



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

## Journal of Experimental Social Psychology

journal homepage: [www.elsevier.com/locate/jesp](http://www.elsevier.com/locate/jesp)

## Report

## I like the sound of your voice: Affective learning about vocal signals

Eliza Bliss-Moreau<sup>a,\*</sup>, Michael J. Owren<sup>c</sup>, Lisa Feldman Barrett<sup>b</sup><sup>a</sup> Department of Psychiatry and Behavioral Sciences, California National Primate Research Center, University of California, Davis, United States<sup>b</sup> Boston College and Psychiatric Neuroimaging Research Program/Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, United States<sup>c</sup> Department of Psychology, Georgia State University, United States

## ARTICLE INFO

## Article history:

Received 11 November 2008

Revised 2 November 2009

Available online 6 January 2010

## Keywords:

Affect

Value

Learning

Vocalization

## ABSTRACT

This paper provides the first demonstration that the content of a talker's speech is sufficient to imbue the acoustics of his voice with affective meaning. In two studies, participants listened to male talkers utter positive, negative, or neutral words. Next, participants completed a sequential evaluative priming task where a neutral word spoken by one of talkers was presented before each target word to be evaluated. We predicted, and found, that voices served as evaluative primes influencing the speed with which participants evaluated the target words. These two experiments demonstrate that the human voice can take on affective meaning merely based on the positive or negative value words it has spoken. Implications for affective processing, the pragmatics of communication, and person perception are discussed.

© 2010 Elsevier Inc. All rights reserved.

## Introduction

Imagine picking up the phone and hearing the warm, soothing, tone of a loved one's voice. The mere sound can make you feel pleasant and happy. Now imagine hearing the voice of a disliked political figure on the radio. Its sound can make you feel so unpleasant that you turn off the radio, even if you want to be informed about the latest political news. Despite the fact that humans are born with the innate capacity to have these sorts of pleasant and unpleasant reactions to objects and people in our environments, we must, for the most part, learn whether another person is friend or foe. In other words, people's voices, like their faces and other defining personal attributes, must acquire affective value in order to have the capacity to influence our internal states. Once voices have acquired affective value, simply hearing those voices is likely sufficient to influence our perception of and interaction with people. Understanding how learning about others proceeds is critical to understanding the mechanisms of person perception. In the present report we provide the first evidence that human vocal acoustics can acquire affective value based on the affectively valued content of speech.

People can quickly identify others by the sounds of their voices. Human voices are richly endowed with acoustic cues to identity, along with other markers for biological sex and age (e.g., Bachorowski & Owren, 1995; Ladefoged & Broadbent, 1957; Owren & Bachorowski, 2003; Owren & Rendall, 1997). Much of this informa-

tion is carried by vowel or vowel-like sounds (Bachorowski & Owren, 1999; Owren & Cardillo, 2006; also see Dudley, 1939; Traunmüller, 1994). According to the affect-conditioning (Owren & Rendall, 1997) or affect-induction model (Owren, Rendall, & Bachorowski, 2005), identity-specific vocal signals acquire meaning for the person hearing the signal (i.e., the perceiver) by influencing the perceiver's affective state. In non-human primates, a distress call from a conspecific does not intrinsically signal or represent threat, but takes on negative meaning (e.g., the presence of a threat) because the acoustics of call induce a negative affective state in the perceiver. A similar process is presumed to be at work in humans: a human laugh is pleasant because the sound of laughter induces a positive affective state in the perceiver. Similarly, a human shout or yell derives its value as threatening, at least in part, because the acoustics of vocalization induce a negative affective state in the perceiver. In humans, then, it should be possible for a talker's voice to take on negative meaning when it is paired with a negative affective state in a perceiver (a listener).

In this report, we present the first evidence that human voices can acquire affective meaning based on the content spoken by those voices. Despite the large literature on affective learning in humans (for reviews, see De Houwer, Thomas & Baeyens, 2001; Delgado, Olsson, & Phelps, 2006), never before have talkers' unique vocal acoustics served as conditioned stimuli, while the content of their speech (i.e., the words spoken) has served as an unconditioned stimulus. In the present studies, participants listened to male talkers utter positive, negative, or neutral words in a neutral tone. We reasoned that listening to a talker utter a series of valenced words should be sufficient to induce a similarly valenced affective state in the listener (e.g., hearing a positive word, such as "lucky", would be enough to induce a mild positive change in the

\* Corresponding author. Address: Department of Psychiatry and Behavioral Sciences, California National Primate Research Center, University of California, Davis, One Shields Avenue, Davis, CA 95616, United States.

