Developing Voices for Smart Homes:
User Personality and Synthetic Voice Interaction

Greta R. Boye

Master of Science

Speech and Language Processing

The University of Edinburgh

2007

Abstract

Participants (n = 40) aged 27 through 76 years (M = 57.7, Mdn = 59.0) took part in a 2 (voice emotion) x 2 (message content) within subjects factorial design experiment with the between-subject covariates agreeableness and emotional stability. Additional exploratory analysis included age and gender as between-subject factors. Analysis indicated that voice emotion level does not interact with message content level to determine user-reported message suitability ratings, with or without age and gender. No known experiments have been conducted (a) concerning the simultaneous examination of our three experimental factors, (b) in this population of seniors, and (c) through the World Wide Web. We intended, therefore, with this research design to make a unique contribution to HCI and speech technology.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text. This work has not been submitted for any other degree or professional qualification except as specified.

(Greta R. Boye)

Acknowledgements

My gratitude and thanks go to those individuals who guided this research, commented and gave suggestions on the drafts, and provided the encouragement to complete this project. In particular, my sincere appreciation goes to Dr. Maria Wolters for serving as my adviser. Her generosity in time, energy, and resources made this research project a success. I also thank Dr. Jon Oberlander for his initial guidance for this project, and for introducing me to the field of human-computer interaction. Also I would like to express appreciation to Joao Cabral for synthesizing the audio files used in the experiment, and to Benjamin Duffin and Neil Mayo for programming the experiment. Thank you to Dr. Robert Clark, my program adviser, and the graduate students who befriended and encouraged me during my studies in Edinburgh. Most importantly, I would like to thank my close friends, especially Diane Scott and Nick Rocco, and my family, especially Mark Boye and Kristina Secnik-Boye, for their unending love, support, and patience during this most challenging year.

Table of Contents

1 Introduction ... 1
1.1 Motivation and overview ... 1
1.2 Research methodology and questions ... 2
1.3 Boundaries of the dissertation ... 3
1.4 Structure of the dissertation ... 3
2 Background ... 4
2.1 Personality ... 4
2.1.1 Personality models for human behavior ... 4
2.1.2 Personality research in HCI ... 6
2.2 Emotion ... 7
2.2.1 Emotion models for human behavior ... 8
2.2.2 Emotion research in speech science and HCI ... 9
2.3 Information content research in human-computer interaction ... 13
2.4 Selected variables for dissertation ... 13
2.4.1 Selected personality model and traits for dissertation ... 13
2.4.2 Selected emotion model and emotions for dissertation ... 15
2.4.3 Selected information content for dissertation ... 15
2.5 Considerations for older users ... 15
3 Methods ... 17
3.1 Data ... 17
3.1.1 Study population and sample ... 17
3.1.2 Data collection ... 18
3.2 Instrumentation ... 19
3.2.1 Overview ... 19
3.2.2 Likert scales ... 20
3.2.3 Demographics ... 20
3.2.4 Rating of synthetic voices ... 20
3.2.5 Personality quiz ... 21
3.3 Study objectives and research questions ... 22
3.4 Description of design and statistical procedures ... 23
3.4.1 Design ... 23
3.4.2 Data analysis ... 23
4 Results ... 25
4.1 Survey response ... 25
4.2 Study sample ... 25

4.2.1 Sample description 25
4.2.2 Demographics 26
4.2.3 Personality scores 27
4.2.4 Voice-content scores 27
4.3 Responses to research questions 30
4.3.1 Two-way ANOVA 30
4.3.2 Exploratory ANOVA 32
4.3.3 ANOVA summary 35
5 Discussion 36
5.1 Main results 36
5.2 Theoretical and practical implications 37
5.3 Study Limitations 38
5.3.1 Data 38
5.3.2 Instrumentation 39
5.4 Recommendations 40
5.4.1 Data collection and instrumentation 40
5.4.2 Future research 40
Annex A Correspondence with Potential Participants 42
Annex B Experimental Queries: Demographics, Messages, and Personality Quiz 44
Annex C Statistical Appendix 47
References 76

List of Tables

3.1 Example of One of Ten All-Possible 2 x 2 Permutation (combination) of the Study Factors Message Content (Positive or Negative) and Voice Emotion Intonation (Happy or Sad) ... 21

3.2 International Personality Item Pool (IPIP) Example of Items to Measure Agreeableness and Emotional Stability (Neuroticism) According to Positive (+) or Negative (-) Ends of Corresponding Personality Poles ... 22

4.1 Tests of Two-Way Association between Demographic Variables ... 27

4.2 Distribution of Likert Suitability Ratings for Voice and Content Combinations: Percentages in the Analysis Sample ... 28

4.3 Comparison of Suitability Ratings for Most-Preferred and Least-Preferred Voice-Content Combinations, by Age Group ... 29

4.4 Results of the Within-subjects Two Factor Repeated Measures ANOVA Controlling for Personality Covariates ... 31

4.5 Marginal Means of User Suitability Ratings for Main Effects of Voice and Content ... 31

4.6 Results of Exploratory Analysis of Variance Results ... 33

A.1 Initial Correspondence: Introductory Email to Websites, Senior Centers, and Public Libraries ... 42

A.2 Follow-Up Correspondence: Posting to Website Forum ... 43

B.1 Demographic Queries Used in Experiment ... 44

B.2 Synthetic Voice Stimuli Used in Experiment ... 45

B.3 Personality Questions Used in Experiment ... 46

C.1 Gender by Employment Category Cross Tabulation: Analysis Sample ... 47

C.2 Gender by Education Category Cross Tabulation: Analysis Sample ... 47

C.3 Gender by Country-of-youth Cross Tabulation: Analysis Sample ... 48

C.4 Gender by Age Category Cross Tabulation: Analysis Sample ... 48

C.5 Employment by Education Category: Analysis Sample ... 49

C.6 Employment Category by Country-of-youth Cross Tabulation: Analysis Sample ... 50

C.7 Employment Category by Age Category Cross Tabulation: Analysis Sample ... 51

C.8 Education Category by Location-of-youth Cross Tabulation: Analysis Sample ... 52

C.9 Education Category by Age Category Cross Tabulation: Analysis Sample ... 53

C.10 Country-of-youth by Age Category Cross Tabulation: Analysis Sample ... 54

C.11 Personality Traits (Agreeableness and Emotional Stability) by Gender Measures of Central Tendency: Analysis Sample ... 55
C.12 Suitability Ratings for Voice-Content Scores by Gender Measures of Central Tendency: Analysis Sample ... 56
C.13 Distribution of Likert Suitability Ratings for Voice and Content Combinations by Gender: Percentages in the Analysis Sample ... 57
C.14 Suitability Ratings for Voice-Content Scores by Age Measures of Central Tendency: Analysis Sample ... 58
C.15 Distribution of Likert Suitability Ratings for Voice and Content Combinations by Age Group: Percentages in the Analysis Sample ... 59
C.16 Suitability Ratings for Voice-Content Scores by Occupation Measures of Central Tendency: Analysis Sample ... 61
C.17 Distribution of Likert Suitability Ratings for Voice and Content Combinations by Occupation: Percentages in the Analysis Sample ... 62
C.18 Suitability Ratings for Voice-Content Scores by Education Measures of Central Tendency: Analysis Sample ... 65
C.19 Distribution of Likert Suitability Ratings for Voice and Content Combinations by Education: Percentages in the Analysis Sample ... 66
C.20 Suitability Ratings for Voice-Content Scores by Country-of-youth Measures of Central Tendency: Analysis Sample ... 69
C.21 Distribution of Likert Suitability Ratings for Voice and Content Combinations by Education: Percentages in the Analysis Sample ... 70
C.22 Marginal Means of Suitability Ratings for Voice-Content Scores, by Voice Emotion and Age Groups ... 72
C.23 Marginal Means of Suitability Ratings for Voice-Content Scores, by Message Content and Age Groups ... 72
C.24 Marginal Means of Suitability Ratings for Voice-Content Scores, by Voice Emotion and Gender ... 73
C.25 Marginal Means of Suitability Ratings for Voice-Content Scores, by Message Content and Gender ... 73
C.26 Marginal Means of Suitability Ratings for Voice-Content Scores, by Voice Emotion, Gender, and Age Group ... 74
C.27 Marginal Means of Suitability Ratings for Voice-Content Scores, by Message Content, Gender, and Age Group ... 75

1 Introduction

1.1 Motivation and overview

The objective of this dissertation research was to investigate the association between user personality, synthetic voice emotion, and message content through an integrative approach across the disciplines of psychology, speech technology, and human-computer interaction (HCI). The research was conducted through the use of an on-line experiment to examine how older users interacted with synthetic voice messaging in the context of a smart-home environment. During preliminary investigations for this work, we found no recognized literature or research concerning the simultaneous examination of our three experimental factors (personality, voice emotion, and message content). Moreover, no known experiments in this substantive area were conducted in this population –seniors, or through the use of the World Wide Web (WWW) as the data collection method. We intended, therefore, with this research design to make a unique contribution to the fields of HCI and speech technology.

We developed the conceptual framework for this study based on research conducted during the last several decades in psychology and concerning personality, emotion, and human behavior. Importantly, we incorporated recent research from the emergent field of HCI to inform the study of 1) the mapping of the emotion of synthetic voices on to the personalities of computer users, 2) the mapping of synthetic voice emotion on to message content, and 3) the link among user personality, synthetic voice emotion, and message content. Specifically, this research drew from the work of Nass, Moon, Fogg, Reeves, and Dryer (1995), who showed that people responded to computer personalities in the same way that they responded to human personalities, and that of Nass and Lee (2001), who expanded that research to investigate the principle of similarity-attraction with synthesized voices. The conceptual framework of this research was also informed by the investigations conducted by Isbister and Nass (2000), who tested the consistency of personality in interactive characters, and Moon (2002) on the effects of customizing messages according to style for consumers.

The results of this study will contribute information concerning factors that influence older-user synthetic voice preferences. Additionally, this research will help inform the viability of conducting research in this population via the WWW. Lastly, this study will contribute to ongoing research at the Center for Speech and Technology Research concerning 1) the development of voice interface technology that is accessible to older adults, and 2) research on making computers easier to operate. This dissertation will also inform the MATCH project, which is an endeavor to create home care systems that support independent living.

1.2 Research methodology and questions

This investigation employed a multi-factorial experimental design and, as noted, was conducted over the Web. The target audience was senior users, who were recruited through advertisements or contacts made on popular senior web pages, web logs, and user lists hosted in Europe, the United States, and Canada. The first experimental factor was the use of a synthetic voice presented to participants to inform them of the result of task or activity completion that mainly concerned the operation of household appliances, or the execution of typical actions that would likely occur during an ordinary day at home. The experimental stimuli included the message presented by a synthetic male voice in a 'happy' or 'sad' style corresponding to successful or unsuccessful task completion. For example, a happy voice might have notified that 'the gas has been turned off' and a sad voice might have announced that 'the alarm system could not be activated.' The voice emotion was randomized with the message content, with a possibility of four voice-content combinations.

The second experimental factor was the characterization of the participants’ personality through self-completion of a web-administered personality inventory. The instrument consisted of several items used to form a composite score mapped to dichotomous characterization of two selected personality traits, agreeableness and emotional stability (also referred to as neuroticism), according to the 'Big Five' personality model. The three main research questions that this study intended to address were as follows.

Research question 1:
Does message content affect user rating of voice emotion suitability?

Research question 2:
Does user personality affect user rating of voice emotion suitability (i.e., whether s/he prefers a happy or a sad voice to deliver a message)?

Question 3:
Does the user's personality affect which type of voice is preferred for which content?

We also conducted additional exploratory research to examine the association of age and gender with preference ratings, based on these research questions.

1.3 Boundaries of the dissertation

As noted, the recurrent themes of user personality, synthetic voices, and message content on which this dissertation is based draw from psychology, speech technology, and HCI. While dependent on theory, we limited our analysis to the background considered necessary to understand each of those components. For example, we reviewed and identified appropriate personality traits for this study, yet we did not address the psychology of personality. Likewise, while we addressed voice emotion within HCI and chose certain emotions for the experimental work to support our hypotheses, we discounted voice perception within speech technology. Nonetheless, in this dissertation we studied how a user's personality influenced his or her preference for a certain voice style to deliver a message having certain content. Although we designed the experiment within a simulated 'smart-home' environment for older people, we view the results as relevant to other situations involving the interaction of humans and computers, which is becoming commonplace in the daily lives of people of all ages.

1.4 Structure of the dissertation

The dissertation is structured as follows. Chapter 2 provides a literature review of personality, emotion, and information content and raises special considerations for older users. Chapter 3 describes the experimental methodology, including a description of the experimental design and statistical procedures. Chapter 4 reports the survey response in terms of the study sample, ratings of the experimental stimuli, the results of the personality quiz, and the responses to the research questions. In the final chapter we present a discussion of the study results and provide recommendations for future research.

2 Background

The objectives of this literature review are to strengthen the rationale for the study and to provide background information on the topics of interest. We first review personality models for human behavior and personality research in HCI, and then do the same for emotion. We then present relevant research on information content. We identify the personality and emotion models on which our study is based, and select the personality traits, emotions, and information content used in the experimental research. Finally, we discuss considerations for older users within the context of HCI.

2.1 Personality

2.1.1 Personality models for human behavior

While there are many definitions of personality, from the viewpoint of psychometrics it can be defined as the total of all variables, other than those of ability, on which individuals differ from one another (Kline, 2000; Wiggins, 1979). Empirical research on personality in the field of psychology was introduced mainly at the turn of the 20th century, and uses the notion of 'trait' to conceptualize personality. The distinction between personality traits and moods is quite clear: personality traits are stable and of long duration while moods are transitory (Kline, 2000). For example, one might describe himself or herself as a generally 'positive' person yet as being in a 'bad' mood during a certain length of time.

The objective of personality questionnaires, also referred to as inventories, is to define personalities as they relate to behaviors (Cattell, 1973, as reported in Matthews et al., 2003). For example, if a person enjoys parties or meeting people in other social activities, we intuitively describe him as an 'extravert'. That extraversion personality trait is often associated with an aspect of sociability or dominance. A person who scores high on the extraversion scale (+E) might be characterized as being sociable, energetic, and outgoing, and as one who prefers to be around people. In contrast, if a person scores low on extraversion scale (-E), he or she might be characterized in a manner opposite to +E such as anti-social, lethargic, shy, and as one who prefers to be alone.

A wide variety of personality questionnaires has been developed over the years. One of the earliest pioneers was the psychometrician Cattell (1973, as reported in Matthews et al., 2003) who created the Sixteen Personality Factor Questionnaire (16PF). Despite criticisms, Eysenck and Eysenck used it as the basis for the so-called EPQ (Eysenck Personality Questionnaire) and its many revisions that culminated in the EPQ-R (Eysenck and Eysenck, 1991, as reported in Matthews et al., 2003). The Revised NEO-Personality Inventory (NEO-PI-R) developed by McCrae and Costa (1992, as reported in Matthews et al., 2003) is also of importance. In terms of inventories targeted at specific purposes, the Minnesota Multiphasic Personality Inventory (MMPI) and its off-shoots are reportedly the most widely used in published clinical papers, and the California Psychological Inventory (CPI) in industry (Matthews et al., 2003).

Nowadays, there is consensus that the Five-Factor Model, also known as the Big Five, best describes personality; it is also considered the most influential model of the last two decades (Goldberg, 1993; Kline, 2000). Its main proponents, McCrae and Costa (1992, as reported in Matthews et al., 2003), argued that the five broad factors or dimensions can be found in the majority of personality questionnaires including the 16PF5 (an extension of the 16PF), the NEO-PI-R, the MMPI, and the Myers-Briggs Type Indicator. Furthermore, they claimed that the five factors were intended to provide a scientific framework that lends itself to organizing the multitude of differences that characterize any one person. Other key advantages of the Big Five are that (1) its model facilitates communication among personality researchers due to its wide array of personality constructs, (2) it provides a basis for research on personality and other phenomena due to its comprehensiveness, and (3) its five scores are efficient (McCrae & John, 1992).

Two classification systems are widely used to number and label the Big Five factors, according to McCrae and John (1992). One system is based on natural language trait terms and typically uses the following numbers and labels: Factor I, extraversion or surgency; Factor II, agreeableness; Factor III, conscientiousness; Factor IV, emotional stability; and Factor V, intellect. The other system is based on questionnaires and typically uses the following dimensions: extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. One of the two differences between the two nomenclatures lies in emotional stability and neuroticism. The relationship between these two traits, or dimensions, is inverse, viz., high neuroticism corresponds to low emotional stability. Hence, researchers usually refer to this dimension as emotional stability or, conversely, as neuroticism. The other difference is with intellect and openness to experience. Openness to experience is a variant of the intellect factor and in fact was originally labeled culture before the 1990s when variables directly related to intellect were omitted from the model. Regardless of the system, each factor incorporates a multitude of individual traits. For example, for the agreeableness factor, the Big Five model contrasts traits such as kindness, trust, and warmth with hostility, selfishness, and distrust. For the emotional stability (neuroticism) trait it includes such traits as nervousness, moodiness, and temperamentality (Goldberg, 1993).

2.1.2 Personality research in HCI

Unlike the long history and breadth of personality research in psychology, personality as it relates to HCI is a new area of study, especially when one considers that voice technology, and speech synthesis in particular, appeared on the commercial market and in academia in the late 1980s (Black, Taylor, & Macon, 2000). To date, experimental research in this field has tended to isolate the personality of the user or of the interface character and not on crossing personality with the emotion of a user, character, or synthetic voice. However, user personality has been crossed with message content.

Research that is of particular relevance to this dissertation is that of Nass, Moon, Fogg, Reeves, and Dryer (1995), who showed that people prefer to interact with others who are similar in personality, which is the basis for the similarity-attraction hypothesis. The researchers tested the effects of matching a single personality dimension, extraversion (as it ranges from dominance to submissiveness), to determine whether the hypothesis held in HCI. They chose that personality dimension from the Big Five model because other researchers have found it to underlie interpersonal behavior in a variety of situations (Wiggins, 1979; Horowitz, 1979, as reported in Nass et al., 1995).[1] Participants completed the Bem Sex Role Inventory personality test (Bem, 1974) and then were paired with a computer having either a 'dominant' or 'submissive' personality to complete the Desert Survival Problem (DSP) using pen and paper.[2] Experimental results confirmed the similarity-attraction hypothesis and showed that the manipulation of personality in computers is not difficult and powerful. Even though the experiment excluded voice technology to manipulate computer personality, results indicated that user personality is a factor in the perception and interaction with computer systems.

