tapraid1/emo-emo/emo-emo/emo00405/emo0081d05g | yodert | S=4 | 11/8/05 | 12:18 | Art: E02-037R4

Emotion
2005, Vol. 5, No. 4, 000–000

1528-3542/05/$12.00 DOI: 10.1037/1528-3542.5.4.000

Affective Speech Elicited With a Computer Game

Tom Johnstone
University of Western Australia and University of Geneva

Carien M. van Reekum
University of Geneva

Kathryn Hird
Curtin University of Technology

Kim Kirsner
University of Western Australia

Klaus R. Scherer
University of Geneva

To determine the degree to which emotional changes in speech reflect factors other than arousal, such as valence, the authors used a computer game to induce natural emotional speech. Voice samples were elicited following game events that were either conducive or obstructive to the goal of winning and were accompanied by either pleasant or unpleasant sounds. Acoustic analysis of the speech recordings of 30 adolescents revealed that mean energy, fundamental-frequency level, utterance duration, and the proportion of an utterance that was voiced varied with goal conduciveness; spectral energy distribution depended on manipulations of pleasantness; and pitch dynamics depended on the interaction of pleasantness and goal conduciveness. The results suggest that a single arousal dimension does not adequately characterize a number of emotion-related vocal changes, lending weight to multidimensional theories of emotional response patterning.

Keywords: emotion, speech, prosody, voice, arousal

Anyone who has felt his or her voice quaver and throat constrict while nervously giving a public speech is well aware of the effects that emotions can have on our speech. Often these effects are only obvious to the speaker, although sometimes the outward signs of this anxiety can be evident to empathetic listeners, who are sensitive to the high-pitched, trembling voice of an anxious presenter. What, then, is known about how our emotions affect the way we produce speech? The answer is relatively little, particularly compared with facial expression of emotion, which has been the subject of much scrutiny since the pioneering work by Ekman and Izard (e.g., Ekman, 1972; Izard, 1971).

Tom Johnstone, Department of Psychology, University of Western Australia, Perth, Australia, and Department of Psychology, University of Geneva, Geneva, Switzerland; Carien M. van Reekum and Klaus R. Scherer, Department of Psychology, University of Geneva; Kathryn Hird, School of Psychology, Curtin University of Technology, Perth, Australia; Kim Kirsner, Department of Psychology, University of Western Australia.

Carien M. van Reekum is now at the Department of Psychology, University of Wisconsin—Madison.

This study was conducted as part of Tom Johnstone's doctoral dissertation. Preliminary results from the study were presented at the 4th International Conference on Spoken Language Processing in Philadelphia. The study was funded by Swiss National Science Fund Grants 1114-037504.93 and 2151-049685.96. We thank Rainer Banse, Alex Etter, and Thomas Wehrle for their assistance with this study. AQ: 3

Correspondence concerning this article should be addressed to Tom Johnstone, who is now at the Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin—Madison, 1500 Highland Avenue, T–127, Madison, WI 53705. E-mail: itjohnstone@wisc.edu

Reviews of the literature on emotional speech have concluded that at least for acted vocal expressions of emotion, recognition rates are comparable, though slightly lower, than they are for facial expressions (Scherer, 1999). Acted vocal expressions of emotion are also recognized well across cultures (Scherer, Banse, & Wallbott, 2001; van Bezooijen, 1984), indicating that the vocal expression of emotion reflects processes that function largely independently of the mechanisms for production of a given spoken language. Attempts to identify the specific vocal characteristics that are used by listeners to infer emotional states of speakers (see reviews by Johnstone & Scherer, 2000; Scherer, 1986) indicate that many measured differences in acoustical patterns across emotions are consequences of the level of physiological arousal (in the sense of an excitation of the sympathetic branch of the autonomic nervous system) that accompanies each emotion. Thus, emotions such as anger, fear, and joy are all characterized by raised fundamental frequency (f_0) and high intensity, whereas emotions such as sadness and boredom are expressed with low f_0 and low intensity. However, Scherer (1986) pointed out that the ability of judges to accurately judge expressed emotions means that parameters that differentiate emotions with similar arousal levels must exist (at least in acted speech), and suggested that a broader set of acoustic parameters would need to be analyzed in future research. More recent studies that included an extensive acoustic analysis of emotional speech, including a variety of spectral parameters, allowed consistent differentiation of emotions with similar arousal levels (e.g., Bachorowski & Owren, 1995b; Banse & Scherer, 1996; Sobin & Alpert, 1999; see also Juslin & Laukka, 2002).