E-mail address: [eblissmoreau@ucdavis.edu](mailto:eblissmoreau@ucdavis.edu) (E. Bliss-Moreau).

affective state of the perceiver). As the talker utters many valenced words, the acoustics of his or her voice will become associated with that affective state. Over a number of repetitions, the voice should come to acquire positive (or negative) value, such that subsequent vocal signals from that talker will induce a positive (or negative) affective state in the listener. In this way, the talker's vocal acoustics come to acquire affective value, such that his or her voice will be experienced as pleasant (or unpleasant) on future occasions. Importantly, once a person's unique voice has acquired affective value its vocal acoustics will be valued even if the semantic content uttered by the person is devoid of affective meaning (i.e., neutral words).

To assess affective learning for voices, we examined the extent to which talkers' voices served as affective primes in a sequential evaluative priming paradigm. In this procedure, participants categorize a series of target stimuli (e.g., words) as being positive or negative. Prior to the presentation of each target word, another stimulus (the prime) is quickly presented. Valenced, as opposed to neutral, primes influence the speed with which targets are evaluated. Two patterns of response latencies are typically observed—assimilation effects and reverse priming effects. Most typically, primes that match the target in valence (e.g., positive prime, positive target) speed the target judgment, while primes that do not match the target in valence (e.g., positive prime, negative target) slow the target judgment (e.g., Bargh, Chaiken, Gollwitzer, & Pratto, 1992; Fazio, Sanbonmatsu, Powell, & Kardes 1986). These are called assimilation priming effects. In some cases, however, reverse priming effects emerge such that primes that match the target in valence (e.g., positive prime, positive target) slow the target judgment, while primes that do not match the target in valence (e.g., positive prime, negative target) speed the target judgment (for a review and discussion, Glaser, 2003; Klauer, Teige-Mocigemba & Spruyt, 2009). Reverse priming effects are thought to emerge when the influence of the prime falls outside the window during which the target is evaluated but is close enough in temporal proximity that it is able to engender an evaluative stance (Klauer et al., 2009). Given that the time between prime and target onset was necessarily long in this experiment, we thought it was possible that we might observe reverse rather than assimilation priming effects. Whether assimilation or reverse priming effects emerged, the degree to which the prime stimulus served to influence the speed of target judgments essentially indexes the prime's affective value (as positive or negative). Using this logic, we reasoned that if talkers' voices acquired affective value because they spoke evocative words at an earlier point in time, then the mere sound of their voices (when later uttering neutral words) would serve to change the speed of target evaluation during a subsequent evaluative priming task.

We present two experiments to test the hypothesis that human voices can acquire affective meaning based on the content of the words spoken by those voices. Target talkers uttered positive, negative, or neutral words during an initial learning phase. Following learning, the vocal acoustics present in neutral words uttered by the same target talkers modulated the speed with which participants evaluated the valence of positive and negative targets during a sequential evaluative priming procedure. In essence, the acoustics of a human voice acquired affective potency by virtue of the words it uttered.

## Experiment 1

### Method

#### Participants

Participants were 39 Boston College students (19 male, 20 female) who completed the study for either course credit or \$15.

Two additional participants were excluded because they did not comply with the task instructions insofar as they responded randomly during the sequential evaluative priming phase (e.g., pressing one key for most of or the entire test phase).