Nass and Lee (2001) expanded the concept developed by Nass et al. (1995) that people will respond to computer personalities in the same way that they respond to human personalities by extending it to voices. Those researchers tested whether particular settings in a text-to-speech (TTS) engine would lead individuals to identify and respond to a computer-generated voice as if it had a personality. To test their hypotheses, they conducted two experiments

---

[1] Agreeableness was also shown to be important in interpersonal interaction, but was omitted because the study was constrained to one personality dimension.

[2] The DSP requires participants to rank in order of importance items they might need in a desert survival situation, such as water, mirror, compass, salt tablets, and map, for example.

requiring participants to complete the Web-based short form of the Myers-Briggs Type Indicator personality inventory (Murray, 1990, as reported in Nass and Lee, 2001) and the Wiggins (1979) personality test. The first experiment required participants to evaluate on-line a book and indicate an intention to buy that book (similar to Amazon). The web page included book details and a link to an audio file containing the book description as read by a synthesized voice. Results upheld the similarity-attraction hypothesis. The second experiment required users to participant in an on-line auction (similar to e-Bay) that included the item description in an audio file that was classified as extravert or introvert and read by a voice that was also classified as extravert or introvert. The researchers tested the consistency-attraction hypothesis, which states that users like and prefer behavioral consistency because it leads to predictability and lightens cognitive load (Lee & Nass, 2003), and found that the results strongly supported this hypothesis. In addition to showing that individuals confirmed the similarity-attraction and consistency-attraction hypotheses, the experiments provide important insights on the interaction between user personality and synthetic voice personality. For example, results indicated that paralinguistic cues were relevant to users' responses to synthetic voices and that synthetic voices influenced perceptions of message content.

Guided by the ways in which people interpret personality in others and principles of user-interface design, Isbister and Nass (2000) designed an experiment to determine whether users would successfully label introverted and extraverted verbal and non-verbal cues from interactive characters just as they identified these types of cues with textual cues (Moon & Nass, 1996). Participants first completed portions of the Myers-Briggs personality inventory and the Wiggins interpersonal adjective set to determine his or her level of extraversion or introversion, and then did the DSP using pen and paper. The researchers then introduced the participants to an on-screen computer character that attempted to convince him or her to change the ranking of items through variations in verbal cues (text phrasing, which acted as a 'voice') and non-verbal cues (character posture). The main results of the study indicated a user preference for character consistency, and a user preference for a character's personality that complemented his or her personality rather than one that was similar. The implications for speech technology are that human-human social expectations will affect the evaluation of interactive characters, which include verbal cues (whether or not spoken), and that an important part of those expectations is associated with consistency.

2.2 Emotion

Emotion is widely accepted as being different than personality. Emotions are usually considered short-lived, thereby implying a transitory period; when one or more emotion is sustained, they are referred to as 'moods'. In contrast, personality is composed of traits that are part of one's make-up, and therefore is long-lasting. Despite this distinction, researchers in the field of psychology often link personality and emotion within the framework of a personality theory of individuals. Kellerman (1980), for example, studied personality at different levels that included emotion, defense, diagnosis, dreams, and nightmares. In contrast, Plutchik (1980) put forth the interpretation of personality traits as derivatives of emotion. For purposes of this dissertation, personality and emotion remain separate because personality is relatively permanent, and it is easier to adapt to known, permanent user characteristics than to detect emotion dynamically.

2.2.1 Emotion models for human behavior

a. Contemporary perspective

Since the 1950s, researchers on emotion theory have extended the classic theories to propose new approaches to understanding emotion. Arnold (1960, as reported in Oatley and Jenkins, 1996), for example, proposed that emotions are based on appraising events, and Tomkins (1962, as reported in Oatley and Jenkins, 1996) introduced research on emotional expressions of the face. Arnold's underlying approach focused largely on inputs and therefore on perception; in contrast, that of Tomkins relied on bodily feedback and outputs, and thereby took into account motor effects. Their overlap stemmed from the belief that emotion was central to normal functioning, which in turned spawned a new era of research involving experimental psychology in emotions led by Isen (for an overview of her work in emotion thought processes beginning in 1970, see Isen, 2000). Others, such as Averill (1980) adopted the so-called constructivist view that interprets emotions as social constructions, not biological givens. Scherer (1984, as reported in Johnstone & Scherer, 2000), Frijda (1986, as reported in Johnstone & Scherer, 2000), and Lazarus (1991, as reported in Johnstone & Scherer, 2000), have led a consensus that emotion needs to be viewed as a multi-component entity involving neurophysiological response patterns, motor expression (in face, voice, and gesture), and feelings.

The link between personality and emotion, as it pertains to this dissertation, can best be captured in Gray's (1981) theory as summarized by Larsen and Ketelaar (1991). Gray argued that there are two neurologically-based motivational systems responsible for many of the observed behavioral and emotional differences between extraverts and neurotics. He suggested that individuals differ in the relative strengths of these two signal-sensitivity systems, with extraversion relating to signals of reward and neuroticism relating to signals of punishment. Similar theoretical foundations of the relationship between affect and personality were put forth by Eysenck (1987) and Strelau (1987), who hypothesized that extraverts should be more susceptible to positive affect than introverts. Likewise, neurotics should be more susceptible to negative affect than emotionally stable or non-neurotic individuals. Larsen and Ketelaar confirmed this hypothesis in experimental work and concluded that, given the appropriate stimuli, the personality dimensions of extraversion and neuroticism are related to differential manifestation of positive and negative affect. In contrast, they pointed out that McCrae and Costa (1992) found that both the conscientiousness and agreeableness personality traits promote positive affect and minimize negative affect.

b. Emotion definition and categories

Given the multitude of emotion models, it should come as no surprise that there is no consensus in the literature about a single definition for emotion or how to categorize it. In fact, according to Averill (1980), the English language has 550 concepts that refer to emotions. Nonetheless, there seems to be agreement that some emotions involve less cognitive processing (e.g., disgust) and structure than others (e.g., shame). In his master's dissertation on emotion in speech technology, Hofer (2004) adopted the so-called componential view to define emotion as having the following characteristics, which embody most of the facets of the various emotion theories described above: (i) evaluation or appraisal of the antecedent event that acts as the stimulus for the individual; (ii) physiological change (e.g., heart palpitations, blushing); (iii) action tendencies, such as 'fight or flight'; (iv) subjective feelings; and (v) verbal (voice) and non-verbal (facial) expressive behavior.

In terms of categorization, Plutchik (1980) was the first to classify emotions along the lines of biological and evolutionary variables. Others, such as Ekman, Friesen, and Ellsworth (1982), used facial variables as classifiers; Roseman (1984) represented emotion in terms of cognitive variables; and De Rivera (1977) presented it according to phenomenal variables. Regardless of the classification scheme, most theorists agreed that the plethora of emotions needed simplification and reduced them to a small number (usually less than ten) of so-called basic or primary emotions (Ortony, Clore, & Collins, 1988).

2.2.2 Emotion research in speech science and HCI

Without a doubt, humans depend heavily on vocal cues to perceive speech that in turn lead them to process messages, make decisions, and take actions according to their interpretation. Nass and Gong (2000) described the remarkable 'cocktail skill' of humans, whereby we are able to tune in and process one voice among an array of simultaneous voices, even if the person is moving throughout the room. This skill also explains why we are able to 'multi-task' while being influenced by various vocal stimuli around us, such as listening to the radio or television while carrying on a conversation. Banse and Scherer (1996) reported in a review of studies covering 60 years on voice discrimination that listeners generally had an accuracy of 50 percent in inferring affective state and speaker attitude from vocal expression. Even though that research focused on five basic emotions (e.g., anger, fear, sadness, joy, disgust) and accuracy rates varied among the individual emotions, the overall rate was reportedly four to five times higher than that expected by chance. The study provides evidence that people are able to differentiate emotions on the basis of valence or quality cues, independently of arousal level or intensity.

a. Emotion research in speech science

The study of voice emotion in HCI is dependent on speech and communication science, psychology, and psycholinguistics. Speech science has had a particular influence on encoding and decoding voice emotion in HCI applications involving sound. Unlike applications that use text, critical differences in voice applications occur in the paralinguistic domain that is concerned with factors of how words are spoken. Speech scientists agree that the main challenge lies in creating natural-like voices, a large part of which rests on the appropriate expression of emotion.

Researchers have been working hard to improve voice emotion in speech synthesis since the early 1990s (for a detailed review, see Schroeder, 2004). Although certain acoustic measures used in experiments are fairly standard, the exact number and parameters for each measure are unknown, thereby making it difficult to perfect voice emotion in a standardized way. Overall phrase duration and word rate carry affective information, but these parameters vary. Likewise, the overall acoustic energy of the speech signal (perceived as loudness) often varies when expressing emotion, as does the fundamental frequency, which determines the perceived pitch of the voice. Voice quality is another important acoustic measure, yet is difficult to capture in emotional speech synthesis despite its presence in the speech signal (Hofer, 2004).

In his doctoral dissertation, Schroeder (2004) reformulated research results from an extensive literature review and data analysis to improve the German MARY (Modular Architecture for Research on speech sYnthesis) TTS system. Despite the fact that Schroeder confirmed his hypothesis that emotional states similar to the intended state would be rated as more consistent than more distant states, the results were somewhat disappointing. He attributed the unexpected results to the criterion used to measure the quality of the synthesized emotional speech, difficulties encountered by participants in the written part of the evaluation, and the testing methodology that allowed participants to adjust their positions on the answer scale relative to other stimuli.

Hofer (2004) used Schroeder's research as a springboard to develop a new system to improve voice emotion in synthesized speech. His system was based on a type of unit-selection speech synthesis that did not explicitly model the prosodic realization of an emotion, yet enabled the modeling of varying degrees of an emotion (known as 'blending', as based on Black, 2003). Actors and amateurs did recordings in neutral, angry, and happy voices. The final system was evaluated through a formal perceptual test whereby participants rated utterances spoken in different emotions along a continuous scale. Hofer confirmed the hypothesis that emotions are perceived as more intense when more emotional units are included in the utterance. Although participants generally noted very good voice quality, Hofer would like to improve database coverage and include additional emotions to the synthesizer. Despite its drawbacks, Hofer concluded that as long as the lack of a clear understanding persists of the parameters involved emotional speech synthesis, unit selection remains the best option.

Cabral (2006) also has made important contributions to voice emotion in synthetic speech by developing the EmoVoice system (for details, see http://www.l2f.inesc-id.pt/~jpcabral/EmoVoice/). His research objective was to develop a method to modify the relevant features of the voice source from natural speech with minimal degradation of the signal quality, and to use that technique to modify the parameters of an emotionally neutral speech signal so that it was perceived as if it was uttered with an emotion. In an experiment to test how seven basic emotions (anger, happiness, sadness, fear, surprise, boredom, and disgust) compared with each other, Cabral discovered that, generally, listeners reliably recognized the simulations of anger, happiness, sadness, and fear. More specifically, listeners found it easy to distinguish between happiness and sadness with nearly 100 percent recognition rates between those two cases. The worst recognition rate obtained for sadness was quite high at 82 percent, when compared with boredom. Listeners found it more difficult to recognize happiness when it was compared with anger and surprise, with recognition rates of around 60 percent for these cases. Although he met his research goal, Cabral raised issues similar to those brought up by Hofer (2004), namely that further research is required on acoustic correlates of emotion, especially features related to voice quality due to their difficulty in measurement and modification.

b. Emotion research in HCI

While extensive research has been undertaken on the subject of emotion in voice within the realm of affective science and emotional speech synthesis, work on when to use which emotions in the field of HCI has been limited. In their recent work on the design of voice interfaces, Nass and Brave (2005) discussed important effects of affect such as attention, performance, and judgment resulting from the interaction between user emotion and interface emotion. Likewise, they found connections between interface emotion and message content, summarized as follows.

The experiment on driving simulation (Jonsson, Nass, Harris, & Takayama, 2005) investigated what would happen when the emotion of the voice interface and the emotion of the driver were mismatched (e.g., when an upset driver heard and upbeat voice). Drivers were first induced to be happy or sad and then were accompanied by a virtual passenger, which was simulated through a recorded happy or sad female voice. In their discussion, Jonsson and her colleagues emphasized that the matching of the voice of the car to the drivers' emotions was so powerful that neither driver emotion nor gender had the expected effect on driving performance. Drivers who interacted with voices that matched their emotional states had less than one-half as many accidents on average as drivers who interacted with mismatched voices. Finding the appropriate in-car voice for the driver's emotion, the researchers discovered, was the most critical factor in enabling a safe driving experience. Moreover, from a business viewpoint the costs involved in influencing the driver from a psychological angle are less than those involving technological changes to improve the safety of the car.

Nass, Foehr, Brave, and Somoza (2001, as reported in Nass & Brave, 2005) conducted a telephone-based experiment that involved mixing voice and content emotion to test whether people would assign emotion to machine-based voices and whether those assignments would influence their perceptions. Participants listened to one happy version and one sad version of a news story, movie descriptor, and health story read by a synthetic male voice. Those who heard matched voice and content (e.g., happy content read in a happy voice) perceived the stories to be happier (or less sad) than stories read with mismatched voice and content. Nass and his colleagues concluded that paralinguistic cues used to evoke happiness and sadness in a voice are integrated with the perceived meaning of spoken words into a message. They claimed that matching the voice to the content is just as important, if not more important, than matching the voice to the emotion of the user.

2.3 Information content research in human-computer interaction

The influence or impact of the design of information content is widely accepted as being a key element in effective marketing or media communication, and has been extended to research in HCI. In a recent study on consumer behavior, Moon (2002) investigated whether messages can be more persuasive when customized according to the personality of the user or recipient. Moon chose extraversion and introversion given the absence of attention to those traits in consumer marketing literature and their prominence in contemporary research using personality theories. The primary hypothesis was that messages would be more effective at generating attitude change when their style of presentation 'matched' the personality style of the recipient. Message personality was accomplished through the use of word choice, not voice style, and within the context of buying a car. Experimental results supported the acceptance of the main hypothesis, and researchers elaborated on the consistency of findings with the similarity-attraction effect discussed above. The study has important implications for HCI researchers as well as marketing practitioners. As Moon notes, in the business world message customization typically focuses on what is marketed to whom, rather than on changing the message style. This study points up the need to consider other variables that can be customized, such as the manner in which the computer addresses the user, the level of confidence with which the computer claims to know what the user wants, and the language used to convey the computer's suggestions, all of which can be extended to the use of voice in HCI research.

2.4 Selected variables for dissertation

2.4.1 Selected personality model and traits for dissertation

Based on the current consensus of personality psychologists, the Big Five personality model was chosen as the experimental basis for the personality portion of this study. The Big Five model is well-operationalized and well-researched. Moreover, questions used in Big Five personality tests, background information, and scoring methods are available publicly, unlike the Myers-Briggs test or the Bem Sex Role Inventory used in other research aligned with this study.

The selection of personality traits for the dissertation was not as straightforward as the personality model. As shown above, the extraversion trait has been widely studied in HCI, due in part to its recognizable verbal and non-verbal cues and also because of its significance within the context of interpersonal behavior (Nass et al., 1995, Nass & Lee, 2001). Nass et al.

(1995) argued that in addition to extraversion, agreeableness is the most meaningful personality trait of the Big Five dimensions in the context of interpersonal interaction. It has also been shown by personality theorists that agreeableness, compared with the other four dimensions, is the personality trait that best reflects interpersonal interaction in a variety of situations, especially those involving social perception and cognition (Graziano, Jensen-Campbell, & Hair, 1996; Horowitz, 1979; Wiggins, 1979). Within that context, it was assumed that agreeable people are better able to control anger and negative affect in situations involving frustration (Graziano et al., 1996).

Other personality theorists such as Eysenck (1987) considered that, in addition to extraversion, the dimension of emotional stability has the most relevance for social aspects. Furthermore, Wilson (1967, as reported in Costa & McCrae, 1980) and Bradburn (1977, as reported in Costa & McCrae, 1980) have associated positive and negative affect with emotional stability. In recent HCI research on personality and human language production, Gill, Harrison, and Oberlander (forthcoming) tested the influence of extraversion and neuroticism on language production in interpersonal interactive situations. They conducted a syntactic priming experiment to test whether those two personality traits would have an effect on priming behavior, and surprisingly found that neuroticism influenced syntactic priming and that extraversion did not. Their research has important implications for computer interface design, namely that if an 'unintelligent computer' were to project personality, one could expect it to vary its degree of priming in addition to its lexicon depending upon the sort of personality it wished to project.

Given that most personality studies to date in the field of HCI have focused uniquely on the extraversion trait and because it is considered distinct from other dimensions (Matthews et al., 2003), we omitted extraversion from this study and chose agreeableness and emotional stability for several reasons. Agreeableness is one of the least explored of the components of the Big Five personality model (Graziano et al., 1996) and to date, like emotional stability, has not been the focus of HCI research concerning personality, emotion, or message content. Moreover, we chose the combination of agreeableness and emotional stability because past research has identified agreeableness as being influential in interpersonal communication and emotional stability as having an impact on affect.[3]

---

[3] Due to the possible negative connotation associated with the word 'neuroticism' when communicating with audiences outside personality psychology, we will rephrase the dimension as 'emotional stability' that in turn will emphasize social desirability.

