# THE AUDIO IMPLICIT ASSOCIATION TEST:

# HUMAN PREFERENCES AND IMPLICIT ASSOCIATIONS CONCERNING MACHINE VOICES

Wade Joseph Mitchell

Submitted to the faculty of the School of Informatics
in partial fulfillment of the requirements
for the degree of
Master of Science in Human-Computer Interaction,
Indiana University

May 2009

Accepted by the Faculty of Indiana University,
in partial fulfillment of the requirements for the degree of Master of Science
in Human-Computer Interaction

**Master's Thesis**
**Committee**

Karl F. MacDorman, Ph.D., Chair

Leslie Ashburn-Nardo, Ph.D.

Alexander Y. Fedorikhin, Ph.D.

Dedicated to my family and friends, whose faith in me far exceeds my ability.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

I thank the members of my committee: My advisor Dr. Karl MacDorman for his never-ending devotion to his students, for his wealth of knowledge, and for his invaluable persistence, Dr. Leslie Ashburn-Nardo for lending her expertise in the Implicit Association Test and psychology, and Dr. Alexander Fedorikhin for his assistance in choosing the psychological indices. Their assistance in more ways than I could list has been invaluable. I would also like to thank the other members of Dr. MacDorman's research group: Chin-Chang Ho for his assistance in analyzing the vast amount of data collected into interpretable results, Sandosh Vasudevan and Himalaya Patel for their assistance with creating the Implicit Association Test website, and all who have been involved in both criticizing and improving this research. I am proud to call them all associates and friends. Finally, I wish to thank those who provided the voices for the stimuli needed to conduct this study.

ABSTRACT

Wade Joseph Mitchell

# THE AUDIO IMPLICIT ASSOCIATION TEST: HUMAN PREFERENCES AND IMPLICIT ASSOCIATIONS CONCERNING MACHINE VOICES

Auditory human-machine interfaces are becoming ubiquitous. Interactive voice response systems, navigation systems, socially assistive robots, and smart houses are just a few examples of technologies that support auditory interactions. This study uses the implicit association test (IAT) to measure participants' associative strength between human and machine voices and pleasant or unpleasant attributes. To accomplish this, the IAT needed to be validated using audio stimuli and the associative strength of secondary features of stimuli, that is, features other than their semantic content. Six IAT experiments were conducted to test the ability of the IAT to measure association strengths of the target concepts of audio stimuli and an attribute dimension in addition to target concepts of secondary features and an attribute dimension. Results support the effectiveness of an audio IAT, an IAT for secondary features, and an IAT that combines audio with secondary features. Results also show that participants had a stronger association between human voices and pleasant attributes than machine voices and pleasant attributes.

# CHAPTER ONE: INTRODUCTION

## 1.1 Interaction Design for Audio Interfaces

Human-machine interaction is an important field of study. Better interface design allows more users to interact with machines and for users to receive more benefit from these interactions. Creating better interaction designs does not always mean better visual interfaces or better hands-on input devices. Auditory human-machine interaction provides an interface that can be used by a wide range of users in a variety of contexts.

Considering the overlap of the fields of psychology and human-machine interaction (Card et al., 1984), it is not surprising that several tools and evaluative methods from psychology have been adapted to research in human-machine interaction. This study focuses on a tool commonly used in social psychology, the implicit association test (IAT), and employs it to analyze human-machine interaction.

## 1.2 The Importance of Using the Implicit Association Test to Investigate Audio Interfaces

Audio interfaces are growing in ubiquity. Many designers are incorporating them in situations where hands-free interactions are desirable, such as driving. Improvements in audio interactions will spur the continued growth of this modality.

The IAT provides an implicit measure of positive and negative associations between a human-human interaction and a human-machine interaction. Comparing these interactions could provide principles for designers to use in creating technologies that more closely approximate successful human-human interactions, increasing their usability. Although the IAT is not a new methodology, it is new to the study of human-machine interaction, and this study helps establish the value of the IAT as a methodology for investigating human-machine interaction.

