A laboratory-based procedure for measuring emotional expression from natural speech

ALEX S. COHEN, KYLE S. MINOR, GINA M. NAJOLIA, AND S. LEE HONG
Louisiana State University, Baton Rouge, Louisiana

Despite dramatic advances in the sophistication of tools for measuring prosodic and content channels of expression from natural speech, methodological issues have limited the simultaneous measurement of those channels for laboratory research. This is particularly unfortunate, considering the importance of emotional expression in daily living and how it can be disrupted in many psychological disorders (e.g., schizophrenia). The present study examined the Computerized assessment of Affect from Natural Speech (CANS), a laboratory-based procedure that was designed to measure both lexical and prosodic expression from natural speech across a range of evocative conditions. The verbal responses of 38 male and 31 female subjects were digitally recorded as they reacted to separate pleasant, unpleasant, and neutral stimuli. Lexical and prosodic expression variables significantly changed across these conditions, providing support for using the CANS in further laboratory research. The implications for understanding the interface between lexical and prosodic expressions are also discussed.

Recent years have seen dramatic advances in the sophistication of measures of emotion for psychological research. Of particular note, a number of computerized technologies have been developed to objectify the human expression of emotion. This is a critical scientific advance, because expression of emotion is integral to social behavior (Decety & Lamm, 2006; Gross, 2002; LeDoux, 2000), and because it is compromised in a number of neuropsychiatric disorders, such as schizophrenia (Cohen et al., 2005; Cohen & Minor, in press), depression (Leventhal, Chasson, Tapia, Miller, & Pettit, 2006), and autism (Matese, Matson, & Sevin, 1994; South et al., 2008). As is discussed below, however, the application of these measures for laboratory study has been complicated by methodological and procedural limitations. The present study reports on a laboratory procedure for evoking and measuring behavioral expression of emotion using natural speech.

Natural speech has been an attractive medium for understanding emotional expression, because it contains a wealth of information. This information is conveyed across multiple domains, including those in *prosodic* and *content* channels. Prosodic channels involve the nonverbal aspects of spoken communication, which can be measured through acoustic analysis of the communication's physical properties (e.g., inflection, emphasis, and speech rate; Alpert, Merewether, Homel, Marz, & Lomask, 1986). Conversely, content channels involve the semantic aspects of communication, which are assessed using a variety of content-analytic procedures (e.g., McAdams, 2001; Pennebaker, Booth, & Francis, 2007). It is worth noting that computerized methods of measuring prosody and content from speech have existed since the 1960s (e.g., General Inquirer; Stone, Dunphy, Smith, & Ogilvie, 1966) and have allowed for efficient, sensitive, and well-validated

measures of emotional expression (Pennebaker & King, 1999; Pennebaker, Mehl, & Niederhoffer, 2003).

Despite voluminous literature on studies that have employed prosodic and content-analytic methodologies, few studies have examined these modes of communication simultaneously. In large part, this reflects striking disparities between the methodologies that have been used to study these phenomena. Prosodic analysis is generally conducted within the context of relatively brief verbal expressions that are uttered under highly controlled emotion-induction procedures. For example, the classic Velten (1968) methodology, used in over 100 published peer-reviewed studies to date, involves having subjects read standardized emotionally valenced scripts (e.g., "There is no hope," and "Nothing can bum me out now."). Although the Velten procedure produces robust changes in prosody across different emotionally valenced conditions (see Scherer, 2003, for a review), these procedures do not allow for content analysis, since content is not self-generated. Moreover, questions have been raised about whether the prosodic changes are genuine in nature, or whether they reflect "posed" emotions that are evoked by demand characteristics of the text (see, e.g., Westermann, Spies, Stahl, & Hesse, 1996). In contrast, content analysis is usually conducted on lengthier narratives that are either written or spoken. The methodologies that have been employed in laboratory studies have generally required subjects to produce freely generated speech in response to standardized stimuli such as nondescript probes (e.g., "Tell me about bad memories from your life"; Cohen & Docherty, 2004, 2005; Hagenaars & van Minnen, 2005; Nelson & Horowitz, 2001), film clips (Kahn, Tobin, Massey, & Anderson, 2007), picture stills (Tolkmitt & Scherer, 1986), or Thematic Apperception Test cards (Markel, Bein, & Phillis, 1973; Rosenberg, Schnurr, & Oxman, 1990). Many of these procedures

A. S. Cohen, acohen@lsu.edu

are limited by their dependence on laboratory assistants, who can profoundly influence a subject's speech, and by their use of ambiguous stimuli that allow for considerable variability in interpretation across subjects. We are aware of only two studies that employed prosodic and content methodologies simultaneously. Both of these studies examined free-speech samples in patients with schizophrenia and found that prosody and content of speech were unrelated (Alpert, Rosenberg, Pouget, & Shaw, 2000; Cohen, Alpert, Nienow, Dinzeo, & Docherty, 2008); thus, the procedures in current use remain untested for their feasibility in both prosodic and content analysis.

We have developed a procedure, called the Computerized assessment of Affect from Natural Speech (CANS), that attempts to bridge the gap between the two aforementioned methodologies. Specifically, the CANS was developed to (1) use a highly controlled laboratory procedure employing well-validated stimuli; (2) genuinely modulate emotion; and (3) allow for free verbal expression. During the CANS, subjects are asked to discuss their personal reactions to a series of standardized picture stills. Administration is automated, thus reducing potential influences from laboratory assistants. Additionally, the valence, intensity, and modality of the emotion-induction stimuli that are used in the CANS can be varied, thus allowing for a wide range of applications.