tapraid1/emo-emo/emo-emo/emo00405/emo0081d05g | yodert | S=4 | 11/8/05 | 12:18 | Art: E02-037R4

Given the predominant use of acted speech in most studies of vocal emotion expression, however, the question remains as to whether the effects of real or naturally occurring emotion on vocal production are distinguishable on the basis of acoustic cues. The presence of emotion-specific acoustic patterns in acted speech, as have been found in previous research, might be attributable to the strategic adoption by the actors of speaking styles that serve to send a signal to cohorts (Russell, Bachorowski, & Fernandez-Dols, 2003). More direct effects of emotion on speech, termed *push* effects by Scherer (1985), presumably reflect the relatively uncontrolled changes to the underlying physiology of speech production that accompany an emotion.

A small number of studies have attempted real emotion induction or have measured "real-life" emotional speech recordings (e.g., Alpert, Kurtzberg, & Friedhoff, 1963; Bachorowski & Owren, 1995b; Duncan, Laver, & Jack, 1983; Simonov & Frolov, 1973). These studies have, however, used predominantly bipolar inductions, such as high–low stress. It is thus not surprising that the results obtained could be explained in terms of a single dimension of arousal. Scherer (1986) has suggested, however, that differences in the vocal characteristics of emotional speech should reflect the three dimensions of emotional response frequently reported (i.e., arousal, valence, and potency). The little empirical evidence for the existence of three dimensions in emotional speech (Green & Cliff, 1975) indicates that of the three, arousal and valence are more easily identifiable in the acoustics of speech than is potency. There is, thus, a clear need for further studies that induce emotional states that vary on dimensions other than just arousal, such as valence.

A number of techniques have been used by emotion researchers to induce emotions in the laboratory (e.g., Gerrards-Hesse, Spies, & Hesse, 1994). In this study, we adopted the approach of inducing emotion-related vocal responses in the laboratory using a computer game (see Kappas & Pecchinenda, 1999; MacDowell & Mandler, 1989). We sought to identify affective changes to the acoustics of speech that could be attributed to a valence-associated response dimension by analyzing responses to events in a computer game intended to be pleasant or unpleasant and obstructive or conducive to performing well in the game.

Method

Participants

Thirty-three volunteers between the ages of 13 and 15 years (27 boys, 6 girls) were recruited from schools in the Geneva area (data from 3 of the boys were not analyzed because of technical problems with the speech recordings). The schools and parents of all children gave fully informed written consent for their children's participation. Participants were reimbursed SFr. 15 (U.S. $10). Adolescents were chosen because they were considered likely to be familiar with, and get emotionally involved in, video games.

Equipment

Participants were fitted with an AKG C420 headset condenser microphone, 20 cm below and to the side of the mouth, connected to a Sony TCD-D8 DAT recorder. Speech was recorded digitally at a sampling rate of 44.1 kHz. Subjects were also fitted with electrodes for the recording of a variety of physiological measures as part of a concurrent study (van Reekum et al., 2004).

Description of the Game

The game, XQuest (Mackey, 1994), situates the player's space ship in a galaxy filled with crystals, mines, and enemies. The player uses a mouse to accelerate and to fire at the enemies. The goal is to gather all the crystals in each galaxy, after which the player proceeds to the next galaxy (i.e., game level). Points and extra ships are awarded for crystals gathered, enemies destroyed, and rapid completion of a game level. After losing all of the ships, the player starts a new game at the first game level.