2.4.2 Selected emotion model and emotions for dissertation

The multi-component view is the underlying approach to emotion adopted in this dissertation. It is based on the understanding that emotions are capable of invoking a neurophysiological response (e.g., when one experiences fear, the heart rate can increase or one can sweat), changes in motor expressions, and changes in feelings. We will test whether voice emotion affects listeners' attitudes towards messages. Secondary effects such as neurophysiological and motor responses will not be recorded since we are not interested in the emotion that a message elicits in the listener, only in listeners' responses to voice emotion. Some emotions are expressed and perceived more or less easily in the voice than others. Fear and alarm are among the most clearly vocally expressed emotions. Disgust, in contrast, is considered difficult to detect in the voice. Other emotions, such as stress, anger, sadness, and joy, are commonly studied. Sadness and happiness were chosen for this dissertation because, in addition to being able to be synthesized without extensive manipulation, they are easily recognized in the voice (Johnstone & Scherer, 2000).

2.4.3 Selected information content for dissertation

We designed the message content for the present research to be clear and concise, and also to be as unambiguous as possible. Messages summarized the outcome of a smart home command, which was either successful (positive content) or unsuccessful (negative content).

2.5 Considerations for older users

Interface designers face a multitude of challenges when considering technology for older users. As one ages, biological and social changes influence behavior that in turn impact on behavioral processes and psychological functions. Birren and Schaie (2001) address these issues and many more in their handbook on psychology and aging. Their summary of research trends in vision and hearing (see Chapter 10 of that handbook) highlights that important scientific discoveries have been made that link behavioral slowing and cognitive decline to changes in both vision and hearing. The authors likewise point out new ways that are being developed to improve vision and hearing, such as training older persons to use contextual information to improve perceptual performance. In hearing, digital processing and transmission technology have improved the quality of assistive hearing devices.

This research has important implications for HCI. For vision, current design principles recommend simplicity, clarity and consistency in text layout, bold search cues, customization of fonts, and designs that use depth perception to convey information, which should go beyond basic design principles applicable to users of any age. For hearing, system designers need to consider that older people face difficulty in hearing tones above 4000 hertz. This distinction is important for attention-getting systems (such as alarms) and also conversation-based systems since in certain languages like English, some consonants are high-pitched (e.g., f, s, t, z). The resulting selective loss of high frequency hearing means that parts of speech are not heard, causing the listener to guess at meanings (Hawthorn, 2000). Smither (1993, as reported in Hawthorn, 2000) found that part of speech understanding shown by elderly people may be due to unconscious lip reading and other contextual clues, which can be helped by systems that include brief spoken messages.

Attention within the framework of cognitive aging is another important consideration for older technology users, especially selective attention or inhibition of irrelevant information (see Chapter 11 of the Birren and Schaie handbook for background). Hasher and Zacks (1988, as reported in Birren and Schaie, 2001) suggested that an age-related difference in inhibition could account for a wide variety of variations in other aspects of cognition, especially in working memory. These differences will be better understood though ongoing research involving neuroscience, which will also have an impact on understanding attention in the normal aging process and dementia. Results of this research have already extended to HCI. Most recently, Czaja and Lee (2007) discussed implications of age-related changes in cognition for the design of computers and technical systems. They found that especially effective systems reduce the working memory demands of users and their navigational requirements, while improving design and taking into account lower processing speeds experienced by older adults (e.g., avoidance of pop-up screens or system queries). Czaja (2001) has studied aging and work performance, and has determined that while there are age-related declines in most aspects of functioning, the decline is gradual and most jobs do not demand constant performance at maximum capacity levels. She emphasized that further research is critically needed to fit job needs with the capabilities of older workers, and to identify interventions that enhance the ability of older people to function effectively in work environments. Other areas of interest to HCI and aging are motor control, memory, and language production and comprehension, all of which are discussed from a psychological perspective in Birren and Schaie (2001, Chapters 13, 14, and 15, respectively).

3 Methods

This chapter describes the experimental methodology. First we describe the study population, study sample, and the data collection strategy. Next, we enumerate the research questions and hypotheses proposed to address the purpose and objectives of this study. Lastly, we provide a description of the experimental design and statistical procedures.

3.1 Data

3.1.1 Study population and sample

Previous studies in HCI concerning personality and emotion have been conducted using data from samples with potentially limited generalizability including groups of university students selected on the basis of convenience (e.g., Nass et al., 1995; Isbister & Nass, 2000; Nass & Lee, 2001; Moon, 2002) and focus groups recruited with a specific task in mind (e.g., Jonsson et al., 2005). Data thus derived are frequently homogeneous with respect to age, education, employment, and life history and experiences.

A study population comprised of university students allows the possibility for a significant sample size due to the availability of participants and their willingness to take part in experimental work for course credit or financial reward. Nonetheless, this type of sample has drawbacks. First, the students in these samples are more or less the same age (viz., usually in their 20s), have the same level of education and employment status, and often have similar backgrounds in terms of country where they spent their youth.

Focus group research designs enable researchers to explore, in depth, key questions or areas of interest. Researchers conducting focus groups assemble participants hoped to be representative of the target population. Notwithstanding this intent, not all studies invest in defining a sampling frame and then implementing a probability sample to then make scientific inference beyond study participants.

We conducted this research to examine the study questions in a group of mature adults --a population not well-represented by the existing literature in the field of HCI. Our study participants were either 1) personal contacts aged at least 40 years, or 2) computer users aged at least 50 years who frequented selected websites aimed specifically that age group. We collected all data from the experiment web interface and included 1) participant self-report of demographics, 2) participant completion of a validated personality questionnaire, and 3) participant evaluation of 40 voice messages with randomly varying content and emotional intonation --ten items were tested for each of the four possible permutations of the two evaluated message content levels and the two voice emotion levels.

3.1.2 Data collection

We solicited participants by first identifying 21 general interest websites frequented by computer users aged at least 50 years, one on-line education and innovative technology website, and 13 physical (brick and mortar) facilities (i.e., centers and public libraries in Southern California offering computing classes to seniors).[4] The websites and public libraries were identified through the Google search engine. Additionally, we compiled a list of 55 personal contacts that included friends, family members, and former colleagues who were aged at least 40 years.

Correspondence with potential participants was initiated by an email to introduce the experiment. For the administrators of the websites and for those in charge of computing classes at public libraries or senior centers, the introduction of the email explained the author's affiliation with the university (Master's student) and the purpose of the experiment (Annex A, Table A.1). We placed emphasis on the brevity of the experiment, followed by the provision of information concerning financial incentive through a lottery, the experiment's hyperlink, and a request to post notice to the website, member forum, or at the library or senior center. Finally, we made reference to research underway that is being conducted by the two dissertation supervisors and to a related project concerning home care systems to support independent living, MATCH. Care was taken to provide an email address specifically established for this experiment and to provide comment indicating that the author would answer all queries personally. No admission-to-study criteria were included, such as no known cognitive impairment, adequate hearing, or prior exposure to the Web.

We omitted the details concerning thesis supervisors and the related research project within our introductory email to personal contacts, reasoning this level of credibility was not necessary or helpful to this audience. Once responses were received from the website administrators or from the appropriate contact persons at senior centers or public libraries, the author posted a short message that summarized the lengthier introductory message described

[4] In most cases, senior are considered those who are 60 years and older. Southern California was chosen, and specifically the Los Angeles area, due to the concentration of seniors in the locale of the author; the author's proximity also allowed any in-person assistance, if needed.

above in order to communicate information about the study in a more succinct and direct message (Annex A, Table A.2). We provided no follow-up to those websites whose administrators did not respond to initial communication or to those websites where messages were posted successfully. Details on the response rate are included in Chapter 4 of this dissertation.

3.2 Instrumentation

3.2.1 Overview

The experiment was designed to run over the WWW and to tap into a much more diverse pool of study participants compared with a traditional laboratory experiment. Additionally, global accessibility allowed the participation of personal contacts located around the world. The study web site was developed using WebExp2 software from the University of Edinburgh (http://www.webexp.info/), which was written in Java and designed for conducting online psychological experiments. Neil Mayo and Benjamin Duffin did the programming in WebExp2 for this experiment. Maria Wolters transformed the WebExp data into files that were ready for analysis. Instrument completion time was estimated to be ten to fifteen minutes with the deliberate intent to maintain brevity, thereby lessening study dropout and facilitating the provision of complete study data. To further reduce study attrition and incomplete data, participants were induced by a chance to win at a lottery for five $25 Amazon gift certificates, redeemable world-wide. As noted by Buchanan, Johnson, and Goldberg (2005), attrition in Web-based experiments is an important issue because of selective dropout.

According to the experiment design, all participants completed three distinct experimental phases. The first phase involved responding to questions on demographics; the second phase involved rating synthetic voices; and the third phase required participants to take a personality quiz. Apart from possible keyboard input for certain questions on demographic information, the participants used only the mouse. Therefore, the experiment required a minimum level of subject-computer interaction and basic computer skills. No time limits were imposed on the experiment. Participants were able to listen to sound files as many times as desired. Once initiated, however, the participant could not skip backwards or forwards —which would cause the experiment to abort, thereby losing all data. Technical details about the type of computer used, its operating system and speed of connection, and whether or not participants used headphones, were not solicited and therefore remain unknown. Likewise, details concerning the listening environment (e.g., level of noise) were not captured.

3.2.2 Likert scales

We used Likert scales to rate the audio messages and to derive scores for the two personality traits under review. We chose these scales because the computation of results is straight forward and scoring keys for scales, such as those used in the IPIP personality quiz, are publicly available. Moreover, Likert scales are widely used by researchers, including those whose work was reviewed as part of this dissertation. Therefore, the experiment developed as part of this dissertation used this scale type due to the ease in application to a Web-based environment, calculation and interpretation of results, and acceptability in technique.

3.2.3 Demographics

The first phase of the experiment required participants to complete six demographic questions (for details, see Annex B, Table B.1). The queries on gender and age (items 1 and 2) gathered basic information. Information on native language and the country where the participant spent his or her youth were captured in items 3 and 4, respectively. Item 5 sought educational level. Lastly, employment status information was gathered through item 6.

3.2.4 Rating of synthetic voices

Subjects completed 40 response items consisting of synthetic voice stimuli stating the successful or unsuccessful execution of a command to operate a household appliance or of an action that could occur during an ordinary day at home. To eliminate order effects, we programmed the random presentation ordering of the 10 items for each of the four experimental factor permutations: two 2-level experimental factors (message content and voice emotion). The ratings were classified on a four-point Likert scale according to the suitability of the voice emotion to the message content, viz., not suitable at all, unsuitable, suitable, and highly suitable. Power analysis was used to determine the number of items required. The experiment was designed to show effects larger than one standard deviation of the mean with a power of 0.8 and a significance level of 0.05.

Listed below (Table 3.1) is an example of items that were included in the experiment according to possible combinations (i.e., permutations) of voice emotion intonation (happy or sad) and message content (positive or negative). The complete set of voice stimuli is presented in Annex B, Table B.2.

Table 3.1.

Example of One of Ten All-Possible 2 x 2 Permutation (combination) of the Study Factors Message Content (Positive or Negative) and Voice Emotion Intonation (Happy or Sad)

Voice Emotion/Message Content	Stimulus Example
Happy/positive	The oven has been switched off.
Happy/negative	The oven could not be switched off.
Sad/positive	The oven has been switched off.
Sad/negative	The oven could not be switched off.

Note. The difference between happy and sad messages is reflected in the tone of the synthetic voice, which is discernible by listening to the audio files.

The voice was selected from five male voices based on sound quality and naturalness of voice --a male native English speaker (EM001) that was built for the Blizzard Challenge 2007 (http://festvox.org/blizzard/). Joao Cabral used the Festival system (http://festvox.org/) to synthesize the speech. He generated the speech with emotions through the use of several prosodic and voice quality (glottal source) parameters that were transformed on the neutral synthetic speech. For the happy voice, he increased the fundamental frequency (F0) or pitch by 18 percent in all voice regions of the speech file. For the sad voice, he decreased the fundamental frequency by 16 percent. The technical specifications for the happy voice used for this experiment were: (i) mean F0, 0.18; (ii) F0 range, 0.3; (iii) duration, -0.05; (iv) energy, 0.3; (v) jitter, 0.03; (vi) return quotient, 0.5. For the sad voice the specifications were: (i) mean F0, -0.16; (ii) F0 range, -0.38; (iii) duration, 0.2; (iv) energy, -0.5.

3.2.5 Personality quiz

We selected the 50-item International Personality Item Pool (IPIP) (Goldberg et al., 2006; http://ipip.ori.org/) representation of the two personality traits of the Five Factor model that are under review in this study, viz., agreeableness and emotional stability. This instrument was chosen based on the IPIP's thorough development and documentation of the multi-item assessment of the agreeableness and emotional stability personality traits. Additionally, the questionnaire was easily duplicated within the WWW environment and was therefore well-suited for this online experiment.

Participants described themselves on each of the 40 descriptions for behavior (20 descriptions for each of the two personality traits) using a five-step rating scale that included the self-descriptors of 'strongly disagree', 'disagree', 'neither disagree or agree' 'agree', and 'strongly agree'. Scores for an individual's personality traits were arrived at by applying a scoring key developed by the IPIP (for details, see http://ipip.ori.org/ newScoringInstructions.htm). The scoring key was based on numbers assigned to whether a behavior was classified by the IPIP as being on the positive (+) or negative (-) ends of the corresponding pole of the personality dimension. Examples of items used to measure the two assessed personality traits are shown in Table 3.2. All questions used in this instrument are listed in Annex B, Table B.3.

Table 3.2

International Personality Item Pool (IPIP) Example of Items to Measure Agreeableness and Emotional Stability (Neuroticism) According to Positive (+) or Negative (-) Ends of Corresponding Personality Poles

Item	Corresponding Personality Trait
Inquire about others' well-being.	Agreeableness (+)
Am indifferent to the feelings of others	Agreeableness (-)
Am relaxed most of the time.	Emotional Stability (+)
Worry about things.	Emotional Stability (-)

Note. The positive (+) and negative (-) signs are part of the IPIP scoring key.

For those behavioral descriptions labeled '+', the number 1 was assigned to 'strongly disagree', 2 was assigned to 'disagree', 3 was assigned to 'neither disagree or agree', 4 was assigned to 'agree', and 5 was assigned to 'strongly agree'. For those behavioral descriptions labeled '-', the scoring was the opposite (viz., 5 was assigned to 'strongly disagree', 4 was assigned to 'disagree', etc). Maria Wolters wrote a script to automatically score the results and sum them, which resulted in individual scores ranging between 0 and 100. The way in which the items were scored takes into account the inverse relationship of emotional stability and neuroticism (viz., -E equates to +N and vice versa), as explained in Chapter 2 (section 2.1.1) of this dissertation.

3.3 Study objectives and research questions

The objective of this dissertation is to examine the association of synthetic voice emotion and message content, conditioned on user personality, with participant rating of the suitability of a synthetic voice-delivered message. The results of the study will provide researchers and interface designers with information concerning attributes useful to consider when constructing a synthetic voice to deliver messages to a user with a known or estimated personality type. Through its design and analysis, this study attempted to answer the following research questions. Since there is very little previous work on the link between user personality and preferred emotion, all research questions are phrased as two-tailed hypotheses.

Research question 1:
Does message content affect user rating of voice emotion suitability?

Research question 2:
Does user personality affect user rating of voice emotion suitability (i.e., whether s/he prefers a happy or a sad voice to deliver a message)?

Question 3:
Does the user's personality affect which type of voice is preferred for which content?

We conducted additional exploratory research to examine the association of age and gender with preference ratings, based on these research questions.

3.4 Description of design and statistical procedures

3.4.1 Design

Participants were involved in a 2 (voice emotion) x 2 (message content) within subjects (or repeated measures) factorial design experiment with the between-subject covariates agreeableness and emotional stability. Each of those experimental factors consisted of two levels: happy or sad for voice emotion and positive or negative for message content. The main covariates examined included the personality variables of agreeableness and emotional stability (or neuroticism). These variables represent personality traits that do not constitute study 'treatment'. Ancillary covariates include the demographic variables of age and gender. The dependent variable was participant preference for a certain synthetic voice emotion-message combination (e.g., preference for a message with positive content presented in a happy voice).

3.4.2 Data analysis

Exploratory analysis was undertaken first to determine the data distribution of certain variables (viz., age and the personality variables of agreeableness and emotional stability). Tests for skewness and kurtosis were then performed to verify further normal data distribution of those variables. Frequency data were then analyzed according to the demographic variables (age, gender, native language, country-of-youth, education, and occupation). Other descriptive statistics were generated depending on the appropriateness for the variable. Lastly, differences among respondents grouped by demographic characteristics were tested using the two-tailed Fisher's exact test due to the small numbers anticipated in some groups. A *p* value of .05 or less was taken to indicate significance in all tests.

Analysis of variance (ANOVA) was conducted to provide information on how the study factors and covariates interacted with each other and to what extent the main and interaction terms had on the dependent variable. Following the experimental design set out in section 3.4.1 above, we conducted two-way factorial ANOVAs with covariates. Three main effects for personality, voice emotion, and message content were tested. Interaction effects were then tested for (1) personality and voice emotion, (2) personality and message content, (3) voice emotion and message content, and (4) personality, voice emotion, and message content. We then undertook additional ANOVAs with gender and age as the between-subjects factors. These analyses, however, were explanatory due to unequal numbers in the study sample for gender and age.

4 Results

We present our research results as three chapter sections: 1) survey response; 2) sample characteristics including self-reported scores of the stimuli included in the experiment and the results of the personality quiz, and 3) participant responses to the research questions. All analyses were conducted using the software program Statistical Package for the Social Sciences (SPSS) 15.0 for Windows (SPSS, 2006).

4.1 Survey response

We recorded 50 data submissions during our 22-day data collection period that extended from June 27, 2007 to July 18, 2007. We sought to implement no accounting of visitors to the frequently-visited internet sites upon which we posted notice of this research. Instead, we actively recorded the number of actual survey submissions.

Nonetheless, we did monitor and track responses to the introductory email notices. One-third (7/21) of the website administrators to whom we sent introductory emails responded to our notice; seven administrators posted a notice of the experiment in fora concerning news, research or general discussion. Two of these seven websites maintained statistics concerning the number of forum members who viewed the notice: during the data collection period, one forum recorded 86 viewings and the other recorded 68 viewings. Only one message was received from a viewer who was unable to load the experiment due to an inadequate version of Java. The administrator for the one on-line education and innovative technology site posted on that site a newspaper-style article announcing the experiment. One of the 13 public computing facilities (i.e., senior centers and public libraries) responded to the introductory email and indicated that the experiment would be used for training purposes in the next computing class for seniors, to take place in September 2007. Sixty-four percent (32/50) of the respondents provided email addresses; the majority of these supplied-email addresses (26/32) were recognized from the author's personal contact list.