The present study is an initial investigation into the feasibility and validity of the CANS that is aimed at addressing three questions. First, we were interested in determining the extent to which extended verbalization influences emotion induction. Our prediction was that verbalization would increase the depth of semantic processing of the stimuli, and that this in turn would enhance its evocative effects. The impact of deep semantic processing on recall of stimuli is well documented in cognitive psychology (see Craik & Tulving, 1975), and we reasoned that it might improve the affective experience as well. We had alternate concerns that verbalization could attenuate emotion-induction effects, on the basis of evidence that expressing emotional states ameliorates unpleasant experiences following stressful laboratory tasks (e.g., Zech, 1999). To examine this issue, we compared subjective emotional and arousal states while subjects either verbalized their responses or silently processed evocative picture stills. Second, we sought to determine the extent to which emotional expression, defined in terms of prosodic and content variables, could be successfully manipulated using the CANS procedure. Subjects were exposed to separate unpleasant, pleasant, and neutral emotional stimuli to arouse emotions. We paid particular attention to potential gender effects during these analyses, given the well-documented differences in prosodic expression in females compared with males (Bachorowski & Owren, 1995; Scherer, 2003). Finally, we examined the associations between prosodic, content, and subjective experiential variables across subjects.

METHOD

Subjects

Subjects (38 male and 31 female) were recruited from Louisiana State University. The sample comprised 61 Caucasians, 5 African-Americans, 2 Asian-Americans, and 1 Hispanic ($M = 21.28$ years, $SD = 5.89$). All subjects were fluent in English and reported having vision that was correctable to 20/20. Subjects received course credit for participating in this experiment. This study was approved by the appropriate institutional review board, and all subjects provided written informed consent prior to beginning the study.

Procedures

Subjects were seated in front of a computer monitor and were out of view of laboratory assistants. The experiment was run using E-Prime software (Version 1.0; Psychology Software Tools, Pittsburgh, PA). Subjects were given counterbalanced *silent* and *voiced* emotion-induction conditions that were separated by an hour epoch. For each condition, subjects were asked to view three separate blocks of five affectively positive (picture numbers 2080, 5910, 2360, 7325, 4643, 4626, 7502, 7330, 1710, 2391), five affectively negative (picture numbers 9800, 9570, 9592, 6350, 6821, 9810, 6540, 9571, 6242, 9594), and five affectively neutral (picture numbers 7496, 7595, 7002, 7037, 7057, 7004, 7056, 7495, 7546, 7620) pictures from the International Affective Picture System (Lang, Bradley, & Cuthbert, 2005). The pictures in the affectively positive and negative conditions were selected for their relatively extreme valence ratings.

Picture display was 40 sec, an amount of time that we decided was adequate for subjects to produce appropriate speech for content analysis, on the basis of our earlier research, in which both healthy adults and patients with schizophrenia generated more than 1.5 words per sec (Cohen, Alpert, et al., 2008). Pennebaker, Booth, and Francis (2007) recommended a minimum of 100 words for content analysis, a criterion that should be met by most subjects during a 200-sec speaking condition (40 sec each for five stimuli). Block order was random, as was picture order within each block. No stimulus was presented more than once to any individual subject in the experiment. Blocks were separated by a 30-sec interval during which subjects were asked to "relax and breathe deeply." On the basis of prior studies reporting that individuals return to electrophysiological baseline after processing IAPS stimuli within 1 sec of stimulus offset (see Lang, Bradley, & Cuthbert, 1999, for a review of this methodology), we expected this epoch to be sufficient to facilitate a return to baseline emotion levels. Before and after each picture block, subjects rated their emotion and arousal levels using the Self-Assessment Manikin (SAM; Lang et al., 2005), an analogue scale ranging from 1 (*pleasant emotion* and *high arousal*) to 9 (*aversive emotion* and *low arousal*). SAM ratings are based on the circumplex model of emotion (Larsen & Diener, 1992; Watson, Wiese, Vaidya, & Tellegen, 1999), which posits that various emotional states reflect the input of two orthogonal valence (from pleasant to unpleasant) and arousal (from low arousal to high arousal) dimensions. Administration time was approximately 15 min. The SAM was selected because it has shown promise as a state measure of emotional experience that can be used repeatedly during emotion-induction studies (see, e.g., Backs, da Silva, & Han, 2005; Cohen, Minor, Baillie, & Dahir, 2008; Gomez & Danuser, 2004).

During the silent condition, subjects were given the following instructions by the experimenter: "In a moment . . . I will show you a series of pictures for 3 minutes. Maintain your focus on the pictures as you silently watch them. We will begin shortly."

During the voiced condition, subjects were asked to verbalize their thoughts about the picture, especially how it made them feel and what memories it conjured. Their speech was digitally recorded at 16 bits/sec at a sampling frequency of 44100 Hz using a headset microphone. Recordings contained the subject's verbalized response to all five pictures in a block (lasting 200 msec total). Prior to presenting each block of pictures, the experimenter read the following instructions:

In a moment . . . I will show you a series of pictures for 3 minutes. While you are focusing on these pictures, I want to record you as you talk. I am curious about how the picture relates to

you. I want you to talk about what the picture means to you, what it reminds you of, and how the picture makes you feel. Each picture will be up for about 40 seconds and it is important that you talk for the full time that the pictures are being displayed. Please maintain your focus on the picture as you talk for the full time.

Prosodic Analysis

The digitized recordings were analyzed using Praat (Boersma, 2001), a program that has been used extensively in speech pathology and linguistic studies. The Praat system organizes the sound file into "frames" for analysis, which for the present study were set at a rate of 20 frames/sec. Analysis was conducted using scripts (www.ling.ohio-state.edu/~welby/praat.html). The entire speech sample, comprising 20,000 frames, was examined. We computed *inflection* (the variability in pitch using information entropy; Shannon, 1948), *amplitude* (the mean volume in decibels), *emphasis* (the variability in volume using information entropy computations), and *vocal output* (the total percentage of frames that were voiced). The information entropy analyses yielded indices of speech amplitude variability that were expressed in bits (binary units). These analyses were conducted by undergraduate research assistants who were well trained in our laboratory procedures. Samples were processed using desktop computers with single Pentium 4 chips, which took approximately 2–3 min per 200-sec recording.