Procedure

After a general introduction to the computer game, participants watched an instructional demonstration of the game. Players were shown how to respond to the emotion-rating screen and verbal-report screen. They then practiced the game for 20 min, during which time they were given extra instruction when necessary. All participants demonstrated sufficient proficiency by reaching at least the fourth game level during this practice period. Physiological sensors and the microphone were then attached, and the microphone recording level was adjusted. After 2.5 min of relaxation, the participants played the game for 45 min.

Selected Game Events

We sought to examine emotional changes to speech production that were associated with specific game events that were manipulated because of their putative positive or negative valence. In the context of the game, completing a game level is conducive and losing a ship is obstructive to the pursuit of gaining points and progressing through game levels. The variable goal conduciveness was thus operationalized by the selection of situations in which a game level was successfully completed (goal conducive) or the player's ship was destroyed (goal obstructive). We also directly manipulated the intrinsic pleasantness of the two game events by concurrently playing 1-s valenced (i.e., pleasant and unpleasant) synthesized sounds equal in mean acoustic intensity. Both sounds had approximately flat spectra to 2 kHz. The pleasant sound had slightly less energy, in the range of 2 kHz–5 kHz, than the unpleasant sound. The sounds had been independently rated by 15 judges on a 7-point scale from -3 (*very unpleasant*) to $+3$ (*very pleasant*). The mean rating for the unpleasant sound was -2.3 ($SD = 1.1$), and the mean rating for the pleasant sound was 2.2 ($SD = 0.8$). Intrinsic pleasantness was manipulated orthogonally to goal conduciveness, leading to a 2 (intrinsic pleasantness) $\times$ 2 (goal conduciveness) within-subject design.

Vocal Reports

Speech was elicited with a vocal-report screen that requested a report of the immediately preceding game events whenever an experiment-relevant event (i.e., loss of ship or new level) occurred. To maintain the continuity of the game, the screen appeared, at most, every 2 min. Players were requested to respond to the screen by pronouncing aloud a seven-character alphanumeric identification code, choosing the reason that matched most closely their explanation of the preceding event (chosen from four short-sentence alternatives), and estimating the percentage chance that they would be successful in the following game level. The screen provided both strings of isolated letters and connected phrases to be pronounced by the participant.

Emotion Self-Report

Emotion self-reports were obtained using a screen that displayed a popular French comic-strip character (Gaston Lagaffe), expressing eight

tapraid1/emo-emo/emo-emo/emo00405/emo0081d05g | yodert | S=4 | 11/8/05 | 12:18 | Art: E02-037R4

different emotions ("interest", "joy", "surprise", "anger", "shame", "pride", "tenseness", and "helplessness"). The images were accompanied by the corresponding emotion labels, each with a 100-point graphic scale on which the participants could use the mouse to indicate the felt intensity of each. The rating screen was presented immediately after a random sample of experiment-relevant events, but not more often than once every 4 min.

Acoustic Analyses

Acoustic analyses of each speech recording were carried out using Kay Computer Speech Laboratory 5300B speech analysis hardware and software (Kay Elemetrics, xxxx), with acoustic parameters chosen on the basis of their utility in previous research on emotional speech (see Johnstone & Scherer, 2000).

Fundamental frequency (f_0). For each speech file, Kay Computer Speech Laboratory software was used to mark the onset of each pitch period, with the constraint that F_0 was between 150 Hz and 400 Hz (the vocal folds of adolescents of the age range studied here typically vibrate at between 150 and 400 times/s). Obvious errors in the pitch extraction were manually corrected. For 5 participants for whom there were many errors, the minimum and maximum allowed f_0 values were adjusted on the basis of visual inspection of the speech waveform, and the speech files were reanalyzed and pitch periods reinspected. A single adjustment of allowed f_0 values was sufficient to ensure accurate calculation of f_0 in all cases.

