agency (Russell, 2003; Wundt, 1897). Despite prior focus on the pronounced enhancements for events that elicit arousal (e.g., Kensinger & Corkin, 2003; Talmi & Moscovitch, 2004), accumulating evidence has indicated that even when arousal is controlled, affective valence (whether it is positive or negative) can impact the details remembered (see reviews by Kensinger & Schacter, 2008; Kensinger, 2009a, 2009b). Negative valence leads to more focused attention on local details and enhances memory accuracy (see review by Kensinger, 2007) whereas positive valence leads to a broadening of attention and to a focus on heuristics (e.g., Fredrickson & Branigan, 2005; Gasper & Clore, 2002; Rowe, Hirsh, & Anderson, 2007). Therefore it is important to separately investigate the effect of valence and arousal on memory, no matter what type of stimuli and what modality it is in.

Second, all prior studies of affective prosody's influence on memory only used neutral-semantic material spoken in different prosody. None has taken into account the *congruency* between prosody valence and semantic valence of the content. However, research in language processing has demonstrated faster and more accurate response to congruent stimuli (e.g., positive words spoken in positive prosody) compare to

incongruent stimuli (e.g., positive words spoken in negative prosody), known as the congruency effect (Nygaard & Queen, 2008; Schirmer & Kotz, 2003; Schirmer, Zysset, Kotz, & Von Cramon, 2004). Faster response indicated faster encoding that might provide more time for consolidation that benefit memory. More accurate response may also lead to more accurate encoding that facilitates memory; therefore, it is highly possible that the effect of affective prosody on memory also depends on its congruency with the valence of the speech content.

Moreover, coherent sentences have better ecological validity than independent words, but no study has investigated the *intentional* memory of *sentence-length* speech. Sentence level memory study can be more difficult to conduct because of the need to control for various factors such as the average frequency and familiarity of words in the entire sentence. Although an earlier study tested the surprise recall of sentences spoken in different prosody under heavy and light cognitive load, the effects of affective prosody could be different when full attention is given towards the encoding process.

Furthermore, previous studies have not examined which acoustic parameters mediated the effect of affective prosody on memory, which is important for areas such as speech synthesis and practical applications. An investigation of the specific impact of F_0 level on memory for speech found that both high- and low-F_0, voices led to better long-term memory than medium-F_0 voices (Helfrich & Weidenbecher, 2011). Although not directly testing the effect of *affective* prosody on memory, this study sheds light on the possible mediating effects of acoustic features on the effect of prosody on memory. Therefore, a detailed analysis of acoustic parameters that might mediate the effects of prosody valence on memory would be informative.

The Present Study

We built upon previous findings to better understand affective prosody's effects on memory by conducting a set of experiments with a specific focus on the effect of prosody *valence* (controlling for the arousal level) on *intentional recall* using *sentence-long speech* in *different semantic valence*. We have three research questions in the present study.

The first question is that when the arousal level is controlled, whether we will still find an effect of prosody valence on the intentional recall for spoken sentences. We predicted that the effect of prosody valence on memory still existed after we controlled the arousal level, but the effect can be either beneficial or detrimental, depending on how relevant the prosody condition was to the memory task. Both enhancing and impairing effects of affect on cognitive processes have been found, from lower level such as perceptual processes, to higher level such as mnemonic and executive processes (see review by Dolcos & Denkova, 2014). On one hand, affective stimuli can benefit from enhanced perceptual processing due to their ability to "capture attention" (Chun & Turk-Browne, 2007), and hence through prioritized processing they can be better encoded and remembered (LaBar & Cabeza, 2006; Vuilleumier, Armony, Driver, & Dolan, 2001). On the other hand, when an affective stimulus is task-irrelevant, it may lead to increased distraction and impaired cognitive process including perceptual (Pessoa, McKenna, Gutierrez, & Ungerleider, 2002) and working memory (Anticevic, Barch, & Repovs, 2010; Iordan, Dolcos, & Dolcos, 2013). Therefore, we predicted that when the speech content was neutral, either positive or negative prosody would be irrelevant to the neutral

facts and become a distraction from remembering the facts and impair memory performance.

The second question concerns whether the effect of prosody valence on memory varies by different semantic valence of the speech. When affective prosody is the same as the semantic content, the spoken sentences are congruent stimuli. As mentioned above, congruent stimuli were responded to faster and more accurately as compared to incongruent stimuli (Schirmer & Kotz, 2003). The retrieval of word information from semantic memory was also facilitated for congruous relative to incongruous prosodic and verbal affective states (Schirmer et al., 2004). From this perspective, the prediction would be that the congruence between prosodic and semantic valence enhance memory (prosody being relevant to the memory task), whereas incongruent prosodic and semantic valence impair memory (prosody being irrelevant to the memory task). However, when it comes to more complex situations, incongruence could also lead to better memory. For example, participants remembered better the behaviors that were incongruent with a given personality description of a target than for those that were congruent with the personality (Srull, Lichtenstein, & Rothbart, 1985; Srull & Wyer, 1989). In these cases, individuals were especially motivated to resolve any inconsistencies in the situation, such as why a kind person would push over an old lady; then, memory became better for incongruent than for congruent materials. Therefore it is also possible that when the incongruence between prosody and semantic valence is unexpected instead of just being irrelevant and distracting, it could elicit motivation and draw more cognitive resources to solve the conflict, which lead to more elaboration and enhanced memory performance.

We thus predicted that the effect of prosody valence on memory would vary among different semantic valence conditions.

The third question is about whether certain acoustic parameters mediate the effect of prosody valence on memory. Although prior studies showed that arousal often masked or obfuscated the effects of valence (Banse & Scherer, 1996) and the associations between valence and acoustic parameters have been less clear-cut than arousal (Bachorowski, 1999; Laukka et al., 2005), after controlling for the arousal level, valence related acoustic parameters have been found (e.g., Busso & Rahman, 2012; Goudbeek & Scherer, 2010). We selected 6 parameters that have been found to differ among positive, neutral, and negative valence prosody: speech rate, standard deviation of F_0, mean of intensity, standard deviation of intensity, the proportion of energy below 500 Hz, and spectral slope[1] (Aguert, Laval, Le Bigot, & Bernicot, 2010; Goudbeek & Scherer, 2010; Laukka et al., 2005; Laukka, Neiberg, Forsell, Karlsson, & Elenius, 2011; Rodway & Schepman, 2007; see Table S5 in supplementary material for description of acoustic measurements). We predicted that one or several of these acoustic parameters might mediate the effects of prosody valence on memory.

We conducted four experiments: a Stimulus Preparation Study and three studies of the effect of prosody valence on memory for positive content (Sample 1), negative content (Sample 2), and neutral content (Sample 3) in speech. The goal was to investigate the effects of prosody valence (i.e., positive, neutral, and negative), controlling for the arousal level, on memory for speech of congruent vs. incongruent semantic valence (i.e., positive-content speech, neutral-content speech, and negative-content speech). Moreover, we sought to explore if acoustic features that differentiate

prosody valence also mediate memory for speech content. To these ends, we made audio recordings of short narratives adapted from standardized tests of declarative memory (Randt, Brown, & Osborne, 1981). The content of the stories, delivered in positive, neutral, and negative prosody, served as memory-test stimuli. To assess participants' capability of retrieving information stored in memory, we used recall tests, both immediately and after a 10-minute delay.

Stimulus Preparation Study

We created the stimulus set by recording participants reading out aloud sentences with different semantic valence in content (positive, neutral, negative) in different affective prosody (positive, neutral, negative). The recordings were rated by a separate group of participants for their perceived valence and arousal levels. We controlled arousal level by (1) instructing speakers to maintain a medium level of arousal during production of spoken sentences in different affective states, and (2) selecting from our database the spoken sentences in different valence but with comparable arousal ratings. Acoustic parameters that have been previously found to correlate with the valence dimension were extracted from each recording and examined for their prosody-valence discrimination ability.

Participants. One hundred and two students (42 male; M_{age} = 18.91 years old, SD_{age} = 1.21, range = [18, 24]) were recruited to produce the narrative stimuli. Twenty students (7 male; M_{age} = 19.68 years old, SD_{age} = 1.32, range = [18, 24]) rated the arousal and valence level of the written narratives. Thirty-seven students (12 male; M_{age} = 19.35 years old, SD_{age} = 1.51, range = [18, 25]) were recruited to rate the valence and arousal level of the recorded narratives. Participants were all native English speakers with

normal hearing and received one departmental research credit or $5 for each half hour of participation.

Materials. Written memory narratives used in the stimulus recording study were adapted from a standardized test of declarative memory called "NYU stories". There were six narratives for each semantic valence (Appendix A). Six negative-content stories were from the original NYU stories. Replacing the negative key words with neutral words matched in word length, word frequency, and familiarity, resulted six neutral-content stories (Kensinger, Anderson, Growdon, & Corkin, 2004). We designed the six positive-content stories by replacing the negative/neutral key words with positive words matched in word length, word frequency, and familiarity (Kuchera & Francis, 1967). The written narratives were rated for their valence and arousal level on a scale from 1 to 7 (see supplemental material Table S1).

Stimulus recording. Narratives were recorded inside a quiet testing room (sound level below 25 dB) with a SHWH30XLR WH30 Head-worn Condenser Vocal Performance Microphone and encoded in mono (one-channel recording) directly onto a computer's hard disk at 44.1 kHz sampling rate and 16-bit quantization. The microphone was placed ½ inch from the right corner of the participant's mouth throughout the recording process. Each speaker recorded all 18 narratives.

In an attempt to compare the effect of two of the most commonly used methods of producing affective vocal recordings (induced affective expression and simulated/portrayed affective expression), we employed both methods (referred to as induction and portrayal) to obtain vocal recordings. In the induction method, we used a 3-minute affect induction video (Zhang, Yu, & Barrett, 2014) to induce positive, neutral,

or negative mood in speakers before they read the written narratives aloud. In the portrayal method, speakers were given specific instructions to speak as if they were in a certain affective state and then read the written narratives aloud.

Stimulus ratings. Because all speakers were undergraduate students without formal training in vocal recording, and despite instructions and practice, there were still some errors in their recordings. We listened to all the recordings and excluded those that had errors such as unclear sound, long pauses, wrong pronunciation, etc. Fifteen speakers' recordings were excluded from further analyses, leaving clear recordings from 87 speakers. We then randomly selected a subset of recordings (from 60 speakers) among the clear recordings and asked another group of participants (n=37) to rate perceived valence and arousal of the recordings on a scale of 1 to 7 (1 = extremely low valence/arousal, 7 = extremely high valence/arousal). No significant difference was found in our stimuli between the induction and portrayal methods, although previous research demonstrated that speech segments extracted from acted and authentic expressions differed in their voice quality, and the play-acted speech tokens revealed a more variable F0-contour (Jürgens et al., 2011). This could be because our portrayals were produced by college students instead of trained actors. Therefore the variations in F0-contour and voice quality in college student may not be as exaggerated as those produced by professional actors.

Stimulus selection. Given our purpose of examining the effect of prosody valence on memory performance, we aimed to select recordings that were rated differently for valence but comparable in arousal. For Sample 1, recordings were selected from both induced and portrayed method. Comparison of the two method

revealed no significant difference and thus for Studies 2 and 3, recordings were selected only from the portrayed method. We could have used just one or two "ideal" speakers whose utterances were rated clearly different in valence but comparable in arousal. But using such a restricted sample might induce bias from the particular voice quality or speaking style of a single speaker. To exclude such possible bias, we used multiple speakers' recordings (6 male and 6 female; each contributing 3 distinctive narratives in positive, neutral, and negative valence). Therefore, a final group of 36 recordings was selected for each study respectively. In the final selection of 36 recordings for each semantic valence, the main effects of valence were significant. The main effects of arousal were non significant, indicating that the arousal levels were comparable among different prosody valence conditions (see supplemental Tables S2, S3, and S4).

Acoustic analyses. Acoustic analyses of the selected recordings were conducted with Praat 5.2 (Boersma & Weenink, 2012), automated using GSU Praat Tools 1.9 scripts (Owren, 2008). Each recording comprised a single file, and was first rescaled to the full 16-bit amplitude range available. We checked every F_0 contour in order to manually correct for outliers. Then we proceeded to feature extraction with scripts. To examine if the acoustic parameters could successfully discriminate prosody valence, we first computed their standardized z-scores so that their scales and variances were comparable, and then used discriminant function analysis (the success of which was tested with subsequent cross-classification). Among the six acoustic parameters that we selected, two parameters were highly collinear (mean intensity and the proportion of energy below 500 Hz). To prevent further multicollinearilty, we used five parameters for the following analyses by dropping the proportion of energy below 500 Hz (see supplementary material

Table S7-9 for descriptive statistics of the acoustic stimuli). In general, the correct classifications for prosody valence across three semantic contents were better than chance (33%), ranging from 50% to 61.1% (see supplementary material Text S2 for details) [1].

Memory Studies

The set of memory studies examined the effect of prosody valence on memory for neutral-content narratives (Sample 1), for positive-content narratives (Sample 2), and for negative-content narratives (Sample 3). In each study, participants listened to three narratives spoken in neutral, positive, and negative prosody, and were tested for both free recall (immediately and 10-minute delay) and multiple-choice recognition. Due to the possible influence from recall tasks, recognition memory related methods and results are reported in supplementary materials (Text S4). Therefore, below we report the results of four memory measurements: immediate verbatim recall, immediate gist recall, 10-min delayed verbatim recall, and 10-min delayed gist recall. Accumulating evidence has indicated that even when arousal is controlled, affective valence (whether it is positive or negative) can impact the details remembered (see reviews by Kensinger & Schacter, 2008; Kensinger, 2009a, 2009b). Therefore we predicted that affective prosody would have an impact on recall performance. Further, negative valence leads to more focused attention on local details and enhances memory accuracy (see review by Kensinger, 2007), whereas positive valence leads to a broadening of attention and to a focus on heuristics (e.g., Fredrickson & Branigan, 2005; Gasper & Clore, 2002; Rowe, Hirsh, & Anderson, 2007). Given that the spoken narratives used in the present study were descriptions of an event with different valence, we predicted that the influence of prosody valence on memory would vary by and semantic valence of the narrative. Also, it is

possible that participants would have broader attention for the positive narratives (Sample 2) and a more focused attention to details for the negative narratives (Sample 3). Therefore gist recall might be a better indicator for positive narratives, whereas verbatim recall might be a better indicator for negative narratives. Delayed recall, no matter verbatim or gist, is predicted to be worse than immediate recall, but still preserve the effect of prosody valence, as prior research indicated the a sustained effect of affect on memory in delayed interval from 10 minutes to 24 hours (e.g., Kensinger et al., 2004). For each study, we also tested the mediation effects of the acoustic parameters.

Method

Participants. All participants were native English speakers with normal hearing. Participants received one departmental research credit or $10 for participating.

Sample 1: Neutral sentences. Participants were 50 undergraduates (25 female; M_{age} = 19.92 years old, SD_{age} = 1.38, *Range* = [18, 25]). One participant did not have any valid memory data for the first narrative because he accidently pulled off the headphones, and the missing data points were replaced by mean scores of their group.

Sample 2: Positive sentences. Participants were 65 undergraduates (40 female; M_{age} = 18.89, SD_{age} = 1.19, *Range* = [18, 23]). One participant's data were excluded from further analyses due to a seeming lack of concentration (as reported by a research assistant, before any data analysis had taken place).

Sample 3: Negative sentences. Participants were 70 undergraduates (38 female; M_{age} = 19.25, SD_{age} = 1.23, *Range* = [18, 24]). No participants were excluded from further analysis, but 5 missing data points were replaced by mean scores of their group. One participant did not remember much of one narrative (among three narratives)

because that narrative included personally relevant information and he wasn't able to attend to other portions of the narrative. For another participant, a computer problem led to no data recorded for one narrative (among three narratives).

Experiment design. Prosody valence was tested in a within-subject design with three levels (i.e., positive, neutral and negative) in each of three studies. The only difference across the three studies was the valence of the narrative content: Sample 1 used neutral-content narratives; Sample 2 used positive-content narratives; and Sample 3 used negative-content narratives. In each study, each participant listened to 3 narratives, with each narrative originally produced in a different Prosody Valence. The order of the narratives was counterbalanced for each listener. Dependent variables were 5 memory measures, described below in detail.

Procedure. Participants were tested using three narratives in each experiment (E-Prime 2.0, Psychology Software Tools, Pittsburgh, PA). For each narrative, participants first listened to the recording and were immediately asked to recall as much of the verbal description as possible. A research assistant checked off a score sheet for answers verbatim, and also wrote down words that were similar to but not an exact match to the original words in the narrative. In order to prevent rehearsal during the 10-min delay period, participants completed a 10-min filler task of three Sudoku puzzles. The set of Sudoku puzzles had three levels of difficulty: easy, challenging, and difficult. So that participants fully focused on the puzzles, we told them that they would receive a prize if they completed all 3 puzzles. Due to the time limit and difficulty of the puzzles, none of the participants finished all 3 puzzles. After 10 min, participants were asked again to recall as much of the narrative as possible. Next, they completed a computer-

administered 14-item recognition test in which they had to choose the correct answer among three distracters by pressing a number key corresponding to the answer. After these three blocks of narrative memory tests, participants were given a demographic form to complete. They were then fully debriefed about the study.

Memory measures. In all three studies, we measured free recall memory performance at two time points, immediate and 10-min delayed, as well as recognition memory performance after all recall tasks. Due to possible influence from recall tasks, the recognition memory method and results were reported in supplementary materials (see Text S4). The recall tests were adapted from the NYU memory test (Randt, Brown, & Osbourne, 1981), and included two method of scoring: verbatim recall and gist recall. For verbatim recall, one point was given for every principal word that was correctly recalled from the narrative (maximum 20 points). For gist recall, each part of the narrative was separated into 10 "idea units". These "idea units" were words or phrases that corresponded to a person, place, or day in the narrative. One point was given for each of the 10 idea units (maximum 10 points). Correct proportion percentage was computed to make all measurements comparable.

Data analysis. For all memory measures, missing data points were replaced with mean scores of their group (4 for Sample 1, 2 for Sample 2, and 10 for Sample 3). For both verbatim recall and gist recall scores, we conducted repeated measures ANOVAs with *Prosody Valence* (positive, neutral, and negative) and *Time* (immediate, delayed) as within-subject factors. We also examined the effect of gender by submitting talker gender and listener gender separately in another set of ANOVAs. All reported *p*-values are two-tailed. We also analyzed all memory data from three studies by entering

semantic valence and prosody valence as independent variables and memory scores as dependent variables. Results revealed a significant effect of semantic valence (all *ps* < 0.01) on all memory measurements and a significant interaction between semantic valence and prosody valence (all *p*s < 0.002).

Mediation analyses were conducted using the method provided by Preacher and Hayes (2008). This method utilizes bootstrapping to generate a reference distribution, which is then used for confidence-interval estimation and significance testing. Bootstrapping overcomes the normality assumptions necessary in other tests of mediation (e.g., Sobel, 1982). This method also improves on the commonly used Baron and Kenny's approach, which has been found to have low statistical power (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). In the present study, mediation models were tested using the SPSS macro PROCESS (Hayes & Preacher, 2014). In our mediation model, the independent variable was prosody valence and the dependent variables were mean aggregated memory scores for each narrative (see Figure 1). We used two sets of coding to explore the mediation effects of the valence conditions. First, to test if there were any mediation effects in the positive or negative condition relative to the neutral condition, we used dummy coding. Second, to explore the contrast between the neutral and valenced conditions, as well as the contrast between positive and negative conditions, we used contrast coding. As shown in Figure 1, D1 and D2 were defined for each condition as D1 = – 0.667, D2 = 0 for neutral prosody condition, D1 = 0.333, D2 = – 0.5 for positive prosody condition, D1 = 0.333, D2 = 0.5 for negative prosody condition. The mediators were chosen based upon the previous discriminant analyses: speech rate for Sample 1, intensity mean and speech rate for Sample 2, speech rate and

spectral slope for Sample 3. Mediation analyses were conducted separately for each study for each memory measurement.