The IAT reduces self-presentational bias, which has been shown to be an issue in human-human interaction. If a strong self-presentational bias should appear in questionnaires that benchmark attitudes toward interaction with machines against attitudes toward interaction with people, this could severely bias their results. It could give the impression that participants strongly prefer interacting with people, when in fact they only feel more hesitant about revealing their unfavorable attitudes about people than revealing their unfavorable attitudes about machines. The comparison of explicit and implicit measures constitutes method triangulation, which can be used to appraise the validity of both explicit and implicit measures. Although the IAT has its own limitations, studies that combine explicit and implicit measures provide a more comprehensive analysis of the interaction.

# CHAPTER TWO: LITERATURE REVIEW

## 2.1 Auditory Human-Machine Interaction

It is hard to avoid auditory interaction with machines. Interactive voice response (IVR) systems are becoming ubiquitous when booking a flight, contacting technical support, checking your bank balance, or performing a myriad of other tasks by phone. In-car navigation systems and interactive systems for people with visual impairments and other disabilities have adopted auditory interactions. Auditory interaction involves a sequence of voice input from a user and audio output from a system. Voice commands and auditory feedback are making their way into every facet of human-machine interaction by providing an additional input and output modality for human-machine interactions, either as a component of a multi-modal system or when other input and output modalities are occupied, unavailable, or not usable (e.g., for users with visual or motor disabilities; Cohen & Oviatt, 1993; Kamm, 1995; Schafer, 1995). The growth of mobile technology, such as smart-phones and personal digital assistants (PDAs), presents issues with small screen space and limited keyboards interfaces, which can be partially alleviated by auditory interfaces. Atkinson, Mayer, and Merrill (2005) showed that in learning exercises participants tried harder and retained more knowledge if they liked the voice of the machine in the interaction. This result is consistent with social agency theory, which posits that the use of verbal and visual social cues by a machine can improve human-machine interaction by approximating human-human social interaction (Mayer et al., 2003; Moreno et al., 2001).

Developing technologies, such as interactive robots and smart houses, certainly benefit from or require auditory interactions (Kawamura et al., 1995). Increasingly complex home technology can be very useful but can also be difficult for nondisabled users to master. Situations where technology could be used to assist the disabled or elderly could be simplified by an audio interface. An auditory interface in a smart house does not need to tether a user to a keyboard, display, or other input or output device (Martin, 1976; Yerrapragada & Fisher, 1993; Ghorbel et al., 2004). The European Union-funded project INSPIRE[1] conducted research in the development of auditory interface systems that not only studied how to control technology systems with voice commands but also how the machine voice should respond (Moller et al., 2004).

Research into auditory human-machine interaction has progressed since the advent of computers. This research has taken several paths relevant to the development of better interaction. A few examples of active research directions include creating more human-sounding synthetic voices, improving prosody, developing empathetic responses, improving natural dialogue systems, reducing comprehension errors, and integrating auditory interaction with other modalities in an interactive system.[2] Research has been performed to study the effects of gender, presumed age, voice quality, and "naturalness" on attitudes toward machine voices. Crabtree, Mirenda, and Beukelman (1990) studied attitudes toward 12 recorded voices, 8 machine and 4 human, in six communicative contexts. Results show that participants preferred the gender and age of a voice to match the associated face (e.g., preferring a female child to have a younger female voice). Couper, Singer, and Tourangeau (2004) studied the responses of 1,396 participants to requests for sensitive information by IVR systems. The study found no significant difference in completion rate, number of answers, or quality of answer whether the

[1] INfotainment management with SPeech Interaction via REmote microphones and telephone interfaces

[2] Some research has even approached vocal interactions through teaching human users a common lexicon rather than developing natural-dialogue interactions (Shriver et al., 2001).

voice was a recording of a male or female human voice, a humanlike text-to-speech (TTS) system, or a machinelike TTS system.