The following f_0-related statistics were derived: mean f_0, standard deviation of f_0, f_0 5th percentile value, and f_0 95th percentile value. They provide a measure of f_0 central tendency, variability, floor, and ceiling, respectively.

Energy. The mean voiced energy of speech was quantified by calculating the root mean square (RMS) value of 15 ms frames of the speech signal, centered about each pitch-period marker (Deller, Proakis, & Hansen, 1993). This 15 ms window was long enough to ensure that energy is averaged over two to three fundamental periods.

Temporal measures. The length of each utterance was quantified as the time from the first pitch-period marker to the last marker in each speech file. Although this estimate ignores unvoiced sounds at the endpoints of each utterance, such unvoiced sounds were not expected to vary greatly between experimental conditions compared with voiced parts of the utterance. The proportion of each speech utterance that was voiced was estimated from the pitch-impulse markers.

Spectral measures. We calculated the average power spectrum of voiced parts of each utterance using a frame size of 512 samples, yielding 256 frequency bins, each one 39.06 Hz in width. The proportions of total energy under 500 Hz and under 1,000 Hz were calculated.

Results

Performance in the Game

On average, participants played 27 full games (SD = 10). All participants attained the 5th game level at least once.

Emotion Reports

Detailed results for the emotion reports are provided in van Reekum et al. (2004; Table 1). In summary, joy and pride were significantly higher in conducive conditions (mean rating 20 and 24, respectively) than in obstructive conditions (mean rating 14 and 9, respectively), whereas anger and surprise were significantly higher in obstructive conditions (mean rating 15 and 9, respectively) than in conducive conditions (mean rating 5 and 5, respectively). Thus, as intended, the two game events differed in elicited valence, with obstructive events leading to reports of greater negative emotion (anger) and conducive events leading to reports of more positive emotion (joy and pride). Surprise was also higher following events accompanied by pleasant sounds than those accompanied by unpleasant sounds. T1

We calculated correlation coefficients between the mean ratings and mean acoustic measures for each experimental condition after partialing the main random effect of participants. Reported anger was correlated negatively with utterance duration ($r = -.27$, $p < .01$), and positively with f_0 ceiling ($r = .19$, $p < .05$). Utterance duration was longer for game events in which participants reported more joy ($r = .20$, $p < .05$) and pride ($r = .34$, $p < .01$). The percentage of each utterance that was voiced was negatively associated with reports of joy ($r = -0.24$, $p < .01$) and pride ($r = -0.28$, $p < .01$). Mean acoustic energy was lower for game events that elicited more joy ($r = -0.20$, $p < .05$). Reports of helplessness were correlated negatively with the proportion of spectral energy below 500Hz ($r = -0.20$, $p < .05$). These outcomes show that some correlations exist between reported positive and negative

Table 1
Means of Measured Acoustic Parameters for the Four Experimental Conditions, for Both High f_0 and Low f_0 Subgroups

	Conducive pleasant		Conducive unpleasant		Obstructive pleasant		Obstructive unpleasant		
Subgroup	Low	High	Low	High	Low	High	Low	High	SE^a
EQ: 2 Mean f_0 (Hz)	140.77	236.85	140.19	235.67	140.46	235.34	141.23	237.87	0.65
f_0 floor (Hz)	109.19	199.97	108.69	199.04	110.81	201.41	113.31	200.17	0.86
f_0 SD (Hz)	21.19	24.11	20.58	23.66	19.35	22.25	19.54	25.24	0.40
f_0 ceiling (Hz)	170.73	274.80	172.77	271.35	166.35	270.28	168.46	279.07	1.47
Mean energy (dB)	68.76	70.36	68.20	70.45	69.01	70.52	69.67	70.75	0.13
Energy < 500 Hz (%)	95.00	91.62	93.74	91.14	94.48	93.28	94.44	91.63	0.003
Energy < 1000 Hz (%)	99.20	99.40	98.68	99.38	99.18	99.57	99.23	99.29	0.001
% of utterance voiced	27.96	30.07	28.77	29.20	30.73	34.16	34.89	34.32	0.22
Utterance duration (s)	10.12	10.48	9.68	10.41	9.11	9.53	9.05	8.68	0.63

Note. f_0 = fundamental frequency.
[a] Values are within-subject standard errors.

emotions and some of the vocal parameters measured, although such correlations are modest.