Results

Memory performance summary. As predicted, after controlling for the arousal level, prosody valence had a significant effect on the recall memory for the speech content and the effect differed by the relationship between semantic valence and prosody valence. Specifically, participants in Sample 1, who heard neutral sentences, recalled more details when sentences were spoken in neutral prosody than when spoken with affective prosody (either positive or negative); thus, memory for neutral content was impaired by affective prosody. In contrast, participants in Samples 2 and 3, who heard positive and negative sentences, respectively, recalled more details when sentences were spoken with an incongruent prosody; that is, memory for positive content was improved by negative prosody (Sample 2), whereas negative content was improved by positive prosody (Sample 3).

Memory performance of neutral-content narratives (Sample 1). Overall, neutral prosody was associated with better memory of neutral-content narratives compared to either positive- or negative- prosody narratives. For verbatim recall, repeated measures ANOVA revealed a main effect of *Prosody Valence* (F (2, 98) = 7.01, p = .001, η^2 = 0.13) and a main effect of *Time* (F (1, 49) = 52.84, p < .0001, η^2 = 0.52) with no interactions (see Table 1 and Figure 2). As predicted, pair-wise comparisons indicated that verbatim recall for neutral narratives read in a neutral prosody was best among the three narratives (better than those read in a positive valence, p <. 008, and better than those read in a negative valence, p < .008). No significant difference between

the positive and negative prosody conditions was present ($p = 1.0$). Delayed verbatim recall was significantly worse than immediate verbatim recall ($p < .001$). Tests of between-subject effects of gender indicated no significant effect of either talker gender or listener gender.

For gist recall of neutral semantic narratives, repeated measures ANOVA also revealed a main effect of *Prosody Valence* (F (2, 98) = 3.68, $p = .03$, $\eta^2 = 0.07$) and a main effect of *Time* (F (1,49) = 29.81, $p < .001$, $\eta^2 = 0.378$) and no interactions (Table 1). Pair-wise comparisons showed that when measuring recall performance in the gist format, neutral-content narratives read in a neutral prosody were marginally better remembered than those read in a negative prosody ($p = 0.063$), but not significantly different from those read in a positive prosody ($p = 0.163$). No difference between the positive and negative prosody conditions was present ($p = 1.0$). Delayed gist recall was significantly worse than immediate gist recall ($p < .001$). No significant effects of listener gender or talker gender in the gist recall performance for neutral narratives.

Memory performance of positive-content narratives (Sample 2). The most consistent finding among all memory measures was that for positive-content narratives, negative prosody was associated with better memory compared to neutral prosody. For verbatim recall, repeated measures ANOVA revealed a main effect of *Prosody Valence* (F (2, 126) = 7.64, $p < .001$, $\eta^2 = 0.108$) and a main effect of *Time* (F (1,63) = 56.13, $p < .0001$, $\eta^2 = 0.471$) with no interactions (see Table 1). Pair-wise comparison indicated verbatim recall for positive-content narratives read in neutral prosody was worst among the narratives (worse than those read in a positive prosody, $p < .008$; and worse than those read in a negative prosody, $p < .004$). No significant difference between the

positive and negative prosody conditions was present ($p = 1.0$). Delayed verbatim recall was significantly worse than immediate verbatim recall ($p < .0001$). Tests of between-subject effects of gender indicated no significant effect of either talker gender or listener gender.

For gist recall of positive-content narratives, repeated measures ANOVA revealed a main effect of *Prosody Valence* (F (2, 126) = 11.09, $p < .001$, $\eta^2 = 0.15$) and a main effect of *Time* (F (1, 63) = 24.73, $p < .0001$, $\eta^2 = 0.282$) and no interactions (see Table 1). Pair-wise comparisons indicated that gist recall for positive-content narratives read in a negative prosody was remembered best among the three narratives (better than positive prosody, $p < .001$, and better than neutral prosody, $p < .001$; see Figure 3). However, no significant difference between the positive and neutral prosody conditions was present ($p = 1.0$). Delayed verbatim recall was significantly worse than immediate verbatim recall ($p < .0001$). Listener gender had a marginal significant effect when entered as a between-subject factor in ANOVA (F (1,62) = 3.57, $p = .064$, $\eta^2 = 0.054$), whereby female listeners performed slightly better than male listeners. No significant effect of talker gender.

Memory performance of negative-content narratives (Sample 3). Similar to Sample 2, recall for negative-content narratives was best when read in the opposite prosody (positive prosody), better than both neutral and negative prosody. For verbatim recall, repeated measures ANOVA revealed a main effect of *Prosody Valence* (F (2, 138) = 19.59, $p < .001$, $\eta^2 = 0.221$) and a main effect of *Time* (F (1,69) = 101.51, $p < .001$, $\eta^2 = 0.406$) with no interactions (see Table 1). Pair-wise comparison indicated that verbatim recall for negative-content narratives was best when read in a positive prosody

(better than those read in a neutral prosody, $p < .001$, and better than those read in a negative prosody, $p < .001$; see Figure 4). No significant difference between the negative and neutral prosody conditions was present ($p = .097$). Delayed verbatim recall was significantly worse than immediate verbatim recall ($p < .001$).

For gist recall of negative semantic narratives, repeated measures ANOVA revealed a main effect of *Prosody Valence* (F (2, 138) = 9.84, $p < .001$, $\eta^2 = 0.125$), a main effect of *Time* (F (1,69) = 25.87, $p < .001$, $\eta^2 = 0.273$), and no interaction effects (see Table 1). Pair-wise comparisons indicated that gist recall for negative-content narratives was best when read in a positive prosody (better than read in a negative prosody, $p < .003$, and better than read in a neutral prosody, $p < .001$). Furthermore, no significant difference between the negative and neutral prosody conditions was present ($p = 1.0$). Delayed verbatim recall was significantly worse than immediate verbatim recall ($p < .001$). No significant effect of listener gender or talker gender in gist recall for negative-content narratives.

Mediation of acoustic parameters on memory performance. Towards the goal of testing acoustical mediation of the effect of prosody valence on memory performance, we first used Discriminant Function Analysis to identify the acoustic parameters that mattered for prosody valence. Specifically, in the prediction of prosody valence for neutral-content narratives, speech rate loaded on the only significant discriminant function. For positive-content narratives, two discriminant functions were identified: mean intensity loaded on the first function, and speech rate loaded on the second function. For negative-content narratives, the 5 a priori selected acoustic parameters produced two discriminant functions: speech rate loaded on the first function and spectral

slope loaded on the second function (see supplementary material S2 text section for details).

Next, these acoustic parameters were entered as mediators using bootstrapping method for testing for mediation effects separately for each study. No mediation effect was found for Sample 1 or Sample 2. But in Sample 3 (negative-content narratives), 3 of the 4 memory measurements (except delayed gist recall) were mediated by speech rate for (1) the effect of valenced prosody relative to neutral prosody on memory, and (2) the effect of positive prosody relative to neutral prosody, as well as relative to negative prosody on memory. Speech rate did not mediate the effect of negative prosody relative to neutral prosody on memory (see Table 2 for direct, indirect, and total effects of the mediation analysis). As shown in Table 2, all of the significant indirect effects of prosody valence through speech rate on memory were negative. That is, when the variance accounted by speech rate was controlled, the direct effect of prosody valence on memory was larger than the total effect of prosody valence on memory. In summary, as to hypothesis 3 about the mediation of acoustic parameters, we only found speech rate suppressed the effect of prosody valence on memory performance when the narrative content was negative.

Discussion

The present findings showed that the valence of affective prosody influenced people's memory for what was said. Such influence varied by the relationship between the valence of verbal content (neutral, positive, or negative) and the valence of affective prosody (neutral, positive, or negative). Limited support for the mediation by acoustic parameters of prosody's effects on memory was found.

We found that affective prosody influenced memory, which is consistent with our hypothesis and in line with the contention that affective prosody is automatically processed. Although we did not ask participants to attend to prosody, it nonetheless affected memory. The automaticity of affective prosody has previously been found in the domain of visual attention (Brosch, Grandjean, Sander, & Scherer, 2008; Rigoulot & Pell, 2012). Affective vocalizations more effectively cued spatial locations than did neutral vocalizations (Brosch et al., 2008). Implicit processing of affective prosody also systematically influenced gaze to facial affective expressions (Rigoulot & Pell, 2012). Moreover, evidence from neuroimaging and event-related potential studies have suggested that affective prosody recruited more processing resources (e.g., Schirmer, Simpson, & Escoffier, 2007; Wiethoff et al., 2008). This automatic processing has been found to be related to memory benefits, both because attended stimuli are often well remembered (reviewed by Chun & Turk-Browne, 2007) and because the amygdala engagement triggered by affect facilitates perceptual (e.g., Vuilleumier et al., 2001; Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004) and mnemonic processes (reviewed by LaBar & Cabeza, 2006).

But affective prosody does not act alone in the influence prosody valence exerted on memory: whether affective prosody facilitated or impaired speech memory depended on the relationship between content valence and prosody valence. For neutral-content narratives, affective (either positive or negative) prosody impaired memory (Sample 1). Both verbatim and gist recall became worse when the neutral-content narrative was spoken in either positive or negative (incongruent) prosody, as compared to neutral (congruent) prosody. This finding is in line with two previous studies that utilized

different paradigms and stimuli to test the effect of affective prosody on neutral stimuli (Kitayama, 1996; Schirmer et al., 2013). Neutrally-spoken sentences were remembered better than positively or negatively spoken sentences in a surprise recall task (Kitayama, 1996). Neutrally-spoken words were also recognized better than sadly-spoken words (Schirmer et al., 2013). A possible explanation for these findings is that the attention that incongruent affective prosody automatically drew to nonlinguistic aspects of the stimulus left fewer resources available for the encoding and consolidation of the linguistic content.

When it comes to incongruence conditions between prosody valence and semantic valence, the effects of affective prosody further depended on whether being task-irrelevant and distractive to the memory for speech or being unexpected (possibly attracting attention and motivating elaboration) to the speech content. Consistent with our predictions, valenced-content narratives were actually remembered better when spoken in the opposite-valence prosody than when spoken in the same-valence prosody. For instance, gist recall for positive-content narrative was best when spoken in a negative prosody (Sample 2), and both verbatim and gist recall for negative-content narratives were best when read in a positive prosody (Sample 3). This is counter-intuitive at first glance when taking account of the congruency effect in language processing where congruence between affective prosody and word meaning facilitated the linguistic processing of words (e.g. Nygaard & Queen, 2008). But, this pattern is understandable when taking the proposed two-dimensional structure of congruency (relevancy and expectancy; Heckler & Childers, 1992) into consideration. That is, the effect of prosody on memory for affective content depended on whether the incongruence between content and prosody valence came from relevancy or expectancy. In Sample 1, positive and

negative prosody were irrelevant to neutral content and impaired memory. In Sample 2 and 3, opposite-valence prosody was unexpected and dramatic to the speech content and enhanced memory. As discussed in the introduction, past research has indicated that materials that are consistent with an expectation are generally better remembered, but when individuals are especially motivated to resolve whatever inconsistencies exist in the stimulus materials, then memory can become better for inconsistent than for consistent materials (e.g., Srull, Lichtenstein, & Rothbart, 1985; Srull & Wyer, 1989). We speculate that in our study prosody set an expectation that was incongruent with the speech content. One possible explanation for our finding is that the opposite-valence prosody is so much unexpected that it drew even more extra attention to the content comparing to the congruent same-valence prosody. The extra attention that the unexpected conflict between prosody and content may automatically allocate more cognitive resources to the encoding of narrative, and thus resulted in better memory performance. Alternatively, the conflict between prosody and content could elicit motivation to resolve the inconsistency, resulting in more elaboration on the content and better memory.

Mediation by acoustic parameters of prosody's effects on memory was limited to negative-content narratives only (Sample 3). Speech rate mediated the total effect by suppressing the direct effect of prosody valence on recall performance. The mediation effect of speech rate is consistent with a previous finding that faster speech rate was associated with poorer recall for spoken word lists (Nygaard, Sommers, & Pisoni, 1995). However, the general absence of mediation by acoustic parameters in the present research was unexpected and could be due to several reasons. First, there is still limited

systematic research on the acoustic correlates to affective dimensions such as valence and arousal. The absence of mediation could be due to the absence of reliable acoustic parameters that represent prosody valence. Second, we used different narratives in order to test the effect of prosody valence as a within-subject variable. Therefore different linguistic prosodic features could also be confounds, interfering with the effect of affective prosodic features on memory. Third, while most previous research used words or short sentences as material, we used longer narrative with over 20 words in average. The valence-related acoustic features could be different between shorter and longer speech. Therefore, although our use of longer narratives made the present research more ecologically valid, more systematic research is needed to explore the specific pattern and relationship among prosody valence, acoustical features, and memory.

Taken together, these findings demonstrated that "the way you say it" influenced how much your message would be remembered. When delivering a neutral event or fact, such as a simple news or academic knowledge, using an affective (either positive or negative) tone of voice may decrease the details remembered for the message. However, when describing a positive or negative event, employing an opposite affective tone of voice may somehow lead to better remembrance. This possibly relates to the reason for the popularity of TV hosts in some daily shows who are already utilizing such skills. For instance, they sometimes reported a terrible mistake politicians made in an exhilarated voice. The contrast between the event content and the tone of voice possibly attracts more attention from the audience and elicits more elaborations on the event, which leads to better memory of the news. Audiences therefore easily favor shows that make the

news not only funnier but also more easily remembered to be discussed in future conversations.

Limitation and future directions

One limitation of the current study was that we only tested a 10-minute delayed memory for the effects of prosody valence. An earlier study showed that the effect of affective prosody on memory for semantic neutral words could last for 24 hours (Chappuis & Grandjean, 2014). Therefore it would be interesting to explore how long the effect of affective prosody lasts for different valenced words and sentences.

Future studies are also needed to look into the brain mechanisms of how affective prosody works and interacts with semantic valence to influence memory. An earlier study explored the underlying neural mechanism of recognition advantage for neutral prosody comparing with sad prosody using ERP (Schirmer et al., 2013). Results showed that sad prosody elicited a greater P200 than did neutral prosody, and the larger the P200 effect was during listening, more negatively were the word rated subsequently. However, the P200 effect was unrelated to recognition advantage for words spoken in neutral prosody as compared to sad prosody. Hence more research is called for finding the brain mechanisms of how affective prosody works, especially after taking the semantic valence of the material content into account, as suggested by the present study.

Another future direction along this line of research would be how the effects of affective prosody on memory vary across culture. Previous cross-cultural studies exploring the spontaneous attention to word content versus affective prosody found different patterns between independent cultures and interdependent cultures. For instance, Americans (independent culture) showed attention bias toward linguistic

content; whereas both Japanese and Filipinos (independent cultures) showed attention bias toward prosody (Ishii, Reyes, & Kitayama, 2003). Therefore it is highly possible that the effect of affective prosody on memory also differs between independent and interdependent cultures.

Conclusions

The present research provides the first systematic exploration of how prosody valence impacted memory for longer speech with an explicitly control for arousal level in affective prosody as well as a full manipulation of semantic valence of speech. Our results showed that prosody valence influenced the amount of details recalled from speech and the effect depended on the relationship between prosody valence and semantic valence. Specifically, congruence between prosodic and semantic valence influenced memory. When people listened to narratives with neutral content, affective prosody (either positive or negative) impaired both immediate recall memory and 10-minute delayed recall and recognition memory (Sample 1). When listening to positive or negative content, however, incongruent prosody led to better recall (Samples 2 and 3). The present research demonstrates the important role of affective prosody in memory: If you want people to remember of your message, pay attention not only to what you say, but also how you say it.

Endnotes

1. Besides the priori approach, we also used an alternative data-driven approach for acoustic parameter selection and discrimination analyses. Results revealed comparable prosody valence classification rates, raging from 50% to 67.7% (see supplementary material for details).

References

Aguert, M., Laval, V., Le Bigot, L., & Bernicot, J. (2010). Understanding expressive speech acts: The role of prosody and situational context in French-speaking 5- to 9-year-olds. *Journal of Speech, Language, and Hearing Research*, *53*(6), 1629–1641. http://doi.org/10.1044/1092-4388(2010/08-0078)

Anticevic, A., Barch, D. M., & Repovs, G. (2010). Resisting emotional interference: brain regions facilitating working memory performance during negative distraction. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 159–173.

Bachorowski, J.-A. (1999). Vocal Expression and Perception of Emotion. *Current Directions in Psychological Science*, *8*(2), 53–57. http://doi.org/10.1111/1467-8721.00013

Bachorowski, J.-A., & Owren, M. J. (2008). Vocal expressions of emotion. In M. Lewis, J. M. HavilandJones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 196–210). New York: The Guilford Press. http://doi.org/10.2307/2076468

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. http://doi.org/10.1037/0022-3514.70.3.614

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity : Evidence for cue trading relations. *Journal of Memory and Language*, *30*, 644–663.

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. *Glot International*, *5*(9/10), 341–345.

Brosch, T., Grandjean, D., Sander, D., & Scherer, K. R. (2008). Behold the voice of wrath: Cross-modal modulation of visual attention by anger prosody. *Cognition*, *106*(3), 1497–1503. http://doi.org/10.1016/j.cognition.2007.05.011

Busso, C., & Rahman, T. (2012). Unveiling the Acoustic Properties that Describe the Valence Dimension. *Interspeech*, (SEPTEMBER 2012). Retrieved from http://20.210-193-52.unknown.qala.com.sg/archive/archive_papers/interspeech_2012/i12_1179.pdf

Chappuis, C., & Grandjean, D. (2014). When voices get emotional : A study of emotion - enhanced memory and impairment during emotional prosody exposure, *2*(September), 1939–1943.

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*(2), 177–184. http://doi.org/10.1016/j.conb.2007.03.005

Dolcos, F., & Denkova, E. (2014). Current emotion research in cognitive neuroscience : Linking enhancing and impairing effects of emotion on cognition. *Emotion Review*. http://doi.org/10.1177/1754073914536449

Fairbanks, G., & Provonost, W. (1938). Vocal pitch during simulated emotion. *Science*, *88*(2286), 382–383.

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion*, *19*(3), 313–332.

Gasper, K., & Clore, G. L. (2002). Attending to the big picture: Mood and global versus local processing of visual information. *Psychological Science*, *13*(1), 34–40.

Goudbeek, M., & Scherer, K. R. (2010). Beyond arousal: valence and potency/control cues in the vocal expression of emotion. *The Journal of the Acoustical Society of America*, *128*(3), 1322–36. http://doi.org/10.1121/1.3466853

Heckler, S. E., & Childers, T. L. (1992). The Role of Expectancy and Relevancy in Memory for Verbal and Visual Information: What is Incongruency? *Journal of Consumer Research*, *18*(4), 475. http://doi.org/10.1086/209275

Helfrich, H., & Weidenbecher, P. (2011). Impact of voice pitch on text memory. *Swiss Journal of Psychology / Schweizerische Zeitschrift Für Psychologie / Revue Suisse de Psychologie*, *70*(2), 85–93. http://doi.org/10.1024/1421-0185/a000042

Iordan, A. D., Dolcos, S., & Dolcos, F. (2013). Neural signatures of the response to emotional distraction : a review of evidence from brain imaging investigations. *Frontiers in Human Neuroscience*, *7*(June), 1–21. http://doi.org/10.3389/fnhum.2013.00200

Ishii, K., Reyes, J. A., & Kitayama, S. (2003). Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychological Science*, *14*(1), 39–46. http://doi.org/http://dx.doi.org/10.1111/1467-9280.01416

Jürgens, R., Hammerschmidt, K., Fischer, J., Jürgenss, R., Hammerschmidt, K., Fischer, J., … Fischer, J. (2011). Authentic and Play-Acted Vocal Emotion Expressions Reveal Acoustic Differences. *Frontiers in Psychology*, *2*(July), 1–11. http://doi.org/10.3389/fpsyg.2011.00180

Kensinger, E. A. (2007). Negative emotion enhances memory accuracy. *Current Directions in Psychological Science*, *16*(4), 213–218.