Approximating human-human interactions may be a benefit in some ways, but it is also a hindrance in others. Creating human-machine interactions that more closely approximate human-human interactions can induce social desirability effects that bias surveys on sexual health, drug use, or other socially sensitive topics. Research has indicated that when asked questions on these topics, participants were more likely to respond and to respond fully when the machine voice is clearly not human. In these cases the greater the lack of humanness, the less social desirability and thus less moderation by social desirability, which was critical to gathering accurate information (Tourangeau et al., 2003).

These studies have certainly contributed to our understanding of voice variation and human attitudes, but they rely on self-reports and sometimes behavioral outcomes to determine attitudes. Self-reports are susceptible to a number of biases, including self-presentational biases or the flaccid repetition of a pat explanation (Greenwald et al., 1998). Thus, it is useful to complement explicit measures with implicit measures and other sources of information. As auditory human-machine interaction grows, it is important to understand how the voice of the machine affects the user and, in turn, the interaction. Linguistic research has indicated that suprasegmental, prosodic cues in voice (secondary features such as tone, rhythm, and quality across segments of speech) provide much information to the listener such as language identification, emotions of the speaker, and importance of the subject and can affect the listener's associations of the stimulus with various attributes (Ramus & Mehler, 1999).

## 2.2 The Implicit Association Test

Investigating human relations with various aspects of technology can be difficult. In this research it is important to understand human attitudes toward machine voices. However, attitudes are not easy data to gather. Simply asking participants to report their thoughts and attitudes about something can lead to the results that are difficult to replicate (Hewstone et al., 2002). A limitation in using self-reports as a measure of association arises from the influence of implicit cognition. Past experiences, memories, and learned behaviors can affect our associations and behavior even if we cannot explicitly explain how or why (Graf & Schacter, 1985; Greenwald, 1990; Jacoby & Dallas, 1981; Jacoby, Lindsay, & Toth, 1992; Jacoby & Witherspoon, 1982; Kihlstrom, 1990; Roediger, Weldon, & Challis, 1989; Schacter, 1987). Implicit cognition cannot be measured using direct self-reporting, which raises the question of how implicit attitudes can be measured (Greenwald & Banaji, 1995).

Various attempts have been made to measure implicit associations. Gaertner and McLaughlin (1983) used the latency of participants' responses to measure the strength of associations between race and an attribute dimension. Participants were presented with pairs of letter strings, requesting a *yes* judgment if both strings were valid words and *no* judgment otherwise. In reality the "strings of letters" were often words representing one of the two target races. This type of misdirection of the actual experimental target helps to remove explicit cognition concerning the target during the experiment. The study calculated the latency of participants' yes responses to measure strength of existing associations between the two words in a pair. They found that white participants responded reliably faster to white-positive word pairs than to black-positive pairs (e.g., *white-smart* vs. *black-smart).* This difference did not emerge on judgments of negative traits (e.g., *white-lazy* vs. *black-lazy).* The participants were also given explicit questionnaires measuring racial prejudice. Association strength was similar for participants who scored high and for participants who scored low on these self-report measures of race prejudice, showing that self-reporting measures did not provide accurate predictions of association strengths for such socially sensitive topics.

Early experiments in measuring implicit associations showed that timing responses can be a valuable measure of association strength (Dovidio et al., 1986; Gaertner & Dovidio, 1986; Devine, 1989; Gilbert and Hixon, 1991; Jamieson & Zanna, 1989; Klinger & Beall, 1992; Paulhus, Martin, & Murphy, 1992). When measuring the strength of association of a target with an attribute dimension (e.g., pleasant or unpleasant), researchers can get a clearer view of a participant's implicit associations. Greenwald et al. (1998) used latency as a measure of association strength in the development of the implicit association test (IAT). The IAT consists of five blocks for participants to complete where latency is recorded in associating a pair of target concepts with an attribute dimension. This is typically performed by having the participant press the *E* key on a keyboard if the stimulus is associated with the concept or attribute on the left side of the screen and the *I* key if the stimulus is associated with the concept or attribute on the right side of the screen. The latency should be shorter if a more closely associated concept–attribute pair shares the same response key than if a less closely associated concept–attribute pair shares the same response key. For example, an IAT measuring associations with race may have two racial target concepts (e.g., photographs of *black* faces versus *white* faces) and an attribute dimension (e.g., *pleasant* versus *unpleasant*). Block 1 requires participants to discriminate between two targets by categorizing them into concept groups (selecting black if presented with an image of a black male versus selecting white if presented with an image of a white male). Block 2 replaces the concepts in the left and right of the screen with attributes (pleasant or unpleasant). Participants are then presented with written words that are usually associated with one of the attributes (e.g., *hate* or *love*). These two blocks act as practice for the participant.