### Materials and procedure

Participants completed the *affective learning task*, which was implemented in E-Prime Version 1 on a Dell Pentium VI PC using a 17" CRT monitor. Participants wore Sennheiser headphones for the duration of the task and the volume of the auditory stimuli was maintained at an equal level for all participants. During the *learning phase*, participants heard a total of nine European-American male talkers. Each talker uttered only positive, negative, or neutral nouns and adjectives (i.e., three talkers uttered only positive words, three talkers uttered only negative words, and three talkers uttered only neutral words; see Appendix A, Table A1 for words). Spoken words were selected from the Affective Norms for English Words database (ANEW; Bradley & Lang, 1999) based on their valence and were matched on their degree of arousal. Each talker uttered 10 words. All talkers were native speakers of American English. The stimuli were recorded individually in a double-walled sound-deadened room using a Special Projects headworn microphone, and a Tascam DA-P1 digital audio recorder. An experimenter was present during each recording session, coaching the talkers as needed in speaking clearly at an appropriate amplitude level. Talkers were instructed to speak in a neutral tone and the tone of the recorded words was assessed to ensure neutrality by one of the co-authors (MJO) who has expertise in assessing vocal acoustics. Recordings were re-digitized at 44.1 kHz and 16-bit wordwidth using Praat acoustics software (Boersma, 2001), amplitude-scaled to the full representational range, and edited so as to include 100 ms silence at the beginning and end of each file. Talker identity and word valence pairings were counterbalanced across participants. Each spoken word was approximately 1000 ms long with a 1000 ms inter-trial interval between words. Each word was played one time and words were played in a random order. Participants were instructed to listen carefully to the words that were spoken and told that later in the experiment they would be asked to select the words they had heard from a list that also included words they had not heard.

Immediately following the learning phase, participants completed a *sequential evaluative priming phase*. Participants were instructed that they would be judge words presented on the computer screen as being positive or negative. They were told to use two keys (i.e., the 'p' and 'q' on a standard QWERT keyboard) which had been marked with '+' and '-' signs to indicate whether the target words were positive or negative. Response keys were counterbalanced across participants such that for half of participants the 'p' key was used to indicate positive targets and for half of participants the 'q' key was used to indicate negative targets. Participants were also told to keep their fingers on the response keys at all times and to make the most accurate judgments possible, while responding as quickly as possible. Finally, participants were instructed that they would hear a distracter word immediately before a target word, and to ignore the distracter while focusing on categorizing the target word.

The primes were neutral words selected from the ANEW database (Bradley & Lang, 1999) and spoken by the same nine target talkers included in the learning phase (see Appendix A, Table A1, for words). Each prime was played in its entirety (up to 1000 ms), followed by an inter-stimulus interval of 100 ms during which a fixation cross was displayed. Following the inter-stimulus interval, the target word appeared on the computer screen. Targets were selected from the word list in Bargh et al. (1992) and were moderately positive and negative target words for which there was a high degree of consensus (60% or more of respondents

agreed on the valence). These words appeared on the screen until the participant responded, or for a maximum of 3000 ms. If the participant did not register a response within the 3000 ms window, the trial was counted as an error and the program moved on to the next trial. The inter-trial interval was 4000 ms.

The evaluative priming procedure consisted of a total of 180 trials, of which 60 were potential congruent trials where a talker's voice was paired with target words that matched the learning words in valence (e.g., a talker who uttered the word "excitement" in the learning phase now uttered a neutral word "seat" which served as the potential prime to a valence-congruent word "kitten"). There were also 60 potential incongruent trials, in which a talker's voice was paired with target words that were opposite in valence to the words spoken in learning (e.g., "taxes"). Finally, there were 60 baseline trials where talkers who uttered neutral words during learning again uttered neutral words as potential primes for positive and negative target words.

Judgments and response latencies were recorded on each trial. Trials with reaction times less than 300 ms were removed because such fast reaction times are thought to occur largely as a result of anticipatory error (Bargh, Chaiken, Raymond, & Hymes, 1996). Reaction times for correct trials with reaction times greater than 300 ms (95% of all trials) were aggregated by trial type.

Immediately following the evaluative priming phase, participants completed an *explicit evaluation phase*. Participants heard each of the nine talkers from the learning phase utter a sequence of twenty neutral words (the 10 neutral words from the learning phase, plus the 10 neutral prime words). Words were played in a random order. At the end of the sequence of words, participants were asked to rate the speaker's voice on a scale of 1–5 where 1 equaled very unpleasant, 2 equaled unpleasant, 3 equaled neutral, 4 equaled pleasant and 5 equaled very pleasant. Once a rating had been keyed, the program advanced to the next speaker's word series.