4.2 Study sample

4.2.1 Sample description

We excluded all non-native English speaker response that accounted for 20 percent (10/50) of the data submissions. Researchers have shown that compared with native English speakers, non-native English speakers may significantly differ in language perception (Jones, Berry, &

Stevens, 2007). We therefore included this native language criterion in our analysis plan to avoid the possible biasing effect that would have been introduced by the inclusion of non-native English speakers. All experimental submissions included complete data and were therefore usable without having to consider missing data; by using WebExp2 our experiment design enforced complete-data-only submissions.

4.2.2 Demographics

The age distribution of the analysis sample was approximately symmetrical; we cannot reject the hypothesis that age was normally distributed (Shapiro-Wilk test for normal data $W = 0.96$, $p = .18$), at least at the 18% level. The ages ranged from 27 through 76 years ($M = 57.7$, $SD = 11.79$, $Mdn = 59.0$). We present in Table C.4 the distribution of analysis sample respondents by age category (20-39, 40-59, and 60 years or older) and gender; age category and gender were independent in this analysis sample (Fisher’s exact test, $p = .218$).

Overall the analysis sample was approximately balanced by gender (females: 52.5%, 21/40), spent most of their youth in the United States or in the United Kingdom (90%, 36/40), and reported having attended at least some college (92.5%, 37/40).The majority of these respondents (57.5%, 23/40) were either employed for wages (37.5%, 15/40) or self-employed (20%, 8/40). The size of the retired employment category fell between these two employed groups (i.e., employed for wages and self-employed) and accounted for 30% (12/40) of the respondents in the analysis sample.

We present in Annex C, Tables C.1 – C.10 two-way tabulations of the categorical demographic variables including gender, occupation, education, country of youth, and age. The test values for Fisher’s exact tests presented in Table 4.1 indicate education and country of youth were associated ($p = .021$) as were occupation and age ($p = .005$). Compared to expected numbers, fewer respondents who spent their youth in the United Kingdom reported college- or university-only education attainment ($n = 0/9$, 0.00%) and more reported some postgraduate education ($n = 3/9$, 33.3%) whereas more respondents who spent their youth in the United States reported college- or university-only education attainment ($n = 11/27$, 40.7%) (Table C.8). Also associated were occupation and age (Table C.7). Compared to expected numbers, fewer respondents aged 40-49 years were retired ($n = 1/19$, 5.3%) whereas more respondents aged at least 60 years reported living in retirement ($n = 11/19$, 57.9%).

Table 4.1

Tests of Two-Way Association between Demographic Variables

Gender					
Employment	0.435				
Education	0.628	0.066			
Youth Country	1.000	0.173	0.021		
Age group	0.218	0.005	0.954	0.723	
	Gender	Employment	Education	Youth Country	Age group

Note. *p* < .05 indicates a significant association between two demographic variables. All tests were conducted with Fisher's exact two-sided statistics.

4.2.3 Personality scores

The agreeableness and emotional stability personality score distributions of the analysis sample were approximately symmetrical. We cannot reject the hypothesis that agreeableness was normally distributed (Shapiro-Wilk test for normal data $W(40) = 0.989$, $p = .952$) nor can we reject the hypothesis that emotional stability was normally distributed ($W(40) = 0.964$, $p = .232$).

Compared with males, female agreeableness scores tended to be higher ($p = .052$) whereas their emotional stability scores were similar ($p = .042$) (Table C.11). The agreeableness scores ranged from 62-100 and 54-95 for females and males, respectively. Similarly, the emotional stability scores ranged from 36-87 and 55-92 for females and males, respectively. The scales for both personality traits were anchored at 0 and ranged through 100.

4.2.4 Voice-content scores

Of the four voice emotion-message content combinations, our participants scored the happy-positive combination, on average, the highest (*M* = 3.07, *Mdn* = 3.00), and the happy-negative combination the second highest (*M* = 2.89, *Mdn* = 3.00). Our participants rated, on average, the sad-negative combination (*M* = 2.17, *Mdn* = 2.00) higher than the sad-positive combination (*M* = 2.08, *Mdn* = 2.00) (Table C.12).

Provided in Table 4.2 are the percentage of scores for each voice and content combination. For both the happy-negative and happy-positive combinations, study participants recorded the greatest number of scores as 'suitable' (viz., that the voice style was suitable to the message content): 37 percent and 43 percent, respectively, of all scores for those combinations. For both the sad-negative and sad-positive combinations, study participants recorded the greatest number of scores as 'unsuitable' (viz., that the voice style was unsuitable to the message content) --46 percent and 41 percent, respectively-- of all scores for those combinations. These data further support our overall finding that our survey participants preferred a happy voice to deliver both negative and positive messages.

Table 4.2

Distribution of Likert Suitability Ratings for Voice and Content Combinations: Percentages in the Analysis Sample

Rating	Happy-Negative	Happy-Positive	Sad-Negative	Sad-Positive
Not at all suitable (1)	8	8	21	28
Unsuitable (2)	26	14	46	41
Suitable (3)	37	43	28	28
Highly suitable (4)	30	36	5	4
Total	100	100	100	100

Table 4.3 illustrates that participants classified as the youngest (20-39 years) scored the sad-positive combination the highest of the four voice-content combinations, compared with the participants aged 40-59 years who scored the happy-positive combination the highest, and with the oldest participants (60 years and older) who scored the happy-negative combination the highest. All participants scored sad voice-combinations as their least-preferred voice-content combination.

Annex C, Tables C.12-C.21 highlight important differences between the sexes and among age groups, occupation, educational, and location. Even though all survey participants preferred the happy-positive combination, females rated the happy voice higher than males (regardless of message content) and they rated the sad voice lower than males (regardless of message content) (Table C.12). Table C.13, Section B, shows important differences between the sexes. Approximately two-thirds of the females who participated in our study scored the sad voice combinations as 'not at all suitable' compared with approximately one-third of males who rated those combinations the same. In contrast, one-third of the females in our survey scored the happy combinations as 'not at all suitable', compared with two-thirds of the males.

Table 4.3

Comparison of Suitability Ratings for Most-Preferred and Least-Preferred Voice-Content Combinations, by Age Group

	20-39 years old	40-59 years old	60+ years old
Most preferred combination:	Sad-Positive	Happy-Positive	Happy-Negative
M	2.75	2.99	3.19
Mdn	2.50	3.00	3.00
Least preferred combination:	Sad-Negative	Sad-Positive	Sad-Positive
M	2.40	1.94	2.14
Mdn	2.00	2.00	2.00

Note. *M* = mean; *Mdn* = median. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50).

Source: Annex C, Table C.14.

Annex C, Table C.15 illustrates that differences exist among the distribution of suitability ratings scores by the selected age groups. Despite their dislike for certain voice-content combinations, approximately 30 percent of the three age groups scored their least favorite combination as 'suitable'. In terms of occupation, the highest ratings across all four combinations were recorded by the home-maker participant (Tables C.16 and C.17). All participants, regardless of occupation and as previously noted, preferred the sad-positive combination the least. The distributions of suitability ratings for the study participants within employment category varied considerably across most categories, although the distributions of scores for those study participants employed for wages and those who were retired were somewhat similar.

Regarding education (Tables C.18 and C.19), all participants except that in the category of high school (or secondary school equivalent) rated the happy-positive voice the highest of the four voice combinations. Generalizations about the distribution of ratings within voice-content combination and within the total are difficult to make given the absence of an obvious pattern. When analyzed according to country, participants in all countries rated the happy-positive combination the highest (Table C.20). Participants from three of the four countries rated the sad-positive combination the lowest. The distribution of scores within location (Table C.21, Section A) was most similar for the sad combinations across location (viz., the greatest number of scores was recorded as 'unsuitable' for each location).

4.3 Responses to research questions

4.3.1 Two-way ANOVA

We list in Table 4.4 the results of the within-subjects two factor repeated measures ANOVA with covariates. Guided by our conceptual framework, we fit a model to the sample data to examine the factors and covariates, which may have generated the subject-reported message suitability ratings. The model consisted of voice emotion and message content as the within-subjects factors and the personality traits of agreeableness and emotional stability as covariates. Further, we included the interaction effects between (a) personality and voice emotion; (b) personality and message content; (c) voice emotion and message content; and (d) personality, voice emotion, and message content.

The marginal means of voice emotion and message content are shown in Table 4.5. Voice emotion was significantly associated with user-reported message suitability ratings ($F(1,397) = 6.17$, $p = .01$) although the association of user message ratings with message content was not significant ($F(1,397) = 2.15$, $p = .14$) (Table 4.4). Similarly, user personality was associated with user-rated message suitability with significant direct effects for both agreeableness ($F(1,397) = 4.61$, $p = .03$) and emotional stability ($F(1,397) = 14.69$, $p = .00$) (Table 4.4).

Research question 1:
Does message content affect user rating of voice emotion suitability?

Results from the statistical analyses we conducted with these sample data indicate that message content — operationalized as positive or negative statements— did not affect user rating of voice emotion suitability. The two-way interaction term of Voice Emotion X Message Content was not significant, $F(1,397) = 1.71$, $p = .19$.

Research question 2:
Does user personality affect user rating of voice emotion suitability (i.e., whether s/he prefers a happy or a sad voice to deliver a message)?

Results from the statistical analyses we conducted with these sample data indicate that user personality —operationalized as continuous scores for agreeableness and emotional stability— did affect user rating of voice emotion suitability. The two-way interaction term of Voice Emotion X Emotional Stability was significant, $F(1,397) = 24.41$, $p < .001$. The two-way interaction term of Voice Emotion X Agreeableness, however, was not significant, $F(1,397) = 2.76$, $p = .098$.

Table 4.4

Results of the Within-subjects Two Factor Repeated Measures ANOVA Controlling for Personality Covariates

Source	df	*F*	η^2	p
		Between subjects		
Agreeableness (A)	1	4.61	.01	.03*
Emotional Stability (E)	1	14.69	.04	.00***
Error	397	122.54		
		Within subjects		
Voice Emotion (V)	1	6.17	.02	.01*
V X A	1	2.76	.01	.10
V X E	1	24.41	.06	.00***
V error	397	374.67		
Message Content (C)	1	2.15	.01	.14
C X A	1	3.36	.01	.07
C X E	1	.23	.00	.63
C error	397	135.11		
V X C	1	1.71	.00	.19
V X C X A	1	.86	.00	.36
V X C X E	1	3.24	.01	.07
V X C error	397	567.24		

Note. Values in parentheses represent mean square errors. *$p < .05$. ** $p < .01$. ***$p < .001$.

Table 4.5

Marginal Means of User Suitability Ratings for Main Effects of Voice and Content

	Voice Emotion		
Message Content	Happy	Sad	Marginal Means (Message Content)
Negative	2.89	2.17	2.53
Positive	3.07	2.08	2.57
Marginal Means (Voice Emotion)	2.98	2.12	2.55

Note. Covariates in the model are evaluated at the following values: Agreeableness, 78.05; emotional stability, 71.00.

Question 3:
Does user personality affect which type of voice is preferred for which content?

Results, as previously presented under Question 1, from the statistical analyses we conducted with these analysis sample data indicate that there was no significant two-way interaction between voice emotion and message content, $F(1,397) = 1.71$, $p = .19$. Also, the three-way interaction terms of (a) Voice Emotion X Message Content X Agreeableness and (b) Voice Emotion X Message Content X Emotional Stability were not significant (Table 4.4). We therefore conclude that the level of voice emotion (i.e., happy or sad voice) does not interact with the level of message content (i.e., positive or negative message) to determine the user-reported message suitability rating.

4.3.2 Exploratory ANOVA

Table 4.6 presents our results for the additional ANOVA that was undertaken with gender and age group defined as the between-subjects factors, in addition to the preceding estimated ANOVA with (a) the personality traits defined as the covariates and (b) the voice emotion and message content defined as within-subject factors. We considered this within-subjects two factor, between-subjects two factor, repeated measures ANOVA with covariates as exploratory due to the imbalance of subjects by gender and age group.

Research question 1:
Does message content affect user rating of voice emotion suitability when we include age and gender in the model?

Results from the exploratory statistical analyses we conducted with these sample data indicate that message content —operationalized as positive or negative statements— did not affect user rating of voice emotion suitability when age and gender were included as between-subject factors in the estimated model. The two-way interaction term of Voice Emotion X Message Content was not significant, $F(1,393) = 1.25$, $p = .26$. Additionally, no higher order interactions that included the Voice Emotion X Message Content term were significant (Table 4.6).

Table 4.6

Results of Exploratory Analysis of Variance Results

Source	df	F	η^2	p
		Between subjects		
Agreeableness (A)	1	8.54	.02	.00**
Emotional Stability (E)	1	10.83	.03	.00**
Age groups (AG)	2	10.56	.05	.00***
Gender (G)	1	6.66	.02	.01*
AG X G	1	4.20	.01	.04*
Error	393	114.43		
		Within subjects		
Voice (V)	1	9.29	.02	.00**
V X A	1	.21	.00	.65
V X E	1	41.01	.09	.00***
V X G	1	16.60	.04	.00***
V X AG	2	1.88	.01	.15
V X G X AG	1	15.85	.04	.00***
V error	393	339.19		
Content (C)	1	2.60	.01	.11
C X A	1	4.81	.01	.03*
C X E	1	.25	.00	.62
C X G	1	1.00	.00	.32
C X AG	2	.75	.00	.47
C X G X AG	1	.00	.00	.96
C error	393	133.99		
V X C	1	1.25	.00	.26
V X C X A	1	.82	.00	.37
V X C X E	1	1.20	.00	.27
V X C X G	1	.76	.00	.38
V X C X AG	2	2.86	.01	.06
V X C X S X AG	1	.22	.00	.64
V X C error	393	558.52		

Note. Values in parentheses represent mean square errors. *$p < .05$. ** $p < .01$. ***$p < .001$.

Annex C, Tables C.22-C.27 illustrate differences in marginal means for scores by age groups and gender. Cell means in Table C.22 show an increasing score for age for a happy voice compared with a sad voice. Cell means also illustrate that scores for the happy voice stimuli increased as participants grew older. Generalizations could not be drawn from Table C.23 for the different levels of age groups across message content levels. As previously noted, females in our study scored the happy voices higher than males and the sad voices lower than males.

They also scored both message content combinations lower than males. The differences in female versus male cell means according to voice emotion were important (Table C.24); in contrast, the differences in female versus male cell means according to message content were not great (Table C.25). Lastly, Annex C, Tables C.26 and C.27 illustrate that differences existed among the female versus male marginal means according to different levels of age and both voice emotion and message content. We can generalize that as females grow older their scores for both negative and positive message content increase. No such pattern is discernible according for males.

Research question 2:
Does user personality affect user rating of voice emotion suitability (i.e., whether s/he prefers a happy or a sad voice to deliver a message) when we include age and gender in the model?

Results from the exploratory statistical analyses we conducted with these sample data indicate that user personality —operationalized as continuous scores for agreeableness and emotional stability— did affect user rating of voice emotion suitability when age and gender were included as between-subject factors in the estimated model. The two-way interaction term of Voice Emotion X Emotional Stability was significant, $F(1,393) = 41.01$, $p < .001$. The two-way interaction term of Voice Emotion X Agreeableness, however, was not significant, $F(1,397) = .21$, $p = .65$. Nonetheless, results from our study showed significant main effects for the both the emotional stability personality trait, $F(1,393) = 10.83$, $p = .001$ and the agreeableness personality trait, $F(1,393) = 8.54$, $p = .004$.

Question 3:
Does the user's personality affect which type of voice is preferred for which content when we include age and gender in the model?

Results, as stated under exploratory Question 1, from the exploratory statistical analyses we conducted with these sample data indicate that there was no significant two-way interaction between voice emotion and message content when age and gender were included as between-subject factors in the estimated model, $F(1,393) = 1.25$, $p = .26$ (Table 4.6). Moreover, the three-way interaction term of Voice Emotion X Message Content X Agreeableness was not significant, $F(1,393)$, $= .82$, $p = .37$, when age and gender were included as between-subject factors in the estimated model. Similarly, the three-way interaction term of Voice Emotion X Message Content X Emotional Stability was also not significant, $F(1,393) =1.20$, $p = .27$, when age and gender were included as between-subject factors in the estimated model. We therefore conclude that the level of voice emotion (i.e., happy or sad voice) did not interact with the level of message content (i.e., positive or negative message valence) to determine the user-reported message suitability rating when age and gender were included as between-subject factors in the estimated model.

4.3.3 ANOVA summary

To summarize, for the two factor within-subjects repeated measures ANOVA with covariates, our experiment showed significant main effects for voice emotion, agreeableness, and emotional stability and a non-significant main effect for message content. The only significant two-way interaction effect was for Voice Emotion X Emotional Stability. The three-way interactions of Voice Emotion X Message Content X Emotional Stability and Voice Emotion X Message Content X Agreeableness were not significant.

When we incorporated age and gender into the ANOVA as between-subjects factors, our results showed significant main effects for voice emotion, agreeableness, emotional stability, age, and gender. We recorded significant two-way interaction effects for Voice Emotion X Emotional Stability, Voice Emotion X Gender, Message Content X Agreeableness, and Gender X Age. We recorded only one significant three-way effect and that was for the Voice Emotion X Gender X Age term. The low observed power concerning all significant results suggests that more data are needed to better estimate the associations between the user suitability ratings and the between-subjects factors (i.e., gender and age), the within-subjects factors (i.e., voice emotion and message content), and the covariates (agreeableness and emotional stability).

Analyses of marginal means indicated that differences exist among the marginal means for the different levels of age across the voice emotion levels, and to a lesser extent across message content levels. Cell means show an increasing score for levels of age for a happy voice compared with a sad voice. Cell means also illustrate that scores for the happy voice stimuli increased with participant age. Finally, the only generalization that can be gleaned from data for marginal means according to gender and age groups is that as females grow older their suitability rating scores for negative message content increase.