Block 3 combines both concept and attribute. In the *black/white* race IAT the concept *black* and the attribute *pleasant* might be on the left (*E* key) while the concept *white* and the attribute *unpleasant* might be on the right (*I* key). A participant will then be shown the combined stimuli from blocks 1 and 2 and has to decide whether the stimulus is *black/pleasant* or *white/unpleasant*. The concepts from block 1 are then reversed in block 4. If *black* was on the left (*E* key) then it is now on the right (*I* key). This block acts as a practice for reversed concepts. Block 5 then repeats block 3 with the exception that the concept–attribute pairings are now reversed. If in block 3 *black* was paired with *pleasant* then in block 5 *black* is paired with *unpleasant*.

Following the IAT participants are given a questionnaire, so that an index derived from self-reported questions could be compared to the implicit results (Greenwald et al., 1998). Correlation between the IAT *D* measure and the explicit index were used to test the validity of the IAT. If the targets of the IAT were not socially sensitive (e.g., flowers vs. insects), the correlation between the implicit results and explicit results should be strong, because there is no reason for participants to mediate their explicit responses. Further research showed that in IATs that measured associations with respect to socially sensitive topics, such as race and gender, the IAT was more resistant to self-reporting bias than explicit measures (Greenwald et al., 1998; Karpinski & Hilton, 2001). Since its development the IAT has been used in experiments to test associative strength on topics ranging from the mundane to the extremely socially sensitive. In some IAT experiments the target was chosen for expected agreement with explicit reports.

These experiments included presidential candidates and the fear of spiders. Progressively, IATs have been used to measure associative strengths with sensitive topics including gender attitudes, religion, homosexual stereotypes, obesity stereotypes, and attitudes toward various races, including black versus white and American versus Arab (Nosek et al., 2006). The importance of the American versus Arab IAT will be further discussed later in this study. Over time the IAT has shown strong evidence that it can measure associative strength in situations where explicit data may be unreliable (Nosek et al., 2006; Lane et al., 2007).

News items about the IAT that drew participants to an IAT website provided several large data sets for comparison of various scoring algorithms in an attempt to find the most valid. The *Bush/Gore* IAT demonstrated on Harvard University's Project Implicit website,[3] provided data from 8,218 participants used to test six scoring algorithms (Greenwald, Nosek, & Banaji, 2003). The *D* measure had the best correlation with explicit results and was selected for scoring the IAT. The results showed that removing the data from the practice trials of blocks 3 and 5 was counterproductive; the practice trials in fact provided better data than the test trials. The results indicated that the best among the evaluated methods included error latencies, used the data from the practice trials and test trials, and used the *D* scoring algorithm.

It is important to note that attitude and association strength are not necessarily the same thing. Karpinski and Hilton (2001) showed that the IAT is susceptible to *environmental associations* reflecting culturally shared but not necessarily individually accepted information. In the experiment they measured associations between apples and candy bars and found no correlation between the explicit and implicit results. Olson and Fazio (2004) attempted to minimize this extrapersonal information by creating a personal IAT. Heavy drinkers would be assumed to have strong positive associations with alcohol, but Houben and Wiers (2007) found

[3] implicit.harvard.edu

this was not the case. Believing these results to be caused by extrapersonal influences, they applied the personal IAT and discovered positive associations with alcohol.