Univariate Analyses

The variation of the acoustical parameters across experimental manipulations was tested using univariate mixed-model analysis of variance, with conduciveness and pleasantness as fixed factors and participant as a random factor.

Sex effects. The sex of the participants might conceivably influence the way in which vocal production changes with different experimental conditions. Because there were only 6 girls in this study, it is difficult to rigorously assess such sex effects. However, sex showed no significant interactions with conduciveness or pleasantness for any acoustic measure (all p s > 0.1). The boys in this study exhibited great variability in vocal characteristics, particularly those related to f_0, because of variability in their stages of adolescent development. To address this variability, we split the participants into two subgroups: those with high f_0 (mean $f_0 > 180$ Hz) and those with low f_0 (mean $f_0 < 180$ Hz). There were no significant interactions of subgroup with conduciveness or pleasantness for any acoustic measure (all p s > 0.1). Thus, for this sample of adolescent participants, the vocal changes that corresponded to different game events were similar for boys and girls and similar across participants with different levels of mean f_0.

EQ: 1

EQ: 1

Means and standard errors for the different acoustic variables, for both high f_0 and low f_0 subgroups, are provided in Table 1.

Pleasantness × Conduciveness interaction. There was very little interaction between the effects of the two variables on the vocal measures. A weak interaction was observed for f_0 ceiling, $F(1, 30) = 2.9$, $p = .10$. Post hoc comparisons showed that this statistical trend was due to the f_0 ceiling being higher in response to unpleasant than to pleasant sounds that accompanied obstructive events, $F(1, 30) = 4.4$, $p = .04$, with no such difference for conducive events, $F(1, 35) = 0.2$, $p = .64$. An interaction for f_0 standard deviation, $F(1, 30) = 5.0$, $p = .03$, was due to higher f_0 standard deviation in response to unpleasant than to pleasant sounds that accompanied obstructive events, $F(1, 32) = 5.6$, $p = .02$, with no such difference for conducive events, $F(1, 33) = 0.6$, $p = .46$.

Pleasantness. No significant effects of pleasantness on mean energy, $F(1, 31) = 0.4$, $p = .54$; f_0 floor, $F(1, 31) = 0$, $p = 1$; f_0 ceiling, $F(1, 30) = 2.1$, $p = .15$; mean f_0, $F(1, 31) = 0.1$, $p = .7$; or f_0 standard deviation, $F(1, 30) = 1.2$, $p = .28$, were observed. The proportion of energy below 500 Hz was significantly lower for unpleasant than for pleasant sounds, $F(1, 31) = 7.3$, $p = .01$. A similar result was found for the proportion of energy under 1000 Hz, $F(1, 29) = 4.2$, $p = .05$. These differences were due to a small but consistent difference in the spectra at high frequencies, with speech in the unpleasant condition showing greater high-frequency energy than speech in the pleasant condition.

Conduciveness. Mean energy was lower for conducive than for obstructive events, $F(1, 29) = 6.4$, $p = .02$. The f_0 floor was lower for conducive events than for obstructive events, $F(1, 30) = 4.6$, $p = .04$, although we found no effects of conduciveness on f_0 ceiling, $F(1, 30) = 0.1$, $p = .7$; mean f_0, $F(1, 31) = 0.5$, $p = .5$; or f_0 standard deviation, $F(1, 30) = 1.2$, $p = .27$. The proportion of each utterance that was voiced was lower for conducive than for obstructive events, $F(1, 30) = 23.4$, $p < .01$. Utterance duration was higher for conducive events than for obstructive events, $F(1, 30) = 22.0$, $p < .01$. The associations between conduciveness and mean energy, f_0 floor, and the temporal parameters are shown in Figure 1. No significant differences between conducive and obstructive events were found for the proportion of energy under 500 Hz, $F(1, 30) = 1.8$, $p = .19$, nor for the proportion of energy under 1,000 Hz, $F(1, 29) = 1.3$, $p = .27$. AQ: 1 F1