5 Discussion

The purpose of this final chapter is to discuss the results of the dissertation and to provide recommendations for future research. We first begin by briefly restating the main results and then assess how those results relate to the research questions laid out at the beginning of the dissertation. We then connect the findings to those of previous studies in HCI, and discuss theoretical implications. We also raise issues concerning the limitations of the study and assess their likely effects on our results. Finally, we make suggestions for further research.

5.1 Main results

For the two factor within-subjects repeated measures ANOVA with covariates, our experiment showed significant main effects for voice emotion, agreeableness, and emotional stability and a non-significant main effect for message content. The only significant two-way interaction effect was for Voice Emotion X Emotional Stability. When we incorporated age and gender into the ANOVA as between-subjects factors, our results showed significant main effects for voice emotion, agreeableness, emotional stability, age, and gender. We recorded significant two-way interaction effects for Voice Emotion X Emotional Stability, Voice Emotion X Gender, Message Content X Agreeableness, and Gender X Age. We recorded only one significant three-way effect and that was for the Voice Emotion X Gender X Age term.

We now relate these results to our research questions. First, we questioned whether message content (positive or negative statement) affects user rating of voice emotion suitability. We found that it had no effect on user rating of voice emotion suitability. When we included age and gender as between-subjects factors, we also found no effect of message content on user rating of voice emotion suitability.

Second, we questioned whether a user's personality influences whether he or she prefers a happy or sad voice to deliver a message. We found that user personality (agreeableness and emotional stability) did make a difference, with and without the inclusion of the age and gender variables. Moreover, we found that how a person reacts to a voice depends more on their emotional stability than on their agreeableness. Nonetheless, we found that the two-way interactions effects of Voice Emotion X Agreeableness were not significant with or without age and gender included in our model.

Finally, we asked whether the user personality affects which type of voice is preferred for which type of content. As stated, we found no significant two-way interaction between voice emotion and message content, with or without age and gender included in our model. Moreover, we found no significant three-way interaction among (a) Voice Emotion X Message Content X Agreeableness and (b) Voice Emotion X Message Content X Emotional Stability, with or without age and gender. We therefore conclude that the level of voice emotion (i.e., happy or sad voice) does not interact with the level of message content (i.e., positive or negative message) to determine the user-reported message suitability rating.

5.2 Theoretical and practical implications

Our findings are important for theory and practice. At the broad theoretical level concerning the interaction between humans and computers, they provide more data about each of the concepts studied, viz., personality behavior, emotions, and message content, independently of and in combination with each other, and the added variables of age and gender. In particular, they offer a better understanding of two personality traits that have not been widely studied within that context --agreeableness and emotional stability-- and the effects of combining human personality with computer voice-generated messages that carry emotion. The findings also offer insights to the preferences of an older age group that largely has been excluded in previous HCI studies.

Originally we were inspired to undertake this study by Nass and Lee (2001) who extended the similarity-attraction hypothesis first tested by Nass et al. (1995) to synthetic voices. We were also intrigued by Isbister and Nass (2000), who revealed that the consistency-attraction hypothesis was also applicable to human-computer social interactions. Moon's (2002) recent research on message customization according to user personality also piqued our interest. All four studies involved personality, and in particular, the extraversion-introversion dimension. Each one justified the claim that computer personalities can equate to human personalities and more generally, that humans depend heavily on verbal and non-verbal cues to perceive speech, whether those cues come from another human or a computer, which in turn lead them to process messages, make decisions, and take action. However, only one study (Moon, 2002) crossed personality with another variable (viz., content). By virtue of the fact that our study is the first known one to cross personality with synthetic voice emotion and content, it makes an important contribution to HCI.

Overall, those studies first revealed that manipulation of voice emotion and message content in computers is not difficult and is powerful. While we did not manipulate computer personality, we confirmed that different personality traits of a user can lead researchers to draw distinct conclusions and that synthetic voice emotion has important implications for user preferences.

Second, past research (Nass & Lee, 2001) emphasized that paralinguistic cues are relevant to users' responses to synthetic voices, and that synthetic voices in turn influence perceptions of content. Our experimental results support the finding that the paralinguistic cue of voice emotion impacted on users' suitability ratings, and that the personality trait of emotional stability had more of an influence on those ratings than the trait of agreeableness. However, our results do not support their finding that synthetic voice influences perceptions of content. Only when we included age and gender in our analysis was there an interaction between user personality (agreeableness) and message content.

These results raise the question of the extent to which the synthetic voice over-rode the message content. Even though the synthetic voice used in our experiment was intelligible and participants were reminded throughout the experiment to assess its suitability to the content, the voice nonetheless had some misplaced accents and word emphases, and odd intonation and rhythm. As emphasized by Nass and Lee (2001), although users might dismiss these disfluencies as technological constraints, these paralinguistic cues nonetheless play a critical role in human-computer interaction, as they do in human-human interaction, and could have caused the participant to focus more on the characteristics of the voice than on the suitability of the voice to the message.

Finally, past research revealed that people relate to computers in a well-defined way, as projected through a synthetic voice, and that those voices need to be consistent with a user's personality. While matching user and computer personality was not an objective of this study, the results of our user-rated audio scores generally supported the psychological research of Larsen and Ketelaar (1991) and McCrae and Costa (1992) that disclosed that emotionally unstable people are more susceptible to negative affect than those who are not, and that agreeable people tend to prefer positive affect over negative affect.

5.3 Study Limitations

5.3.1 Data

The study sample met our objective of representing a group of mature adults, who up to now have not been well-represented by the existing literature in this emergent research area of HCI. Indeed, all but 2 of our 40 participants were at least 40 years old. However, given the short timeframe permitted to collect data, participants were recruited simultaneously from two sources: personal contacts 40 years and older, and computer users who frequented selected websites aimed specifically at the 50+ age group. Therefore, our study sample effectively resulted in a convenience sample combined with a random sample to overcome the difficulties of obtaining adequate response rates in the short time frame. Had time permitted, additional website administrators could have been contacted, and a sample could have been obtained through a senior organization.

5.3.2 Instrumentation

The choice of WebExp2 software proved to have several advantage and disadvantages. Given that our experiment was Web-based, its administration was automated, which cut back on time and associated costs that otherwise would have been required to manipulate data manually, organize appointments, and track experimental results. It also offered a large pool of potential participants that could be accessed without time constraints, thereby allowing for a potentially high statistical power and correspondingly, a variety of results. Other advantages include the opportunity for volunteer participation and the ease of access for participants, and ecological validity. Finally, the instrumentation followed most of the effective design principles laid out by the Tailored Design Method by Dillman (2007) (e.g., welcome screen with instructions; consistency in color, font, and page design; limited Likert scale choices).

However, because WebExp2 was written in Java, participants' computers required that a recent version of Java (1.5.0 or higher) be loaded and that their browsers be enabled. Even though the introductory screen of the experiment highlighted this requirement, many participants sent emails advising that they could not advance past the first page. Replies were sent with detailed step-by-step instructions that possibly could have been avoided had had they been linked or included on the introductory page of the experiment. Had the instrumentation been written in Hypertext Mark-up Language (HTML), the Java problem could have been avoided all together. Even though the number of failed or aborted attempts to run the experiment is unknown, the extensive correspondence with personal contacts about Java indicates that this technical requirement likely influenced the number of participants in the study sample.

5.4 Recommendations

5.4.1 Data collection and instrumentation

Future improvements in data collection techniques and instrumentation design for Web-based experiments would likely lead to higher response rates and reduced coverage and sampling errors. If time permitted, future research efforts could attempt to randomly select samples off-line in order to reduce sampling error inherent in an on-line survey. Likewise, if participants in a narrowly-defined age group, such as seniors 60 years and older, were the focus, then the data collection process could be designed and implemented within that boundary, which would likely reduce the coverage error. One way to limit questionnaire access would be to require a PIN or identification number. However, the data collection process for that type of instrumentation would likely require many more steps than ours, such as introduction and follow-up emails, thereby raising costs in terms of administration and time.

Regarding the instrumentation, future Web survey designers should carefully consider the use of WebExp2 and choose HTML programming language if possible to avoid likely participant confusion over the JAVA requirement. Facilitating the engagement of the participant is the first step in improving response rates. Keeping the survey short and varied, which were two characteristics of our survey, also will help researchers to ensure that the survey is completed. Finally, the use of HTML will allow for improved design, such as progress bars and the ability to move backwards or forwards without aborting the questionnaire. Regardless of the programming language, a debriefing screen that includes a comment box will provide more information such as possible listener bias towards prosody imbalances of synthetic speech, as suggested by some of our participant comments, or a few questions on ways to improve future research. This type of debriefing would encourage anonymous feedback and avoid forcing the participant to take an extra step and send an email, as our experiment required.

5.4.2 Future research

It is hoped that this study will spark sufficient interest to further investigate emotional speech synthesis and HCI in three areas. First, it would be useful to investigate in more depth the mixing of personality, voice emotion, and message content with a larger study sample to determine the real significance of these effects, especially across age groups and gender. Even though our results showed that content had a limited influence on user preference for voice, more research is needed within the context of voice emotion and personality to understand this finding. Likewise, future work could investigate the effects of using a male or female synthetic voice or a voice with varied accents.

Second, there are many issues that require exploration in the field of aging and computer technology that the present study has begun to address, viz., differences between genders. The ratings on our audio scores suggest that as women grow older, they prefer a happy voice to deliver a message regardless of its content. Further preliminary investigation is required to know more about the profile of users of a smart home system, which was used as the framework for this study. For example, researchers and interface designers would benefit from knowing the typical age of users of this system and whether they have hearing, cognitive, or motor limitations. Likewise, a survey of attitudes of potential and current users towards smart homes would help to define direction and allow for customization of such a system within an experimental setting.

Finally, in addition to a need for more research on technology and the aging, this study has uncovered the need to better understand how non-native English speakers perceive synthetic speech compared with native English speakers. To date, only a few studies have been conducted using this study sample. This segment of the population would benefit from inclusion in future research given their growing number in English-speaking countries. They also likely represent a viable market in terms of commercialization of any speech-based system or product. Nonetheless, this topic is more complicated than at face value: cultural issues would likely need to be introduced to this research.

Annex A Correspondence with Potential Participants

Table A.1

Initial Correspondence: Introductory Email to Websites, Senior Centers, and Public Libraries

Dear [website name], I found your site while researching older people and Internet usage and need your help. I am a master's student at the University of Edinburgh in Scotland working towards a degree in speech technology. As part of my thesis, I am conducting an on-line experiment to test how older computer users perceive synthetic voices. This research is very important because there has been very little work done on this subject involving older users. As a user with more 'life' experience (47 years old), I often wish that many computer applications were geared towards an older crowd. The experiment takes only about 10-15 minutes and involves no risks. Participants can enter a free raffle to win a US$25 gift certificate redeemable at an Amazon website anywhere in the world. They can also receive a short summary of results in September once data are analyzed. The first part involves listening to voices generated by a computer and rating them; the second part is a personality test. The experiment is on-line now and only requires JAVA to be installed on the computer and enabled in the browser, and the speakers on. The link is: http://fordyce.inf.ed.ac.uk/webexp2/voices-smart-homes.html I therefore would like to ask whether it will be possible to post a short paragraph inviting [website name] members to participate in this experiment. This is a good opportunity for older computer users to have their voices heard. I welcome all feedback and comments from participants in the experiment and will answer them personally. For more information on making voice interface technology more accessible to older adults, and research on making computers easier to operate, you might want to have a look at the home pages of my two thesis advisors, Dr. Maria Wolters (http://homepages.inf.ed.ac.uk/mwolters) and Professor Jon Oberlander (http://homepages.inf.ed.ac.uk/jon/). This experiment also directly supports the MATCH project (http://www.match-project.org.uk), which is working on home care systems to support independent living. I look forward to hearing from you at your earliest convenience. Many thanks, Greta Boye MSc Speech and Language Processing, 2007 University of Edinburgh greta.boye.experiment @googlemail.com

Table A.2

Follow-Up Correspondence: Posting to Website Forum

University of Edinburgh urgent research request I am a 'mature' (47 years old) Master's student at the University of Edinburgh in Scotland and need your help with an on-line experiment to test how older computer users perceive synthetic voices. Little work has been done on this subject, so results will provide more information on making voice interface technology more accessible to older adults, and support research on making computers easier to operate. Participants can receive a short summary of results in September once data are analyzed, and enter a raffle to be one of five persons to win a $25 Amazon gift certificate redeemable world-wide. The research supports the University's ongoing research at its Center for Speech Technology Research and the UK-funded MATCH Program. Copy and paste http://fordyce.inf.ed.ac.uk/webexp2/voices-smart-homes.html into your browser and help seniors have their voices heard. Thank you!!

Annex B Experimental Queries: Demographics, Messages, and Personality Quiz

Table B.1

Demographic Queries Used in Experiment

Ref.	Query
1	What is your age in years? ___
2	Gender? male___ female___
3	What is your native language? English___ Other ___
4	Where did you spend most of your youth? USA ___ UK ___ Other______
5	What is the highest level of education you have completed? ___Primary/Grammar school ___High school/Secondary school equivalent ___ Vocational/Technical school or college ___Some college/University ___College/University graduate ___Some postgraduate education ___Postgraduate/Professional degree (e.g., PhD, MD)
6	Which of these best describes your main current occupational status? ___Employed for wages ___Self-employed ___Unemployed ___Home-maker ___Retired ___Student ___ Unable to work for health or other reasons

Table B.2

Synthetic Voice Stimuli Used in Experiment

Ref.	Stimulus	Content	Voice
1	The watering system did not turn on.	negative	sad
2	The blinds could not be opened.	negative	happy
3	The blinds have been opened.	positive	happy
4	The house alarm could not be turned on.	negative	sad
5	The oven has been switched off.	positive	sad
6	The iron has been switched on.	positive	sad
7	The oven has been switched off.	positive	happy
8	The security company could not be called.	negative	sad
9	The patio door cannot be closed.	negative	happy
10	The iron could not be switched off.	negative	happy
11	The television has been switched off.	positive	sad
12	The iron has been switched on.	positive	happy
13	The blinds could not be opened.	negative	sad
14	The watering system did not turn on.	negative	happy
15	The computer could not be turned off.	negative	happy
16	The iron could not be switched off.	negative	sad
17	The oven could not be switched off.	negative	happy
18	The watering system has been automatically turned on.	positive	happy
19	The house alarm has been turned on.	positive	happy
20	The fax machine could not be switched on.	negative	sad
21	The oven could not be switched off.	negative	sad
22	The patio door has been closed.	positive	happy
23	The house alarm could not be turned on.	negative	happy
24	The television could not be switched on.	negative	happy
25	The fax machine could not be switched on.	negative	happy
26	The patio door has been closed.	positive	sad
27	The security company has been called.	positive	sad
28	The computer has been turned off.	positive	sad
29	The television has been switched off.	positive	happy
30	The watering system has been automatically turned on.	positive	sad
31	The television could not be switched on.	negative	sad
32	The patio door cannot be closed.	negative	sad
33	The fax machine has been switched on.	positive	happy
34	The security company could not be called.	negative	happy
35	The security company has been called.	positive	happy
36	The blinds have been opened.	positive	sad
37	The computer could not be turned off.	negative	sad
38	The fax machine has been switched on.	positive	sad
39	The house alarm has been turned on.	positive	sad
40	The computer has been turned off.	positive	happy

Table B.3

Personality Questions Used in Experiment

Ref.	Characteristic	Personality Type (+/-)
1	Am interested in people.	Agreeableness (+)
2	Sympathize with others' feelings.	Agreeableness (+)
3	Have a soft heart.	Agreeableness (+)
4	Take time out for others.	Agreeableness (+)
5	Feel others' emotions.	Agreeableness (+)
6	Make people feel at ease.	Agreeableness (+)
7	Inquire about others' well-being.	Agreeableness (+)
8	Know how to comfort others.	Agreeableness (+)
9	Love children.	Agreeableness (+)
10	Am on good terms with nearly everyone.	Agreeableness (+)
11	Have a good word for everyone.	Agreeableness (+)
12	Show my gratitude.	Agreeableness (+)
13	Think of others first.	Agreeableness (+)
14	Love to help others.	Agreeableness (+)
15	Insult people.	Agreeableness (-)
16	Am not interested in other people's problems.	Agreeableness (-)
17	Feel little concern for others.	Agreeableness (-)
18	Am not really interested in others.	Agreeableness (-)
19	Am hard to get to know.	Agreeableness (-)
20	Am indifferent to the feelings of others.	Agreeableness (-)
21	Am relaxed most of the time.	Emotional Stability (Neuroticism) (+)
22	Seldom feel blue.	Emotional Stability (Neuroticism) (+)
23	Am not easily bothered by things.	Emotional Stability (Neuroticism) (+)
24	Rarely get irritated.	Emotional Stability (Neuroticism) (+)
25	Seldom get mad.	Emotional Stability (Neuroticism) (+)
26	Get stressed out easily.	Emotional Stability (Neuroticism) (-)
27	Worry about things.	Emotional Stability (Neuroticism) (-)
28	Am easily disturbed.	Emotional Stability (Neuroticism) (-)
29	Get upset easily.	Emotional Stability (Neuroticism) (-)
30	Change my mood a lot.	Emotional Stability (Neuroticism) (-)
31	Have frequent mood swings.	Emotional Stability (Neuroticism) (-)
32	Get irritated easily.	Emotional Stability (Neuroticism) (-)
33	Often feel blue.	Emotional Stability (Neuroticism) (-)
34	Get angry easily.	Emotional Stability (Neuroticism) (-)
35	Panic easily.	Emotional Stability (Neuroticism) (-)
36	Feel threatened easily.	Emotional Stability (Neuroticism) (-)
37	Get overwhelmed by emotions.	Emotional Stability (Neuroticism) (-)
38	Take offense easily.	Emotional Stability (Neuroticism) (-)
39	Get caught up in my problems.	Emotional Stability (Neuroticism) (-)
40	Grumble about things.	Emotional Stability (Neuroticism) (-)

Note: The positive (+) and negative (-) signs are part of the IPIP scoring key. For details, see http://ipip.ori.org/ newScoringInstructions.htm

Source: http://ipip.ori.org/newBigFive5broadKey.htm

Annex C Statistical Appendix

Table C.1
Gender by Employment Category Cross Tabulation: Analysis Sample

	Employed	Home-maker	Retired	Self-employed	Student	Employed	Total
Female (count)	7	2	7	3	2	—	21
w/in sex (%)	33.33	9.52	33.33	14.29	9.52	—	100.00
w/in employment (%)	46.67	100.00	58.33	37.50	100.00	—	52.50
w/in total (%)	17.50	5.00	17.50	7.50	5.00	—	52.50
Male (count)	8	—	5	5	—	1	19
w/in sex (%)	42.11	—	26.32	26.32	—	5.26	100.00
w/in employment (%)	53.33	—	41.67	62.50	—	100.00	47.50
w/in total (%)	20.00	—	12.50	12.50	—	2.50	47.50
Total	15	2	12	8	2	1	40
w/in sex (%)	37.50	5.00	30.00	20.00	5.00	2.50	100.00
w/in employment (%)	100.00	100.00	100.00	100.00	100.00	100.00	100.00
w/in total (%)	37.50	5.00	30.00	20.00	5.00	2.50	100.00

Note. Fisher's exact test = 0.44. Dashed cells (—) indicate that no respondents self-reported along these column/row dimensions. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50).