It is worth expounding on our use of entropy statistics, given that they are rarely used for psychological research yet provide a more sensitive measure of signal variability than do variance and standard deviation scores (Lai, Mayer-Kress, Sosnoff, & Newell, 2005). To compute the entropy statistics, we first removed all unvoiced frames (i.e., pitch = 0 Hz). Each data set was then statistically normalized (mean subtracted, divided by standard deviation). As a result, the pitch and intensity for every trial possessed a mean of 0 and a variance of 1. A frequency histogram of the data was obtained using equally sized bins. The number of bins was set at $N/30$, with N being the total number of data points within a trial in which speech was produced. A probability distribution was then obtained for the data set by dividing the frequency of occurrence of data points within each bin by N. This was done to ensure that the sensitivity of the analysis remained constant across subjects, minimizing the potential for bias from the data range and amount of speech that was produced.

The information entropy, H, for each data set was calculated as

$$H = -\sum p_i \log_2 p_i, \qquad (1)$$

where p is the probability that a data point occurs within the ith bin. This provides a measure of uncertainty that is contained within the probability distribution as measured in bits of information. Higher information-entropy values indicated a more evenly distributed data set across the bins, whereas lower information-entropy values indicated a more peaked distribution. An example of the distributions from a single subject can be seen in Figure 1. For a simplified illustration of entropy computations, see the Appendix.
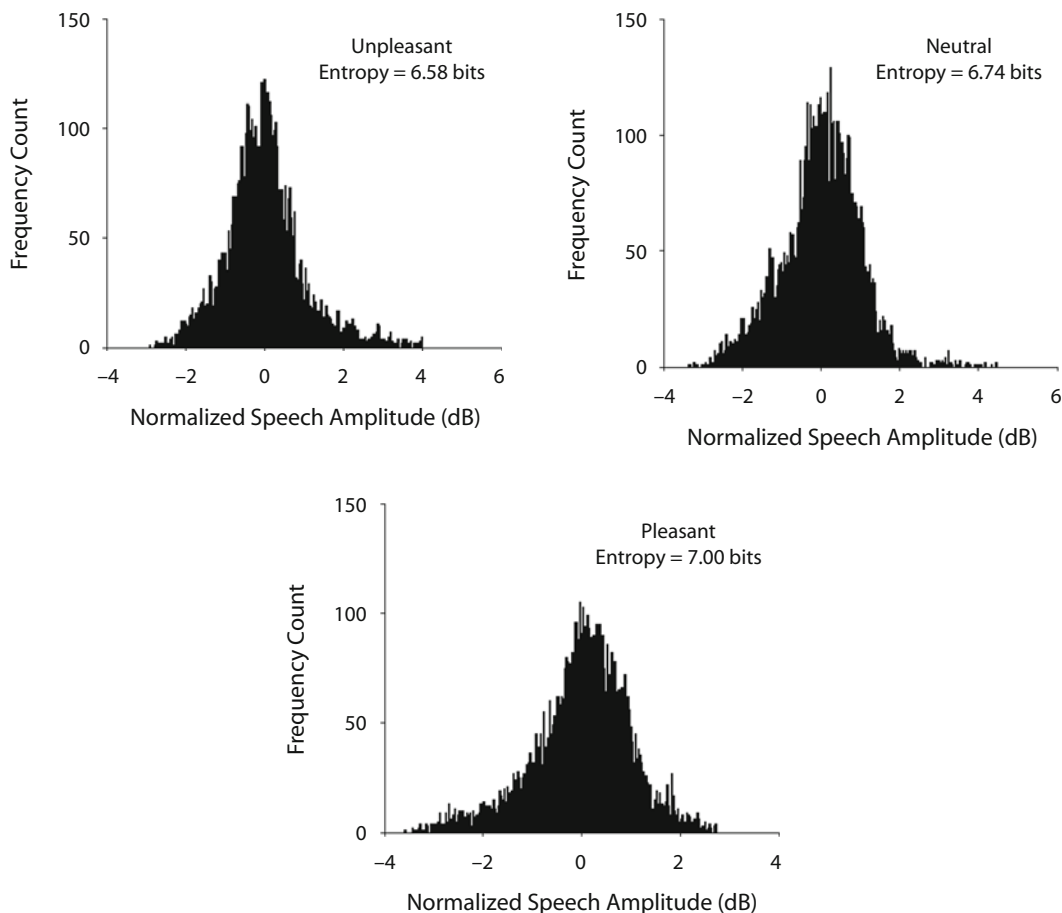

Figure 1. Exemplar frequency histograms of speech amplitude in decibels (dB) from a single subject during the different stimulus conditions. The entropy scores that were obtained for the emphasis data are included. The ranges on the horizontal axes vary as a result of the statistical normalization procedure, and they are maintained in order to preserve the shape of the probability distributions.

Content Analysis

The digitally recorded narratives were carefully transcribed by trained research assistants, were double-checked for accuracy, and were analyzed using the Linguistic Inquiry and Word Count (LIWC) software (Pennebaker et al., 2007). The LIWC program processes text files one word at a time, matching the base form of each word to a dictionary of over 2,290 word stems. Word stems are organized into 83 categories. A frequency count of the total number of instances of target words from each category is yielded, and this count is then divided by the total number of words that are in the text to control for individual differences in verbosity. Scores thus reflect a percentage of word matches in that category. We were primarily interested in the positive and the negative emotional categories, which comprise words that relate to emotional processes (e.g., *happy*, *sorrow*). Subjects averaged over 400 words for each of the neutral, pleasant, and unpleasant conditions. The LIWC has been used extensively in psychology research, and it has been validated for analysis of emotion in a wide range of applications. Of note, positive and negative lexical expression during laboratory procedures has been the focus of a number of studies from our lab (Cohen, Alpert, et al., 2008; Cohen & Minor, in press; Cohen, St-Hilaire, Aakres, & Docherty, in press), and they have shown significant associations with both trait measures of emotionality (Cohen & Minor, in press) and subjective emotion-rating scales (Kahn et al., 2007).