### 2.3 The Audio IAT

The methodology of the IAT has been well investigated, but it has always used visual stimuli. Greenwald et al. (1998) used text for the target stimuli when creating the IAT. Later research has used either text or images (Houben & Wiers, 2007; Lane et al., 2007; Ashburn-Nardo et al., 2003; MacDorman et al., 2008). Nosek, Greenwald, and Banaji (2007) stated that the stimuli for an IAT could be text, images, sound, or a combination of modalities; however, experiments using audio stimuli have not yet been attempted.

An Audio IAT may produce more accurate results than a visual IAT because audio stimuli are processed faster than visual stimuli. Research has confirmed that simple reaction time to audio stimuli is faster than reaction to visual stimuli, with mean auditory reaction times being 140-160 ms and visual reaction times being 180-200 ms (Brebner and Welford, 1980; Welford, 1980; Woodworth and Schlosberg, 1954). This may be because an auditory stimulus takes 8-10 ms to reach the brain (Kemp *et al.*, 1973), but a visual stimulus takes 20-40 ms (Marshall *et al.*, 1943).

The IAT is a form of choice reaction time (CRT) measure. In choice reaction time experiments, the user must give a response that corresponds to the stimulus, such as pressing a key corresponding to a letter if the letter appears on the screen (Luce, 1986). Hick's Law states that in choice reaction time experiments, response is proportional to log($N$), where $N$ is the number of different possible stimuli (Hick, 1952). Thus, due to the cognitive processing required to make the choice, the latency is greatest in CRT. However, CRT with audio stimuli is still consistently lower than the CRT for visual stimuli.

This research study uses audio stimuli, because it explores audio interaction and, in particular, examines participants' association strength between voice stimuli and an attribute dimension. To do so required creating an IAT whose categorization task is not applied to the semantic content of the stimuli but to a secondary feature, such as an accent. To provide evidence that the IAT can be used to test target concepts that are the secondary features of the target stimuli, six IATs were developed and tested. In the *male voice–pleasant* IAT, the association in question was not the word being spoken but the gender of the speaker. In the *US-accented speech–pleasant* IAT, the association in question was not the words spoken or the gender of the speaker but the speaker's accent.

A traditional-style IAT was used as a baseline. The written common *US names–pleasant* IAT was conducted and the results used as the baseline for the following IATs. To validate the IAT using secondary features of the target stimuli, this study used the *Western font–pleasant* IAT. In this experiment 16 neutral phrases were used to represent the targets. The test of the phrases was presented to participants in four distinct typefaces. Representing Western typefaces were Harrington TTF and Old English TTF. Representing Arabic typefaces were Arab Dances TTF and Arabian Onenightstand TTF. The results section shows evidence that the IAT was effectively used to test associations with secondary features of the targets in the *US names–pleasant* and in the *Western fonts–pleasant* IATs.

To verify that the IAT could be used with audio stimuli, this study used the spoken *US names–pleasant* IAT. In this experiment the targets were common United States male first names and common Arabic Muslim male first names spoken by two female native English speakers who have US accents.  This experiment was designed to test whether results would be similar to the more typical written *US names–pleasant* IAT.  The results show similar *D* scores in both the audio and visual versions of this IAT.

Once the experiments showed evidence that the IAT could be used to test association strength with secondary target features (the visual IATs) and with audio stimuli (the spoken *US names–pleasant* IAT), the two variables were combined.  The *US-accented speech–pleasant* IAT used the same 16 neutral phrases as the *Western fonts–pleasant* IAT, but the phrases were spoken by four distinct male voices:  Two were US-accented native English speakers, and the other two voices were Arabic-accented English-as-a-second-language speakers.  Results demonstrated that the IAT could be used to test target concepts that were secondary features of both audio and visual stimuli.

The *female voice–pleasant* IAT used the same 16 neutral phrases for the target stimuli as the *Western font–pleasant* and the *US-accented speech–pleasant* IATs, but eight distinct voices were used to generate the stimuli in this experiment:  two male US-accented native English speakers, two female US-accented native English speakers, two male US-accented machine voices, and two female US-accented machine voices.  Previous experiments concerning gender have shown participants prefer women to men (Greenwald et al., 2000).  The present study showed similar results even though the gender was a secondary feature of the target stimuli. Using the eight audio target stimuli from the *female voice–pleasant* IAT, the *human voice–pleasant* IAT measured participants' associative strength with human voices versus machine voices.