Table C.2
Gender by Education Category Cross Tabulation: Analysis Sample

	College/ Univ. Graduate	High School/ 2ndary School Equiv.	Postgrad./ Prof. Degree	Some College/ Univ.	Some Postgrad. Educ.	Vocat./ Tech. School	Total
Female (count)	7	1	7	2	3	1	21
w/in sex (%)	33.30	4.76	33.33	9.52	14.29	4.76	100.00
w/in education (%)	58.30	100.00	38.89	100.00	60.00	50.00	52.50
w/in total (%)	17.50	2.50	17.50	5.00	7.50	2.50	52.50
Male (count)	5	-	11	-	2	1	19
w/in sex (%)	26.32	-	57.89	-	10.53	5.26	100.00
w/in education (%)	41.67	-	61.11	-	40.00	50.00	47.50
w/in total (%)	12.50	-	27.50	-	5.00	2.50	47.50
Total	12	1	18	2	5	2	40
w/in sex (%)	30.00	2.50	45.00	5.00	12.50	5.00	100.00
w/in education (%)	100.00	100.00	100.00	100.00	100.00	100.00	100.00
w/in total (%)	30.00	2.50	45.00	5.00	12.50	5.00	100.00

Note. Fisher's exact test = .63. Dashed cells (—) indicate that no respondents self-reported along these column/row dimensions. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50).

Table C.3
Gender by Country-of-youth Cross Tabulation: Analysis Sample

	Australia	Canada	New Zealand	UK	USA	Total
Female (count)	1	1	—	5	14	21
Within female (%)	4.76	4.76	—	23.81	66.67	100.00
Within employment (%)	100.00	100.00	—	55.56	51.85	52.50
Within total (%)	2.50	2.50	—	12.50	35.00	52.50
Male (count)	1	—	1	4	13	19
Within male (%)	5.26	—	5.26	21.05	68.42	100.00
Within employment (%)	100.00	—	100.00	44.44	48.15	47.50
Within total (%)	2.50	—	2.50	10.00	32.50	47.50
Total	1	1	1	9	27	40
Within location (%)	100.00	100.00	100.00	100.00	100.00	100.00
Within total (%)	2.50	2.50	2.50	22.50	67.50	100.00

Note. Fisher's exact test = 1.00. Dashed cells (—) indicate that no respondents self-reported along these column/row dimensions. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50).

Table C.4
Gender by Age Category Cross Tabulation: Analysis Sample

	20-39 years old	40-59 years old	60+ years old	Total
Female (count)	—	9	12	21
Within female (%)	—	42.86	57.14	100.00
Within employment (%)	—	47.37	63.16	52.50
Within total (%)	—	22.50	30.00	52.50
Male (count)	2	10	7	19
Within male (%)	10.53	52.63	36.84	100.00
Within employment (%)	100.00	52.63	36.84	47.50
Within total (%)	5.00	25.00	17.50	47.50
Total	2	19	19	40
Within Age Category (%)	100.00	100.00	100.00	100.00
Within total (%)	5.00	47.50	47.50	100.00

Note. Fisher's exact test = .22. Dashed cells (—) indicate that no respondents self-reported along these column/row dimensions. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50).

Table C.5
Employment by Education Category: Analysis Sample

	College/ Univ. Graduate	High School/ 2ndary School Equiv.	Postgrad./ Prof. Degree	Some College/ Univ.	Some Postgrad. Education	Vocat./ Tech. School	Total
Employed for wages (count)	1	—	9	—	3	2	15
Within employed for wages (%)	6.67	—	60.00	—	20.00	13.33	100.00
Within education category (%)	8.33	—	50.00	—	60.00	100.00	37.50
With total (%)	2.50	—	22.50	—	7.50	5.00	37.50
Home-maker (count)	1	—	—	1	—	—	2
Within home-maker (%)	50.00	—	—	50.00	—	—	100.00
Within education category (%)	8.33	—	—	50.00	—	—	5.00
Within total (%)	2.50	—	—	2.50	—	—	5.00
Retired (count)	6	1	4	1	—	—	12
Within retired (%)	50.00	8.33	33.33	8.33	—	—	100.00
Within education category (%)	50.00	100.00	22.22	50.00	—	—	30.00
Within total (%)	15.00	2.50	10.00	2.50	—	—	30.00
Self-employed (count)	4	—	3	—	1	—	8
Within self-employed (%)	50.00	—	37.50	—	12.50	—	100.00
Within education category (%)	33.00	—	16.67	—	20.00	—	20.00
Within total (%)	10.00	—	7.50	—	2.50	—	20.00
Student (count)	—	—	1	—	1	—	2
Within student (%)	—	—	50.00	—	50.00	—	100.00
Within education category (%)	—	—	5.56	—	20.00	—	5.00
Within total (%)	—	—	2.50	—	2.50	—	5.00
Unemployed (count)	—	—	1	—	—	—	1
Within unemployed (%)	—	—	100.00	—	—	—	100.00
Within education category (%)	—	—	5.56	—	—	—	2.50
Within total (%)	—	—	2.50	—	—	—	2.50
Total (count)	12	1	18	2	5	—	40
Within education category (%)	100.00	100.00	100.00	100.00	100.00	—	100.00
Within total (%)	30.00	2.50	45.00	5.00	12.50	—	100.00

Note. Fisher's exact test = .07. Dashed cells (—) indicate that no respondents self-reported along these column/row dimensions. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50).

Table C.6
Employment Category by Country-of-youth Cross Tabulation: Analysis Sample

	Australia	Canada	New Zealand	UK	USA	Total
Employed for wages (count)	—	1	1	6	7	15
Within employed for wages (%)	—	6.67	6.67	40.00	46.67	100.00
Within education category (%)	—	100.00	100.00	66.67	25.93	37.50
With total (%)	—	2.50	2.50	15.00	17.50	37.50
Home-maker (count)	1	—	—	—	1	2
Within home-maker (%)	50.00	—	—	—	50.00	100.00
Within education category (%)	100.00	—	—	—	3.70	5.00
Within total (%)	2.50	—	—	—	2.50	5.00
Retired (count)	—	—	—	1	11	12
Within retired (%)	—	—	—	8.33	91.67	100.00
Within education category (%)	—	—	—	11.11	40.74	30.00
Within total (%)	—	—	—	2.50	27.50	30.00
Self-employed (count)	1	—	—	1	6	8
Within self-employed (%)	2.50	—	—	12.50	75.00	100.00
Within education category (%)	100.00	—	—	11.11	22.22	20.00
Within total (%)	2.50	—	—	2.50	15.00	20.00
Student (count)	—	—	—	1	1	2
Within student (%)	—	—	—	50.00	50.00	100.00
Within education category (%)	—	—	—	11.11	3.70	5.00
Within total (%)	—	—	—	2.50	2.50	5.00
Unemployed (count)	—	—	—	—	1	1
Within unemployed (%)	—	—	—	—	100.00	100.00
Within education category (%)	—	—	—	—	3.70	2.50
Within total (%)	—	—	—	—	2.50	2.50
Total (count)	1	1	1	9	27	40
Within education category (%)	100.00	100.00	100.00	100.00	100.00	100.00
Within total (%)	5.00	2.50	2.50	22.50	67.50	100.00

Note. Fisher's exact test = .17. Dashed cells (—) indicate that no respondents self-reported along these column/row dimensions. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50).

Table C.7
Employment Category by Age Category Cross Tabulation: Analysis Sample

	20-39 years old	40-59 years old	60+ years old	Total
Employed for wages (count)	1	9	5	15
Within employed for wages (%)	6.67	60.00	33.33	100.00
Within education category (%)	50.00	47.37	26.32	37.50
With total (%)	2.50	22.50	12.50	37.50
Home-maker (count)	—	2	—	2
Within home-maker (%)	—	100.00	—	100.00
Within education category (%)	—	10.53	—	5.00
Within total (%)	—	5.00	—	5.00
Retired (count)	—	1	11	12
Within retired (%)	—	8.33	91.67	100.00
Within education category (%)	—	5.26	57.89	30.00
Within total (%)	—	2.50	27.50	30.00
Self-employed (count)	1	5	2	8
Within self-employed (%)	12.50	62.50	25.00	100.00
Within education category (%)	50.00	26.32	10.53	20.00
Within total (%)	2.50	12.50	5.00	20.00
Student (count)	—	2	—	2
Within student (%)	—	100.00	—	100.00
Within education category (%)	—	10.53	—	5.00
Within total (%)	—	5.00	—	5.00
Unemployed (count)	0	0	1	1
Within unemployed (%)	0.00	0.00	100.00	100.00
Within education category (%)	0.00	0.00	5.26	2.50
Within total (%)	0.00	0.00	2.50	2.50
Total (count)	2	19	19	40
Within education category (%)	100.00	100.00	100.00	100.00
Within total (%)	5.00	47.50	47.50	100.00

Note. Fisher's exact test = .01. Dashed cells (—) indicate that no respondents self-reported along these column/row dimensions. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50).

Table C.8
Education Category by Location-of-youth Cross Tabulation: Analysis Sample

	Australia	Canada	New Zealand	UK	USA	Total
College/university graduate (count)	1	—	—	—	11	12
Within college/university graduate (%)	8.33	—	—	—	91.67	100.00
Within country (%)	100.00	—	—	—	40.74	30.00
Within total (%)	2.50	—	—	—	27.50	30.00
High school/secondary equiv. (count)	—	—	—	1	—	1
Within high school/2nd equivalent (%)	—	—	—	100.00	—	100.00
Within country (%)	—	—	—	11.11	—	2.50
Within total (%)	—	—	—	2.50	—	2.50
Post graduate/professional (count)	1	0	1	4	12	18
Within post graduate/professional (%)	5.56	0.00	5.56	22.22	66.67	100.00
Within country (%)	100.00	0.00	100.00	44.44	44.44	45.00
Within total (%)	2.50	0.00	2.50	10.00	30.00	45.00
Some college/university (count)	—	—	—	—	2	2
Within some college/university (%)	—	—	—	—	100.00	100.00
Within country (%)	—	—	—	—	7.41	5.00
Within total (%)	—	—	—	—	5.00	5.00
Some post graduate (count)	—	—	—	3	2	5
Within some post graduate (%)	—	—	—	60.00	40.00	100.00
Within country (%)	—	—	—	33.33	7.41	12.50
Within total (%)	—	—	—	7.50	5.00	12.50
Vocational/technical school (count)	—	1	0	1	—	2
Within vocational/technical school (%)	—	50.00	0.00	50.00	—	100.00
Within country (%)	—	100.00	0.00	11.11	—	5.00
Within total (%)	—	2.50	0.00	2.50	—	5.00
Total (count)	2	1	1	9	27	40
Within country (%)	100.00	100.00	100.00	100.00	100.00	100.00
Within total (%)	5.00	2.50	2.50	22.50	67.50	100.00

Note. Fisher's exact test = .02. Dashed cells (—) indicate that no respondents self-reported along these column/row dimensions. The analysis sample consisted of only native English-speaking respondents (n = 40) versus the entire study sample that included native and non-native English speakers (N = 50).

Table C.9
Education Category by Age Category Cross Tabulation: Analysis Sample

	20-39 years old	40-59 years old	60+ years old	Total
College/university graduate (count)	1	4	7	12
Within college/university graduate (%)	8.33	33.33	58.33	100.00
Within country (%)	50.00	21.05	36.84	30.00
Within total (%)	2.50	10.00	17.50	30.00
High school/secondary equivalent (count)	—	—	1	1
Within high school/secondary equivalent (%)	—	—	100.00	100.00
Within country (%)	—	—	5.26	2.50
Within total (%)	—	—	2.50	2.50
Post graduate/professional (count)	1	10	7	18
Within post graduate/professional (%)	5.56	55.56	38.89	100.00
Within country (%)	50.00	52.63	36.84	45.00
Within total (%)	2.50	25.00	17.50	45.00
Some college/university (count)	—	1	1	2
Within some college/university (%)	—	50.00	50.00	100.00
Within country (%)	—	5.26	5.26	5.00
Within total (%)	—	2.50	2.50	5.00
Some post graduate (count)	—	3	2	5
Within some post graduate (%)	—	60.00	40.00	100.00
Within country (%)	—	15.79	10.53	12.50
Within total (%)	—	7.50	5.00	12.50
Vocational/technical school (count)	—	1	1	2
Within vocational/technical school (%)	—	50.00	50.00	100.00
Within country (%)	—	5.26	5.26	5.00
Within total (%)	—	2.50	2.50	5.00
Total (count)	2	19	19	40
Within country (%)	100.00	100.00	100.00	100.00
Within total (%)	5.00	47.50	47.50	100.00

Note. Fisher's exact test = .95. Dashed cells (—) indicate that no respondents self-reported along these column/row dimensions. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50).

Table C.10
Country-of-youth by Age Category Cross Tabulation: Analysis Sample

	20-39 years old	40-59 years old	60+ years old	Total
Australia	—	2	—	2
Within Australia (%)	—	100.00	—	100.00
With age category (%)	—	5.26	—	2.50
Within total (%)	—	2.50	—	2.50
Canada	—	—	1	1
Within Canada (%)	—	—	100.00	100.00
Within age category (%)	—	—	5.26	2.50
Within total (%)	—	—	2.50	2.50
New Zealand	—	1	—	1
Within New Zealand (%)	—	100.00	—	100.00
Within age category (%)	—	5.26	—	2.50
Within total (%)	—	2.50	—	2.50
UK	1	4	4	9
Within UK (%)	11.11	44.44	44.44	100.00
Within age category (%)	50.00	21.05	21.05	22.50
Within total (%)	2.50	10.00	10.00	22.50
USA	1	12	14	27
Within USA (%)	3.70	44.44	51.85	100.00
Within age category (%)	50.00	63.16	73.68	67.50
Within total (%)	2.50	30.00	35.00	67.50
Total	2	19	19	40
Within age category (%)	100.00	100.00	100.00	100.00
Within total (%)	5.00	47.50	47.50	100.00

Note. Fisher's exact test = .72. Dashed cells (—) indicate that no respondents self-reported along these column/row dimensions. The analysis sample consisted of only native English-speaking respondents (n = 40) versus the entire study sample that included native and non-native English speakers (N = 50).

Table C. 11

Personality Traits (Agreeableness and Emotional Stability) by Gender Measures of Central Tendency: Analysis Sample

	Personality trait	
	Agreeableness[a]	Emotional Stability[b] (Neuroticism)
Female	*M* = 81.24 *Mdn* = 82.00 *SD* = 10.69	*M* = 69.57 *Mdn* = 70.00 *SD* = 13.07
Male	*M* = 74.53 *Mdn* = 76.00 *SD* = 10.46	*M* = 72.58 *Mdn* = 70.00 *SD* = 9.91
Total	*M* = 78.05 *Mdn* = 77.50 *SD* = 10.98	*M* = 71.00 *Mdn* = 70.00 *SD* = 11.63

Note. *M* = mean; *Mdn* = median; *SD* = standard deviation. The analysis sample consisted of only native English-speaking respondents (n = 40) versus the entire study sample that included native and non-native English speakers (N = 50). The agreeableness and personality scales were anchored at 0 (low agreeableness and low emotional stability) and ranged through 100 (high agreeableness and high emotional stability).

[a] Equal variance (Levine F (1, 38) = .193, p = .663) two-tailed test of agreeableness for unpaired female versus male groups (t(38) = 2.03, p = .052). [b] Equal variance (Levine F (1, 38) = .538, p = .468) two-tailed test of emotional stability for unpaired female versus male groups (t(38) = -.81, p = .421).