Kensinger, E. A. (2009a). Remembering the Details: Effects of Emotion. *Emotion Review*, *1*(2), 99–113. http://doi.org/10.1177/1754073908100432

Kensinger, E. A. (2009b). What Factors Need to be Considered to Understand Emotional Memories ? *Emotion Review*, *1*(2), 120–121. http://doi.org/10.1177/1754073908100436

Kensinger, E. A., Anderson, A., Growdon, J. H., & Corkin, S. (2004). Effects of Alzheimer disease on memory for verbal emotional information. *Neuropsychologia*, *42*(6), 791–800. http://doi.org/10.1016/j.neuropsychologia.2003.11.011

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words : Are emotional words more vividly remembered than neutral words ?, *31*(8), 1169–1180.

Kensinger, E. A., & Schacter, D. L. (2008). Memory and emotion. *Handbook of Emotions*, *3*, 601–617.

Kitayama, S. (1996). Remembrance of Emotional Speech: Improvement and Impairment of Incidental Verbal Memory by Emotional Voice. *Journal of Experimental Social Psychology*, *32*(4), 289–308. http://doi.org/10.1006/jesp.1996.0014

Kuchera, H., & Francis, W. N. (1967). Computational analysis of present-day American English.

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews. Neuroscience*, *7*(1), 54–64. http://doi.org/10.1038/nrn1825

Laukka, P., Juslin, P., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition & Emotion*, *19*(5), 633–653. http://doi.org/10.1080/02699930441000445

Laukka, P., Neiberg, D., Forsell, M., Karlsson, I., & Elenius, K. (2011). Expression of affect in spontaneous speech: Acoustic correlates and automatic detection of irritation and resignation. *Computer Speech and Language*, *25*(1), 84–104. http://doi.org/DOI 10.1016/j.csl.2010.03.004

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 1017–1030. http://doi.org/10.1037/0096-1523.34.4.1017

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics*, *57*(7), 989–1001. http://doi.org/10.3758/BF03205458

Owren, M. J. (2008). GSU Praat Tools: scripts for modifying and analyzing sounds using Praat acoustics software. *Behavior Research Methods*, *40*(3), 822–829. http://doi.org/10.3758/BRM.40.3.822

Pessoa, L., Mckenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences*, *99*(17), 11458–11463.

Randt, C. T., Brown, E. R., & Osborne, D. P. (1981). Randt memory test. New York University, Department of Neurology.

Rigoulot, S., & Pell, M. D. (2012). Seeing emotion with your ears: Emotional prosody implicitly guides visual attention to faces. *PLoS ONE*, *7*(1), 19–26. http://doi.org/10.1371/journal.pone.0030740

Rodway, P., & Schepman, A. (2007). Valence specific laterality effects in prosody: Expectancy account and the effects of morphed prosody and stimulus lead. *Brain and Cognition*, *63*(1), 31–41. http://doi.org/10.1016/j.bandc.2006.07.008

Rowe, G., Hirsh, J. B., & Anderson, A. K. (2007). Positive affect increases the breadth of attentional selection. *Proceedings of the National Academy of Sciences*, *104*(1), 383–388.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–72. http://doi.org/10.1037/0033-295X.110.1.145

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1-2), 227–256. http://doi.org/10.1016/S0167-6393(02)00084-5

Schirmer, A. (2010). Mark My Words: Tone of Voice Changes Affective Word Representations in Memory. *PLoS ONE*, *5*(2), e9080. http://doi.org/10.1371/journal.pone.0009080

Schirmer, A., Chen, C.-B., Ching, A., Tan, L., & Hong, R. Y. (2013). Vocal emotions influence verbal memory: neural correlates and interindividual differences. *Cognitive, Affective & Behavioral Neuroscience*, *13*(1), 80–93. http://doi.org/10.3758/s13415-012-0132-8

Schirmer, A., & Kotz, S. a. (2003). ERP evidence for a sex-specific Stroop effect in emotional speech. *Journal of Cognitive Neuroscience*, *15*(8), 1135–1148. http://doi.org/10.1162/089892903322598102

Schirmer, A., Simpson, E., & Escoffier, N. (2007). Listen up! Processing of intensity change differs for vocal and nonvocal sounds. *Brain Research*, *1176*(1), 103–112. http://doi.org/10.1016/j.brainres.2007.08.008

Schirmer, A., Zysset, S., Kotz, S. a., & Von Cramon, D. Y. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage*, *21*(3), 1114–1123. http://doi.org/10.1016/j.neuroimage.2003.10.048

Srull, T. K., Lichtenstein, M., & Rothbart, M. (1985). Associative storage and retrieval processes in person memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(2), 316.

Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review*, *96*(1), 58–83. http://doi.org/10.1037/0033-295X.96.1.58

Talmi, D., & Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Memory & Cognition*, *32*(5), 742–751. http://doi.org/10.3758/BF03195864

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. *Neuron*, *30*(3), 829–841. http://doi.org/10.1016/S0896-6273(01)00328-2

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience*, *7*(11), 1271–1278. http://doi.org/10.1038/nn1341

Wennerstrom, A. (2001). *The music of everyday speech: Prosody and discourse analysis*. New York, NY: Oxford University Press.

Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., & Ethofer, T. (2008). Cerebral processing of emotional prosody—influence of acoustic parameters and arousal. *NeuroImage*, *39*(2), 885–893. http://doi.org/10.1016/j.neuroimage.2007.09.028

Wundt, W. M. (1897). *Outlines of psychology*. Bristol, UK: Thoemmes Press.

Zhang, X., Yu, H. W., & Barrett, L. F. (2014). How does this make you feel? A comparison of four affect induction procedures. *Frontiers in Psychology*, *5*(1975), 1–10. http://doi.org/10.3389/fpsyg.2014.00689

Appendix

Appendix A. Written Narratives

Please take your time to get familiar with each narrative first. You could read the narrative through your head if it helps. Then, please practice reading them out loud smoothly without any stuttering, unusual pause or unclear pronunciation.

Neutral narratives

On Monday, March 4th, in Denver, Colorado, a tourist group visited the Hackett Ski Resort on Billings Road, consisting of sixty visitors who rented eight snowboards plus twenty pairs of skis.

On Tuesday, June 3rd, in St. Paul, Minnesota, a lukewarm rain soaked the Milton Township on Carleton Lake, watering thirteen gardens and dampening nine boys plus fifteen friends.

On Wednesday, July 9th, in Newport, California, a blazing sun warmed the Seamount Ferry off Jackson Sound, tanning seven crewmen and drying ten overcoats plus sixteen sneakers.

On Thursday, May 6th, in Mobile, Alabama, a large corporation reopened the Carson Warehouse on Harvey Harbor, employing nineteen workmen and housing two watchmen plus fourteen architects.

On Friday, April 5th, in Cincinnati, Ohio, a haggard animal entered the Belmont Hotel on Windy Street, approaching fourteen guests and sniffing four plants plus eighteen suitcases.

On Saturday, August 2nd, in Seattle, Washington, a newspaper reporter visited the theater district on Sterling Avenue, observing twenty performers and interviewing five actors plus twelve spectators.

Positive narratives

On Monday, March 4th, in Denver, Colorado, a fluffy snow covered the Hackett Ski Resort on Billings Road, exciting sixty skiers and pleasing eight teachers plus twenty pupils.

On Tuesday, June 3rd, in St. Paul, Minnesota, a bright rainbow dazzled the Milton Township on Carleton Lake, surprising thirteen rowers and delighting nine swimmers plus fifteen residents.

On Wednesday, July 9th, in Newport California, a wedding proposal excited the Seamont Ferry off Jackson Sound, surprising seven bystanders and delighting ten family members plus sixteen crewmen.

On Thursday, May 6th, in Mobile, Alabama, a birthday party lit up the Carson Warehouse on Harvey Harbor, entertaining nineteen employees and celebrating two twins plus fourteen family members.

On Friday, April 5th, in Cincinnati Ohio, the Easter Bunny visited the Belmont Hotel on Windy Street, bringing fourteen baskets and delivering four chocolates plus eighteen eggs.

On Saturday, August 2nd, in Seattle, Washington, an amazing musical hit the theater district on Sterling avenue, receiving twenty awards and delighting five critics plus twelve celebrities.

Negative narratives

On Monday, March 4th, in Denver, Colorado, a raging blizzard buried the Hackett Airport on Billings Road, stranding sixty travelers and trapping eight children plus twenty performers.

On Tuesday, June 3rd, in St. Paul, Minnesota, a torrential rain flooded the Milton Township on Carleton Lake, swamping thirteen families and marooning nine adults plus fifteen animals.

On Wednesday, July 9th, in Newport, California, a hurricane wind grounded the Seamount Ferry off Jackson Sound, drowning seven crewmen and sparing ten passengers plus sixteen rescuers.

On Thursday, May 6th, in Mobile, Alabama, a large explosion destroyed the Carson Warehouse on Harvey Harbor, blasting nineteen workmen and burning two watchmen plus fourteen bystanders.

On Friday, April 5th, in Cincinnati, Ohio, a four-alarm fire gutted the Belmont Hotel on Windy Street, killing fourteen guests and injuring four firemen plus eighteen residents.

On Saturday, August 2nd, in Seattle, Washington, a shattering earthquake struck the theater district on Sterling Avenue, trapping twenty performers and smothering five actors plus twelve spectators.

Tables

Table 1

Memory performance as a function of prosody valence conditions.

Semantic Valence	Prosody Valence	Immediate verbatim		Immediate gist		Delayed verbatim		Delayed gist	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Neutral content (Sample 1)									
	Positive	40.2	1.5	56.2	11.9	35.2	14.6	52.8	14.1
	Neutral	46.9	13.8	60.6	12.6	39.9	15.8	56.4	14.2
	Negative	38.9	13	55.7	15.2	33.9	11.2	50.6	14.3
Positive content (Sample 2)									
	Positive	37.9	11.5	49.5	12. 4	32.2	13.5	45	12.6
	Neutral	33.2	13.8	48.4	16	27.5	14	43.6	17.7
	Negative	38.3	13	55.9	15.1	32.9	14.6	51.9	17.1
Negative content (Sample 3)									
	Positive	40.7	13.6	58.6	14.2	37.1	14.8	55.7	17.7
	Neutral	32.3	11.0	52.8	13.4	26.6	12.6	52.9	15.7
	Negative	35.6	14.9	52.9	15.7	30.0	14.9	48.3	14.7

Note. Only correct proportion (%) of the memory performance is included. Correct proportion = raw score / total score

Table 2.

Total, direct, and indirect effects of prosody valence through speech rate on memory performance of negative-content narratives.

Memory measure	Coding	Total Effect			Direct Effect			Indirect Effect (Bootstrap %95 CI)			
		B	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*LL*	*UL*
Immediate	Ind D1	.627	.164	.000	.969	.183	.000	-.342	.102	-.550	-.151
Verbatim	Ind D2	.248	.164	.131	.216	.159	.177	.033	.045	-.043	.139
	Cont D1	.438	.142	.002	.593	.144	.000	-.155	.053	-.276	-.067
	Cont D2	.378	.164	.022	.753	.188	.000	-.375	.117	-.621	-.160
Immediate	Ind D1	.406	.167	.016	.600	.191	.002	-.194	.099	-.398	-.005
Gist	Ind D2	.011	.167	.945	-.007	.166	.965	.019	.028	-.021	.101
	Cont D1	.209	.144	.150	.296	.150	.049	-.087	.048	-.197	-.006
	Cont D2	.395	.167	.019	.607	.196	.002	-.212	.110	-.437	-.004
Delay	Ind D1	.706	.162	.000	.964	.184	.000	-.258	.093	-.457	-.089
Verbatim	Ind D2	.228	.162	.161	.203	.160	.205	.025	.035	-.030	.113
	Cont D1	.467	.141	.001	.584	.144	.000	-.117	.046	-.228	-.043
	Cont D2	.478	.162	.004	.760	.189	.000	-.283	.106	-.515	-.096
Delay Gist	Ind D1	.499	.166	.003	.626	.190	.001	-.126	.089	-.310	.042
	Ind D2	.047	.166	.778	.035	.209	.835	.012	.021	-.012	.083
	Cont D1	.273	.143	.058	.330	.149	.028	-.057	.041	-.149	.015
	Cont D2	.453	.166	.007	.591	.195	.003	-.138	.010	-.348	.045

Note. 10,000 bootstrap samples. CI = confidence interval; LL = lower limit; UL = upper limit; Ind = Indicator coding; Cont = Contrast Coding; Ind D1 (positive vs. neutral) = positive prosody (1), neutral prosody (0), negative prosody (0); Ind D2 (negative vs. neutral)= positive prosody (0), neutral prosody (0), negative prosody (1); Cont D1 (contrast between valenced prosody and neutral prosody) = positive prosody (-1/3), neutral prosody (2/3), negative prosody (-1/3); Cont D2 (contrast between positive and negative prosody) = positive prosody (1/2), neutral prosody (0), negative prosody (-1/2).

Figures

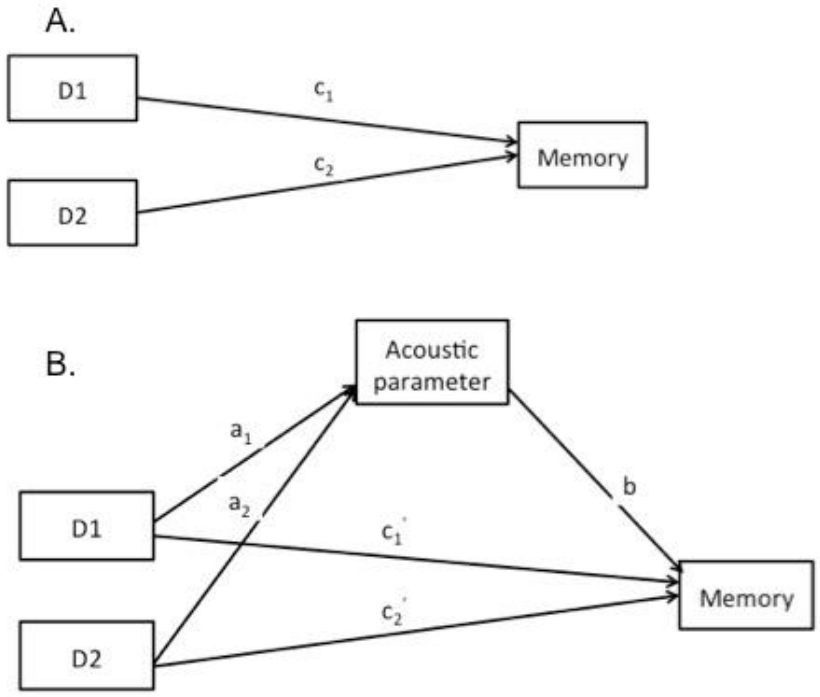

Figure 1.

Mediation of acoustic parameters on the effects of prosody valence on memory. D1 and D2 are dummy codes that represent either positive and negative prosody or valenced and neutral prosody. c_1 and c_2 quantifies the total effects of prosody on memory. a_1 and a_2 quantifies differences between D1 and D2 on Mediator, c_1' and c_2' quantifying differences between D1 and D2 on dependent variable (Y: memory) holding M (acoustic parameters) constant, and b estimating the effect of M on Y while statistically equating the groups on average on X. The direct effect of X on Y is captured in the estimates of c_1' and c_2' and the indirect effect of X on Y through M is estimated by the products a_1b and a_2b. Evidence that at least one relative indirect effect is different from zero supports the conclusion that M mediates the effect of X on Y.

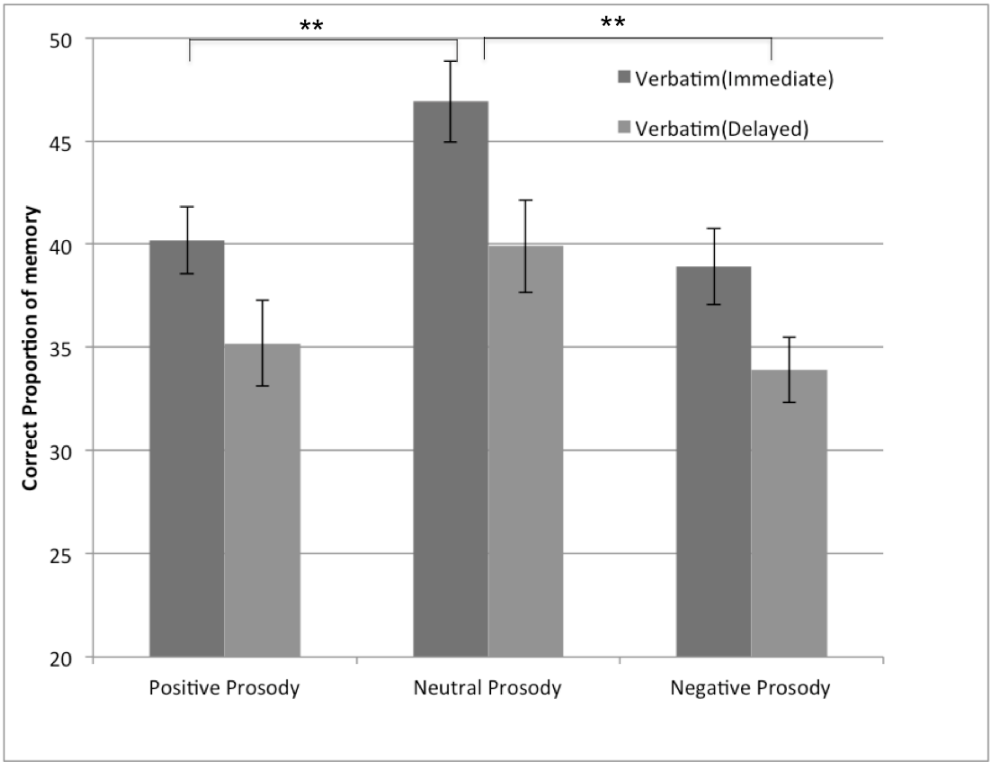

Figure 2.

Mean correct proportion of verbatim memory in each prosody condition in neutral-content narratives. Error bars indicate the standard errors of the means. ** indicates significant level $p < .001$.

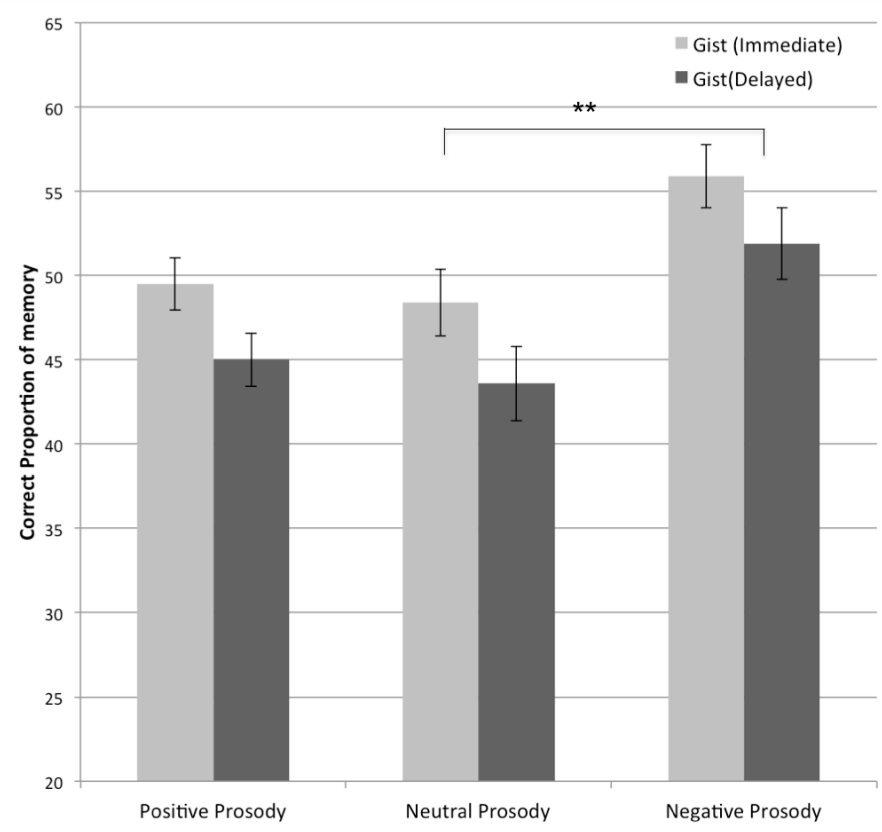

Figure 3.