Discussion

Speech following obstructive events was higher in energy and had a higher f_0 level, as indicated by f_0 floor, than speech following conducive events. These results suggest that physiological arousal was higher following the destruction of a ship than following the completion of a game level. This interpretation is supported by measurements of skin conductance (a measure that reflects sympathetic autonomic nervous system arousal), taken in a concurrent study, which were higher following obstructive events than following conducive events (van Reekum et al., 2004). No significant spectral differences were measured between conducive and obstructive events.

In contrast to the conduciveness findings, the significant effects of the pleasantness manipulation on the acoustic speech signal were limited to the distribution of energy in the spectrum, with a greater proportion of energy in higher frequencies being measured after unpleasant sounds than after pleasant sounds. In addition, for obstructive events only, f_0 had a more reduced dynamic range for pleasant than for unpleasant sounds. This result is difficult to explain in terms of either an arousal or a valence dimension.

In summary, this experiment revealed that variations in the intrinsic pleasantness of an event cause changes to spectral energy distribution, but not to overall energy, f_0, or the measured temporal parameters, and that changes to the conduciveness of an event are associated with changes to the latter set of variables but not to spectral energy distribution. This pattern of results suggests that emotional changes to the voice reflect two or more dimensions, presumably reflecting two or more underlying mechanisms. This general outcome is consistent with both evidence and theory that posits that at least two, and possibly three, dimensions characterize emotional responses: activation, valence, and potency/power (e.g., Davitz, 1964; Osgood, Suci, & Tannenbaum, 1957; Russell, 1980; Wundt, 1909).

Despite the widespread acceptance of a two- or three-dimensional description of emotional responses, there is very little empirical evidence supporting such a view with respect to putative push effects on the voice. Recall that Green and Cliff (1975) arrived at a three-dimensional description of vocal changes in acted emotional speech. The results from this experiment also provide support for two of these dimensions: Changes in f_0 and mean energy could clearly be related to activation. In addition, the spectral distribution of energy, which varied significantly with manipulations of intrinsic pleasantness, seems to match the description given by Scherer (1986) for a hedonic valence dimension. The other acoustic variables that showed differences across experimental conditions—f_0 dynamics, utterance duration, and proportion of speech that was voiced—do not clearly map on to either dimension.