Analyses

We conducted the data analysis in four phases. First, we compared self-reported emotion and arousal ratings in the voiced condition with those in the nonvoiced condition to determine the extent to which verbalization affected the emotion-induction effects. We hypothesized that the magnitude of valence and arousal change (measured comparing precondition ratings with postcondition ratings) would be stronger in the voiced condition. Second, we compared prosody and content variables across the pleasant, unpleasant, and neutral conditions of the voiced condition, with the expectation that there would be significant change in these variables across the three conditions. As part of this analysis, we included gender as a between-groups factor. Third, we computed zero-order correlations between the prosodic and content variables in order to better understand the interrelationships between these variables. Finally, we examined the relationship between the prosodic and content variables and the self-report ratings using partial correlations (controlling for gender). All significance tests that are reported here are two-tailed, and all variables are normally distributed unless otherwise noted.

RESULTS

Analysis Set 1: Emotional Experience During the Voiced Versus Silent Conditions

Change scores were computed by subtracting baseline valence and arousal ratings (i.e., the precondition ratings) from the postcondition ratings. Positive valence-change scores reflected increasingly unpleasant emotional states, whereas negative valence-change scores reflected increasingly pleasant emotional states. Positive arousal-change scores reflected decreased arousal, and negative arousal-change scores reflected increased arousal. The means and standard deviations for these scores are presented in Table 1. Separate repeated measures ANOVAs were computed to determine within-condition changes in valence and arousal. These analyses revealed significant changes for both valence and arousal ratings across the voiced [valence ratings, omnibus $F(3,66) = 67.29$, $p < .05$; arousal ratings, omnibus $F(3,66) = 6.75$, $p < .05$] and silent [valence ratings, omnibus $F(3,66) = 102.78$, $p < .05$; arousal ratings, omnibus $F(3,66) = 16.78$, $p < .05$] conditions, suggesting

Table 1
Means and Standard Deviations for Precondition–Postcondition Changes in Valence and Arousal for the Neutral, Pleasant, and Unpleasant Conditions

Affective Condition	Voiced Procedure		Silent Procedure	
	M	SD	M	SD
Valence Ratings				
Neutral	0.27	1.09	0.28	1.47
Pleasant	−1.09	1.43	−1.37	1.29
Unpleasant	1.77	1.75	2.39	1.73
Arousal Ratings				
Neutral	0.25	1.45	0.71	1.75
Pleasant	−0.42	1.43	−0.85	1.80
Unpleasant	−0.54	1.76	−1.08	2.43

that successful emotion induction was achieved for both voiced and nonvoiced procedures. Follow-up repeated measures *t* tests were then used to determine the nature of these changes. These results indicated that valence ratings became significantly more pleasant for the pleasant condition than for the neutral and unpleasant conditions in both the voiced and silent procedures, and they became significantly more unpleasant for the unpleasant condition than for the neutral and pleasant conditions in both procedures. With

Table 2
Means and Standard Deviations of Prosodic and Content Variables, With Test Statistics From Repeated Measures ANOVAs

Affective Condition	Males		Females	
	M	SD	M	SD
Inflection (bits)				
condition $F = 3.56^*$, sex $F = 15.88^*$, interaction $F = 0.39$				
Neutral	4.93	0.65	5.59	0.53
Pleasant	5.03	0.88	5.56	0.67
Unpleasant	4.82	0.76	5.42	0.74
Amplitude (in decibels)				
condition $F = 5.92^*$, sex $F = 0.24$, interaction $F = 0.30$				
Neutral	57.17	5.31	56.78	5.16
Pleasant	57.69	5.73	57.00	4.81
Unpleasant	58.14	5.35	57.42	4.67
Emphasis (bits)				
condition $F = 1.75$, sex $F = 1.36$, interaction $F = 1.00$				
Neutral	6.61	0.72	6.89	0.51
Pleasant	6.79	0.70	6.84	0.67
Unpleasant	6.59	0.81	6.75	0.81
Speech Output (% of total time voiced)				
condition $F = 3.63^*$, sex $F = 4.09^*$, interaction $F = 2.15$				
Neutral	30.87	10.66	37.51	8.38
Pleasant	32.31	10.16	36.24	9.74
Unpleasant	30.60	11.15	33.53	11.00
Positive Emotion Words (% of total words)				
condition $F = 114.96^*$, sex $F = 0.03^*$, interaction $F = 1.86$				
Neutral	2.21	1.09	2.16	1.01
Pleasant	4.72	2.39	5.23	2.06
Unpleasant	2.09	1.22	1.76	0.84
Negative Emotion Words (% of total words)				
condition $F = 227.60^*$, sex $F = 10.98^*$, interaction $F = 6.04^*$				
Neutral	0.67	0.50	0.85	0.72
Pleasant	0.50	0.49	0.57	0.56
Unpleasant	3.16	1.45	4.28	1.58

$^*p < .05$.

Table 3
Zero-Order Partial Correlations (Controlling for Effects of Gender)
of Prosodic and Content Variables

Variables	% Positive Words	% Negative Words	Inflection	Amplitude	Emphasis
		Neutral Condition			
% positive words	1.00	–	–	–	–
% negative words	.16	1.00	–	–	–
Inflection	.04	.01	1.00	–	–
Amplitude	−.05	−.13	.08	1.00	–
Emphasis	.02	.00	.39*	.07	1.00
Speech output	.02	.06	.60*	.16	.67*
		Pleasant Condition			
% positive words	1.00	–	–	–	–
% negative words	−.13	1.00	–	–	–
Inflection	−.04	−.08	1.00	–	–
Amplitude	.17	−.10	.20	1.00	–
Emphasis	−.06	−.30*	.56*	.27*	1.00
Speech output	−.07	−.38*	.65*	.25*	.75*
		Unpleasant Condition			
% positive words	1.00	–	–	–	–
% negative words	.09	1.00	–	–	–
Inflection	.08	−.06	1.00	–	–
Amplitude	.06	−.10	.32*	1.00	–
Emphasis	.06	−.05	.67*	.32*	1.00
Speech output	−.05	−.12	.74*	.36*	.82*

*$p < .05$.

respect to the arousal ratings in both the silent and voiced procedures, the pleasant and unpleasant conditions elicited more arousal than did the neutral conditions, but the two did not differ significantly from each other.