Mean correct proportion of gist memory in each prosody condition in positive-content narratives. Error bars indicate the standard errors of the means. ** indicates significant level $p < .001$.

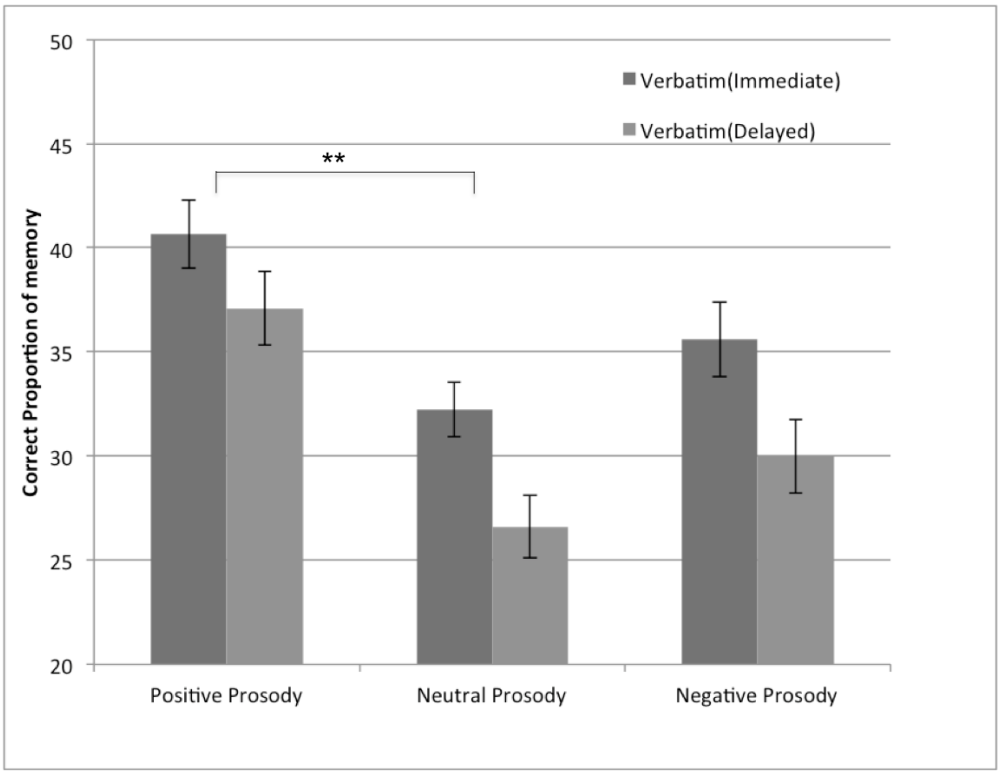

Figure 4.

Mean correct proportion of verbatim memory in each prosody condition in negative-content narratives. Error bars indicate the standard errors of the means. ** indicates significant level $p < .001$.

Supplementary Materials

How Much Does Your Tone of Voice Matter?

Effects of Prosody Valence on Memory for Speech

Supplementary Materials-Texts

Text S1. Affective Ratings of Spoken Narrative Recordings

Affective ratings of the selected narrative recordings were first aggregated from the 37 participants' ratings for valence ratings and arousal ratings separately for each of the 36 recordings in each study. The aggregated valence and arousal ratings were then subjected to ANOVAs and t-tests to examine their differences among three affective prosody conditions.

Affective ratings of neutral content narrative recordings. ANOVAs conducted on affective ratings of the neutral-content narratives spoken in different affective prosody indicated a significant effect of affective prosody on valence ratings (F (2,33) = 7.17, p = .003), but non-significant effect on arousal ratings (F (2,33) = 0.79, p = .401; Table S2). Post hoc Bonferroni tests showed that semantically neutral narratives spoken in a positive tone were rated significantly higher (more positive) in valence than those spoken in a negative tone (p = .002), and marginally higher than those spoken in a neutral tone (p = *.077)*; no significant difference was found between neutral and negative prosody readings (p > 0.1).

Affective ratings of positive content narrative recordings. One-way ANOVA conducted on affective ratings indicated a significant difference for valence ratings (F

(2,33) = 7.49, p = .002, see Table S3). Post hoc Bonferroni tests showed that for semantically positive content narratives, the positively spoken utterances were rated significantly higher in valence ratings than those spoken in a negative state (p = .002). There was no significant difference between positive and neutral states, or between neutral and negative states (*ps* > *.05*). A second one-way ANOVA indicated a non-significant difference for arousal ratings (F (2,33) = 0.99, p = .382).

Affective ratings of negative content narrative recordings. One-way ANOVA conducted on affective ratings indicated a significant difference for valence ratings (F (2,33) = 3.61, p = .038; see Table S4). Post hoc Bonferroni tests showed that for negative content narratives, when spoken in a negative prosody, they were rated significantly lower in valence than those spoken in a neutral prosody (*p = .034)*; there was no significant difference between positive and neutral tone, or between positive and negative (*ps* > 0.1). A second one-way ANOVA indicated a non-significant difference for arousal ratings (F (2,33) = 2.46, p = .101).

Text S2. Discrimination of Prosody Valence via Acoustic Parameters

For neutral-content narratives, the 5 a priori selected acoustic parameters produced one significant discriminant function in the prediction of valence categories of the affective prosody, Wilks' Lambda = .759, $\chi 2(2) = 9.091$, p = .011. The discriminant function had an Eigenvalue of .317 and explained 100% of variance (Canonical correlation = .491). Speech rate loaded on the function (1.000). The discriminant function achieved a level of 50% correct classification (also 50% with "leave-one-out" cross-validation). Cross-validated classification results were: 25% of negative, 50% of neutral, and 75% of positive.

For positive-content narratives, the 5 a priori selected acoustic parameters produced two discriminant functions. The first discriminant function was statistically significant, Wilks' Lambda = .370, $\chi 2(4) = 32.35$, $p < .001$, but the second was not, Wilks' Lambda = .989, $\chi 2(1) = .360$, $p = .548$. The first function had an Eigenvalue of 1.676 and explained 99.3% of variance (Canonical correlation = .791). The second function had an Eigenvalue of .011 and explained 0.7% of variance (Canonical Correlation = .105). Mean intensity loaded on the first function (0.766), and speech rate loaded on the second function (0.879). The discriminant functions achieve a level of 61.1% correct classification (also 61.1% with "leave-one-out" cross-validation). Cross-validated classification results were: 83.3% of negative, 41.7% of neutral, and 58.3% of positive.

For negative-content narratives, the 5 a priori selected acoustic parameters produced two discriminant functions. The first discriminant function was statistically significant, Wilks' Lambda = .511, $\chi 2(4) = 21.808$, $p < .001$, but the second was not, Wilks' Lambda = .993, $\chi 2(1) = .219$, $p = .64$. The first function has an Eigenvalue of .943 and explained 99.3% of variance (Canonical correlation = .697). The second function has an Eigenvalue of .07 and explained .7% of variance (Canonical Correlation = .082). Speech rate loaded on the first function (.706) and spectral slope loaded on the second function (.708). The discriminant functions achieve a level of 61.1% correct classification (50% with "leave-one-out" cross-validation). Cross-validated classification results were: 16.7% of negative, 50% of neutral, and 83% of positive.

Text S3. Data-driven Approach in Acoustic Parameter Selection

In the data driven approach, we used 31 acoustic parameters commonly measured in affective prosody studies. First, to control for variations due to inter-individual

differences, we used z-transformations to make all acoustics parameters independently standardized within speaker. Then, to reduce multicollinearity in subsequent analyses, we selected a smaller set of parameters on the basis of an exploratory principal component analysis of the 31 extracted acoustic parameters. Lastly, we used multiple regressions to assess the extent to which the affective valence ratings can be predicted by the set of selected acoustic parameters. All analyses were done separately for each of the 3 studies. Table S5 provided a summary of acoustic parameter selection and Table S6 provided a comparison of the classification result for prosody valence from both a-priori-based and data-driven approach.

Acoustic parameters used in data-driven approach. The 31 acoustic parameters focused on duration, pitch, intensity, and voice quality (spectral balance and vocal perturbation). Duration measures included the overall duration of each narrative recording and speech rate (defined as the number of syllables per second). We also further divided the overall duration into the duration of the voiced part, the duration of the unvoiced part, and the duration of the silent part.

Pitch measures included mean, minimum, maximum, and standard deviation of fundamental frequency (MeanF0, MinF0, MaxF0, StDevF0). Pitch range (F0Range) was also computed from MaxF0 and MinF0. We checked all the F0 contours before proceeding to feature extraction to manually correct for outliers. All F0 measures were computed over the voiced parts of the utterance. Intensity measures included mean, minimum, maximum, and standard deviation of intensity (MeanInt, MinInt, MaxInt, StDevInt). Intensity range (IntRange) was also computed from MaxInt and MinInt.

Spectral measures included mean and standard deviation of spectrum, spectral kurtosis, spectral skewness, spectral slope, the Hammarberg index, the proportion of energy below 500 Hz, and the proportion of energy below 1000 Hz. Spectral kurtosis and skewness are used to describe the shape of the spectrum. The spectral kurtosis is a measure for how much the shape of the spectrum around the centre of gravity (how high the frequencies in a spectrum are on average) is different from a Gaussian shape (Boersma and Weenink, 2012). The skewness of the spectrum was defined as the extent to which the spectrum skews around its mean. Spectral slope is a measure for how quickly the spectrum of an audio sound tails off towards the high frequencies. The Hammarberg index (Hammarberg et al., 1980) characterized the spectral balance by comparing the energy maxima in the 0–2000 Hz range and the 2000–5000 Hz range. The proportions of the energy below and above 500 and 1000 Hz attempt to divide the signal into a part related to F0 energy and/or vowel expression, as well as the high frequency parts of the spectrum, well-known in the field of vocal expression of emotion (Van Bezooijen, 1984).

Vocal perturbation measures included percentage of voicing, (%Voicing, computed based on 100-ms frames), jitter (mean cycle-to-cycle, F0 variation across voiced frames), shimmer (mean cycle-to-cycle, F0 amplitude variation across frames), and harmonic-to-noise ratio (HNR, relative proportion of periodic versus aperiodic energy in voiced frames), which included MaxHNR, MeanHNR, and StDevHNR. Jitter and shimmer are voice quality parameters that reflect small variations in pitch and intensity respectively. HNR reflected the proportion of periodicity that is present in the

sound expressed in dB (an HNR of 0 dB indicates an equal amount of noise and periodicity in the signal).

Lastly, pitch and intensity entropy were also computed using information entropy (Shannon, 1948) to measure the variability in pitch and intensity, as the "inflection" and "emphasis" variables in Cohen et al.'s (2009) "Laboratory-based Procedure for Measuring Emotional Expression from Natural Speech". Pitch entropy and intensity entropy (*H*Pitch and *H*Int) were derived after first removing all unvoiced frames from the passage, computing F0 and intensity (in dB) for each frame, and then normalizing the resulting datasets by subtracting their respective means from each value and dividing by the standard deviation. Resulting values were tabulated as frequency distributions using *N*/30 equally sized bins, with *N* representing the total number of frames in the file. A probability distribution was then created by dividing the number of cases in each bin by *N*. Entropy (*H*) was calculated for each probability distribution as:

$$H = -\sum p_i \log_2 p_i,$$

where p was the probability a given data point occurring in the ith bin. H thus measured uncertainty within the probability distribution in bits. Higher H values indicated a more even distribution across bins, while lower H values indicated a more peaked distribution. The H value computation was scripted into the GSU Praat Tools "quantifyEmotion" (developed by co-author M.O.).

Text S4. Recognition Memory

With regard to recognition performance, we predicted that the memory pattern would be less consistent in recognition performance comparing to recall performance. Research has shown that affective events and stimuli (e.g., words, sentences, pictures,

and narrated slide shows) were usually recalled at higher rates than were neutral events and stimuli (see reviews by Buchanan, 2007; Hamann, 2001). However, effects of affect on recognition were less consistent (see reviews by Kensinger & Schacter, 2008). For instance, while rates of "remembering" tend to be much higher for affective stimuli than neutral stimuli when it comes to vividly remembering the item's prior presentation, the overall recognition is equivalent for affective and neutral information (e.g., Kensinger & Corkin, 2003; Sharot, Delgado, & Phelps, 2004). Therefore, we predicted and found that the effects of valence on recall would be more consistent across different measures and conditions, whereas the effects on recognition may not be manifested in particular condition.

Recognition memory measurement. The recognition test was adapted from Kensinger et al. (2004) and consisted of 14 multiple-choice questions, each with 3 distractors and 1 correct answer. For example, "On what day of the week did the event occur? 1) Friday, 2) Thursday, 3) Tuesday, 4) Sunday". The recognition score was computed from the correctness of the 14 recognition questions (maximum 14 points). No point was given when the reaction time of answering a recognition question was less than 1000 ms, in which case participants probably pressed the wrong button or skipped the question. Recognition performance was measured once after the delayed recall.

Sample 1 recognition results. As predicted, repeated measure ANOVA with Prosody Valence as the within-subject factor revealed a main effect of prosody on neutral content narratives ($F(2, 98) = 5.76$, $p = .004$, $\eta 2 = .105$; see Table S10 for descriptive statistics). As shown in Figure S1, pairwise comparisons indicated that the specific content of neutral narratives read in a neutral prosody was better recognized than those

read in a positive prosody (p = .007) and negative prosody (p = .025). The results were similar to the recall performance for neutral semantic narrative.

Sample 2 recognition results. A repeated measure ANOVA revealed a main effect of prosody valence on positive-content narratives (F (2, 126) = 8.27, $p < .001$, $\eta 2 = .116$; see Table S10 for descriptive statistics). As shown in Figure S2, pairwise comparisons indicated recognition for positive-content narratives read in a negative prosody was better than those read in a neutral prosody ($p < .001$). However, no significant differences for other pairwise comparisons were present ($ps > .1$).

Sample 3 recognition results. For negative-content narratives, repeated measure ANOVA with Prosody Valence as within-subject factor revealed that the effect of prosody valence was not significant on recognition ($p = .725$; see Table S10 and Figure S3). It is possible that the negative events described in the narratives elicited more focused attention which led to more accuracy in recognition performance, similar to previous research findings (see review by Kensinger, 2007). Another possibility is that the retrieval and consolidation of memory in the recall tasks before the recognition task, and the recognition tests were generally easy and therefore recognition performance reached a ceiling effect.

Text S5. Mediation of Acoustic Parameters on Memory Performance: Dummy coding and contrast coding

First, we used dummy coding to test if there were any mediation effects in positive or negative condition relative to neutral condition. Then, we explore the contrast between neutral and valenced condition, as well as the contrast between positive and negative condition, we used contrast indicators to code the different affective prosody

conditions. If we use well-disseminated rules for the construction of contrasts (see, for example, Keppel & Wickens, 2004; Rosenthal & Rosnow, 1985), the codes corresponding to the first contrast would be – 2, 1, and 1 for the neutral, positive, and negative prosody conditions, respectively. For the second contrast the codes would be 0, – 1, and 1. Hayes and Preacher (2014) recommend a transformation of the contrast codes so that the largest and smallest codes in a set differ by only one unit, which was accomplished by dividing each of the codes in the k – 1 sets by the absolute value of the difference between the largest and smallest contrast codes. This scaled all relative direct, indirect, and total effects on a mean difference metric. In this case, the first set contained three codes (– 2, 1, 1) the largest and smallest which differ by 3 units, and the second set contained codes with a maximum absolute difference of 2. Thus, the resulting transformed codes became – 2/3, 1/3, and 1/3 for the first set and 0, – 1/2, and 1/2 for the second set. Therefore, D1 and D2 were defined for each condition as D1 = – 0.667, D2 = 0 for neutral prosody condition, D1 = 0.333, D2 = – 0.5 for positive prosody condition, D1 = 0.333, D2 = 0.5 for negative prosody condition.

References

Buchanan, T. W. (2007). Retrieval of emotional memories. *Psychological Bulletin*, *133*(5), 761–779. doi:10.1037/0033-2909.133.5.761

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences*, *5*(9), 394–400. doi:10.1016/S1364-6613(00)01707-1

Kensinger, E. a. (2007). Negative emotion enhances memory accuracy. *Current Directions in Psychological Science*, *16*(4), 213–218. doi:10.1111/j.1467-8721.2007.00506.x

Kensinger, E. a, & Corkin, S. (2003). Memory enhancement for emotional words: are emotional words more vividly remembered than neutral words? *Memory & Cognition*, *31*(8), 1169–1180. doi:10.3758/BF03195800

Kensinger, E. A., & Schacter, D. L. (2008). Memory and emotion. *Handbook of Emotions*, *3*, 601–617.

Sharot, T., Delgado, M. R., & Phelps, E. A. (2004). How emotion enhances the feeling of remembering. *Nature Neuroscience*, *7*(12), 1376–1380.

Supplementary Material - Tables

Table S1.

Affective ratings of the written narratives in different content valence.

	Valence Ratings		Arousal Ratings	
	M, SD	95% CI	*M, SD*	95% CI
Positive content	5.07, 0.27	[4.98, 5.16]	3.74, 0.29	[3.64, 3.83]
Neutral content	4.12,0.38	[3.99, 4.25]	3.29, 0.52	[3.11, 3.46]
Negative content	1.91,0.33	[1.80, 2.02]	5.37, 0.34	[5.25, 5.48]

Notes: Valence ratings ranged from 1 (extremely negative) to 7 (extremely positive). Arousal ratings ranged from 1 (extremely calming or soothing) to 7 (extremely exciting or agitating).

Table S2.

Affective ratings of the neutral content narrative recordings (Sample 1) as a function of prosody condition.

	Valence Ratings		Arousal Ratings	
	M, SD	95% CI	*M, SD*	95% CI
Positive prosody	4.92, 0.71	[4.48, 5.37]	3.59, 1.04	[2.93, 4.25]
Neutral prosody	4.30, 0.59	[3.93, 4.67]	3.92, 0.88	[3.37, 4.48]
Negative prosody	3.92, 0.67	[3.50, 4.35]	4.09, 0.82	[3.57, 4.61]

Notes: Valence ratings ranged from 1 (extremely negative) to 7 (extremely positive). Arousal ratings ranged from 1 (extremely calming or soothing) to 7 (extremely exciting or agitating).

Table S3.

Affective ratings of the positive content narrative recordings (Sample 2) as a function of prosody condition.