Table C.12

Suitability Ratings for Voice-Content Scores by Gender Measures of Central Tendency: Analysis Sample

	Voice-content			
	Happy-Negative	Happy-Positive	Sad-Negative	Sad-Positive
Female	*M* = 2.99 *Mdn* = 3.00	*M* = 3.13 *Mdn* = 3.00	*M* = 2.07 *Mdn* = 2.00	*M* = 1.97 *Mdn* = 2.00
Male	*M* = 2.78 *Mdn* = 3.00	*M* = 3.01 *Mdn* = 3.00	*M* = 2.28 *Mdn* = 2.00	*M* = 2.20 *Mdn* = 2.00
Total	*M* = 2.89 *Mdn* = 3.00	*M* = 3.07 *Mdn* = 3.00	*M* = 2.17 *Mdn* = 2.00	*M* = 2.08 *Mdn* = 2.00

Note. M = mean; *Mdn* = median. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

Table C.13

Distribution of Likert Suitability Ratings for Voice and Content Combinations by Gender: Percentages in the Analysis Sample

A. Within gender	Happy-Negative			Happy-Positive			Sad-Negative			Sad-Positive		
	Female	Male	Total	Female	Male	Total	Female	Male	Total	Female	Male	Total
Not at all suitable (1)	5	11	8	5	11	8	25	16	21	36	19	28
Unsuitable (2)	27	25	26	17	10	14	45	48	46	33	49	41
Suitable (3)	34	41	37	39	48	43	28	28	28	30	26	28
Highly Suitable (4)	35	24	30	40	32	36	2	8	5	1	6	4
Total [a]	100	100	100	100	100	100	100	100	100	100	100	100
B. Within voice-content combination	Happy-Negative			Happy-Positive			Sad-Negative			Sad-Positive		
	Female	Male	Total [a]	Female	Male	Total [a]	Female	Male	Total [a]	Female	Male	Total [a]
Not at all suitable (1)	33	67	100	33	67	100	64	36	100	68	32	100
Unsuitable (2)	54	46	100	65	35	100	51	49	100	43	57	100
Suitable (3)	48	52	100	47	53	100	52	48	100	56	44	100
Highly Suitable (4)	61	39	100	58	42	100	25	75	100	20	80	100
Total	53	47	100	53	47	100	53	47	100	53	47	100
C. Within total	Happy-Negative			Happy-Positive			Sad-Negative			Sad-Positive		
	Female	Male	Total	Female	Male	Total	Female	Male	Total	Female	Male	Total
Not at all suitable (1)	3	5	8	3	5	8	13	8	21	19	9	28
Unsuitable (2)	14	12	26	9	5	14	24	23	47	18	23	41
Suitable (3)	18	19	37	21	23	44	15	14	29	16	12	28
Highly Suitable (4)	18	12	30	21	15	36	1	4	5	1	3	4
Total	53	47	100	53	47	100	53	47	100	53	47	100

Note. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

[a] May not sum to 100 due to rounding.

Table C.14

Suitability Ratings for Voice-Content Scores by Gender Measures of Central Tendency: Analysis Sample

	Voice-content			
	Happy-Negative	Happy-Positive	Sad-Negative	Sad-Positive
20-39 years old	*M* = 2.60 *Mdn* = 3.00	*M* = 2.65 *Mdn* = 3.00	*M* = 2.40 *Mdn* = 2.00	*M* = 2.75 *Mdn* = 2.50
40-59 years old	*M* = 2.77 *Mdn* = 3.00	*M* = 2.99 *Mdn* = 3.00	*M* =2.09 *Mdn* = 2.00	*M* = 1.94 *Mdn* = 2.00
60+ years old	*M* = 3.04 *Mdn* = 3.00	*M* = 3.19 *Mdn* = 3.00	*M* = 2.23 *Mdn* = 0.79	*M* = 2.14 *Mdn* = 2.00
Total	*M* = 2.89 *Mdn* = 3.00	*M* = 3.07 *Mdn* = 3.00	*M* = 2.17 *Mdn* = 2.00	*M* = 2.08 *Mdn* = 2.00

Note. M = mean; *Mdn* = median. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

Table C.15
Distribution of Likert Suitability Ratings for Voice and Content Combinations by Age Group: Percentages in the Analysis Sample

A. Within age group	Happy-Negative				Happy-Positive			
	20-39 years old	40-59 years old	60+ years old	Total	20-39 years old	40-59 years old	60+ years old	Total
Not at all suitable (1)	20	11	3	8	25	10	3	8
Unsuitable (2)	15	20	25	26	5	13	15	14
Suitable (3)	50	36	36	37	50	44	42	43
Highly Suitable (4)	15	31	35	30	20	33	41	36
Total [a]	100	100	100	100	100	100	100	100
	Sad-Negative				Sad-Positive			
	20-39 years old	40-59 years old	60+ years old	Total	20-39 years old	40-59 years old	60+ years old	Total
Not at all suitable (1)	10	26	16	21	0	34	25	28
Unsuitable (2)	50	42	51	46	50	39	42	41
Suitable (3)	30	28	27	28	25	27	29	28
Highly Suitable (4)	10	4	6	5	25	1	5	4
Total [a]	100	100	100	100	100	100	100	100
B. Within Combination	Happy-Negative				Happy-Positive			
	20-39 years old	40-59 years old	60+ years old	Total [a]	20-39 years old	40-59 years old	60+ years old	Total [a]
Not at all suitable (1)	13	67	20	100	17	63	20	100
Unsuitable (2)	3	51	47	100	2	46	52	100
Suitable (3)	7	47	47	100	6	49	46	100
Highly Suitable (4)	3	41	56	100	3	43	54	100
Total	5	48	48	100	5	48	48	100

	Sad-Negative				Sad-Positive			
	20-39 years old	40-59 years old	60+ years old	Total [a]	20-39 years old	40-59 years old	60+ years old	Total [a]
Not at all suitable (1)	2	60	37	100	0	58	42	100
Unsuitable (2)	5	43	52	100	6	45	48	100
Suitable (3)	5	48	46	100	5	26	50	100
Highly Suitable (4)	10	35	55	100	33	7	60	100
Total	5	48	48	100	5	48	48	100

C. Within Total	Happy-Negative				Happy-Positive			
	20-39 years old	40-59 years old	60+ years old	Total	20-39 years old	40-59 years old	60+ years old	Total
Not at all suitable (1)	1	5	2	8	1	5	2	8
Unsuitable (2)	1	13	12	26	0	6	7	14
Suitable (3)	3	17	17	37	3	21	20	43
Highly Suitable (4)	1	12	17	30	1	16	19	36
Total	5	48	48	100	5	48	48	100

	Sad-Negative				Sad-Positive			
	20-39 years old	40-59 years old	60+ years old	Total	20-39 years old	40-59 years old	60+ years old	Total
Not at all suitable (1)	1	13	8	21	0	16	12	28
Unsuitable (2)	3	20	24	46	3	19	20	41
Suitable (3)	2	14	13	28	1	13	14	28
Highly Suitable (4)	1	2	3	5	1	0	2	3
Total	5	48	48	100	5	48	48	100

Note. The analysis sample consisted of only native English-speaking respondents (n = 40) versus the entire study sample that included native and non-native English speakers (N = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

a May not sum to 100 due to rounding.

Table C.16
Suitability Ratings for Voice-Content Scores by Occupation Measures of Central Tendency: Analysis Sample

	Voice-content			
	Happy-Negative	Happy-Positive	Sad-Negative	Sad-Positive
Employed for wages	*M* = 2.01 *Mdn* = 3.00	*M* = 3.21 *Mdn* = 3.00	*M* = 2.05 *Mdn* = 2.00	*M* =1.96 *Mdn* = 2.00
Home-maker	*M* = 3.75 *Mdn* = 4.00	*M* = 3.90 *Mdn* = 4.00	*M* = 2.80 *Mdn* = 3.00	*M* = 2.80 *Mdn* = 3.00
Retired	*M* = 3.02 *Mdn* = 3.00	*M* = 3.12 *Mdn* = 3.00	*M* = 2.26 *Mdn* = 2.00	*M* = 2.21 *Mdn* = 2.00
Self-employed	*M* = 2.55 *Mdn* = 3.00	*M* = 2.63 *Mdn* = 3.00	*M* = 2.03 *Mdn* = 2.00	*M* = 1.83 *Mdn* = 2.00
Student	*M* = 2.45 *Mdn* = 2.00	*M* = 2.80 *Mdn* = 3.00	*M* = 2.60 *Mdn* = 3.00	*M* = 2.45 *Mdn* = 3.00
Unemployed	*M* = 3.00 *Mdn* = 3.00	*M* = 2.00 *Mdn* = 3.00	*M* = 2.10 *Mdn* = 2.00	*M* = 2.00 *Mdn* = 2.00
Total	*M* = 2.89 *Mdn* = 3.00	*M* = 3.07 *Mdn* = 3.00	*M* = 2.17 *Mdn* = 2.00	*M* = 2.08 *Mdn* = 2.00

Note. M = mean; *Mdn* = median. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

Table C.17
Distribution of Likert Suitability Ratings for Voice and Content Combinations by Occupation: Percentages in the Analysis Sample

A. Within occupation	Happy-Negative							Happy-Positive						
	Employed for wages	Self-employed	Unemployed	Home-maker	Retired	Student	Total	Employed for wages	Self-employed	Unemployed	Home-maker	Retired	Student	Total
Not at all suitable (1)	5	24	0	0	3	5	8	3	23	0	0	6	5	8
Unsuitable (2)	26	23	0	5	28	55	26	11	14	0	0	15	40	14
Suitable (3)	43	29	100	15	34	30	37	49	43	100	10	41	25	43
Highly Suitable (4)	26	25	0	80	35	10	30	37	21	0	90	38	30	36
Total [a]	100	100	100	100	100	100	100	100	100	100	100	100	100	100

	Sad-Negative							Sad-Positive						
	Employed for wages	Self-employed	Unemployed	Home-maker	Retired	Student	Total	Employed for wages	Self-employed	Unemployed	Home-maker	Retired	Student	Total
Not at all suitable (1)	24	29	0	0	20	0	21	27	43	0	0	27	20	28
Unsuitable (2)	53	46	90	20	39	40	46	54	39	100	20	28	15	41
Suitable (3)	17	19	10	80	36	60	28	14	13	0	80	43	65	28
Highly Suitable (4)	6	6	0	0	5	0	5	5	6	0	0	3	0	4
Total [a]	100	100	100	100	100	100	100	100	100	100	100	100	100	100

B. Within combination	Happy-Negative							Happy-Positive						
	Employed for wages	Self-employed	Unemployed	Home-maker	Retired	Student	Total [a]	Employed for wages	Self-employed	Unemployed	Home-maker	Retired	Student	Total [a]
Not at all suitable (1)	24	64	0	0	10	3	100	13	60	0	0	23	3	100
Unsuitable (2)	38	18	0	1	33	11	100	32	20	0	0	33	15	100
Suitable (3)	44	16	7	2	28	4	100	42	20	6	1	28	3	100
Highly Suitable (4)	33	17	0	13	35	2	100	39	12	0	13	32	4	100
Total	38	20	3	5	30	5	100	38	20	3	5	30	5	100
	Sad-Negative							Sad-Positive						
	Employed for wages	Self-employed	Unemployed	Home-maker	Retired	Student	Total [a]	Employed for wages	Self-employed	Unemployed	Home-maker	Retired	Student	Total [a]
Not at all suitable (1)	43	28	0	0	29	0	100	37	31	0	0	29	4	100
Unsuitable (2)	43	20	5	2	25	4	100	50	19	6	3	21	2	100
Suitable (3)	22	13	1	14	38	11	100	19	9	0	14	46	12	100
Highly Suitable (4)	45	25	0	0	30	0	100	47	33	0	0	20	0	100
Total	38	20	3	5	30	5	100	38	20	3	5	30	5	100

C. Within total	Happy-Negative							Happy-Positive						
	Employed for wages	Self-employed	Unemployed	Home-maker	Retired	Student	Total	Employed for wages	Self-employed	Unemployed	Home-maker	Retired	Student	Total
Not at all suitable (1)	2	5	0	0	1	0	8	1	5	0	0	2	0	8
Unsuitable (2)	10	5	0	0	9	3	27	4	3	0	0	5	2	14
Suitable (3)	16	6	3	1	10	2	37	18	9	3	1	12	1	43
Highly Suitable (4)	10	5	0	4	11	1	30	14	4	0	5	12	2	36
Total	38	20	3	5	30	5	100	38	20	3	5	30	5	100
	Sad-Negative							Sad-Positive						
	Employed for wages	Self-employed	Unemployed	Home-maker	Retired	Student	Total	Employed for wages	Self-employed	Unemployed	Home-maker	Retired	Student	Total
Not at all suitable (1)	9	6	0	0	6	0	21	10	9	0	0	8	1	28
Unsuitable (2)	20	9	2	1	12	2	46	20	8	3	1	9	1	41
Suitable (3)	6	4	0	4	11	3	28	5	3	0	4	13	3	28
Highly Suitable (4)	2	1	0	0	2	0	5	2	1	0	0	1	0	4
Total	38	20	3	5	30	5	100	38	20	3	5	30	5	100

Note. The analysis sample consisted of only native English-speaking respondents (n = 40) versus the entire study sample that included native and non-native English speakers (N = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

[a] May not sum to 100 due to rounding.

Table C.18
Suitability Ratings for Voice-Content Scores by Education Measures of Central Tendency: Analysis Sample

	Voice-content			
	Happy-Negative	Happy-Positive	Sad-Negative	Sad-Positive
College/ Univ. Graduate	*M* = 3.36 *Mdn* = 4.00	*M* = 3.39 *Mdn* = 4.00	*M* = 2.13 *Mdn* = 2.00	*M* =2.06 *Mdn* = 2.00
High School/ 2nd. School Equiv.	*M* = 2.40 *Mdn* = 2.00	*M* = 2.50 *Mdn* = 2.50	*M* = 2.60 *Mdn* = 3.00	*M* =2.60 *Mdn* = 3.00
Postgrad./ Prof. Degree	*M* = 2.68 *Mdn* = 3.00	*M* = 2.93 *Mdn* = 3.00	*M* = 2.21 *Mdn* = 2.00	*M* =2.08 *Mdn* = 2.00
Some College/Univ.	*M* = 3.15 *Mdn* = 4.00	*M* = 3.35 *Mdn* = 4.00	*M* =3.10 *Mdn* = 3.00	*M* = 3.05 *Mdn* = 3.00
Some Postgrad. Educ.	*M* = 2.52 *Mdn* = 3.00	*M* = 2.72 Mdn= 3.00	*M* = 1.86 *Mdn* = 2.00	*M* = 1.70 *Mdn* = 2.00
Vocat./Tech. school	*M* = 2.90 *Mdn* = 3.00	*M* = 3.30 *Mdn* = 3.00	*M* = 1.75 *Mdn* = 2.00	*M* = 1.80 *Mdn* = 2.00
Total	*M* = 2.89 *Mdn* = 3.00	*M* = 3.07 *Mdn* = 3.00	*M* = 2.17 *Mdn* = 2.00	*M* = 2.08 *Mdn* = 2.00

Note. M = mean; *Mdn* = median. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

Table C.19
Distribution of Likert Suitability Ratings for Voice and Content Combinations by Education: Percentages in the Analysis Sample

A. Within education	Happy-Negative							Happy-Positive						
	College/ Univ. Graduate	High School/ 2ndary School Equiv.	Postgrad./ Prof. Degree	Some College/ Univ.	Some Postgrad. Education	Vocat./ Tech. School	Total	College/ Univ. Graduate	High School/ 2ndary School Equiv.	Postgrad./ Prof. Degree	Some College/ Univ.	Some Postgrad. Education	Vocat./ Tech. School	Total
Not at all suitable (1)	0	0	0	3	26	7	8	0	0	5	4	24	7	8
Unsuitable (2)	60	35	40	10	18	34	26	50	15	15	5	8	18	14
Suitable (3)	40	40	5	34	34	43	37	50	40	20	38	40	50	43
Highly Suitable (4)	0	25	55	53	22	16	30	0	45	60	53	28	25	36
Total [a]	100	100	100	100	100	100	100	100	100	100	100	100	100	100

	Sad-Negative							Sad-Positive						
	College/ Univ. Graduate	High School/ 2ndary School Equiv.	Postgrad./ Prof. Degree	Some College/ Univ.	Some Postgrad. Education	Vocat./ Tech. School	Total	College/ Univ. Graduate	High School/ 2ndary School Equiv.	Postgrad./ Prof. Degree	Some College/ Univ.	Some Postgrad. Education	Vocat./ Tech. School	Total
Not at all suitable (1)	0	30	0	23	30	19	21	0	35	0	35	38	24	28
Unsuitable (2)	40	65	5	44	54	48	46	40	50	10	28	54	48	41
Suitable (3)	60	5	80	29	16	26	28	60	15	75	33	8	24	28
Highly Suitable (4)	0	0	15	3	0	7	5	0	0	15	4	0	4	4
Total [a]	100	100	100	100	100	100	100	100	100	100	100	100	100	100

B. Within combination	Happy-Negative							Happy-Positive						
	College/ Univ. Graduate	High School/ 2ndary School Equiv.	Postgrad/ Prof. Degree	Some College/ Univ.	Some Postgrad. Education	Vocat./ Tech. School	Total [a]	College/ Univ. Graduate	High School/ 2ndary School Equiv.	Postgrad ./ Prof. Degree	Some College/ Univ.	Some Postgrad. Education	Vocat./ Tech. School	Total [a]
Not at all suitable (1)	0	0	0	13	43	43	100	0	0	3	17	40	40	100
Unsuitable (2)	6	7	8	12	9	59	100	9	6	6	11	7	61	100
Suitable (3)	3	5	1	28	12	52	100	3	5	2	27	12	52	100
Highly Suitable (4)	0	4	9	53	9	24	100	0	6	8	44	10	32	100
Total [a]	3	5	5	30	13	45	100	3	5	5	30	13	45	100

	Sad-Negative							Sad-Positive						
	College/ Univ. Graduate	High School/ 2ndary School Equiv.	Postgrad./ Prof. Degree	Some College/ Univ.	Some Postgrad. Education	Vocat./ Tech. School	Total [a]	College/ Univ. Graduate	High School/ 2ndary School Equiv.	Postgrad ./ Prof. Degree	Some College/ Univ.	Some Postgrad. Education	Vocat./ Tech. School	Total [a]
Not at all suitable (1)	0	7	0	34	18	41	100	0	6	0	38	17	39	100
Unsuitable (2)	2	7	1	29	15	47	100	3	6	1	21	17	53	100
Suitable (3)	5	1	14	31	7	41	100	5	3	14	35	4	40	100
Highly Suitable (4)	0	0	15	20	0	65	100	0	0	20	33	0	47	100
Total [a]	3	5	5	30	13	45	100	3	5	5	30	13	45	100

C. Within total	Happy-Negative							Happy-Positive						
	College/ Univ. Graduate	High School/ 2ndary School Equiv.	Postgrad./ Prof. Degree	Some College/ Univ.	Some Postgrad. Education	Vocat./ Tech. School	Total	College/ Univ. Graduate	High School/ 2ndary School Equiv.	Postgrad ./ Prof. Degree	Some College/ Univ.	Some Postgrad. Education	Vocat./ Tech. School	Total
Not at all suitable (1)	0	0	0	1	43	3	8	0	0	0	1	3	3	8
Unsuitable (2)	2	2	2	3	9	15	26	1	1	1	2	1	8	14
Suitable (3)	1	2	0	10	12	19	37	1	2	1	12	5	23	43
Highly Suitable (4)	0	1	3	16	9	7	30	0	2	3	16	4	11	36
Total [a]	3	5	5	30	13	45	100	3	5	5	30	13	45	100

	Sad-Negative							Sad-Positive						
	College/ Univ. Graduate	High School / 2ndary School Equiv.	Postgrad./ Prof. Degree	Some College/ Univ.	Some Postgrad. Education	Vocat./ Tech. School	Total	College/ Univ. Graduate	High School/ 2ndary School Equiv.	Postgrad ./ Prof. Degree	Some College/ Univ.	Some Postgrad. Education	Vocat./ Tech. School	Total
Not at all suitable (1)	0	2	0	7	4	9	21	0	2	0	11	5	11	28
Unsuitable (2)	1	3	0	13	7	22	46	1	3	1	9	7	22	41
Suitable (3)	2	0	4	9	2	12	28	2	1	4	10	1	11	28
Highly Suitable (4)	0	0	1	1	0	3	5	0	0	1	1	0	2	4
Total [a]	3	5	5	30	13	45	100	3	5	5	30	13	45	100

Note. The analysis sample consisted of only native English-speaking respondents (n = 40) versus the entire study sample that included native and non-native English speakers (N = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

[a] May not sum to 100 due to rounding.