Our a priori hypothesis was that voices acquired affective value after speaking valenced words corresponded to a test of the interaction between prime and target valence. We conducted a set of doubly centered contrasts that allowed us to examine the specific effect of the interaction, independent of main effects (see *Ableson & Prentice, 1997*). A doubly centered interaction contrast removes the contribution of main effects and the grand mean from the cell means to specifically examine the pattern of the interaction. This is comparable to creating an interaction term from centered variables in multiple regression (*Aiken & West, 1991*). We conducted doubly centered contrasts on the reaction-time data that compared value congruent prime-target trials (i.e., prime faces paired with positive information during learning-positive primes; prime faces paired with negative information during learning-negative primes) to va-

lue incongruent prime-target trials (i.e., prime faces paired with negative information during learning-negative primes; prime faces paired positive information during learning). To assess positive and negative affective learning specifically, we ran two similar sets of doubly centered contrasts (one set for positive affective learning and one set for negative affective learning) that compared all three valenced primes (positive, negative, neutral) for a given target type.

Results and discussion

Raw aggregated reaction times, doubly centered mean reaction times, and standard errors from the evaluative priming procedure are presented in *Table 1*.

The pattern of reaction times indicated that reverse priming occurred. Participants were significantly faster to evaluate target words after hearing a voice speaking neutral words when that voice had spoken oppositely valenced words during the learning phase,  $F(1, 76) = 18.13, p < .01$ . For example, for voices that spoke the word "excitement" during learning, speaking the word "seat" slowed the judgment of "kitten" as "positive" compared to its effect on the judgment of "taxes" as "negative."

This reverse priming effect was driven primarily by negative affective learning. Voices that previously uttered negative words during learning slowed the judgment of negative targets when compared to the influence of voices that previously spoke neutral or positive words, and significantly sped the evaluation of positive targets compared to those other voices,  $F(1, 76) = 33.20, p < .01$ . Voices that spoke positive words during learning did not differentially influence the judgment of positive targets when compared to the effect of voices that spoke neutral or negative words during learning, however,  $F(1, 76) = 2.60, ns$ , suggesting that talkers' vocal acoustics did not acquire positive affective value by speaking a series of positive words.

There were no differences in the explicit valence ratings of voices,  $F(2, 76) = .14, ns$ . In fact, voices that spoke positive words during learning ( $M = 2.98, SD = .65$ ), voices that spoke negative words during learning ( $M = 3.05, SD = .59$ ), and voices that spoke neutral words during learning ( $M = 3.01, SD = .57$ ) were all rated as being neutral (i.e., not different from "3" using one-sample *t*-tests); all  $t_s < .5$ . These findings suggest that the evaluative priming findings were not the result of strategic, explicit responding.

These findings demonstrate that talkers' vocal acoustics acquired affective value based on the semantic content of their speech. This learning occurred despite the fact that pre-testing showed that listeners could not distinguish among the talkers with reliability and all voices were judged to be neutral during the explicit evaluation phase.

**Table 1**  
Experiment 1 – raw aggregated reaction time data and double centered reaction time data.

		Prime valence			Target valence marginal means
		Positive	Neutral	Negative	
<i>Raw aggregated reaction time means</i>					
Target valence	Positive	854.94 (39.50)	851.24 (40.12)	860.08 (40.00)	855.42 (39.48)
	Negative	851.53 (40.33)	842.63 (38.21)	871.81 (42.79)	855.32 (39.86)
Prime valence Marginal means		853.24 (39.32)	846.94 (38.23)	865.94 (40.20)	<i>Grand mean</i> 855.37 (39.04)
<i>Doubly centered reaction times</i>					
Target valence	Positive	1.65	4.26	-5.91	
	Negative	-1.65	-4.26	5.91	

Note: standard errors are presented in parentheses under means.

**Table 2**  
Experiment 2 – raw aggregated reaction time data and double centered reaction time data.

		Prime valence			Target valence marginal means
		Positive	Neutral	Negative	
<i>Raw aggregated reaction time means</i>					
Target valence	Positive	772.64 (40.17)	754.50 (39.22)	755.55 (39.85)	760.90 (39.12)
	Negative	750.35 (35.68)	762.35 (38.15)	777.05 (39.71)	763.25 (37.37)
Prime valence		761.50	758.43	766.30	<i>Grand mean</i>
Marginal means		(37.37)	(38.07)	(38.98)	762.07 (37.86)
<i>Doubly centered reaction times</i>					
Target valence	Positive	12.32	–2.75	–9.57	
	Negative	–12.32	2.75	9.57	

Note: standard errors are presented in parentheses under means.

