Abstract To determine how sad affect (or brief sad mood) interacts with paralinguistic aspects of speech, we investigated the effect of a happy or sad mood induction on speech production in 49 healthy volunteers. Several speech parameters measuring speech rate, loudness and pitch were examined before and after a standardized mood-induction procedure that involved viewing facial expressions. Speech samples were collected during the self-initiated reading of emotionally “neutral” sentences; there was no explicit demand to produce mood-congruent speech. Results indicated that, after the mood induction, the speech of participants in the sad group was slower, quieter and more monotonous than the speech of participants in the happy group. This speech paradigm provides a model for studying how changes in mood states interact with the motor control of speech.

Keywords Affect · Speech production · Mood induction

Introduction

In recent years, much research has focused on the brain mechanisms underlying major depression. The most robust finding, obtained from brain imaging studies of depressed patients scanned in a resting state, has been that of hypoperfusion and hypometabolism in the prefrontal cortex and the anterior cingulate cortex (Baxter et al. 1985; Bench et al. 1992; Drevets 2001; Mayberg et al. 1997). Knowledge of this apparent “hypofrontality” in depression led to the development of repetitive transcranial magnetic stimulation (rTMS) as a treatment option for drug-resistant depressed patients. Repetitive TMS is a relatively non-invasive and painless method of stimulating the cortex. To treat depression, rTMS is typically applied over the prefrontal cortex (for 2–12 sessions) and mood change is monitored by the use of a subjective rating scale (see George et al. 1997; Grunhaus et al. 2000). In this study, we hoped to facilitate future investigations of rTMS-induced mood change by developing an objective measure to quantify the influence of affect (i.e., brief mood states) on a particular motor behavior, namely, paralinguistic aspects of speech. Whereas much research has examined the acoustical properties of emotional speech and the discrimination of emotion through speech perception (for a review of this research area see Pittam and Scherer 1993), less is known about how changes in affect influence speech production. Cues about affective state are often revealed through paralinguistic aspects of speech production such as pitch, loudness and speech rate (Pittam and Scherer 1993; Scherer 1989). Many researchers have taken advantage of this phenomenon to study both normal affect as well as depression. Using quantitative analyses, it is commonly observed that the speech of depressed patients is slow, quiet and monotonous (Flint et al. 1993; Garcia-Toro et al. 2000; Godfrey and Knight 1984; Nilsonne 1988; Stassen et al. 1991; Talavera et al. 1994). The relationship between non-depressed sad affect and changes in speech quality is less known.

The speech parameters that have been most often studied during sad affect as well as depression are speech rate, loudness and pitch. Fundamental frequency ($F_0$), the dominating frequency of the sound produced by the vibration of the vocal folds, is a major contributor to perceived vocal pitch. It is a complicated physiological mechanism involving intrinsic and extrinsic laryngeal muscles, muscles involved in respiration (e.g., muscles of the chest cavity and abdomen), as well as feedback mechanisms involving sensory receptors (Larson 1998). $F_0$ variation across a speech sample reflects the amount of intonation in speech. An acoustical correlate of loudness, modulated primarily through the control of respiratory muscles, is root-mean-square amplitude (RMS-amplitude). Measuring the range of amplitude provides information about the variability of loudness throughout an
utterance. Finally, speech rate is typically measured by the speed at which utterances are produced or the length of pauses between utterances and syllables.

A number of studies have investigated the relationship between affect and paralinguistic aspects of speech, observing that speech during sad affect is often slower (i.e., lower speech rate) and/or quieter (i.e., lower RMS-amplitude) and/or more monotonous (i.e., lower range of F0) than speech during happy affect (Banse and Scherer 1996; McMenna and Lewis 1994; Scherer 1989; Scherer et al. 1991; Sobin and Alpert 1999). Inferences about the interaction between affect and speech based on this research are limited, however, by a number of important methodological issues. Firstly, most studies examining the relationship between affect and speech have studied speech samples of actors portraying sadness and happiness (for a recent example see Banse and Scherer 1996). While this method is informative as to why we perceive speech as being sad, happy or fearful, or how people may produce such speech, it does not allow for an understanding of how the actual experience of sad affect influences speech production. Secondly, a few studies have collected speech samples from affectively-laden utterances used to induce specific affect (for a recent example see Sobin and Alpert 1999). This strategy could make the demand for mood-congruent speech explicit or capture changes in speech production that are simply related to the presence of affective stimuli.