Next, t tests were used to address whether the voiced and silent procedures differed in either valence or arousal. Subjects in the silent procedure had significantly more dysphoric emotion for the unpleasant condition than did those in the voiced procedure $[t(67) = 2.29, p < .05]$, but no other comparisons were statistically significant. Overall, the differences between the two procedures were relatively minor.

Analysis Set 2: Prosodic and Content Variables Across Emotion Conditions

Repeated measures ANOVAs (condition × sex) were computed to determine changes in prosodic and content-analytic variables across the three conditions (see Table 2). Significant omnibus F values for condition were observed for each of the prosodic and content variables except emphasis. Follow-up contrasts revealed that subjects' speech during the pleasant condition, compared with that during the unpleasant condition, was characterized by increased inflection $[t(67) = 2.54, p < .05]$, decreased speech amplitude $[t(67) = 2.23, p < .05]$, increased speech output $[t(67) = 2.49, p < .05]$, and more positively $[t(67) = 11.76, p < .05]$ and less negatively $[t(67) = 15.42, p < .05]$ valenced word use. Speech in the unpleasant condition, compared with that in the neutral condition, was characterized by less inflection $[t(67) = 2.05, p < .05]$, higher speech amplitude $[t(67) = 3.23, p < .05]$, less speech output $[t(67) = 2.20, p < .05]$, and more negatively $[t(67) = 15.12, p < .05]$ valenced word use. Subjects used more positively valenced words in the pleasant than in the neutral condition $[t(67) = 10.84, p < .05]$, but

there were no other significant differences between these conditions for any other variables. As can also be seen in Table 2, female subjects showed significantly higher pitch, more inflection, higher speech output, and more negatively valenced words than did male subjects. The general lack of interaction effects suggested that male and female subjects were relatively similar in how their prosodic characteristics changed across the neutral, pleasant, and unpleasant conditions, although it is noteworthy that female subjects showed a more dramatic increase in negative word use from the neutral and pleasant conditions to the unpleasant condition.

Analysis Set 3: Interrelationships Between Prosodic and Content Variables

Partial correlations (controlling for sex) were computed between the prosodic and content analytic variables (see Table 3). There are several noteworthy findings here. First, the prosodic and content variables showed few significant correlations with each other, suggesting that they tap into distinct expressive channels. Second, the prosodic variables showed modest intercorrelation with each other. Increased inflection, emphasis, and speech rate were each significantly correlated across each condition. Third, the magnitude of the intercorrelations between the prosodic variables tended to increase more in the pleasant and unpleasant conditions than in the neutral condition.

Relationship Between Speech-Analytic Variables and Subjective Emotion

Change scores were computed by statistically regressing the baseline arousal scores from the postcondition scores separately for the neutral, pleasant, and unpleasant conditions. The resulting standardized scores were structured so that increasing scores reflected lower levels

Table 4
Partial Correlations (Controlling for Effects of Gender) Between Emotional
Experience Change Scores (Valence and Arousal Change) and Prosodic and Content
Variable Scores, Computed for the Neutral, Pleasant, and Unpleasant Conditions

Variables	Valence Change			Arousal Change		
	Neutral	Pleasant	Unpleasant	Neutral	Pleasant	Unpleasant
Prosodic						
Inflection	−.04	−.03	−.01	−.24*	−.12	−.27*
Amplitude	.12	−.01	.07	.00	−.19	−.19
Emphasis	−.09	−.07	−.29*	−.36*	−.26*	−.33*
Speech output	.07	.05	.16	−.37*	−.29*	−.24*
Lexical						
Positive words	−.21	−.03	−.11	.01	−.05	−.12
Negative words	.01	.10	−.12	.23	.05	−.14

Note—Increasing scores indicate more euphoric emotion or decreasing arousal. Positive correlations reflect increasing prosody and decreasing arousal, whereas negative correlations reflect decreasing prosody and increasing arousal (or more dysphoric emotion). $^*p < .05$.

of arousal and higher levels of euphoric emotion. Partial correlations (controlling for sex) between the subjective change scores and the prosodic scores for the neutral, pleasant, and unpleasant conditions were then computed (see Table 4). There are several noteworthy findings from these analyses. First, subjects who were aroused during the emotion-induction conditions tended to show greater inflection, talk louder, and produce more speech. They did not show any concomitant changes in the emotional content of their speech. Second, subjects' valence scores were relatively unrelated to the prosodic or content variables. These findings suggest that prosodic variables are a function of emotional arousal rather than emotional valence.

DISCUSSION

Significant changes in prosodic and lexical variables as well as in phenomenological states were observed across the emotionally valenced conditions in this study, providing support for our laboratory-based procedure as a method for both evoking emotional states and generating free speech for computerized analysis. We believe that the CANS procedure reflects an important potential advance in the pursuit of understanding normal emotive processes, because it allows for simultaneous analysis of both prosodic- and content-communicative channels. Moreover, it holds promise for clarifying the abnormalities that are involved in various forms of psychopathology. Our own research is focused on understanding how emotional expression is attenuated in schizophrenia, since there has been limited research applying computerized measures in this regard (for elaboration, see Alpert et al., 1986; Cohen & Minor, in press; Cohen, St-Hilaire, et al., in press). The chief obstacle in this line of research is the lack of validated procedures for securing speech, an obstacle that the CANS can potentially help researchers overcome.