	Valence Ratings		Arousal Ratings	
	M, SD	95% CI	*M, SD*	95% CI
Positive prosody	5.49, 0.92	[4.91, 5.08]	4.07, 0.42	[3.80, 4.33]
Neutral prosody	4.77, 0.70	[4.33, 5.21]	3.84, 0.38	[3.60, 4.08]
Negative prosody	4.25, 0.74	[3.78, 4.72]	3.86, 0.50	[3.77, 4.07]

Notes: Valence ratings ranged from 1 (extremely negative) to 7 (extremely positive). Arousal ratings ranged from 1 (extremely calming or soothing) to 7 (extremely exciting or agitating).

Table S4.

Affective ratings of the negative content narrative recordings (Sample 3) as a function of prosody condition.

	Valence Ratings [a]		Arousal Ratings [b]	
	M, SD	95% CI	*M, SD*	95% CI
Positive prosody	2.11, 0.65	[1.70, 2.52]	4.23, 0.50	[3.91, 4.55]
Neutral prosody	2.47, 0.41	[2.21, 2.72]	3.72, 0.49	[3.41, 4.03]
Negative prosody	1.86, 0.59	[1.49, 2.23]	3.90, 0.70	[3.46, 4.34]

Notes: [a] Valence ratings ranged from 1 (extremely negative) to 7 (extremely positive).[b] Arousal ratings ranged from 1 (extremely calming or soothing) to 7 (extremely exciting or agitating).

Table S5

Selected acoustic parameters from Principal Component Analyses of 31 acoustic parameters in the data-driven approach for Studies 1, 2 and 3

Acoustic Parameter	Description	Sample 1 Neutral Content	Sample 2 Positive Content	Sample 3 Negative Content
Overall Duration	Duration of the entire vocal recording	X		X
Speech Rate	**Speech rate by number of syllables**		**X**	
F0 Min	Minimum of fundamental frequency	X		
F0 SD	**Standard deviation of F0**	**X**		
F0 Range	The range of fundamental frequency		X	
Intensity Mean	**Mean of intensity**		**X**	
Intensity Max	Maximum Intensity		X	X
Intensity SD	**Standard deviation of intensity**		**X**	**X**
Intensity Range	The range of intensity	X	X	X
Intensity<1000	The proportion of energy below 1000 Hz			X
Intensity<500	The proportion of energy below 500 Hz	X		
Spectrum SD	Standard deviation of spectrum	X		
Spectral Skewness	The extent to which the spectrum skews around its mean	X	X	
Spectral Slope	**How quickly the spectrum of a sound tails off towards the high frequencies**			**X**
% Voiced Frames	%Voicing, computed based on 100-ms frames			X
HNR SD	Standard Deviation of HNR			X
Mean HNR	Relative proportion of periodic versus aperiodic energy in voiced frames		X	
Entropy of Pitch	Variability in F0 using information entropy	X		

Note. **Bolded** indicated the acoustic parameters that were selected in the a priori approach.

Table S6

Summary of acoustic parameter selection and classification results from both a-priori-based and data-driven approach.

	Sample 1 Neutral Content	Sample 2 Positive Content	Sample 3 Negative Content
A priori approach	Speech rate Wilks' Lambda = .759, $\chi^2(2) = 9.091$, $p = .011$ 50% correct classification 50% "leave-one-out" cross-validation 25% of negative prosody 50% of neutral prosody 75% of positive prosody	Intensity mean Speech rate Wilks' Lambda = .370, $\chi^2(4) = 32.35$, $p < .001$ 61.1% correct classification 61.1% "leave-one-out" cross-validation 83% of negative prosody, 42% of neutral prosody 58% of positive prosody	Speech rate Spectral slope Wilks' Lambda = .511, $\chi^2(4) = 21.808$, $p < .001$ 61.1% correct classification 50% "leave-one-out" cross-validation 17% of negative prosody, 50% of neutral prosody 83% of positive prosody
Data-driven approach	Overall duration Wilks' Lambda = .707, $\chi^2(4) = 11.419$, $p = .003$ 50% correct classification 50% "leave-one-out" cross-validation 83% of negative prosody, 8% of neutral prosody 58% of positive prosody	Intensity mean Speech rate Wilks' Lambda = .370, $\chi^2(4) = 32.35$, $p < .001$ 61.1% correct classification 61.1% "leave-one-out" cross-validation 83% of negative prosody, 42% of neutral prosody 58% of positive prosody	Overall duration Spectral slope % voiced frame Wilks' Lambda = .433, $\chi^2(6) = 26.767$, $p < .001$ 66.7% correct classification 61.1% "leave-one-out" cross-validation 67% of negative prosody, 58% of neutral prosody 75% of positive prosody

Table S7.

Speaker gender, induction method, and values of the 6 acoustic parameters in Memory Study Sample 1 (neutral content).

Speaker#	Gender	Induced/Portrayal	Prosody Valence	Speech Rate	F0 SD	Intensity Mean	Intensity SD	Below 500	Spectral Slope
1	F	I	1	0.256	26.69	69.3	12.61	75.55	-23.12
		I	3	0.227	24.36	69.5	12.51	75.52	-21.97
		I	2	0.2174	19.62	69.6	12.47	76.56	-24.91
2	F	I	1	0.2313	58.47	69	12.57	74.36	-14.51
		I	3	0.2257	37.23	69.4	12.02	74.60	-16.18
		I	2	0.2005	38.78	70.5	11.53	75.39	-18.47
3	F	P	2	0.2027	11.45	69.9	10.95	75.49	-21.05
		P	1	0.2151	35.97	71.3	10.97	76.42	-19.23
		P	3	0.2127	12.61	67.2	10.95	73.04	-20.46
4	F	I	1	0.2515	58.04	68.9	10.06	74.35	-18.21
		I	3	0.2327	33.47	70.1	9.96	75.36	-19.27
		I	2	0.2243	46.28	67.8	9.97	73.60	-20.73
5	M	I	3	0.2116	13.3	69.7	11.1	75.10	-20.56
		I	2	0.1919	14.17	67.6	11.51	73.78	-21.48
		I	1	0.2189	20.24	66.9	12.03	73.92	-22.37
6	F	P	2	0.2223	32.49	70.7	10.57	75.19	-16.06
		P	1	0.2227	45.91	72.2	9.31	75.28	-15.50
		P	3	0.1917	41.02	71.7	10	75.59	-17.79
7	M	I	1	0.2598	22.55	67	12.07	74.03	-23.36
		I	3	0.2257	11.88	69.6	11.69	74.86	-19.03
		I	2	0.2146	12	69.5	12.03	75.88	-21.53
8	M	P	3	0.1914	26.99	64.6	13.11	72.67	-16.86
		P	2	0.2127	19.37	64	12.59	71.65	-16.71
		P	1	0.2369	25.63	66.9	12.04	73.31	-17.10
9	M	I	2	0.1845	8.38	66.6	10.34	72.81	-18.66

		I	1	0.1946	11.68	66.8	10.43	73.17	-18.09
		I	3	0.1959	10.59	66.7	12.59	73.99	-17.67
10	M	P	2	0.1971	5.41	70.4	8.69	74.26	-17.69
		P	1	0.1947	8.11	72.4	8.64	75.97	-18.74
		P	3	0.1752	11.27	72.3	8.19	75.65	-17.79
11	M	P	2	0.2085	6.65	68.2	12.18	75.03	-21.16
		P	1	0.224	10.92	67.8	12.04	74.41	-18.49
		P	3	0.195	5.68	66.9	13.08	74.80	-21.29
12	F	I	3	0.2136	35.79	68.5	10.61	74.23	-20.29
		I	2	0.1922	36.57	67.6	10.67	73.68	-21.20
		I	1	0.2206	50.87	68.2	11.5	74.47	-21.48

Note. For prosody valence, 1 = positive prosody; 2 = neutral prosody; 3 = negative prosody.

Table S8.

Speaker gender and values of the 6 acoustic parameters in Memory Study Sample 2 (positive content)

Speaker#	Gender	Prosody Valence	Speech Rate	F0 SD	Intensity Mean	Intensity SD	Below 500	Spectral Slope
1	M	3	0.2071	6.43	69.7	10.84	74.907	-19.24
		2	0.1927	5.56	71.2	9.85	75.6161	-21.94
		1	0.2012	12.42	70.1	12.23	76.1217	-18.73
2	F	2	0.2279	20.09	72.7	10.89	76.9559	-18.90
		1	0.2155	28.32	70.3	11.4	74.6331	-15.40
		3	0.2515	24.93	69.9	12	74.974	-16.07
3	M	2	0.2082	16.94	70	9.57	74.5311	-21.27
		1	0.2027	25.53	71.2	10.44	76.1675	-20.96
		3	0.2473	27.61	69.8	10.22	75.1005	-20.07
4	M	3	0.2256	14.38	70.2	13.47	78.0918	-25.23
		2	0.207	14.63	72.2	11.56	77.677	-25.51
		1	0.2053	13.01	72.1	12.55	78.4454	-24.96
5	F	2	0.2046	16.83	71.6	11.19	76.7954	-22.34
		1	0.2053	16.94	71	11.41	76.319	-21.71
		3	0.223	16.75	68.2	11.15	74.1416	-24.20
6	M	2	0.2062	9.69	69.2	10.91	74.2113	-18.87
		1	0.1877	23.03	73.1	10.55	77.9046	-20.62
		3	0.2177	9.73	66.9	10.15	72.4295	-18.28
7	F	3	0.2056	29.2	65.5	10.94	71.8985	-17.90
		2	0.1976	29.67	70.4	11.15	75.906	-19.83
		1	0.1971	39.78	72.4	11.55	77.5606	-19.87
8	F	1	0.2031	62.04	70.1	10.03	74.8392	-17.63
		3	0.2274	64.86	69.4	9.71	74.9061	-17.82
		2	0.2103	35.38	70.5	10.29	75.8491	-22.75
9	M	1	0.2196	18.54	70.7	11.51	76.539	-20.10

		3	0.2035	12.58	67.8	11.68	73.791	-20.94
		2	0.2235	9.65	69.4	10.49	74.9048	-21.69
10	F	3	0.1918	34.87	67.2	12.18	72.8875	-16.36
		2	0.2035	38.03	71.1	12.34	76.5977	-20.15
		1	0.1895	35.79	70.2	11.54	75.5859	-20.50
11	M	2	0.2097	6.45	70.4	11.43	75.8964	-20.13
		1	0.1895	10.77	70.8	11.21	76.5927	-19.72
		3	0.2164	7.13	67.3	11.93	74.1592	-20.97
12	F	1	0.2088	24.95	69.8	11.13	74.3166	-17.62
		3	0.2329	23.18	66.1	13.21	71.6541	-16.41
		2	0.2104	22.45	68.6	12.51	73.1938	-16.36

Note. For prosody valence, 1 = positive prosody; 2 = neutral prosody; 3 = negative prosody.

Table S9.

Speaker gender and values of the 6 acoustic parameters in Memory Study Sample 3 (negative content)

Speaker#	Gender	Prosody Valence	Speech Rate	F0 SD	Intensity Mean	Intensity SD	Below 500	Spectral Slope
1	M	3	0.2192	7	70.6	11.39	76.0421	-20.18
		2	0.2077	5.42	69.9	10.01	74.27	-20.86
		1	0.2088	17.62	68.4	12.89	74.3819	-18.00
2	F	2	0.2398	17.98	71.8	11.34	76.2984	-17.61
		1	0.2139	20.02	70.6	11.29	74.0632	-14.18
		3	0.2528	13.87	71.5	12.11	76.5805	-17.39
3	M	2	0.2141	14.32	69.4	9.46	74.0949	-21.64
		1	0.2029	26.44	69	11.65	74.051	-18.06
		3	0.2456	14.4	70.3	9.7	75.6091	-21.94
4	M	3	0.2195	7.8	69.2	13.34	76.796	-26.20
		2	0.2272	8.47	71.7	12.24	77.8433	-24.16
		1	0.1989	10.46	70.9	12.08	77.0597	-22.95
5	F	2	0.2173	14.79	72.8	11.12	77.579	-21.56
		1	0.2109	20.7	69.5	12.38	75.6058	-19.47
		3	0.2184	13.79	70.5	11.31	76.2664	-23.02
6	M	2	0.2194	7.32	71.1	10.59	76.1323	-19.61
		1	0.1797	17.87	68.4	10.98	73.3137	-17.62
		3	0.2148	9.02	68.7	10.85	74.2132	-19.99
7	F	3	0.1981	20.93	71.2	10.03	76.3662	-19.08
		2	0.2174	33.75	73.1	10.62	78.4753	-19.79
		1	0.1972	34.17	70.1	12.31	75.3181	-17.28
8	F	1	0.2075	66.2	68.8	10.44	74.0652	-17.23
		3	0.2462	42.42	69.3	9.67	74.6696	-19.74
		2	0.2241	29.24	70.4	9.46	75.4061	-21.40
9	M	1	0.2008	12.27	68.8	11.91	74.5016	-19.64

		3	0.2016	10.48	68.2	10.81	74.753	-22.26
		2	0.2136	9.32	70.3	11.53	76.0022	-22.18
10	F	3	0.2018	24.71	67	11.65	73.0052	-17.50
		2	0.2017	22.63	69.2	10.59	73.7357	-21.44
		1	0.1707	27.83	71.7	11.16	75.502	-18.19
11	M	2	0.2186	7.17	70	11.42	76.0189	-20.49
		1	0.1931	11.54	68.3	12.22	74.823	-17.71
		3	0.2102	6.99	69.2	12.19	75.9684	-20.69
12	F	1	0.2177	25.2	63.7	12.57	68.7313	-15.83
		3	0.2327	20.8	68.3	12.71	72.7306	-15.72
		2	0.227	17	68	11.09	71.7821	-16.51

Note. For prosody valence, 1 = positive prosody; 2 = neutral prosody; 3 = negative prosody.

Table S10

Recognition performance as a function of prosody valence conditions.

Prosody Valence	Sample 1 (Neutral content)		Sample 2 (Positive content)		Sample 3 (Negative content)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Positive	64.7	15.3	60.7	19.4	61.7	12.6
Neutral	72.9	12.6	50.7	15.4	62.4	14.1
Negative	66	13.2	65.5	13.6	63.4	15.5

Note. Only correct proportion (%) of the recognition scores is included. Correct proportion = raw score/ total score.

Supplementary Material – Figures

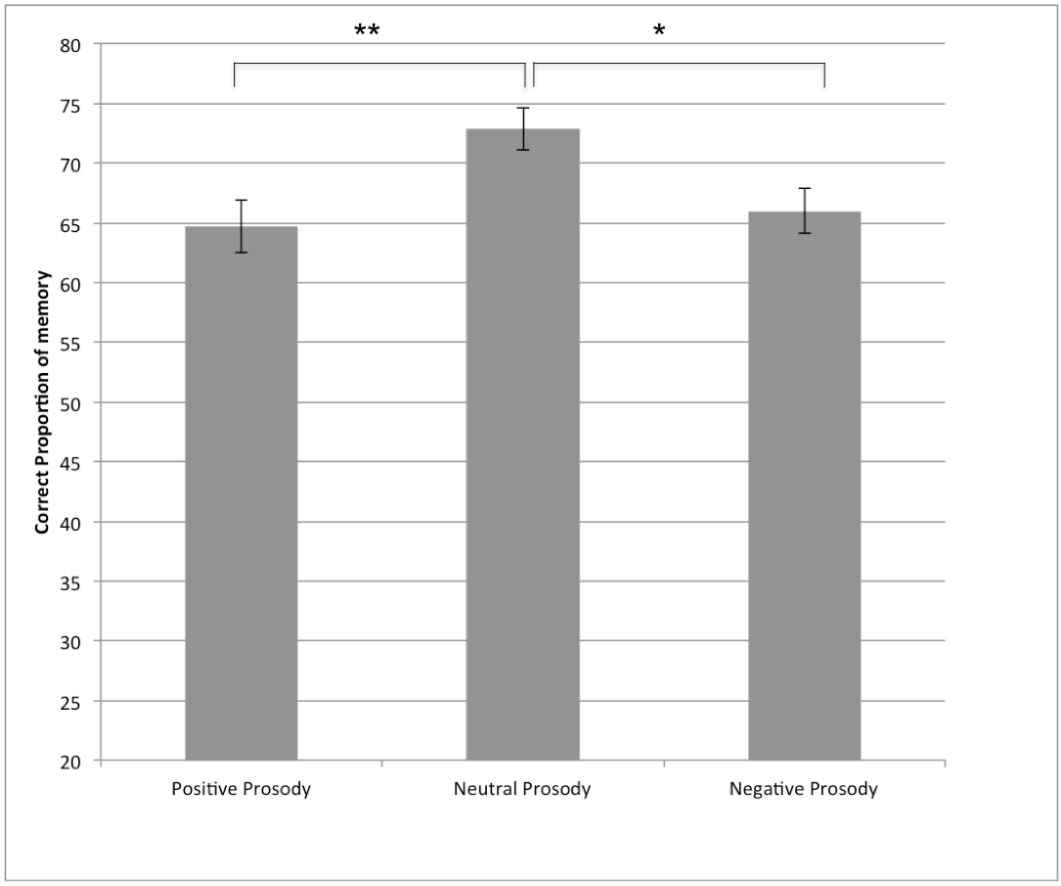

Figure S1.

Mean correct proportion of recognition memory in each prosody condition for neutral-content narratives. Error bars indicate the standard errors of the means. ** indicates significant level $p < .001$. * indicates significant level $p < .05$.

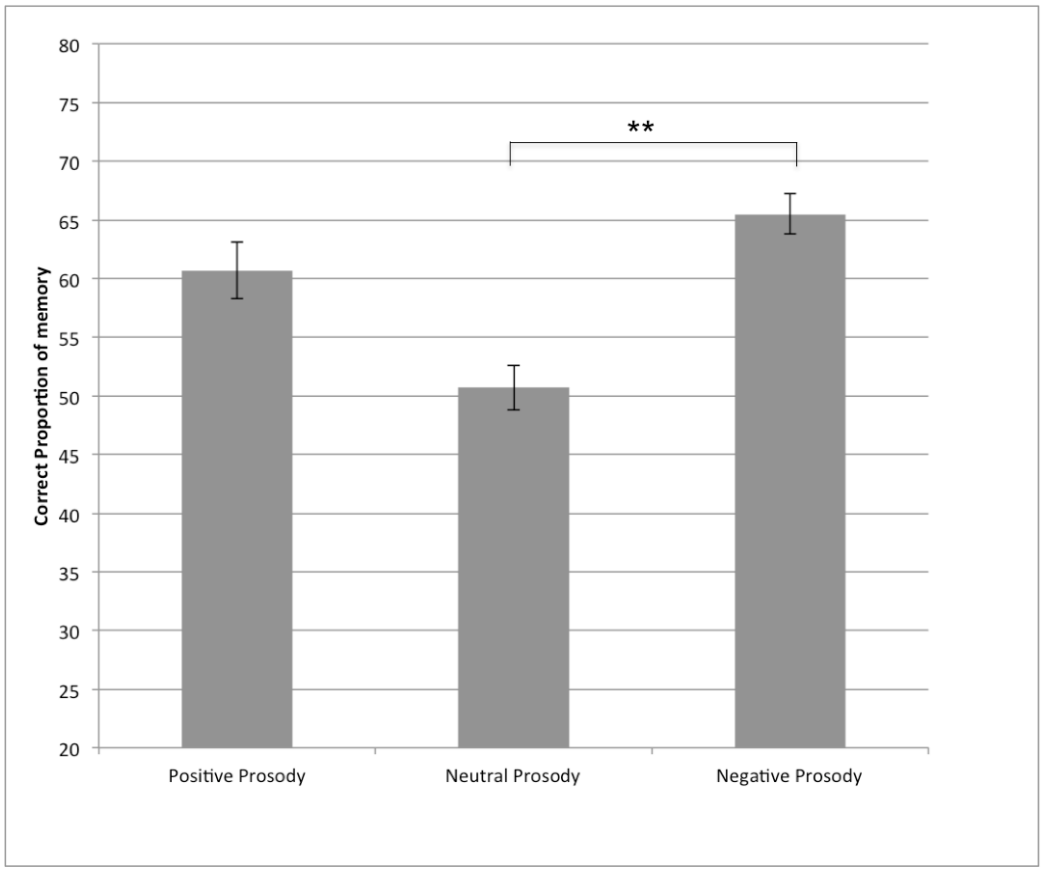

Figure S2.