To address these issues, we developed a speech task that was administered in conjunction with a standardized happy or sad mood induction, or with a control (i.e., “neutral”) procedure. In order to prevent possible carry-over effects across the three experimental conditions, we used a between-subjects design with the following groups: happy, sad and control. The speech task was performed before, in the middle of, and after undergoing either a happy or sad mood induction. The speech task began with the presentation of a blank screen and, immediately after the cover story, the computer keyboard was positioned to allow the participant to imagine what would make the person in the picture express that emotion. For example, the face of an actor’s faces presented on a computer screen and were asked to use SuperLab Pro, v. 2.0 (2000; Cedrus, San Pedro, Calif., USA) on a desktop computer. Stimuli (200x200 cm) were presented on a 45 cm color monitor (1024x768 pixels, 85 Hz), approximately 90 cm away from where the participant was seated (stimulus size 12“x12”). The computer keyboard was positioned to allow the easy pressing of the space bar. All happy faces were viewed in the happy induction and all sad faces were viewed in the sad induction. To help participants achieve the desired mood, they were instructed to imagine what would make the person in the picture express that emotion, or to think of a personal event or memory that made them feel like the person in the picture. Each face was viewed for as much time as necessary to feel the desired emotion. When they were finished viewing the face, participants pressed the space bar to begin viewing the next picture until all 40 faces had been viewed. The speech task was a control measure (i.e., providing a “neutral state” for GSR collection), and that the speech task was to be considered a “break” from the mood procedure. Measures of speech rate (speech initiation and speech duration), loudness (RMS-amplitude), variation in loudness (range of amplitude), pitch (mean F0) and variation in pitch (range and standard deviation of F0) were obtained from the sentences.

Materials and methods

Overview

Participants read sets of sentences, out loud and at their own pace, before, in the middle of, and after undergoing either a happy or sad mood induction or a control procedure. To deter participants from changing the quality of their speech to conform to the demands of the mood procedure, the study was described as an investigation of galvanic skin response (GSR) and mood. Participants were told that the speech task was a control measure (i.e., providing a “neutral state” for GSR collection), and that the speech task was to be considered a “break” from the mood procedure. Measures of speech rate (speech initiation and speech duration), loudness (RMS-amplitude), variation in loudness (range of amplitude), pitch (mean F0) and variation in pitch (range and standard deviation of F0) were obtained from the sentences.

Participants

Sixty-three undergraduate students at McGill University participated in the study (56 women, 7 men). Ages ranged from 17 to 31 years, with a mean age of 19.48 (SD =1.68). Seventy-percent of participants reported English as their first language and all were fluent English speakers. Participants were recruited through a website organized by the Department of Psychology. For their participation, students received extra-credit in their psychology courses. The procedures of this experiment were approved by the Research Ethic Committee of the Montreal Neurological Institute and Hospital.

Affect questionnaire

Participants completed an affect questionnaire before and after the mood induction. The questionnaire was designed to assess levels of comfort, fatigue, irritation, mood, anxiety and emotional arousal. Ratings were made on a seven-point Likert scale, with –3 indicating the highest negative level and 3 indicating the highest positive level for each affective state. For example, the mood rating ranged from I feel very sad (–3) to I feel very happy (3) and the fatigue rating ranged from I feel very fatigued (–3) to I feel very rested (3). The questionnaire was used as a general indicator of how the participant was feeling upon arrival at the laboratory and to assess affect after the mood induction and control procedure.

Mood induction

Each participant underwent either a happy or sad mood induction or a control procedure. The mood induction was adapted from the face stimuli and procedure standardized by Schneider et al. (1994). This procedure has been shown to elicit temporary states of sadness and happiness in people not suffering from affective disturbance. Participants viewed 40 monochrome front-view photographs of actor’s faces presented on a computer screen and were asked to use SuperLab Pro, v. 2.0 (2000; Cedrus, San Pedro, Calif., USA) on a desktop computer. Stimuli (200x200 cm) were presented on a 45 cm color monitor (1024x768 pixels, 85 Hz), approximately 90 cm away from where the participant was seated (stimulus size 12“x12”).

Speech task

The speech task, where participants read 12 sentences out loud, was administered immediately before the mood induction, after viewing 20 faces during the mood induction and, immediately after the mood induction. The task began with the presentation of a blank