In the present study, emotion was conceptualized using the circumplex model (Larsen & Diener, 1992; Watson et al., 1999), a "dimensional" model of emotion that has considerable empirical support in basic emotion research and that has been used in prior studies of prosody (e.g., Laukka, 2005). It is worth briefly considering the present findings within the framework of a competing model of emotion, one that involves "categories" or distinct kinds of emotion (see Barrett, 2006, for further discussion), on the basis of evidence that different kinds of emotions produce different prosodic profiles. For example, anger and fear have been associated with increases in inflection, amplitude, and emphasis, whereas sadness and disgust have been associated with declinations in at least some of these variables (Scherer, 2003; Sobin & Alpert, 1999; Ververidis & Kotropoulos, 2006). In the present sample, unpleasant emotion contributed to decreased inflection, increased amplitude, and decreased speech production, a pattern consistent with expression of disgust (see Sobin & Alpert, 1999; Ververidis & Kotropoulos, 2006). This is not surprising, given that the negative pictures predominantly contained disgusting and violent scenes. With an eye toward future CANS studies, it is important to acknowledge that negative emotion is not an isomorphic construct. An important next step would thus be to adopt a categorical approach to understanding how prosodic expression varies across conditions, particularly across disgust, happiness, sadness, surprise, fear, and anger states (Barrett, 2006).

Insofar as very few studies to date have simultaneously employed content and prosodic analysis, the present findings offer support for their simultaneous use in a laboratory setting. It is interesting that prosodic and content variables were relatively unrelated in this study, perhaps revealing that they reflect different communicative channels. In contrast with findings from Kahn et al. (2007), lexical-analysis variables were not significantly related to the subjective ratings, raising questions as to what the content-analysis variables were capturing. When interpreting this finding, it is worth noting that only a handful of studies to date have employed lexical analysis during laboratory procedures, so this methodology is in need of further validity studies. It could be the case that the lexical dictionaries that were used in the present study were not precise enough to capture emotion states accurately, and that current dictionaries need to be adapted for laboratory use. Pessimism should be tempered by the relatively large literature employing the LIWC to understand emotional processes, and on the basis of the finding that lexi-

cal expression changed dramatically across the evocative conditions.

Changes in vocal prosody were associated with subjective arousal but not with valence for each of the neutral, pleasant, and unpleasant conditions. This finding may reflect the involvement of common neurobiological underpinnings in both subjective arousal and prosodic systems. The amygdala and other basal ganglia structures, which have been consistently linked to subjective and physiological arousal levels, are likely candidates, particularly given recent research that has implicated the basal ganglia in prosodic expression (see, e.g., Van Lancker, Sidtis, Pachana, Cummings, & Sidtis, 2006). It is possible that the nonsignificant correlations between subjective valence and prosody reflect low variability in valence scores across subjects, since emotion-induction states were achieved by most subjects. Currently, the link between subjective and expressive emotional systems is poorly understood, and further investigation of the shared systems underlying emotional arousal and prosody using the CANS seems a promising avenue.

Overall, verbal expression during the emotion-induction condition appeared to have little effect on subjective experience of emotion. This finding is encouraging, given our concern that articulation of emotional states could attenuate the magnitude of experience (see, e.g., Zech, 1999); however, it is also important to note that subjective emotion did not intensify when it was verbalized, a prediction we made on the basis of research from cognitive psychology that found that depth of semantic processing improves recall (Craik & Tulving, 1975). Several recent studies have demonstrated that, under some circumstances, verbal expression can inhibit depth of processing through verbal overshadowing, presumably because individuals attend to more superficial aspects of the stimuli (Lane & Schooler, 2004), so it could be the case that subjects focused minimally on the evocative features of the picture still. This seems unlikely to explain the present results, considering the dramatic change in lexical expression across the conditions. Regardless, the fact that the voiced condition still produced measurable changes in subjective emotion is encouraging for the use of the CANS for laboratory research.

Several limitations warrant mention here. First, prosodic changes were relatively benign when pleasant emotion was aroused, compared with those in the neutral condition. Although some have noted that unpleasant emotion is often easier to arouse in laboratory settings than is pleasant emotion (Wiseman & Levin, 1995), the CANS procedure appears to have a limited application for generating pleasant emotion states at present. Second, our measure of subjective emotions was probably not immune to the demand characteristics of the task. Nonetheless, the SAM has shown promise as a state measure of emotion in prior studies (Backs et al., 2005; Cohen & Minor, in press; Gomez & Danuser, 2004), and the arousal ratings showed convergent validity with the prosodic measures in the present study. Third, the present study was largely exploratory in nature and did not control for multiple comparisons, so some of the present findings could be mis-

leading. Finally, many of the effects that were observed in this study were relatively modest. Future research might improve on these effects by varying stimulus intensity, subject instructions, and experimental procedures (e.g., length of stimulus display). The present project serves as an important platform for refining emotion-induction procedures for speech analysis.

In sum, the present study found encouraging support for the CANS procedure as a method for procuring speech samples for prosodic and content analysis. Moreover, this procedure effectively produced meaningful changes in emotional states. The development of this procedure will further facilitate understanding of the emotion system as well as how individual differences in expression relate to pathological states.

REFERENCES

Alpert, M., Merewether, F., Homel, P., Marz, J., & Lomask, M. (1986). Voxcom: A system for analyzing natural speech in real time. *Behavior Research Methods, Instruments, & Computers*, **18**, 267-272.

Alpert, M., Rosenberg, S. D., Pouget, E. R., & Shaw, R. J. (2000). Prosody and lexical accuracy in flat affect schizophrenia. *Psychiatry Research*, **97**, 107-118.

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.

Backs, R. W., da Silva, S. P., & Han, K. (2005). A comparison of younger and older adults' self-assessment manikin ratings of affective pictures. *Experimental Aging Research*, **31**, 421-440.

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, **1**, 28-58.

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

Cohen, A. S., Alpert, M., Nienow, T. M., Dinzeo, T. J., & Docherty, N. M. (2008). Computerized analysis of negative symptoms in schizophrenia. *Journal of Psychiatric Research*, **42**, 827-836.

Cohen, A. S., Dinzeo, T. J., Nienow, T. M., Smith, D. A., Singer, B., & Docherty, N. M. (2005). Diminished emotionality and social functioning in schizophrenia. *Journal of Nervous & Mental Disease*, **193**, 796-802.