We further hypothesized that we observed stronger negative, as compared to positive, affective learning because the negative words spoken during the learning phase were more likely to be experienced as self-relevant to the university students who were our participants (e.g., hangover, smoking, cavities, divorce, taxes) and therefore subserved robust negative affective learning; in contrast, the positive words were less self-relevant (e.g., kitten, butterfly, strawberries, money, silk). In Experiment 2, we therefore recorded talkers uttering positive, negative, and neutral adjectives that were all likely to be experienced as self-relevant by the listeners because they can be used to describe characteristics of people. In addition, we increased the number of words each talker uttered from 10 to 20 in order to increase the likelihood that at least a number of words spoken in each valence type would be perceived as highly self-relevant by participants. We also increased the number of times each word was spoken during learning (from 1 to 4) to increase the number of experience the listeners had with each talker's voice.

## Experiment 2

### Method

#### Participants

Participants were 29 Boston College students (10 male, 19 female) who completed the experiment for either course credit or \$15. Two additional participants were excluded because of technical problems and one additional participant was excluded because that person did not comply with the task instructions.

#### Materials and procedure

The materials and procedure for Experiment 2 were nearly identical to Experiment 1 with exceptions as indicated. During the *learning phase*, participants heard a total of six European-American male talkers. Each talker uttered only positive, negative, or neutral adjectives referring to personal characteristics (i.e., two talkers uttered only positive descriptive adjectives, two talkers uttered only negative descriptive adjectives and two talkers uttered only neutral descriptive adjectives; see Appendix A for words). Each talker uttered twenty words. Each word was played four times. Immediately following the acquisition phase, participants completed a *sequential evaluative priming phase* which was identical to the sequential evaluative priming procedure from Experiment 1, with one exception. The prime stimuli were neutral words spoken by the same six talkers who uttered either positive, negative, or neutral words during the affective learning task (see Appendix A, Table A2 for words). As in Experiment 1, reaction

times for correct trials with reaction times greater than 300 ms (93% of all trials) were aggregated by trial type. No explicit evaluations of the voices were collected.

### Results and discussion

As in Experiment 1, we used a series of doubly centered contrasts to evaluate the reaction-time data. Raw aggregated reaction times, doubly centered mean reaction times and standard errors are presented in Table 2. As in Experiment 1, we observed reverse priming effects. Participants were significantly slower to evaluate target words after hearing a voice that had spoken a similarly valenced word during the learning phase when compared to hearing a voice that had spoken an oppositely valenced words during learning,  $F(1, 56) = 191.65, p < .01$ , indicating that voices acquired affective value based on the valence of the words spoken during the learning phase.

Further contrasts indicated that both negative and positive affective learning occurred. Voices that previously spoke negative words during learning slowed the judgment of negative targets when compared to voices that previously spoke neutral or positive words, and significantly sped the evaluation of positive targets compared to those other voices,  $F(1, 56) = 109.90, p < .01$ . Voices that spoke positive words during learning slowed the judgment of positive targets when compared to voices that spoke neutral or negative words during learning, and significantly sped the evaluation of negative targets compared to those other voices,  $F(1, 56) = 182.11, p < .01$ . Experiment 2 thus demonstrated that hearing talkers utter positive and negative adjectives was sufficient to imbue their voices (specifically, their utterances of neutral words) with positive and negative meaning.<sup>1</sup>

### General discussion

In two experiments, we demonstrated that a human voice becomes affectively potent and acquires the capacity to influence behavior simply based on the content of the words spoken by that voice on a previous occasion. These studies demonstrate, for the first time, that a perceiver's (or listener's) responses to a particular talker can be influenced by the literal sound of that talker's voice based on the affective connotations of the words that talker produced in the past.

<sup>1</sup> To assess whether there was variation in the extent to which the individual voices acquired affective value, we considered talker identity as a within subjects factor within a series of subsequent ANOVAs and follow-up *t*-tests. There were no significant effects of talker identity on response time to judge target words.