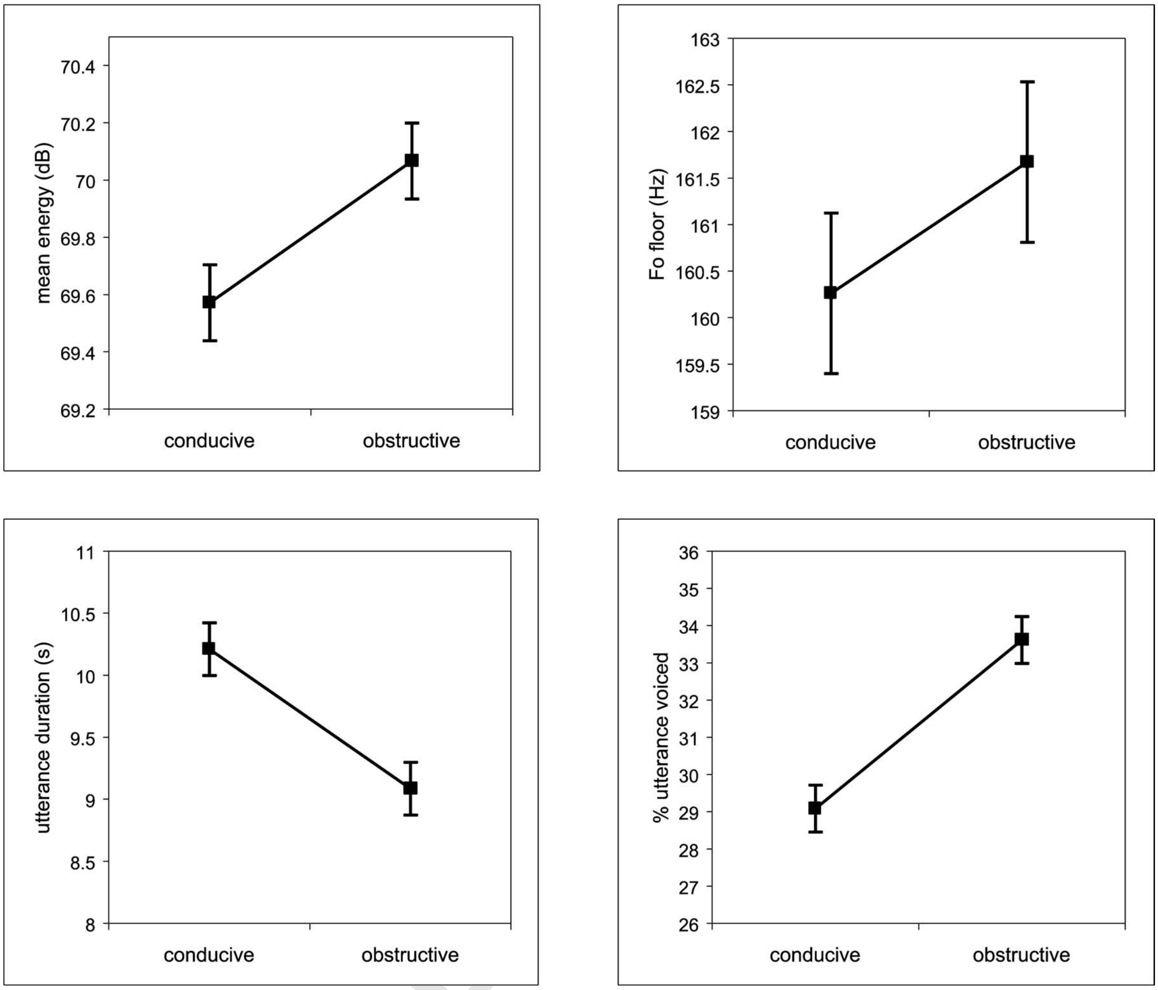

Figure 1. Mean voiced energy in decibels (top left), F_0 floor (top right), mean utterance duration (bottom left), and the percentage of each utterance that was voiced (bottom right) as a function of conduciveness. Error bars represent $\pm$ 1 standard error.

The results from this experiment indicate the potential for computer games to induce measurable emotional changes to the acoustic properties of speech. It is not clear, however, whether the measured acoustic differences in speech between the experimental conditions, which in this experiment were small, would be perceptible. Nor can it be completely ruled out that players' speech was influenced to some extent by social context. It would be preferable in future studies of this type to measure the acoustic changes to speech during, rather than following, game events, and under different social and experiment contexts (e.g., Stemmler, 1992). In addition, measuring relevant physiological variables, such as subglottal and supraglottal air pressure and laryngeal and articulator muscular tension and dynamics, would aid in the interpretation of acoustic measurements. Although such methods might preclude realistic induction of a range of emotional responses, the use of noninvasive devices, such as the electroglottograph, to obtain some indication of changes to vocal-fold function in naturally occurring or induced emotional states is currently feasible. Such studies of induced or natural emotional speech will provide valuable complementary information about the causes and functions of emotional expression.

References

Alpert, M., Kurtzberg, R. L., & Friedhoff, A. J. (1963). Transient voice changes associated with emotional stimuli. *Archives of General Psychiatry, 8,* 362–365.