Cohen, A. S., & Docherty, N. M. (2004). Affective reactivity of speech and emotional experience in patients with schizophrenia. *Schizophrenia Research*, **69**, 7-14.

Cohen, A. S., & Docherty, N. M. (2005). Effects of positive affect on speech disorder in schizophrenia. *Journal of Nervous & Mental Disease*, **193**, 839-842.

Cohen, A. S., & Minor, K. S. (in press). Emotional experience in schizophrenia revisited: Meta-analysis of laboratory studies. *Schizophrenia Bulletin*.

Cohen, A. S., Minor, K. S., Baillie, L. S., & Dahir, A. (2008). Clarifying the linguistic signature: Measuring personality from natural speech. *Journal of Personality Assessment*, **90**, 559-563.

Cohen, A. S., St-Hilaire, A., Aakres, J. M., & Docherty, N. M. (in press). The emotional underpinnings of anhedonia in schizophrenia: Lexical analysis of natural speech. *Cognition & Emotion*.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, **104**, 268-294.

Decety, J., & Lamm, C. (2006). Human empathy through the lens of social neuroscience. *Scientific World Journal*, **6**, 1146-1163.

Gomez, P., & Danuser, B. (2004). Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology*, **53**, 91-103.

GROSS, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, **39**, 281-291.

HAGENAARS, M. A., & VAN MINNEN, A. (2005). The effect of fear on paralinguistic aspects of speech in patients with panic disorder with agoraphobia. *Journal of Anxiety Disorders*, **19**, 521-537.

KAHN, J. H., TOBIN, R. M., MASSEY, A. E., & ANDERSON, J. A. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *American Journal of Psychology*, **120**, 263-286.

LAI, S.-C., MAYER-KRESS, G., SOSNOFF, J. J., & NEWELL, K. M. (2005). Information entropy analysis of discrete aiming movements. *Acta Psychologica*, **119**, 283-304.

LANE, S. M., & SCHOOLER, J. W. (2004). Skimming the surface: Verbal overshadowing of analogical retrieval. *Psychological Science*, **15**, 715-719.

LANG, P. J., BRADLEY, M. M., & CUTHBERT, B. N. (1999). *International Affective Picture System: Instruction manual and affective ratings* (Tech. Rep. No. A-4). Gainesville: University of Florida, Center for Research in Psychophysiology.

LANG, P. J., BRADLEY, M. M., & CUTHBERT, B. N. (2005). *International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual* (Tech. Rep. No. A-6). Gainesville: University of Florida.

LARSEN, R. J., & DIENER, E. (1992). Promises and problems with the circumplex model of emotion. In M. E. Clark (Ed.), *Review of personality & social psychology: 13. Emotion* (pp. 25-59). Thousand Oaks, CA: Sage.

LAUKKA, P. (2005). Categorical perception of vocal emotion expressions. *Emotion*, **5**, 277-295.

LEDOUX, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, **23**, 155-184.

LEVENTHAL, A. M., CHASSON, G. S., TAPIA, E., MILLER, E. K., & PETTIT, J. W. (2006). Measuring hedonic capacity in depression: A psychometric analysis of three anhedonia scales. *Journal of Clinical Psychology*, **62**, 1545-1558.

MARKEL, N. N., BEIN, M. F., & PHILLIS, J. A. (1973). The relationship between words and tone-of-voice. *Language & Speech*, **16**, 15-21.

MATESE, M., MATSON, J. L., & SEVIN, J. (1994). Comparison of psychotic and autistic children using behavioral observation. *Journal of Autism & Developmental Disorders*, **24**, 83-94.

MCADAMS, D. P. (2001). The psychology of life stories. *Review of General Psychology*, **5**, 100-122.

NELSON, K. L., & HOROWITZ, L. M. (2001). Narrative structure in recounted sad memories. *Discourse Processes*, **31**, 307-324.

PENNEBAKER, J. W., BOOTH, R. J., & FRANCIS, M. E. (2007). Linguistic inquiry and word count (LIWC 2007): A text analysis program. Austin, TX: www.liwc.net.

PENNEBAKER, J. W., & KING, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality & Social Psychology*, **77**, 1296-1312.

PENNEBAKER, J. W., MEHL, M. R., & NIEDERHOFFER, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, **54**, 547-577.

ROSENBERG, S. D., SCHNURR, P. P., & OXMAN, T. E. (1990). Content analysis: A comparison of manual and computerized systems. *Journal of Personality Assessment*, **54**, 298-310.

SCHERER, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, **40**, 227-256.

SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379-423.

SOBIN, C., & ALPERT, M. (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research*, **28**, 347-365.

SOUTH, M., OZONOFF, S., SUCHY, Y., KESNER, R. P., MCMAHON, W. M., & LAINHART, J. E. (2008). Intact emotion facilitation for nonsocial stimuli in autism: Is amygdala impairment in autism specific for social information? *Journal of the International Neuropsychological Society*, **14**, 42-54.

STONE, P. J., DUNPHY, D. C., SMITH, M. S., & OGILVIE, D. M. (1966). *The general inquirer: A computer approach to content analysis*. Cambridge, MA: MIT Press.

TOLKMITT, F. J., & SCHERER, K. R. (1986). Effect of experimentally induced stress on vocal parameters. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 302-313.

VAN LANCKER SIDTIS, D., PACHANA, N., CUMMINGS, J. L., & SIDTIS, J. J. (2006). Dysprosodic speech following basal ganglia insult: Toward a conceptual framework for the study of the cerebral representation of prosody. *Brain & Language*, **97**, 135-153.

VELTEN, E., JR. (1968). A laboratory task for induction of mood states. *Behaviour Research & Therapy*, **6**, 473-482.

VERVERIDIS, D., & KOTROPOULOS, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication*, **48**, 1162-1181.

WATSON, D., WIESE, D., VAIDYA, J., & TELLEGEN, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality & Social Psychology*, **76**, 820-838.