Mean correct proportion of recognition memory in each prosody condition for positive-content narratives. Error bars indicate the standard errors of the means. ** indicates significant level $p < .001$.

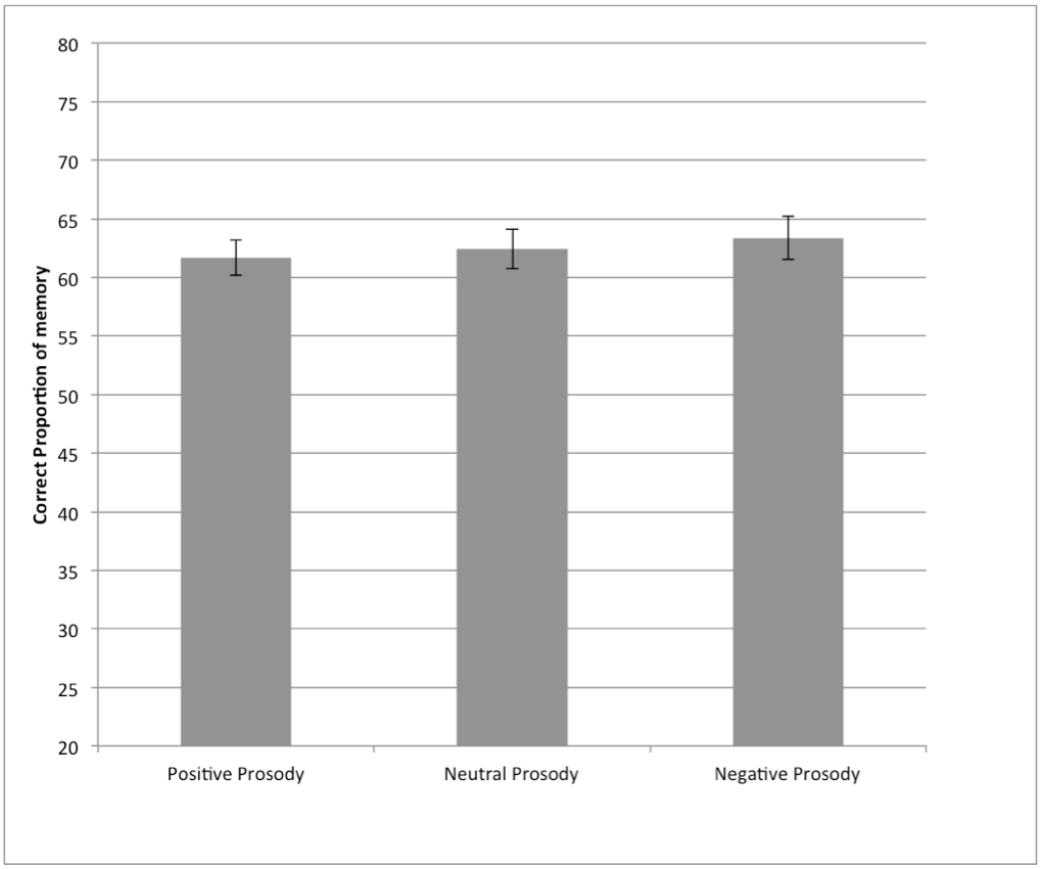

Figure S3.

Mean correct proportion of verbatim memory in each prosody condition in negative-content narratives. No significant difference was found among different prosody valence condition.

CHAPTER 2

Underlying Dimensions of Social Perception from Voice vs. Face

Xuan Zhang[1,2], Michael Owren[3],

Spencer Lynn[2*], Lisa Feldman Barrett[2,4*]

[1]Department of Psychology, 300 McGuinn Hall, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA, USA 02467

[2]Affective Science Institute and Department of Psychology, 125 Nightingale Hall, 360 Northeastern University, 360 Huntington Avenue, Boston, MA, USA 02115

[3]Department of Psychology, Emory University, Clifton Road NE Suite 234, Atlanta, GA, USA 30322

[4]Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA 02129

*shared senior authorship

Correspondence to:

Xuan Zhang

E-mail: xuan.zhang@bc.edu

Abstract

This study examined the impact of modality (voice vs. face) and the gender on dimensions of social perception using an experimental, data-driven approach. Participants listened to voices or viewed faces and freely wrote anything that came to mind about what they think of the person who possesses the voice or face. The generated descriptors were classified into categories among which the most frequently occurring social trait categories were selected for subsequent ratings. A subsequent group of participants separately rated the voices and faces on the selected social traits. For social evaluation of voices, Principal Component Analyses revealed that female voices were evaluated mostly on three dimensions: *attractiveness, trustworthiness, and dominance*; whereas male voices were evaluated mostly on two dimension: *social engagement and trustworthiness*. The dissociation between *attractiveness* and *dominance* dimensions was discussed with respect to gender stereotype of voice: whereas lower voice pitch was found to be perceived as more dominant and less attractive for female, it was perceived as more dominant and more attractive for male. For social evaluation of faces, a two-dimensional structure of *social engagement and trustworthiness* was found for both genders. Our results also suggested that trait dominance was judged as a more negative trait for female but a more positive trait for male when evaluating faces.

Keywords: social perception, underlying dimension, voice, face

Introduction

We make judgments about a person's social and personality traits all the time: from a brief phone conversation with a job candidate to a glance at a profile picture on Facebook. Despite mixed evidence for the accuracy of these social judgments (Gilbert, 1998; Todorov & Porter, 2014; Todorov, Said, Engell, & Oosterhof, 2008), understanding how people make social judgments from thin-sliced information is not only important for the theoretical advancement in social perception, but also meaningful for applications in clinical and engineering contexts (Petrican, Todorov, & Grady, 2014; Polzehl, 2014; Slepian, Bogart, & Ambady, 2014). A data-driven approach has been advocated for its advantage in discovering patterns without strict hypothesis testing and capitalization on rich and large datasets (Adolphs, Nummenmaa, Todorov, & Haxby, 2016; Todorov, Dotsch, Wigboldus, & Said, 2011). The underlying dimensions of face evaluations and their relationship with affective dimensions have been explored with such approach (Oosterhof & Todorov, 2008, 2009; Said, Sebe, & Todorov, 2009; Todorov, Mende-Siedlecki, & Dotsch, 2013). However, to our knowledge, no study has explored the social perception dimensions of voice in a data-driven approach so far.

Previous research has used *pre-selected* social/personality traits for the exploration of voice evaluation dimensions. For example, *sociability/extraversion* and *assertiveness*/*dominance* were found to be able to summarize the judgments of 35 pre-selected personality traits for voice samples from mock-jury deliberation(Scherer, 1972). In another study, *dominance*, *likeability* and *achievement* were identified to be the three key dimensions in personality judgments based on spoken passages (Zuckerman & Driver, 1989). Recently, *valence* and *dominance* were proposed to summarize all traits in

an online study of rating a simple word "Hello" on 10 pre-selected personality traits (McAleer, Todorov, & Belin, 2014).

Data-driven methods have been proposed to be particularly well suited to tackle the often high-dimensional nature of stimulus spaces that characterize social perception (see reviews by Adolphs et al., 2016; Todorov et al., 2011). The data-driven approach is more exploratory due to its attempt to discover patterns often without strict hypothesis testing, and capitalization on rich and large datasets. Previous empirical studies that took this approach have focused on visual stimuli. For instance, a two-dimensional structure of *trustworthiness* and *dominance* was identified for face evaluation using a data-driven approach (Oosterhof & Todorov, 2008). Testing this two-dimensional model on a highly variable sample of 1000 ambient images (images that are intended to be representative of those encountered in everyday life), another study found a third dimension of *youth-attractiveness* in addition to the original two dimensions (Sutherland et al., 2013). Therefore, further empirical studies of face evaluation dimensions are also in need to check which structure the evidence will lend support to.

The current study aimed to employ a data-driven approach to understand the underlying dimensions of social perception from voices as compared to faces. In Study 1 (free-description study), we collected unconstrained descriptions that participants generated of the voices they heard and the faces they saw. The descriptions were then classified into trait categories and the most frequently used social trait categories were selected for the next rating study. In Study 2, a subsequent group of participants rated the voices or faces separately on the selected traits. The trait ratings for voices and faces from the second study were submitted to Principal Component Analysis to identify the

underlying dimensions of voice evaluation and face evaluations respectively. The resulting dimensional structures of social perception from voice and face were discussed with respect to the gender stereotype of how dominance is perceived as a more negative trait for female but a more positive trait for males (Rudman & Glick, 1999; Williams & Tiedens, 2016).

Study 1: Free-description Study

Method

Participants. Participants were 66 students (33 male; M_{age} = 19.7, SD_{age} = 2.62, *Range* = 18- 27 years old). They were native English speakers with normal hearing and received one departmental research credit or $10 for each hour of participation with informed consent (same for trait-rating study).

Material. Experiment stimuli consisted of vocal recordings and static face photos from 64 college students (one recording and one photo from one person). Each voice was reading a neutral-content narrative in a neutral tone of voice. Each face was photographed under the instruction to pose a neutral expression (medium arousal level and medium valence). For the recordings, written narratives with neutral semantic content were adapted from standardized tests of declarative memory (Randt, Brown, & Osborne, 1981). Recording was conducted inside a quite testing room (sound level below 25 dB) with a headset WH30 microphone placed ½ inch from the right corner of the participant's mouth throughout the recording process. Recordings were encoded in mono (one-channel recording) directly onto computer hard disk at 44.1 kHz sampling rate and 16-bit quantization via Praat (Boersma & Weenink, 2012). The final recordings

were between 10 to 12 seconds. Photographs were taken using a Canon PowerShot ELPH 300 digital camera attached to a tripod. The picture was cropped to include only the face area (from shoulder to above head). No glasses or visible jewelry were worn.

Procedure. After the consenting procedure, we explained and asked participants to complete an Affect Grid (Russell, Weiss, & Mendelsohn, 1989) to report their momentary affective experience in terms of valence and arousal. Participants then watched a 3-minute video that served as neutral mood induction to set participants' mood to neutral before formal experiments (Zhang, Yu, & Barrett, 2014). Participants again rated their affective feelings on another Affect Grid. Then they were asked to freely describe their first impression of a person from his/her voice or face. The key instruction was "In the experiment, a voice or a face of a person will be presented to you. Please think about what would come to your mind if you met this person for the first time in your life, and how you would describe him/her to others later. Then type in or write down EVERYTHING that comes to mind about the person." We also instructed participants to indicate when they recognized the voice or face by writing down "I know this person" and to make no further description of that person.

Because the task required participants to freely describe as much as they can about the voices they hear and the faces they see, we wanted to keep their attention focused and have enough patience to give as many descriptions as possible. Hence we constructed five versions of the E-prime experiment that participants were randomly assigned to. Each version of the experiment contained 2 sessions of face description and 2 sessions of voice description, with each session consisting of 6 faces/voices. A practice trial was offered at the start of each testing session to familiarize the participants with the

process. A researcher was present in the room during the practice trials to make sure the participants understood the task. Following the practice trial, participants received four sessions of free description tasks. During each session participant was asked to describe 6 stimuli (voices or faces). We offered optional 2-minute breaks after each session. After the experiment participants were debriefed and their demographic information was collected.

Data analysis. ***Manipulation check.*** We compared participants' affective ratings of their current mood before and after the neutral mood induction slideshow to make sure that the induction was effective and all participants were in a neutral mood before completing further tasks.

Descriptor categorization. We based our classification upon the Merriam-Webster dictionary and Thesaurus for synonyms and antonyms, and also upon our common knowledge for phrases or short sentences (e.g., "flipped out", "easily offended", and "stressed easily" were categorized into 'emotionally stable' as they described emotionally unstable conditions). Each descriptive word or phrase was placed into one of the fourteen categories identified from Oosterhof and Todorov (2008)'s study (attractive, unhappy, sociable, emotionally stable, mean, boring, aggressive, weird, intelligence, confident, caring, egotistic, responsible, trustworthy). These traits were the first set of social trait categories freely generated by participants instead of pre-selected categories thought up by researchers in most standard person-perception studies. When a particular description was unclassifiable, a new category was created by the researchers (X.Z. and L.F.B.) to accommodate the description. Based on this process, we added the following trait categories: conscientious, energetic, neuroticism, motivation, likable/pleasantness,

trusting. There were also descriptors that did not describe social evaluations, and we created the following categories to accommodate them: personal history, physical qualities, social categories (age, sex, occupation), preference, emotional state, and attitude. Some descriptions such as "average", "normal", etc., are too vague or general to be classified into just one category, so we put them in a 'vague' category. Two research assistants were trained and independently classified all the descriptions into one of above categories. A third coder (X.Z.) resolved any different opinions between the two coders. A fourth coder (L.F.B.) double-checked and the final decision was made on any difficult and vague description. Table S1 in supplementary material provided a summary of all categories.

Result

Manipulation check. T-tests were used to check and confirmed that the neutral mood induction procedure was successful in inducing neutral mood in participants. On the scale of 1 to 9, participants' average self-reported *valence ratings* changed from 6.27 (SD = 1.58) before induction to 5.06 (SD = 1.08) after induction; *arousal ratings* changed from 5.53 (SD = 1.64) before induction to 4.91 (SD = .96) after induction. A paired sample t-test showed that the difference between before-induction and after-induction was both significant for valence ($t_{Valence}(65)= 5.97$, $p < .000$) and arousal ($t_{Arousal}(65)= 3.04$, $p = .003$). Another paired sample t-test comparing after-induction ratings to neutral mood rating (5 in the scale of 1 to 9) showed that there were no significant difference ($t_{Valence}(65)= .46$, $p = .65$; $t_{Arousal}(65)= -.77$ $p = .44$). The results indicated that participants were in neutral states before continuing subsequent tasks.

Trait categorization and selection. In total, sixty-six participants freely generated 3960 descriptions for the voice they heard and 4085 descriptions for the face they saw. All descriptions were classified into 29 categories (see Table S1 in Supplementary Material for frequency count and relative proportion of each category). Categories irrelevant to the interest of current study were excluded, such as those that described physical qualities, age, gender, occupation, and etc. Based on the frequency count of descriptors within each category, we selected the most frequently used categories of social judgments for the subsequent trait-rating study. The selected traits included Sociable, Intelligent, Energetic, Boring, Mean, Emotionally Stable, Confident, Happy, Caring, Conscientious, Dominant, Trustworthy, Attractive, and Likeable. The 14 traits accounted for 57.1% of the 3960 descriptors freely generated from voice evaluation and 57.5% of the 4085 descriptors generated from face evaluation (see Table 1 for frequency count and relative proportion of each selected trait).

Study 2: Trait-rating Study

Method

Participants. Participants (raters) were 157 college students (81 male; M_{age} = 20.09, SD_{age} = 3.33, *Range* =18-36 years old). One participant who reported being uncomfortable judging people just by voice or face was excluded from further analyses. Two participants' data were not recorded fully due to technical issues and were also excluded. Further, six participants' data were excluded due to unusual fast responses in many trials; four participants were excluded due to low intra-rater reliability (see Data Analysis section for details on exclusion criteria). The final data for analyses were from

the rest of 144 participants (75 male; M_{age} = 20.10, SD_{age} = 3.19, *Range* = (18, 36) years old).

Material. Rating stimuli were 64 voices and 64 faces (same as Study 1). Four voices and four faces were used for practice. The rest 60 voices and faces were targeting stimuli. Fourteen trait-rating scales went from "1 (not at all) - 5 (moderately) - 9 (extremely)" about how much an individual trait applied to the voice or face shown.

Procedure. Each stimulus was presented twice to increase the inter-rater agreement and the reliability of judgments by reducing the measurement error for each participant. To make sure participants did not lose focus during rating, we created six versions of E-prime tasks. Participants were randomly assigned to the six versions of E-prime task. The sequence of voices and faces were randomized in each session so that the participant did not know which voice belonged to which face. Participants started with two practice sessions. Then they went through 8 sessions of trait ratings with 3 short breaks in between (one break after every two sessions). Each session consisted 10 stimuli (either voice or face) for participants to rate for 7 traits. Participant's task was to "rate each of the voices/faces based on your first impressions". Participants were encouraged to rely on their "gut feelings" when judging the stimuli.

Data analysis. ***Exclusion criteria and reliability test***. We excluded ratings that showed insufficient time for judging or lack of concentration based upon their reaction time (RTs). Six raters who had more than a third trials with RTs less than 500ms were excluded from further analyses. Then within each rater, individual trials with response time (RT) higher or lower than *M*+-2.5*SD* were excluded. Next, we computed the

correlation of the rating responses at Time 1 and Time 2 of the same stimulus, and excluded the raters with correlation (r) lower than 0.2. For reliability analysis of the trait ratings, mean ratings of the two independent ratings of the same stimulus were computed and standardized. Inter-rater reliability was computed for each of the 14 traits.

Underlying dimensions. We first averaged the two ratings for the same stimulus on the same trait from all participants. Then entered the mean ratings of each trait for the 60 voices and faces into Principal Component Analysis (PCA) to identify the underlying dimensions of voice evaluation and face evaluation. Preliminary analysis indicated gender clustering for voice stimuli, consistent with biological differences in female and male voices (e.g. higher average pitch in female voices; Titze, 1989). Thus, separate PCAs were carried out for the different stimuli gender. Although it would be interesting to examine any changes within rater gender, the number of raters within each gender was not enough to make acceptable. Therefore no further analysis by rater gender was carried out.

Result

Reliability. The ratings of the voices and faces on the 14 traits demonstrated good reliability, with alphas above 0.7 (see Table S2 in Supplementary Material). Bartlett's tests of sphericity indicated that the correlations were large enough that factor analyses were appropriate; female voice, $X^2(91) = 664.69$; female face, $p<0.001$; $X^2(91) = 540.12$, $p<0.001$; male voice, $X^2(91) = 566.94$, $p<0.001$; male face, $X^2(91) = 437.47$, $p<0.001$ (see Supplementary Material Table S3 - S6 for the correlational matrix).

Dimensions of social perception from voice. Stimulus gender impacted the dimensions of voice evaluation. Ratings of female voices on the 14 traits revealed three

principal components: *Attractiveness*, *Trustworthiness*, and *Dominance*. Ratings of male voices revealed two principal components: *Socially Engagement* and *Trustworthiness*.

Female voice. Principal Component Analysis revealed three principal components explaining 87.3% of the variance. As shown in the left three columns in Table 2, traits such as Attractive, Likable, and Sociable loaded positively on the first principal component (PC1), whereas Boring loaded negatively on PC1. Trait Conscientious, Caring, Intelligent, and Trustworthy positively loaded on the second principal component (PC2). Trait Dominance, Confident, Energetic, and interestingly, trait Mean, positively loaded on the third principal component (PC3).

To find a trait that best represents principal component, repeated PCAs were performed systematically removing individual traits as likely candidates, and correlating the new PCs to the removed personality scales. A trait is proposed as a suitable summary if it correlates strongly with one PC and weakly with the other. Trait attractive highly correlated with PC1 of all ratings excluding attractiveness ($r_s = 0.86$, $p < 0.001$) but did not significantly correlate with either PC2 ($r_s = 0.10$, *n.s.*) or PC3 ($r_s = 0.02$, *n.s.*). Therefore, the first principal component was summarized to be *attractiveness*. Similarly, the second and third dimension were summarized to be *trustworthiness* and *dominance.* We plotted the 14 traits into a three dimensional space in Figure 1, with different colors for traits that loaded together on the same component. It is worth noting that trait "Mean" loaded on *dominant* dimension with positive loadings, indicating that female voices perceived as more dominant were also rated as more mean.