One possibility is that learning was enhanced by subtle affective cues present in the talker's pronunciation of words during the learning. In this view, when talkers spoke positive words their vocal acoustics were positive in affective tone simply because they were speaking words with positive connotations; these tonal attributes may have served to enhance the degree of positivity of the spoken words. We cannot rule out this possibility because we did not collect affective ratings of the talkers' vocal acoustics prior to learning so as not to bias perception of the voices during learning or test. Nevertheless, even if learning was enhanced by subtle affective cues present in the voices, the words spoken during test were neutral in content and tone. As a result, we can conclude that the priming effects were based on affective value acquired by the identity-linked vocal acoustics of each talker.

Interestingly, the pattern of reaction times was reversed from those most often observed in evaluative priming. There are two possible explanations for this pattern of results: that reverse learning occurred (e.g., hearing a voice speak positive words lead to that voice acquiring negative value) or that reverse priming occurred (e.g., voices that had acquired positive value facilitated the judgment of negative targets). A secondary index of the voices' acquired affective value would be required to rule out the possibility of reverse learning. In Experiment 1, we assessed participants' explicit evaluations of all of the voices but all voices were judged to be neutral. Rather than allow us to rule out the possibility of reverse learning, this finding suggests that the observed evaluative priming effects were not the result of strategic responding based on explicitly recalled valued information about the voices. In future studies, additional measures of voice value should be included in order to rule out the possibility of reverse learning.

Reverse priming, or contrast effects, have been documented in diverse set of evaluative priming experiments (for reviews, Glaser (2003), Klauer et al. (2009)). A new theoretical approach proposed by Klauer and colleagues (2009) suggests that reverse priming effects occur when prime stimuli are presented outside of the time window during which target stimuli are evaluated but in recent temporal proximity to the prime. In other words, there is a critical, early evaluation window during which traditional priming effects emerge (called assimilation effects) wherein the prime occurs in the window during which the target is evaluated. Immediately following the assimilation priming evaluation window, is a temporally later window during which reversed priming is observed. In this window, the prime has already activated an evaluative stance but is too temporally distant to exert direct influence on the evaluation of the target. Finally, after the reverse priming evaluation window, the effect of the prime eventually dissipates all together and no priming effects are observed. The specific timing of the different evaluation windows depends on the properties of the prime and targets and their relative timing (or stimulus onset asynchrony, SOA, the time between the onset of the prime and onset of the target).

One way to vary the position of evaluation windows is by varying the SOA between the prime and the target. Typical SOAs in evaluative priming experiments that result in assimilation priming effects are around 300 ms (e.g., Bargh et al., 1992). A good deal of research has demonstrated priming effects are quite sensitive to the SOA, although the pattern of results is somewhat unclear. In some experiments assimilation priming effects appeared at 300 ms but not 1000 ms (e.g., Hermans, De Houwer, & Eelen, 1994), while in other cases assimilation priming effects emerged at 0 and 100 ms SOAs and reverse priming emerged at 1200 ms SOA (e.g., Klauer, Roßnagel, & Musch, 1997). In the present experiment the SOA was necessarily long because our prime stimuli were full spoken words (around 1000 ms in length) which required a longer time to play than is typical for visually presented prime stimuli; additionally a fixation cross was presented for 100 ms be-

tween the prime and target to cue participants that the target was going to appear.<sup>2</sup> Given this long SOA and the fact that the target evaluation window was likely cued by the fixation cross, it is possible, even probable, that the prime stimulus fell outside of the critical window during which assimilation priming is observed resulting in reverse priming. Regardless of the mechanism underlying the effects, the fact that the voices speaking neutral words served to modulate the speed with which target words were evaluated clearly indicates that the acoustics of those voices acquired affective value.