WESTERMANN, R., SPIES, K., STAHL, G., & HESSE, F. W. (1996). Relative effectiveness and validity of mood induction procedures: A meta-analysis. *European Journal of Social Psychology*, **26**, 557-580.

WISEMAN, D., & LEVIN, I. P. (1995). A new laboratory method for altering positive affect. *Psychological Reports*, **76**, 1103-1106.

ZECH, E. (1999). Is it really helpful to verbalise ones emotions? *Gedrag & Gezondheid: Tijdschrift voor Psychologie en Gezondheid*, **27**, 42-47.

NOTE

1. To illustrate the breadth of prosodic and content-analytic studies to date, we conducted PsycINFO literature searches. The terms "prosod*" or "acoustic" and "emotion" or "express*" yielded 719 peer-reviewed studies, and the terms "content analysis" and "emotion" or "express*" yielded 634 peer-reviewed studies. Once these search terms were combined, however, the search yielded only two peer-reviewed studies (only one of which was relevant to the present study; Alpert et al., 2000). This highlights the dearth of research combining prosodic and content analyses of human expression.

APPENDIX

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i$$

$H = -[10/120 \times \log_2 (10/20) +$
$\quad 20/120 \times \log_2 (20/120) +$
$\quad 30/120 \times \log_2 (30/120) +$
$\quad 30/120 \times \log_2 (30/120) +$
$\quad 20/120 \times \log_2 (20/120) +$
$\quad 10/120 \times \log_2 (10/120)]$
$\quad = \textbf{2.46}$ bits

$H = -[0 + 0 + 120/120 \times \log_2$
$\quad (120/120) + 0 + 0 + 0]$
$\quad = 1 \times 0$
$\quad = \textbf{0.00}$ bits

Zero Entropy

$H = -[20/120 \times \log_2 (20/120)$
$\quad +20/120 \times \log_2 (20/120)$
$\quad +20/120 \times \log_2 (20/120)$
$\quad +20/120 \times \log_2 (20/120)$
$\quad +20/120 \times \log_2 (20/120)$
$\quad +20/120 \times \log_2 (20/120)]$
$\quad = -\log_2 (20/120)$
$\quad = \textbf{2.59}$ bits
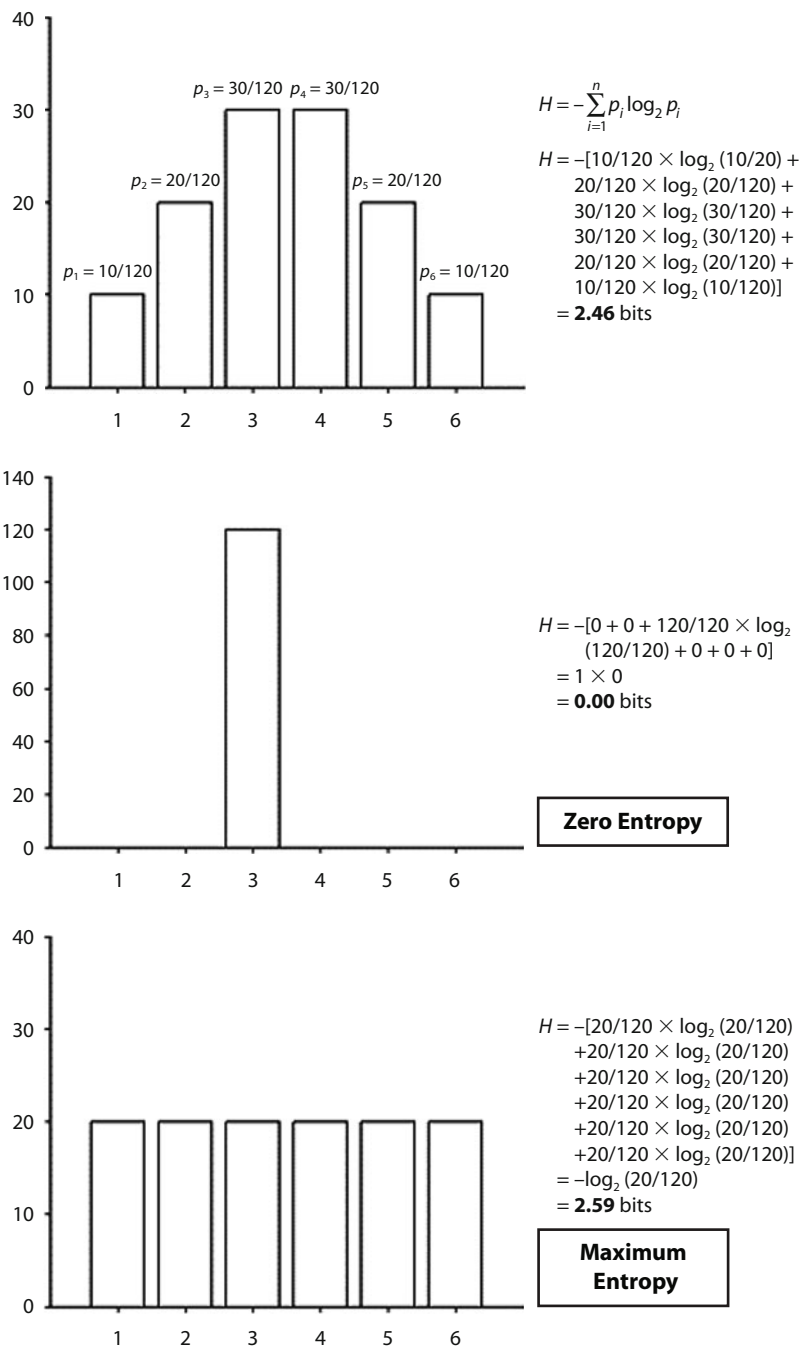
Maximum Entropy

Figure A1. A simplified illustration of the information entropy calculation procedure. This example presents a situation in which a die is rolled 120 times. The upper panel shows how an entropy value is derived from the probability distribution. The lower panels show the extreme cases of zero and maximum entropy.