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

Bachorowski, J. A., & Owren, M. J. (1995b, November). *Vocal expression of emotion is associated with formant characteristics.* Paper presented at the 130th meeting of the Acoustical Society of America, St. Louis, MO.

Banse, R., & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70,* 614–636.

Davitz, J. R. (1964). Auditory correlates of vocal expression of emotional feeling. In J. R. Davitz (Ed.), *The communication of emotional meaning* (pp. 101–112). New York: McGraw-Hill.

Deller, J. R., Proakis, J. G., & Hansen, J. H. L. (1993). *Discrete time processing of speech signals.* New York: Macmillan.

Duncan, G., Laver, J., & Jack, M. A. (1983). *A psycho-acoustic interpretation of variations in divers' voice fundamental frequency in a pressured helium-oxygen environment* (Work in Progress Report Ser. No. 16 (S. 9–16)). Edinburgh, Scotland: University of Edinburgh, Department of Linguistics.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotions. In J. Cole (Ed.), *Nebraska Symposium on Motivation, 1971* (pp. 207–283). Lincoln: University of Nebraska Press.

Gerrards-Hesse, A., Spies, K., & Hesse, F. W. (1994). Experimental inductions of emotional states and their effectiveness: A review. *British Journal of Psychology, 85,* 55–78.

Green, R. S., & Cliff, N. (1975). Multidimensional comparisons of structures of vocally and facially expressed emotion. *Perception and Psychophysics, 17,* 429–438.

Izard, C. E. (1971). *The face of emotion.* New York: Appleton-Century-Crofts.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of emotions,* (2nd ed., pp. 220–235). New York: Guilford Press.

Juslin, P. N., & Laukka, P. (2002). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion, 1,* 381–412.

Kappas, A., & Pecchinenda, A. (1999). Don't wait for the monsters to get you: A video game task to manipulate appraisals in real time. *Cognition and Emotion, 13,* 119–124.

AQ: 2 Kay Elemetrics. (xxxx). Computer Speech Laboratory 5300B [Computer software]. Lincoln Park, NJ: Author.

MacDowell, K. A., & Mandler, G. (1989). Constructions of emotion: Discrepancy, arousal, and mood. *Motivation and Emotion, 13,* 105–124.

Mackey, M. (1994). XQuest [Computer software]. Cambridge, England: Cambridge University.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178.

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

Scherer, K. R. (1985). Vocal affect signalling: A comparative approach. In J. Rosenblatt, C. Beer, M. Busnel, & P. J. B. Slater (Eds.), *Advances in the study of behavior* (pp. 189–244). New York: Academic Press.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*(2), 143–165.

Scherer, K. R. (1999). Universality of emotional expression. In D. Levinson, J. Ponzetti, & P. Jorgenson (Eds.), *Encyclopedia of human emotions* (Vol. 2, pp. 669–674). New York: Macmillan.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*(1), 76–92.

Simonov, P. V., & Frolov, M. V. (1973). Utilization of human voice for estimation of man's emotional stress and state of attention. *Aerospace Medicine, 44,* 256–258.

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

Stemmler, G. (1992). The vagueness of specificity: Models of peripheral physiological emotion specificity in emotion theories and their experimental discriminability. *Journal of Psychophysiology, 6,* 17–28.

van Bezooijen, R. (1984). *The characteristics and recognizability of vocal expression of emotions.* Dordrecht, the Netherlands: Foris.

van Reekum, C. M., Johnstone, T., Banse, R., Etter, A., Wehrle, T., & Scherer, K. R. (2004). Psychophysiological responses to appraisal dimensions in a computer game. *Cognition and Emotion, 18,* 663–688.

Wundt, W. (1909). *Grundriss der psychologie. Achte auflage* [*Outlines of psychology*]. Leipzig, Germany: Wilhelm Engelmann.

Received May 3, 2002
Revision received May 30, 2005
Accepted May 31, 2005 ■