## 2.4 In-group Bias

Intergroup relations are a subdiscipline of social psychology that concerns any thought, action, or feeling taken because of group membership (Sherif, 1966). In-group bias is the tendency to evaluate one's own membership group (the in-group) or its members more favorably than a nonmembership group (the out-group) or its members. Bias can affect both behavior and attitude. The idea of in-group bias has been well researched in social psychology (Mackie & Smith, 1998; Wilder & Simon, 2001). As humans we may identify with many in-groups including gender, age range, nationality, race, and species. Explicit methods have been used to measure in-group bias for decades, but the results of different explicit indices using methods incorporating cognitive, affective, and behavioral components often exhibit only a weak correlation. Implicit measures (which include the IAT) estimate automatic biases activated by the presence of the target stimuli tapping biases that the participant may be unaware of (Hewstone et al., 2002). In-group bias predicts that in each of the IAT experiments in this study, participants will favor the group to which they belong. Thus, in the four IAT experiments comparing *US* stimuli to *Arabic* stimuli, it can be predicted that US participants will have stronger associations with the *US* stimuli and *pleasant* than *Arabic* stimuli and *pleasant*.

The IAT experiment concerning gender may not be as simple to predict, because gender groups have been found to be an exception to this pattern. Women strongly prefer female stimuli when response latency techniques are used; men typically show neutral gender associations (Nosek & Banaji, 2002). Men are much less likely than women to show automatic in-group bias with gender (Rudman & Goodwin, 2004). This predicts that in the *female voice–pleasant* IAT experiment, the participants will have a stronger implicit association with *female* and *pleasant* than with *male* and *pleasant*. Female participants will have a stronger association with *female* and *pleasant* than with *male* and *pleasant* compared to male participants, who will be gender neutral.

In-group bias would seem to predict that in the *human voice–pleasant* IAT experiment, all participants will fall into the same in-group: human. It is probable that the results of this IAT will show a stronger implicit association with *human voices* and *pleasant* than with *machine voices* and *pleasant*. Because the topic of this IAT experiment is not socially sensitive, it is also expected that participants will have an explicit preference for human voices.

## 2.5 Mere Exposure

Mere exposure is the relation between frequency of encounter and stimulus preference. Established by Zajonc (1968), the mere exposure effect simply states that the more a person is exposed to a stimulus, the greater the chance the person will make positive associations about that stimulus. At the time the effect was established, there was no generally accepted explanation for it, but recent findings indicate that mere exposure is an implicit effect. Increased exposure first creates a greater likelihood for association and eventually a greater likelihood for positive association (Greenwald et al., 1998; Greenwald & Banaji, 1995; Fazio & Olson, 2003). Considering this is an implicit effect, it must be acknowledged as a contributing factor to the results of the current experiments in the same way as in-group bias.

Like in-group bias, mere exposure also predicts that in the four IAT experiments comparing *US* stimuli to *Arabic* stimuli, US participants will have stronger associations with the *US* stimuli and *pleasant* than *Arabic* stimuli and *pleasant*. The participants for the experiments in the current study were US residents (see section 3.1 for further participant details), and it is assumed that they have much greater exposure to US stimuli than to Arabic stimuli. This is also assumed to be the case for human voices rather than machine voices. For participants aged 18 years and older, it is assumed that they will have much greater exposure, especially in the formative years, to human voices rather than machine voices.

As our society is not sectioned by gender, it is difficult to predict what effect, if any, mere exposure will have on the *female voice–pleasant* IAT. Previous work suggests that maternal bonding from very early in our lives may be a cause of stronger implicit preference toward women (Rudman & Goodwin, 2004). This indicates that participants in the *female voice–pleasant* IAT will have stronger implicit associations between *female* and *pleasant* than *male* and *pleasant* but does not predict strength or gender differences.