The present findings have important implications for understanding the mechanisms of person perception. Little empirical research has addressed how vocal cues influence complex social judgments or social interactions. One recent study suggests that cues from voices may have a large impact on social perception – women whose voices were judged to be more feminine were associated to a greater degree with highly stereotypic female descriptions (Ko, Judd, & Blair, 2006). We know that much information can be extracted from the human voice – not only can people extract information about a speaker's sex and age (e.g., Bachorowski & Owren, 1995; Ladefoged & Broadbent, 1957; Owren & Bachorowski, 2003; Owren & Rendall, 1997), but also information about a speaker's personality (Scherer, 1979), emotional state (Russell, Bachorowski & Fernandez-Dols, 2003, for a review), attractiveness (Berry, 1992; Zuckerman & Miyake, 1993), maturity (Berry, 1992) and even probable occupation (Yamada, Hakoda, Yuda, & Kusuhara, 2000). Despite evidence that such information can be gleaned from human voices, how vocal acoustics come to have such meaning has not been addressed. While it is possible that some types of human vocal communication are innately pleasant (e.g., the sound of laughter; Owren & Bachorowski, 2003), it is likely that the pairing between particular vocal acoustics and complex social constructs such as occupation are learned through experience. The present findings suggest that such learning is not only possible, but occurs through relatively mild and limited experience (hearing a disembodied voice speak 10 or 20 valued words) and is highly specific (individual speakers' voices could not be explicitly distinguished, yet still acquired specific affective value). How such learning influences subsequent interpersonal interactions should be investigated.

The extent to which the affect-inducing properties of a voice become context-independent, change over time, and explicitly influences person perception are potentially fruitful avenues of future research. For example, if a given talker routinely induces negative affect in others by predominantly discussing topics that are negative or unpleasant, listeners may become more likely to attribute negative intentions to that individual even on occasions when the linguistic content of the speech is neutral or even positive. Conversely, if a talker's voice quality elicits generally positive affective responses through associations with positive content, then listeners will be more likely to make positive attributions. These pragmatic considerations are important in everyday linguistic interactions, where a listener's understanding of the significance of a communicative interaction is highly dependent on perceptions of a talker's communicative intentions. Such pragmatic considerations may become even more important in long-term relationships that are marked by repeated, socially significant interactions between two parties.

#### Acknowledgments

Preparation of this manuscript was supported an NSF Grant (BCS 0527440), an NIMH Independent Scientist Research Award

<sup>2</sup> We added this fixation cross after a pilot during which response latencies were long because participants were looking away from the screen when the target appeared.

(K02 MH001981), an NIH Director's Pioneer Award (DP1OD003312) to Lisa Feldman Barrett; NIMH Prime Award 1 R01 MH65317-01A2, and Subaward 8402-15235-X to Michael Ow-

ren; and by the Center for Behavioral Neuroscience, STC Program of the National Science Foundation under Agreement No. IBN-9876754. The authors wish to thank Ray Becker for his help with the audio recordings, Lauren Brennan and William Shirer for their help with data collection, and Karl Christoph Klauer and Jan De Houwer for comments on an earlier version of this manuscript.

**Table A1**  
Words from the affective learning task in Experiment 1.

Acquisition phase words		
Positive	Negative	Neutral
Cash	Abuse	Elbow
Ecstasy	Anger	Hairdryer
Engaged	Demon	Hairpin
Excitement	Destroy	Hydrant
Joke	Hostage	Kettle
Laughter	Intruder	Lamp
Lucky	Murderer	Locker
Profit	Rage	Paper
Rescue	Terrified	Street
Win	Tornado	Subdued
Sequential evaluative priming phase words		
Spoken primes	Visually presented targets	
	Positive	Negative
Bench	Food	Weeds
Chin	Birthday	Grease
Column	Strawberries	Taxes
Obey	Pizza	Hornet
Pamphlet	Kitten	Divorce
Reserved	Chocolate	Rattlesnake
Seat	Money	Mosquito
Statue	Magazine	Hangover
Table	Swimming	Garbage
Violin	Gold	Litter

**Appendix A**

See Tables A1 and A2.

**References**

Abelson, R. P., & Prentice, D. A. (1997). Contrast tests of interaction hypothesis. *Psychological Methods*, 2, 315–328.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Bachorowski, J.-A., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science*, 6, 219–224.

Bachorowski, J.-A., & Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America*, 106, 1054–1063.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology*, 62, 893–912.

Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology*, 32, 104–128.

Berry, D. (1992). Vocal types and stereotypes: Joint effect of vocal attractiveness and vocal maturity on person perception. *Journal of Nonverbal Behavior*, 16, 41–54.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5, 341–345.

Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. Technical report C-1. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, 127, 853–869.

Delgado, M. R., Olsson, A., & Phelps, E. A. (2006). Extending animal models of fear conditioning to humans. *Biological Psychology*, 73, 39–48.

Dudley, H. (1939). Remaking speech. *Journal of the Acoustical Society of America*, 11, 169–177.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229–238.

Glaser, J. (2003). Reverse priming: Implications for the (un)conditionability of automatic evaluation. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 87–108). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hermans, D., De Houwer, J., & Eelen, P. (1994). The affective priming effect: Automatic activation of evaluative information in memory. *Cognition and Emotion*, 8, 515–533.

Klauer, K. C., Roßnagel, C., & Musch, J. (1997). List context effects in evaluative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 246–255.

Klauer, K. C., Teige-Mocigemba, S., & Spruyt, A. (2009). Contrast effects in spontaneous evaluations: A psychophysical account. *Journal of Personality and Social Psychology*, 96, 265–287.

Ko, S. J., Judd, C. M., & Blair, I. V. (2006). What the voice reveals: Within- and between-category stereotyping on the basis of voice. *Personality and Social Psychology Bulletin*, 32, 806–819.

Ladefoged, P., & Broadbent, D. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29, 98–104.

Owren, M. J., & Bachorowski, J.-A. (2003). Reconsidering the evolution of nonlinguistic communication: The case of laughter. *Journal of Nonverbal Behavior*, 27, 183–200.

Owren, M. J., & Cardillo, G. C. (2006). The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *Journal of the Acoustical Society of America*, 119, 1727–1739.

Owren, M. J., & Rendall, D. (1997). An affect-conditioning model of nonhuman primate vocal signaling. In D. H. Owings, M. D. Beecher, & N. S. Thompson (Eds.), *Perspectives in ethology: Communication* (Vol. 12, pp. 299–346). New York: Plenum.

Owren, M. J., Rendall, D., & Bachorowski, J.-A. (2005). Conscious and unconscious emotion in nonlinguistic vocal communication. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 185–204). New York: Guilford.

Russell, J. A., Bachorowski, J.-A., & Fernandez-Dols, J.-M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology*, 54, 329–349.

**Table A2**  
Words from the affective learning task in Experiment 2.

Acquisition phase words		
Positive	Negative	Neutral
Attractive	Awful	Acceptable
Beautiful	Brutal	Adequate
Brave	Cruel	Anonymous
Brilliant	Deceitful	Beside
Charming	Disgusting	Comparable
Delighted	Dumb	Demonstrative
Excellent	Greedy	Impartial
Friendly	Hostile	Indulgent
Generous	Pathetic	Moderate
Great	Repulsive	Near
Honest	Retarded	Neutral
Inspiring	Rotten	Noisy
Kind	Rude	Perpetual
Intelligent	Scum	Practical
Successful	Selfish	Quiet
Talented	Stupid	Reasonable
Terrific	Terrible	Seated
Triumphant	Ugly	Sedate
Trustworthy	Useless	Standard
Wise	Worthless	Wrapped
Sequential evaluative priming phase words		
Spoken primes	Visually presented targets	
	Positive	Negative
Standing	Food	Weeds
Regular	Birthday	Grease
Nonchalant	Strawberries	Taxes
Malleable	Pizza	Hornet
Formal	Kitten	Divorce
Busy	Chocolate	Rattlesnake
Absolute	Money	Mosquito
Adjacent	Magazine	Hangover
Discreet	Swimming	Garbage
Visible	Gold	Litter

- Scherer, K. R. (1979). Personality markers in speech. In K. R. Scherer & H. Giles (Eds.), *Social markers in speech* (pp. 147–209). Cambridge, UK: Cambridge University Press.
- Traunmüller, H. (1994). Conventional, biological, and environmental factors in speech communication: A modulation theory. *Phonetica*, 51, 170–183.
- Yamada, N., Hakoda, Y., Yuda, E., & Kusuhara, A. (2000). Verification of impression voice in relation to occupational categories. *Psychological Reports*, 86, 1249–1263.
- Zuckerman, M., & Miyake, K. (1993). The attractive voice: What makes it so? *Journal of Nonverbal Behavior*, 17, 119–135.