Male voice. For evaluations of male voices, a two-dimensional solution explained 83.8% of the variance. The first principal component (PC1) accounted for

54.87% of the variance and the second principal component (PC2) accounted for 28.97% of the variance. As shown in the right two columns in Table 2 and Figure 2, traits such as Confident, Dominant, Energetic, and Sociable, loaded positively and highly on PC1, whereas Boring loaded negatively on PC1. For the second component (PC2) of male voice evaluation, trait Trustworthy, Conscientious, and Caring loaded positively on PC2, whereas Mean loaded negatively on PC2.

We interpreted the two principal components as *social engagement* and *trustworthiness*. Trait Dominant highly correlated with PC1 of all ratings excluding Dominant (r_s = 0.86, p <0.001), but did not correlate with PC2 (r_s = 0.26, p = 0.27, *n.s.*). Trait Sociable alos highly correlated with PC1 of all ratings excluding Sociable (r_s = 0.84, p < 0.001) but did not correlate with PC2 (r_s = 0.27, p = 0.19, *n.s.*). Noted that trait Sociable loaded together with trait Attractiveness on PC1 of female voice evaluation, but loaded together with trait Dominance on PC1 of male voice evaluation. Therefore, instead of using *dominance* to summarize PC1, *social engagement* was used for interpreting the first principal component of male voice evaluation. Similar to female voice evaluation, the second principal component of male voice evaluation was interpreted as *trustworthiness*. Trait Attractive loaded together with trait Dominant and Confident, suggesting that for social perception of males, perceived dominance and confidence in voices was perceived to be attractive. This is very different than female voice evaluation, where trait Dominant and Confident loaded together with Mean, not Attractiveness.

Dimensions of social perception from face. Consistent with previous research, ratings of the 14 social traits on faces revealed a two-dimensional structure. Similarly

across two genders, as shown in Table 3, traits such as Sociable, Energetic, Happy, Confident, Attractive, Likable and Emotionally Stable loaded positively on PC1. Trait Boring negative loaded on PC1. To be consistent with voice evaluation dimensions, we interpreted the first dimension as *social engagement*. Trait such as Trustworthy, Caring, and Conscientious loaded positively on PC2, and trait "Mean" negatively loaded on PC2. We interpreted the second dimension as *trustworthiness*.

When analyzed separately for female and male faces, the two-dimensional structures were slightly different between two genders (see Figure 3 and Figure 4 for comparison), and the difference was mainly how trait Dominant is being perceived. Female faces perceived as more dominant were also judged as more mean, less trustworthy and less caring, as trait Dominant loaded closer to Mean, and opposite to Trustworthy and Caring. However, male faces perceived as more dominant were judged as more sociable, more confident, more attractive, and less boring, as trait Dominant loaded closer to Sociable, Attractive, and Confident, and opposite to Boring. This pattern suggested that Dominance was perceived quite differently for female and male faces.

Discussion

Different dimensional structures of the social perception of voice versus face were revealed in the present study utilizing a data-driven approach. For female voices, three dimensions captured the majority of variance: *attractiveness, trustworthiness*, and *dominance*. For male voices, two dimensions summarized the majority of variance: *social engagement* and *trustworthiness*. For social perception of faces, similar two-dimensional structures were found for both gender, *social engagement* and *trustworthiness*, but a

closer look at the relationship between traits within each dimension suggested a difference in how trait "Dominance" is perceived for different genders.

Our results of voice evaluation dimensions suggested *attractiveness* to be a third dimension in the social perception of female voices, in addition to *trustworthiness* and *dominance*. One possibility is that people associate voice pitch with perceived attractiveness and dominance differently for male and female. Sexual dimorphism in voice pitch has been suggested to be molded by sexual selection in human evolution (Collins, 2000; Puts, Gaulin, & Verdolini, 2006; Puts, Hodges, Cardenas, Gaulin, & Cárdenas, 2007). Previous research found that male voice dominance is positively related to voice attractiveness, whereas female voice dominance is negatively related to voice attractiveness (Borkowska & Pawlowski, 2011). Research on voice attractiveness showed that female voices with higher pitch were perceived to be more attractive, whereas male voices with lower voice pitch were perceived to be more attractive (Collins, 2000; Puts, 2005; Zuckerman & Miyake, 1993). For instance, men's preferences for high voices in women have been confirmed in both hunter-gatherer societies (Apicella & Feinberg, 2009) and developed nations (Feinberg, DeBruine, Jones, & Perrett, 2008; Puts, Barndt, Welling, Dawood, & Burriss, 2011). On the other hand, research on the dominant judgments from voice showed that both male and female voices with lowered pitch were perceived to be more dominant than those with raised pitch (Jones, Feinberg, DeBruine, Little, & Vukovic, 2010). Higher levels of testosterone usually make men have lower voice pitch (Dabbs & Mallinger, 1999). Since higher testosterone levels might signal reproductive quality, women might have evolved the preference for lower voice pitch in men (Feinberg, Jones, Little, Burt, & Perrett, 2005). This nicely explained our

finding that trait Attractiveness did not load together with trait Dominance for female voices, but loaded together with Dominance for male voice evaluation. Hence *attractiveness* emerged as a third dimension in female voice evaluation.

In addition to the gender difference in perceiving Dominance as more or less attractive, Dominance also appeared to be judged as a more negative trait for female voices (loaded together with Mean), but a more positive trait for male voices (loaded together with Sociable and Confident). This pattern is consistent with the finding that Dominant behavior was evaluated more negatively when enacted by women versus men (Rudman & Glick, 1999). A meta-analysis of studies on the dominance judgment also revealed that dominance indeed hurts women's, relative to men's, likability (Williams & Larissa, 2016).

The dimensions of face evaluations found in the present study are consistent with the two-dimensional structure of 'valence/trustworthiness/warmth' and 'dominance/competence' proposed in prior research (see reviews by Fiske, Cuddy, & Glick, 2007; Todorov et al., 2008). Specifically, a review of social cognition suggested that perceived warmth and perceived competence were two universal dimensions (Fiske et al., 2007). As mentioned above, also using a data-driven approach, the underlying dimensions of face evaluation were suggested to be trustworthiness and dominance (Oosterhof & Todorov, 2008). Although 'youthful- attractiveness' was suggested to be a third dimension in a study that utilized more ecologically valid stimuli from everyday life (Sutherland et al., 2013), our results did not support the additional dimension of attractiveness in face evaluation.

However, our separate analyses by gender on face evaluation revealed differences in the detailed structures. A recent study found that by averaging the 20 highest rated trustworthy faces among 1000 faces from the internet, the computer generated a female face, whereas averaging the 20 most dominant faces among 1000 faces generated a male face (Sutherland et al., 2013). In other words, the most trustworthy faces were mostly from females, yet the most dominant faces are mostly from males. Moreover, the trustworthiness and dominance continua appeared to change in gender: the averaged low dominance face is female looking, and the averaged high dominance face is male; the averaged low trustworthy face is male looking, yet the averaged high trustworthy face is female looking. This indicates a gender difference in the perception of trustworthiness and dominance from face. It is possible that people in general rate female to be more trustworthy (less dominant) and male less trustworthy (more dominant). It is also possible that people rate trustworthy and dominance on different scales for different genders. Our results supported the third possibility: dominance in women is perceived differently than dominance in men.

To our knowledge, the current study was the first to explore the underlying dimensions of social evaluation of both voices and faces from a data-driven approach. This approach capitalized on the rich dataset and provided more exploration without strict hypothesis. The thousands of descriptions generated by participants in the present study reflected what traits the current college students were frequently using to evaluate others. The gender differences in the dimensional structures were also for the first time clearly revealed from the data-driven approach. Another advantage of the current study was that we induced a neutral state (a medium level of arousal and valence experience) in

participants before they rated any stimulus. Previous studies found that the affective state of the perceivers influenced their social perception (e.g., Anderson, Siegel, & Barrett, 2011; Bliss-Moreau, Owren, & Barrett, 2010). Therefore it is important to control for perceivers' affective state for social perception tasks.

Further research is needed to explore the acoustic parameters correlated to each dimension. For example, a two-dimensional 'social voice space' suggested by earlier study found that each dimension was driven by differing combinations of vocal acoustics (McAleer et al., 2014). Moreover, the relationship between acoustical parameters to social perception of voice also depends on the contexts. When a context makes a specific evaluative dimension relevant (e.g, competence), perceptual judgments and decisions would be most likely influenced by evaluations on this dimension. As proposed by The Conceptual Act Theory, each mental category is populated with a set of variable instances, where the variation is meaningfully tied to the situation (Barrett, Wilson-Mendenhall, & Barsalou, 2014; Barrett, 2014). Situated conceptualization is an act of categorization, during which the utilized conceptual knowledge is tied to the situation and prepares a perceiver for situated action (Barrett, 2006, 2012, 2013; Barrett et al., 2014). Therefore more research into the context effect on social perception is needed for a deeper and fuller understanding of social perception.

Reference

Adolphs, R., Nummenmaa, L., Todorov, A., & Haxby, J. (2016). Data-driven approaches in the investigation of social perception. *Philosophical Transactions of the Royal Society B*.

Apicella, C. L., & Feinberg, D. R. (2009). Voice pitch alters mate-choice-relevant perception in hunter − gatherers Voice pitch alters mate-choice-relevant perception in hunter – gatherers. *Proceedings of the Royal Society B, Biological Sciences*, *276*(March), 1077–1082. http://doi.org/10.1098/rspb.2008.1542

Barrett, L. F. (2006). Solving the emotion paradox: categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*(1), 20–46. http://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F. (2012). Emotions are real. *Emotion*, *12*(3), 413–429. http://doi.org/10.1037/a0027555

Barrett, L. F. (2013). Psychological Construction: The Darwinian approach to the science of emotion. *Emotion Review*, *5*(4), 379–389. http://doi.org/10.1177/1754073913489753

Barrett, L. F. (2014). The Conceptual Act Theory: A Précis. *Emotion Review*, 1–20. http://doi.org/10.1177/1754073914534479

Barrett, L. F., Wilson-Mendenhall, C., & Barsalou, L. (2014). The conceptual act theory: A roadmap. In L. F. Barrett & J. Russell (Eds.), *The Psychological Construction of Emotion* (pp. 83–110). New York: Guilford.

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. *Glot International*, *5*(9/10), 341–345.

Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, *82*(1), 55–59. http://doi.org/10.1016/j.anbehav.2011.03.024

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour*, *60*, 773–780.

Dabbs, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitch among men. *Personality and Individual Differences*, *27*, 801–804.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2008). The relative role of femininity and averageness of voice pitch in aesthetic judgments of women's voices. *Perception*, (37), 615–623.

Feinberg, D. R., Jones, B. C., Little, a. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, *69*(3), 561–568. http://doi.org/10.1016/j.anbehav.2004.06.012

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. http://doi.org/10.1016/j.tics.2006.11.005

Gilbert, D. T. (1998). Ordinary personology. In *The Handbook of Social Psychology* (pp. 89–150).

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, *79*, 57–62.

McAleer, P., Todorov, A., & Belin, P. (2014). How Do You Say "Hello"? Personality Impressions from Brief Novel Voices. *PLoS ONE*, *9*(3), e90779. http://doi.org/10.1371/journal.pone.0090779

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092. http://doi.org/10.1073/pnas.0805664105

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, *9*(1), 128–133. http://doi.org/10.1037/a0014520

Petrican, R., Todorov, A., & Grady, C. (2014). Personality at face value: Facial appearance predicts self and other personality judgments among strangers and spouses. *Journal of Nonverbal Behavior*, *38*(2), 259–277. http://doi.org/10.1007/s10919-014-0175-3

Polzehl, T. (2014). *Personality in Speech: Assessment and Automatic Classification*. Springer.

Puts, D. A. (2005). Mating context and menstrual phase affect women's preferences for male voice pitch. *Evolution and Human Behavior*, *26*, 388–397.

Puts, D. A., Barndt, J. L., Welling, L. L. M., Dawood, K., & Burriss, R. P. (2011). Intrasexual competition among women: vocal femininity affects perceptions of attractiveness and flirtatiousness. *Personality and Individual Differences*, *50*, 111–115.

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, *27*(4), 283–296. http://doi.org/10.1016/j.evolhumbehav.2005.11.003

Puts, D. A., Hodges, C. R., Cardenas, R., Gaulin, S. J. C., & Cárdenas, R. a. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, *28*(5), 340–344. http://doi.org/10.1016/j.evolhumbehav.2007.05.002

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward Agentic women : The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, *77*(5), 1004–1010.

Russell, J. a, Weiss, A., & Mendelsohn, G. a. (1989). Affect Grid: A Single-Item Scale of Pleasure and Arousal. *Journal of Personality and Social Psychology*, *57*(3), 493–502. http://doi.org/10.1037/0022-3514.57.3.493

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, *9*(2), 260–264. http://doi.org/10.1037/a0014681

Scherer, K. R. (1972). Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. *Journal of Personality*, *40*(2), 191–210. http://doi.org/10.1111/j.1467-6494.1972.tb00998.x

Slepian, M. L., Bogart, K. R., & Ambady, N. (2014). Thin-slice judgments in the clinical context. *Annual Review of Clinical Psychology*, *10*, 131–53. http://doi.org/10.1146/annurev-clinpsy-090413-123522

Sutherland, C. a M., Oldmeadow, J. a., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118. http://doi.org/10.1016/j.cognition.2012.12.001

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *Journal of Acoustical Society of America*, *85*(4), 1699–707.

Todorov, A., Dotsch, R., Wigboldus, D. H. J., & Said, C. P. (2011). Data-driven methods for modeling social perception. *Social and Personality Psychology Compass*, *5*(10), 775–791. http://doi.org/10.1111/j.1751-9004.2011.00389.x

Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current Opinion in Neurobiology*, *23*(3), 373–380. http://doi.org/10.1016/j.conb.2012.12.010

Todorov, A., & Porter, J. M. (2014). Misleading first impressions: Different for different facial images of the same person. *Psychological Science*, *25*(7), 1404–1417. http://doi.org/10.1177/0956797614532474

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. http://doi.org/10.1016/j.tics.2008.10.001

Williams, M. J., & Tiedens, L. Z. (2016). The subtle suspension of backlash: A meta-analysis of penalties for women's implicit and explicit dominance behavior. *Psychological Bulletin*, *142*(2), 165–97. http://doi.org/10.1037/bul0000039

Zhang, X., Yu, H. W., & Barrett, L. F. (2014). How does this make you feel? A comparison of four affect induction procedures. *Frontiers in Psychology*, *5*(1975), 1–10. http://doi.org/10.3389/fpsyg.2014.00689

Zuckerman, M., & Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior*, *13*(2), 67–82. http://doi.org/10.1007/BF00990791

Zuckerman, M., & Miyake, K. (1993). The attractive voice: What makes it so? *Journal of Nonverbal Behavior*, *17*(2), 119–135. http://doi.org/10.1007/bf01001960

Tables

Table 1

Frequency, percentage and ranking of the trait categories in Study 1 (unconstrained person descriptions of emotionally neutral voices and faces), and the respective means and standard deviations of trait judgments in Study 2.

	Voice				Face			
Traits	Frequency count	Relative proportion	Ranking	Rating Mean (*SD*)	Frequency count	Relative proportion	Ranking	Rating Mean (*SD*)
Sociable	425	0.107	1	5.68(0.86)	511	0.125	1	5.42(0.84)
Intelligent	305	0.077	2	5.93(0.60)	222	0.054	5	5.55(0.62)
Energetic	240	0.061	3	5.22(1.03)	189	0.046	6	5.11(0.85)
Boring	192	0.048	4	4.58(0.82)	143	0.035	7	4.62(0.61)
Mean	185	0.047	5	3.17(0.53)	258	0.063	3	3.99(0.78)
EmoStable	177	0.045	6	5.76(0.65)	229	0.056	4	4.94(0.63)
Confident	174	0.044	7	5.74(1.08)	128	0.031	8	5.41(0.84)
Happy	159	0.04	8	5.44(0.84)	260	0.064	2	4.91(0.81)
Caring	96	0.024	9	5.54(0.58)	101	0.025	9	5.23(0.63)
Conscientious	92	0.023	10	5.49(0.53)	67	0.016	10	5.18(0.56)
Dominant	70	0.018	11	4.93(0.93)	67	0.016	11	5.11(0.94)
Trustworthy	56	0.014	12	5.60(0.47)	65	0.016	12	5.07(0.63)
Attractive	53	0.013	13	5.28(0.77)	57	0.014	13	4.33(1.01)
Likeable	38	0.01	14	5.70(0.67)	53	0.013	14	5.15(0.71)

Table 2

Loadings of trait judgments of female voices on the three principal components

Trait	Female Voice			Male Voice	
	PC 1	PC 2	PC3	PC1	PC2
Attractive	**0.88**	-0.04	0.06	**0.79**	0.43
Likable	**0.85**	0.41	0.18	**0.75**	0.61
Sociable	**0.77**	0.21	0.55	**0.89**	0.37
Boring	**-0.73**	-0.35	-0.43	**-0.89**	-0.37
Happy	**0.67**	0.47	0.49	**0.85**	0.47
EmoStable	0.65	0.24	0.65	**0.80**	0.40
Conscientious	0.07	**0.89**	-0.09	0.31	**0.74**
Intelligent	0.02	**0.81**	0.39	0.54	0.51
Caring	0.52	**0.78**	-0.03	0.49	**0.73**
Trustworthy	0.48	**0.74**	0.19	0.46	**0.82**
Dominant	0.39	0.09	**0.87**	**0.96**	-0.14
Mean	-0.14	-0.48	**0.77**	0.17	**-0.83**
Confident	0.46	0.41	**0.76**	**0.96**	0.15
Energetic	0.60	0.41	**0.61**	0.92	0.29
Explained variance	33.9%	27.1%	26.3%	54.87%	28.96%
Total explained variance	87.3%			83.83%	

Table 3

Loadings of trait judgments of female and male faces on the first two principal components

Traits	Female Face		Male Face	
	PC 1	PC 2	PC 1	PC 2
Sociable	**0.94**	-0.07	**0.95**	0.04
Energetic	**0.93**	0.14	**0.90**	0.14
Happy	**0.87**	0.34	**0.82**	0.31
Boring	**-0.87**	-0.03	**-0.93**	0.01
Confident	**0.86**	-0.23	**0.95**	-0.08
Attractive	**0.82**	-0.12	**0.81**	0.14
Likable	**0.76**	0.55	**0.76**	0.53
EmoStable	0.66	0.33	**0.91**	0.24
Trustworthy	0.14	**0.88**	0.21	**0.89**
Mean	0.08	**-0.88**	0.17	**-0.83**
Caring	0.33	**0.85**	0.30	**0.89**
Intelligent	-0.13	**0.77**	0.09	**0.90**
Conscientious	0.36	**0.74**	-0.07	**0.75**
Dominant	0.51	**-0.71**	0.69	**-0.60**
Explained variance	44.51%	32.57%	48.92%	31.02%
Total explained variance	77.08%		79.94%	

Figures

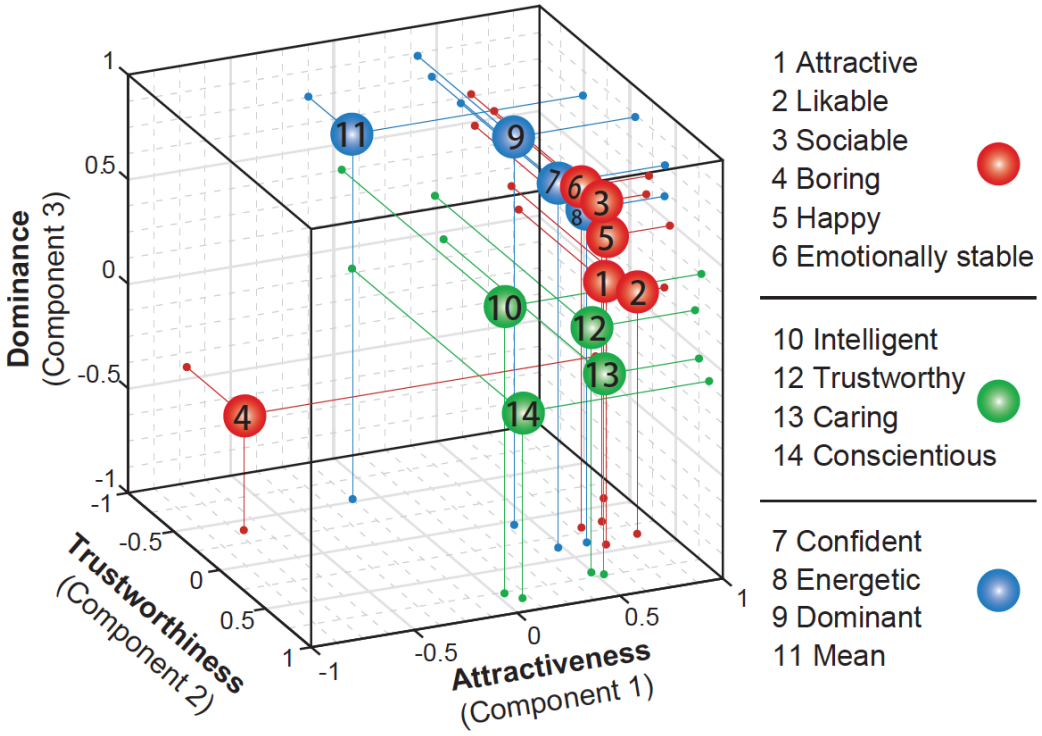

Figure 1. The structure of female voice evaluation. The first component could be interpreted as Attractiveness evaluation. The second component could be interpreted as Trustworthiness evaluation. The third component could be interpreted as Dominance evaluation.