## 2.6 Psychological Indices

Five psychological indices were used to gather further explicit data concerning the target stimuli: reassurance, warmth, competence, pleasure, and arousal. *Pleasure* and *arousal* are from the abbreviated pleasure-arousal-dominance indices (Mehrabian, 1995). The reassurance index was created in the development of the present study. The warmth and competence indices were taken from research in stereotypes from Fiske and colleagues (2002) and adapted into semantic differential format. Explanation of the creation of the reassurance scale and the adaptation of the warmth and competence indices can be found in section 3.4 along with further explanation of all psychological indices used in the present study.

## 2.7 Research Hypotheses

The following hypotheses test the ability of the IAT to measure secondary features of target stimuli and the use of audio stimuli. This methodology will be used to measure implicit associations with human and machine voices.

H1: The IAT can be used to measure the strength of association between a target concept that is a secondary feature of the stimuli and an attribute dimension.

H2: The IAT can be used to measure the strength of association between a target concept that is audio stimuli and an attribute dimension.

H3A: Female participants will have a stronger association between female voices and pleasant attributes than male voices and pleasant attributes compared to male participants.

H3B: Male participants will not have a significantly stronger or weaker association between female voices and pleasant attributes or male voices and pleasant attributes.

H4: Participants will have a stronger association between human voices and pleasant attributes than machine voices and pleasant attributes.

# CHAPTER THREE: METHODS

## 3.1 Participants

A random sample of undergraduate students at the nine campuses of a Midwestern university system was solicited by email. Age 18 or older, born in the United States, and current US residency were additional inclusion criteria. Among the six IAT topics, 1061 IAT sessions and their corresponding explicit sections were completed. Incomplete data was not retained. Some participants partook in more than one IAT; however, they were not permitted to take the same IAT twice.

## 3.2 Materials

The six IAT studies used the same set of 24 stimuli for attribute association. These 24 words were taken from a study of words rated as either high or low in pleasantness by male and female college students (Bellezza et al., 1986).

*Table 1: Attribute association visual stimuli. There are 12 positive terms and 12 negative terms.*

| Positive | Negative |
|---|---|
| 1. Wonderful | 13. Horrible |
| 2. Glorious | 14. Shameful |
| 3. Happy | 15. Sad |
| 4. Love | 16. Hate |
| 5. Good | 17. Evil |
| 6. Pleasure | 18. Pain |
| 7. Success | 19. Failure |
| 8. Peace | 20. Nasty |
| 9. Joy | 21. Awful |
| 10. Laughter | 22. Hurt |
| 11. Affection | 23. War |
| 12. Ecstasy | 24. Agony |

The *US-accented speech–pleasant*, *Male voice–pleasant*, and *Human voice–pleasant* IATs consisted of various distinct vocal sources all saying 16 neutral terms. The *US fonts–pleasant* IAT used the same 16 neutral terms represented by one of the following: Harrington, Old English, Arab Dances, or Arabian Onenightstand.

*Table 2: Target association visual or audio stimuli*

| | |
|---|---|
| 1. Candle holder | 9. Mixing bowl |
| 2. Cardboard box | 10. Peanut butter |
| 3. Ceiling fan | 11. Piano bench |
| 4. Coffee cup | 12. Picket fence |
| 5. Glass bottle | 13. Plastic cup |
| 6. Ironing board | 14. Remote control |
| 7. Living room | 15. Television set |
| 8. Magazine rack | 16. Vacuum cleaner |

For the *US-accented speech–pleasant* IAT the vocal sources were two males from the United States for whom English was the first language and two males for whom Arabic was a first language and English was a second language.

For the *male voice–pleasant* and the *human voice–pleasant* IATs the vocal sources were two male US citizens, two female US citizens, two male machine voices (Microsoft Mike and the ReadPlease Male voice) and two female machine voices (Microsoft Mary and the ReadPlease female voice).