Table C.20
Suitability Ratings for Voice-Content Scores by Country-of-youth Measures of Central Tendency: Analysis Sample

	Voice-content			
	Happy-Negative	Happy-Positive	Sad-Negative	Sad-Positive
Australia	*M* = 2.95 *Mdn* = 3.00	*M* = 3.30 *Mdn* = 3.00	*M* = 2.70 *Mdn* = 3.00	*M* = 2.50 *Mdn* = 2.50
Canada	*M* = 2.30 *Mdn* = 2.00	*M* = 2.70 *Mdn* = 3.00	*M* = 2.00 *Mdn* = 2.00	*M* = 2.30 *Mdn* = 2.00
New Zealand	*M* = 2.10 *Mdn* = 2.00	*M* = 2.80 *Mdn* = 3.00	*M* = 2.30 *Mdn* = 2.00	*M* = 1.90 *Mdn* = 2.00
UK	*M* = 3.01 *Mdn* = 3.00	*M* = 3.26 *Mdn* = 3.00	*M* = 2.22 *Mdn* = 2.00	*M* = 2.00 *Mdn* = 2.00
USA	*M* = 2.90 *Mdn* = 3.00	*M* = 3.02 Mdn= 3.00	*M* = 2.12 *Mdn* = 2.00	*M* = 2.07 *Mdn* = 2.00
Total	*M* = 2.89 *Mdn* = 3.00	*M* = 3.07 *Mdn* = 3.00	*M* = 2.17 *Mdn* = 2.00	*M* = 2.08 *Mdn* = 2.00

Note. M = mean; *Mdn* = median. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

Table C.21
Distribution of Likert Suitability Ratings for Voice and Content Combinations by Education: Percentages in the Analysis Sample

A. Within location	Happy-Negative					Happy-Positive				
	AUS-NZ	Canada	UK	USA	Total	AUS-NZ	Canada	UK	USA	Total
Not at all suitable (1)	10	0	2	9	8	3	0	1	10	8
Unsuitable (2)	33	70	21	25	26	10	30	10	14	14
Suitable (3)	37	30	50	33	37	57	70	51	38	43
Highly Suitable (4)	20	0	27	33	30	30	0	38	37	36
Total [a]	100	100	100	100	100	100	100	100	100	100
	Sad-Negative					Sad-Positive				
	AUS-NZ	Canada	UK	USA	Total	AUS-NZ	Canada	UK	USA	Total
Not at all suitable (1)	0	0	9	28	21	7	0	18	34	28
Unsuitable (2)	50	100	61	39	46	57	70	64	30	41
Suitable (3)	43	0	29	27	28	37	30	18	30	28
Highly Suitable (4)	7	0	1	6	5	0	0	0	6	4
Total [a]	100	100	100	100	100	100	100	100	100	100
B. Within combination	Happy-Negative					Happy-Positive				
	AUS-NZ	Canada	UK	USA	Total [a]	AUS-NZ	Canada	UK	USA	Total [a]
Not at all suitable (1)	10	0	7	83	100	3	0	3	93	100
Unsuitable (2)	10	7	18	65	100	6	6	17	72	100
Suitable (3)	7	2	30	60	100	10	4	27	60	100
Highly Suitable (4)	5	0	20	75	100	6	0	24	70	100
Total [a]	7	3	23	68	100	7	3	23	68	100

	Sad-Negative					Sad-Positive				
	AUS-NZ	Canada	UK	USA	Total [a]	AUS-NZ	Canada	UK	USA	Total [a]
Not at all suitable (1)	0	0	10	90	100	2	0	14	84	100
Unsuitable (2)	8	5	30	57	100	10	4	36	50	100
Suitable (3)	12	0	23	65	100	10	3	14	73	100
Highly Suitable (4)	10	0	5	85	100	0	0	0	100	100
Total [a]	7	3	23	68	100	7	3	23	68	100
C. Within Total	Happy-Negative					Happy-Positive				
	AUS-NZ	Canada	UK	USA	Total	AUS-NZ	Canada	UK	USA	Total
Not at all suitable (1)	1	0	1	6	8	0	0	0	7	8
Unsuitable (2)	3	2	5	17	26	1	1	2	10	14
Suitable (3)	3	1	11	22	37	4	2	12	26	43
Highly Suitable (4)	2	0	6	22	30	2	0	9	25	36
Total [a]	7	3	23	68	100	7	3	23	68	100
	Sad-Negative					Sad-Positive				
	AUS-NZ	Canada	UK	USA	Total	AUS-NZ	Canada	UK	USA	Total
Not at all suitable (1)	0	0	2	19	21	1	0	4	23	28
Unsuitable (2)	4	3	14	26	48	4	2	15	20	41
Suitable (3)	3	0	7	18	28	3	1	4	20	28
Highly Suitable (4)	1	0	0	4	5	0	0	0	4	4
Total [a]	7	3	23	68	100	7	3	23	68	100

Notes. The analysis sample consisted of only native English-speaking respondents (n = 40) versus the entire study sample that included native and non-native English speakers (N = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable). For purposes of this analysis, data for participants from Australia and New Zealand were combined.

[a] May not sum to 100 due to rounding.

Table C.22
Marginal Means of Suitability Ratings for Voice-Content Scores, by Voice Emotion and Age Groups

	Voice Emotion		
	Happy	Sad	Marginal Means (Age group)
20-39 years old	2.71	2.44	2.57
40-59 years old	2.90	1.98	2.44
60+ years old	3.14	2.26	2.70
Marginal Means (Voice Emotion)	2.96	2.18	2.57

Note. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

Table C.23
Marginal Means of Suitability Ratings for Voice-Content Scores, by Message Content and Age Groups

	Message Content		
	Negative	Positive	Marginal Means (Age group)
20-39 years old	2.45	2.69	2.57
40-59 years old	2.43	2.46	2.44
60+ years old	2.68	2.72	2.70
Marginal Means (Message Content)	2.53	2.61	2.57

Note. The analysis sample consisted of only native English-speaking respondents (n = 40) versus the entire study sample that included native and non-native English speakers (N = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

Table C.24
Marginal Means of Suitability Ratings for Voice-Content Scores, by Voice Emotion and Gender

	Voice Emotion Happy	Sad	Marginal Means (Gender)
Female	3.05	1.94	2.49
Male	2.90	2.35	2.62
Marginal Means (Voice Emotion)	2.96	2.18	2.57

Note. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

Table C.25
Marginal Means of Suitability Ratings for Voice-Content Scores, by Message Content and Gender

	Message Content Negative	Positive	Marginal Means (Gender)
Female	2.49	2.49	2.49
Male	2.56	2.69	2.62
Marginal Means (Message Content)	2.53	2.61	2.57

Note. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

Table C.26
Marginal Means of Suitability Ratings for Voice-Content Scores, by Voice Emotion, Gender, and Age Group

	Voice Emotion		
	Happy	Sad	Marginal Means (Gender)
Female			
20-39 years old	—	—	—
40-59 years old	3.10	1.75	2.43
60+ years old	3.00	2.12	2.56
Male			
20-39 years old	2.71	2.44	2.57
40-59 years old	2.70	2.21	2.45
60+ years old	3.29	2.40	2.84
Marginal Means (Voice Emotion)	2.96	2.18	2.57

Note. Dashed cells (—) indicate that no respondents self-reported along these column/row dimensions. The analysis sample consisted of only native English-speaking respondents (n = 40) versus the entire study sample that included native and non-native English speakers (N = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

Table C.27

Marginal Means of Suitability Ratings for Voice-Content Scores, by Message Content, Gender, and Age Group

	Message Content		
	Negative	Positive	Marginal Means (Gender)
Female			
20-39 years old	—	—	—
40-59 years old	2.43	2.43	2.43
60+ years old	2.56	2.56	2.56
Male			
20-39 years old	2.45	2.69	2.57
40-59 years old	2.42	2.48	2.45
60+ years old	2.81	2.88	2.84
Marginal Means (Message Content)	2.53	2.61	2.57

Note. Dashed cells (—) indicate that no respondents self-reported along these column/row dimensions. The analysis sample consisted of only native English-speaking respondents (*n* = 40) versus the entire study sample that included native and non-native English speakers (*N* = 50). Suitability rating were measured with subject self-report on a 4-point Likert scale (1= Not at all suitable, 2 = Unsuitable, 3 = Suitable, 4 = Highly suitable).

References

Arnold, M. B. (1960). *Emotion and personality*. New York: Columbia University Press.

Averill, J. R. (1980). A constructivist view of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience: Volume 1. Theories of emotion* (pp. 305-339). New York: Academic Press.

Banse, R. & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, (70)*3, 614-636.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*, 155-162.

Birren, J. E. & Schaie, K. W. (Eds.). (2001). *Handbook of the psychology of aging*. San Diego: Academic Press.

Black, A. W. (2003). Unit selection and emotional speech. Proceedings of the 8th Conference on Speech Communication and Technology (Eurospeech 2003), Geneva, Switzerland.

Black, A. W., Taylor, P., & Macon, M. (2000). *Speech synthesis in Festival: A practical course on making computers talk* (Ed.1.4.1 for Festival Version 2.0. 8 May 2000). Retrieved May 25, 2007, from Carnegie Mellon University Web site: http://festvox.org/festtut/notes/festtut_toc.html

Blizzard challenge 2007. Retrieved May 1, 2007, from http://festvox.org/blizzard/

Bradburn, N. (1977). The measurement of psychological well-being. In J. Ellinson (Chair), *Health Goals and Health Indicators.* Presented at the meeting of the American Association for the Advancement of Science, Denver, Colorado.

Buchanan, T., Johnson, J. A., & Goldberg, L. R. (2005). Implementing a five-factor personality inventory for use on the Internet. *European Journal of Psychological Assessment, (21)*2, 115–127.

Cabral, J. P. (2006). *Transforming Prosody and Voice Quality to Generate Emotions in Speech*. Unpublished master's thesis. Technical University of Lisbon, Lisbon, Portugal.

Cattell, R. B. (1973). *Personality and mood by questionnaire.* San Francisco, CA: Jossey-Bass.

Costa, P. T., Jr. & McCrae, R. R. (1990). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology, (38)*4, 668-678.

Czaja, S. & Lee, C. C. (2007). The impact of aging on access to technology. *Universal Access in the Information Society (5)*4, 341-349.

Czaja, S. (2001). Technological change and the older worker. In J. E. Birren and K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 547-565). San Diego: Academic Press.

Ekman, P., Friesen, W., & Ellsworth, P. (1982). *Emotions in the human face*. New York: Cambridge University Press.

De Rivera, J. (1977). A structural theory of the emotions. *Psychological Issues, (10)*40, 9-169.

Dillman, D. A. (2007). *Mail and internet surveys: The tailored design method* (2nd ed.). Hoboke, NJ: John Wiley & Sons, Inc.

Eysenck, H. J., & Eysenck. S. B. G. (1991). *The Eysenck Personality Questionnaire-Revised.* Sevenoaks: Hodder and Stoughton.

Eysenck, M. W. (1987). Trait theories of anxiety. In J. Strelau & H. J. Eysenck (Eds.), *Personality dimensions and arousal* (pp. 79-97). New York: Plenum Press.

Festival voice system: http://festvox.org/ accessed June 1, 2007.

Frijda, N. H. (1986). *The emotions*. Cambridge, England: Cambridge University Press.

Gill, A. J., Harrison, A. J., & Oberlander, J. (forthcoming). Interpersonality: Individual differences and interpersonal priming.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist (48)*1, 26-34.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84–96.

Gray, J. A. (1981). A critique of Eysenck's theory of personality. In H. J. Eysenck (Ed.), *A model for personality*. New York: Springer.

Graziano, W. G., Jensen-Campbell, L. A., & Hair, E. C. (1996). Perceiving interpersonal conflict and reacting to it: The case for agreeableness. *Journal of Personality and Social Psychology, (70)*4, 820-835.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. K. Bower (Ed.), *The psychology of learning and motiviation* (Vol. 22, 193-225). San Diego: Academic Press.

Hawthorn, D. (2000). Possible implications of aging for interface designers. *Interacting with Computers, 12*, 507-528.

Hofer, G. (2004). *Emotional Speech Synthesis*. Unpublished master's thesis, University of Edinburgh, Edinburgh, Scotland.

Horowitz, L. M. (1979). On the cognitive structure of interpersonal problems treated in psychotherapy. *Journal of Consulting and Clinical Psychology, 47*, 5-15.

International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences. Accessed June 1, 2007: http://ipip.ori.org/

Isbister, K. & Nass, C. (2000). Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies, 53*, 251-267.

Isen, A. M. (2000). Positive affect and decision making. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed.), 417-435. New York: Guilford Press.

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

Jones, C., Berry, L., & Stevens, C. (2007). Synthesized speech intelligibility and persuasion: Speech rate and non-native listeners. *Computer Speech and Language, 21*, 641–651.

Jonsson, I.-M., Nass, C., Harris, H., & Takayama, L. (2005). Matching in-car voice with driver state: Impact on attitude and driving performance. *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Rockport, Maine, 173-180.

Kellerman, H. (1980). A structural model of emotion and personality: Psychoanalytic and sociobiological implications. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience: Volume 1. Theories of emotion*. New York: Academic Press.

Kline, P. (2000). *A psychometrics primer*. London: Free Association Books.

Larsen, R. J. & Ketelaar, T. (1991). Personality and susceptibility to positive and negative emotional states. *Journal of Personality and Social Psychology,* (*61)*1, 132-140.

Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Lee, K. M & Nass, C. (2003). Designing Social Presence of Social Actors in Human Computer Interaction. *CHI 2003 New Horizons* (5)1, 289-296.

Matthews, G., Deary, I. J., & Whiteman, M. C. (2003). *Personality traits* (2nd ed.). Cambridge: Cambridge University Press.

McCrae, R. R., & Costa, P. T., Jr. (1992). Discriminant validity of NEO P-IR facet scales. *Educational and Psychological Measurement*, *52*, 229-237.

McCrae, R. R. & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*, 175-215.

Moon, Y. (2002). Personalization and personality: some effects of customizing message style based on consumer personality. *Journal of Consumer Psychology, 12*(4), 313-326.

Moon, Y. & Nass, C. (1996). How 'real" are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research*, *23*, 651-674.

Murray, J. B. (1990). Review of research on the Myers-Briggs type indicator. *Perceptual and Motor Skills*, *70*, 1187-1202.

Nass, C., & Brave, S. (2005). Wired for speech: How voice activates and advances the human-computer relationship. Cambridge: The MIT Press.

Nass, C. & Gong, L. (2000). Speech interfaces from an evolutionary perspective. *Communications of the ACM, (43)*9, 36-43.

Nass, C. & Lee, K. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology, Applied*, *7*(3), 171-181.

Nass, C., Foehr, U., Brave, S., and Somoza, M. (2001). The effects of emotion of voice in synthesized and recorded speech. *Proceedings of Emotional and Intelligent II: The Tangled Knot of Social Cognition,* North Falmouth, Massachusetts, November 2-4.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, *43*, 223-239.

Oatley, K. & Jenkins, J. (1996). *Understanding emotions.* Oxford : Blackwell.

Ortony, A., Clore, G., & Collins, A. (1988). *The Cognitive structure of emotions*. Cambridge: Cambridge University Press.

Plutchik, R. (1980). Emotions: A general psychoevolutionary theory. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion*, (pp. 3-33). London: Lawrence Erlbaum Associates.

Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. In P. Shaver (Ed.), *Review of personality and social psychology: Vol. 5. Emotions, Relationships, and Health* (pp. 11-36). Beverly Hills, CA: Sage.

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion.* London: Lawrence Erlbaum Associates.

Smither, J. A. (1993). Short term memory demands in processing synthetic speech by young and old adults. *Behavior and Information Technology, 12,* 330–335.

Schroeder, M. (2004). Speech and Emotion Research: An Overview of Frameworks and a Dimensional Approach to Speech Synthesis. Unpublished doctoral dissertation, Saarland University, Saarbrücken, Germany.

SPSS (2006). Statistical Package for the Social Sciences (SPSS) (Version 15.0 for Windows, Release 15.0.1) [Computer software]. Chicago: SPSS, Inc.

Strelau (1987). Emotion as a key concept in temperament research. *Journal of Research in Personality, 21,* 510-528.

Tomkins, S. S. (1962). Affect, imagery and consciousness: Vol. 1. The positive affects. New York: Springer.

Wiggins, J. S. (1979). Psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology, (37)*3, 395-412.

Wilson, W. (1967). Correlates of avowed happiness. *Psychological Bulletin, 67,* 294-306.