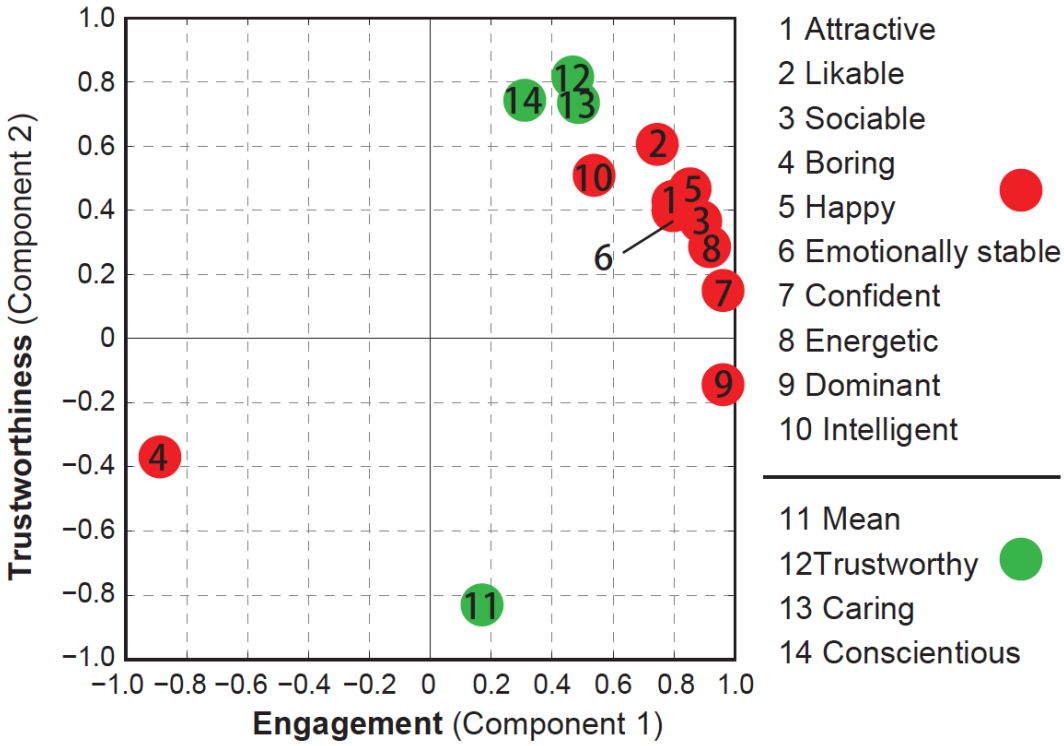

Figure 2. The structure of male voice evaluation. The first component could be interpreted as the evaluation of Socially Engaged-Boring, and the second component could be interpreted as the evaluation of Socially Nice-Mean evaluation.

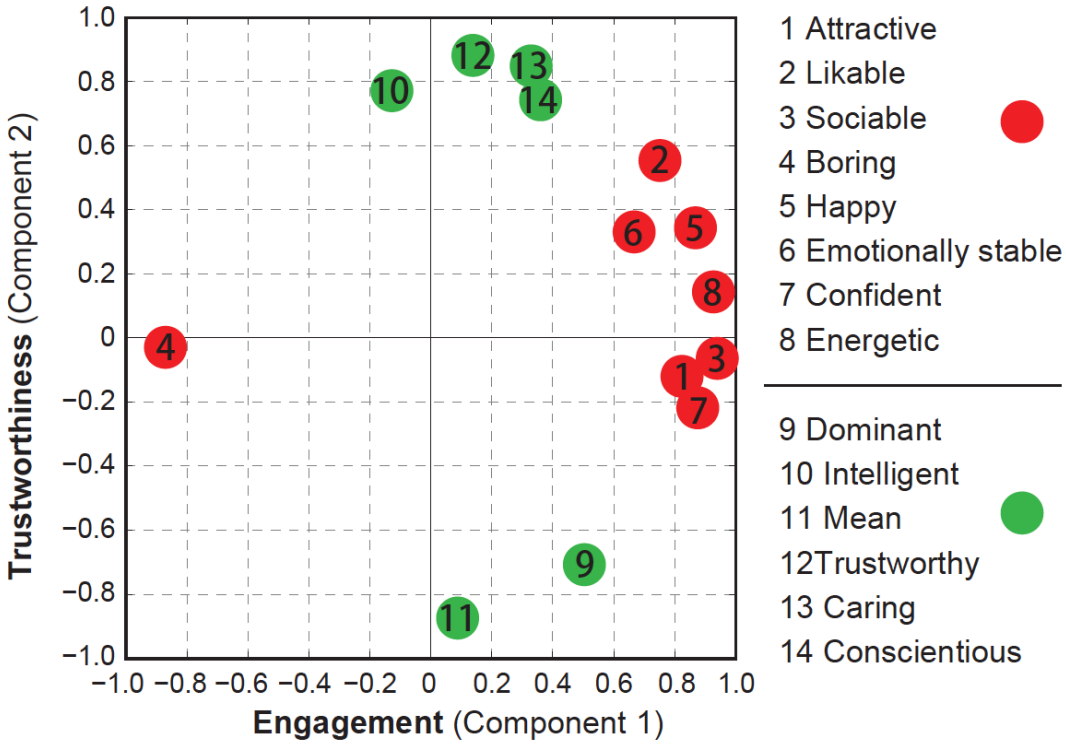

Figure 3. The structure of female face evaluation. The first component could be interpreted as Socially Engaged-Boring evaluation, and the second component could be interpreted as Socially Nice-Mean evaluation.

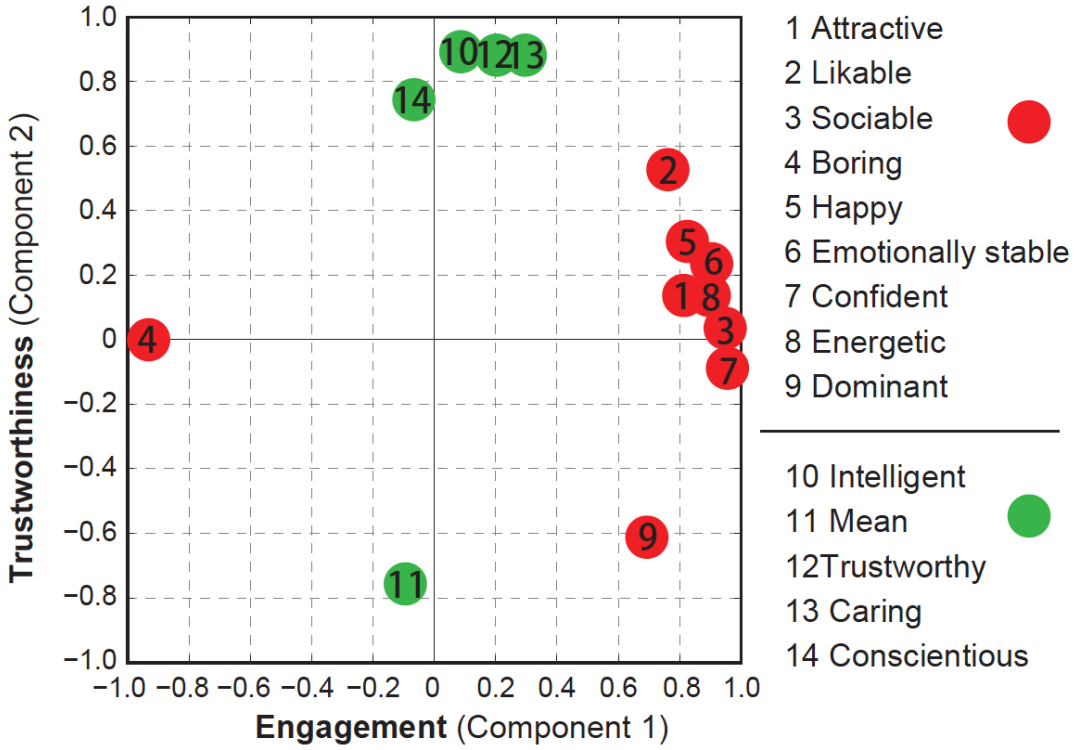

Figure 4. The structure of male face evaluation. The first component could be interpreted as Socially Engaged-Boring evaluation, and the second component could be interpreted as Nice-Mean evaluation.

Supplementary Materials

Table S1

The occurrence frequency and proportion of the descriptions in all 29 categories in the free-description study

	Voice		Face	
	Number	Proportion	Number	Proportion
Sociable	425	0.107	511	0.125
Physical qualities [a]	339	0.086	363	0.089
Social categories [b]	313	0.079	246	0.06
Intelligent	305	0.077	222	0.053
Energetic	240	0.061	189	0.046
Vague	241	0.061	147	0.036
Actions	210	0.053	200	0.049
Boring	192	0.048	143	0.035
Mean	185	0.047	258	0.063
Attitudes	178	0.045	223	0.055
Emotionally stable	177	0.045	229	0.056
Confident	174	0.044	128	0.031
Unhappy	159	0.04	260	0.064
Caring	96	0.024	101	0.025
Emotional states [c]	92	0.023	146	0.036
Conscientious	92	0.023	67	0.016
Preferences [d]	85	0.021	114	0.028
Dominance	70	0.018	67	0.016
Motivation	61	0.015	51	0.012
Egotistic	60	0.015	87	0.021
Trustworthy	56	0.014	65	0.016
Attractive	53	0.013	57	0.014
Likeable	38	0.01	53	0.013
Weird	35	0.009	45	0.011
Aggressive	23	0.006	44	0.011
Responsible	23	0.006	20	0.005
Trusting	17	0.004	28	0.007
Personal History	16	0.004	21	0.005
Neuroticism	5	0	0	0
Total	3960	1	4085	1

Note. [a] Examples of "physical qualities": hair color, height. [b] Examples of "social categories": age, gender, occupation. [c] Examples of "emotional states": angry, sad. [d] Examples of "preferences": like to shop, enjoys music

Table S2

Inter-rater agreement (r) and reliability (Cronbach's alpha) for judgments of voices and faces on 14 social traits

	Voice (Voice1-20/Voice21-40/Voice41-60)		Face (Face1-20/Face21-40/Face41-60)		Sample size
	Inter-rater agreement (*r*)	Cronbach alpha	Interrater agreement (*r*)	Cronbach alpha	
Energetic	0.46/0.47/0.47	0.95/0.95/0.95	0.32/0.25/0.19	0.92/0.89/0.84	24/22/24
Attractive	0.23/0.28/0.29	0.88/0.88/0.89	0.32/0.38/0.35	0.92/0.93/0.92	24/22/24
Boring	0.29/0.28/0.20	0.90/0.88/0.86	0.18/0.38/0.13	0.84/0.93/0.76	24/22/24
Caring	0.15/0.33/0.25	0.78/0.90/0.88	0.24/0.12/0.23	0.89/0.70/0.86	24/22/24
Confident	0.46/0.53/0.49	0.95/0.97/0.95	0.33/0.33/0.22	0.92/0.92/0.86	25/25/24
Conscientious	0.10/0.17/0.09	0.73/0.85/0.70	0.15/0.13/0.16	0.78/0.78/0.80	25/25/24
Dominant	0.34/0.36/0.39	0.92/0.92/0.93	0.26/0.32/0.30	0.89/0.91/0.90	24/22/24
EmoStable	0.27/0.30/0.36	0.89/0.90/0.91	0.33/0.17/0.09	0.92/0.80/0.70	24/22/24
Happy	0.38/0.44/0.44	0.93/0.95/0.95	0.35/0.29/0.29	0.93/0.91/0.89	25/25/24
Intelligent	0.14/0.40/0.12	0.82/0.94/0.75	0.19/0.19/0.24	0.86/0.84/0.86	25/25/24
Likable	0.23/0.35/0.33	0.87/0.92/0.91	0.29/0.17/0.23	0.91/0.83/0.84	25/25/24
Mean	0.22/0.08/0.19	0.89/0.70/0.83	0.24/0.19/0.29	0.87/0.86/0.88	25/25/24
Sociable	0.36/0.36/0.43	0.93/0.93/0.94	0.30/0.41/0.27	0.91/0.94/0.87	25/25/24
Trustworthy	0.14/0.19/0.13	0.80/0.82/0.76	0.22/0.15/0.18	0.87/0.77/0.85	24/22/24

Table S3

Intercorrelations between ratings of female voices

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Energetic	1.00												
2. Attractive	0.50	1.00											
3. Boring	-0.86	-0.71	1.00										
4. Caring	0.64	0.42	-0.63	1.00									
5. Confident	0.90	0.43	-0.78	0.54	1.00								
6. Conscientious	0.32	0.11	-0.34	0.72	0.34	1.00							
7. Dominant	0.83	0.40	-0.69	0.23	0.88	0.04	1.00						
8. EmoStable	0.83	0.60	-0.79	0.47	0.89	0.21	0.83	1.00					
9. Happy	0.93	0.52	-0.80	0.71	0.88	0.38	0.73	0.86	1.00				
10. Intelligent	0.55	0.12	-0.51	0.58	0.63	0.63	0.40	0.48	0.55	1.00			
11. Likable	0.78	0.67	-0.77	0.71	0.70	0.38	0.52	0.80	0.88	0.42	1.00		
12. Mean	0.15	0.03	-0.08	-0.40	0.31	-0.44	0.52	0.28	0.02	-0.13	-0.24	1.00	
13. Sociable	0.87	0.63	-0.83	0.53	0.88	0.20	0.77	0.91	0.90	0.36	0.86	0.20	1.00
14. Trustworthy	0.67	0.38	-0.66	0.87	0.65	0.64	0.41	0.62	0.73	0.61	0.72	-0.21	0.62

Table S4

Intercorrelations between ratings of female faces.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Energetic	1.00												
2. Attractive	0.69	1.00											
3. Boring	-0.82	-0.72	1.00										
4. Caring	0.47	0.14	-0.34	1.00									
5. Confident	0.73	0.64	-0.68	0.04	1.00								
6. Conscientious	0.44	0.20	-0.34	0.74	0.17	1.00							
7. Dominant	0.37	0.52	-0.42	-0.46	0.51	-0.36	1.00						
8. EmoStable	0.66	0.42	-0.45	0.44	0.55	0.36	0.19	1.00					
9. Happy	0.82	0.62	-0.72	0.56	0.70	0.49	0.17	0.86	1.00				
10. Intelligent	-0.10	-0.12	0.13	0.44	-0.25	0.57	-0.45	0.31	0.15	1.00			
11. Likable	0.73	0.61	-0.66	0.69	0.49	0.61	0.02	0.63	0.90	0.42	1.00		
12. Mean	-0.04	0.14	0.00	-0.71	0.36	-0.51	0.62	-0.13	-0.30	-0.66	-0.52	1.00	
13. Sociable	0.87	0.74	-0.77	0.28	0.85	0.36	0.44	0.52	0.82	-0.24	0.66	0.15	1.00
14. Trustworthy	0.30	-0.01	-0.16	0.85	-0.05	0.65	-0.58	0.47	0.35	0.58	0.53	-0.66	0.05

Table S5

Intercorrelations between ratings of male voices.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Energetic	1.00												
2. Attractive	0.85	1.00											
3. Boring	-0.96	-0.90	1.00										
4. Caring	0.69	0.79	-0.74	1.00									
5. Confident	0.93	0.77	-0.88	0.51	1.00								
6. Conscientious	0.51	0.48	-0.51	0.59	0.51	1.00							
7. Dominant	0.83	0.70	-0.80	0.34	0.90	0.19	1.00						
8. EmoStable	0.79	0.80	-0.86	0.64	0.79	0.42	0.72	1.00					
9. Happy	0.94	0.83	-0.94	0.73	0.90	0.60	0.73	0.87	1.00				
10. Intelligent	0.61	0.45	-0.60	0.63	0.66	0.70	0.45	0.61	0.71	1.00			
11. Likable	0.84	0.82	-0.87	0.72	0.81	0.66	0.64	0.87	0.95	0.68	1.00		
12. Mean	-0.09	-0.21	0.19	-0.41	0.05	-0.44	0.22	-0.29	-0.29	-0.17	-0.47	1.00	
13. Sociable	0.92	0.88	-0.90	0.64	0.92	0.56	0.80	0.82	0.93	0.58	0.93	-0.22	1.00
14. Trustworthy	0.64	0.81	-0.71	0.92	0.54	0.70	0.30	0.69	0.73	0.62	0.79	-0.50	0.69

Table S6

Intercorrelations between ratings of male faces.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Energetic	1.00												
2. Attractive	0.67	1.00											
3. Boring	-0.86	-0.77	1.00										
4. Caring	0.35	0.45	-0.28	1.00									
5. Confident	0.82	0.72	-0.84	0.21	1.00								
6. Conscientious	0.09	0.05	0.15	0.62	-0.05	1.00							
7. Dominant	0.50	0.52	-0.58	-0.35	0.72	-0.37	1.00						
8. EmoStable	0.89	0.79	-0.84	0.49	0.81	0.04	0.45	1.00					
9. Happy	0.79	0.52	-0.71	0.44	0.78	0.14	0.37	0.80	1.00				
10. Intelligent	0.23	0.26	-0.07	0.76	0.03	0.71	-0.43	0.30	0.32	1.00			
11. Likable	0.72	0.64	-0.71	0.65	0.66	0.27	0.18	0.81	0.84	0.49	1.00		
12. Mean	-0.17	-0.06	0.15	-0.57	0.04	-0.29	0.44	-0.26	-0.47	-0.65	-0.60	1.00	
13. Sociable	0.79	0.76	-0.86	0.34	0.97	0.03	0.64	0.79	0.79	0.10	0.74	-0.05	1.00
14. Trustworthy	0.26	0.33	-0.15	0.94	0.13	0.65	-0.36	0.41	0.40	0.74	0.58	-0.53	0.26