The visual and audio *US names–pleasant* IATs used eight popular US male names and eight popular Arabic Muslim male names. In the visual version, the names were displayed in either Harrington or Old English. In the audio version the names were spoken by two female US-accented native English speakers.

*Table 3: Target stimuli for visual and audio US Names–Pleasant IATs.*

| | |
|---|---|
| 1. Michael | 9. Muzaffar |
| 2. Christopher | 10. Kashif |
| 3. Matthew | 11. Jamal |
| 4. Andrew | 12. Abdul |
| 5. Justin | 13. Rashid |
| 6. James | 14. Naazir |
| 7. Robert | 15. Kasim |
| 8. John | 16. Abbudin |

In each IAT experiment participants were asked to complete a self-reporting preferences section. This consisted of one question of relative preference concerning the target concept and the five psychological indices.

## 3.3 Procedures

Each of the six IAT experiments was conducted online at a website. Each participant was able to register at the website, provide demographic data (or skip that step), and give consent. Each IAT was followed by a questionnaire comprised of a relative preference question and semantic differential scales for the reassurance, warmth, competence, pleasure, and arousal indices. The order of the items in the semantic differential scales was randomized. The presentation order of the attribute-concept pairings within each IAT was counterbalanced. For example, for half the participants in the human-voice pleasant IAT, *human* was assigned to the same response key as *pleasant words* (vs. *machine* and *unpleasant words*) in block 3 and then, in the reverse combination task of block 5, *machine* was assigned to the same response key as *pleasant* words (vs. *human* and *unpleasant* words). For the other half of the participants, the presentation order of the combination tasks was reversed.

3.4 Creation of Psychological Indices

The reassurance index was created during the design of this study. These questions are semantic differential items ranging from -3 to 3 with 0 as neutral. The negative and positive anchors were *frightening-reassuring, disgusting-appealing, nervous-calm,* and *shocked-indifferent,* respectively. Internal reliability tests for this index showed a Cronbach's $\alpha$ of .77 for primary target concepts (e.g., *human voices, US fonts*) and .67 for secondary target concepts (e.g., *machine voices, Arabic fonts*) . These results show acceptable internal reliability.

Warmth and competence indices were adapted from stereotype research (Fiske et al., 2002). In Fiske and colleagues' (2002) study, warmth and competence were 5-point Likert scales based on the sentence "As viewed by society, how … are members of this group?" The words for warmth were *friendly, well-intentioned, trustworthy, warm, good-natured,* and *sincere*. The words for competence were *competent, confident, capable, efficient, intelligent,* and *skillful*. To stay consistent with methodology in the other psychological indices, the warmth and competence scales were transformed into semantic differentials ranging from -3 to 3 with 0 as neutral. This required assigning a negative counterpart to each positive anchor. The chosen negative-positive anchors for the warmth scale were *hostile-friendly, spiteful-well-intentioned, unreliable-trustworthy, cold-warm, grumpy-good-natured,* and *phony-sincere*. The chosen negative-positive anchors for the competence index were *incompetent-competent, timid-confident, helpless-capable, inefficient-efficient, stupid-intelligent,* and *clumsy-skillful*. Internal reliability for the adapted warmth index showed a Cronbach's $\alpha$ of .91 for the primary target concepts and .89 for the secondary target concepts. The adapted confidence index showed a Cronbach's $\alpha$ of .91 for primary target concepts and .88 for the secondary target concepts. These results indicate that the adaptation from the Likert scale to semantic differential was successful.

The pleasure and arousal indices were designed as semantic differential scales with a negative anchor of -3 and positive anchor of 3 with 0 as neutral. The negative-positive anchor pairs for the pleasure index are *sad-happy, bad-good, terrible-wonderful,* and *annoyed-pleased*. The negative-positive anchor pairs for the arousal index are *relaxed-stimulated, sluggish-frenzied, depressing-upbeat,* and *drowsy-energetic*. The pleasure index showed a Cronbach's $\alpha$ of .89 for primary target concepts and .89 for secondary target concepts. The arousal index showed a Cronbach's $\alpha$ of .78 for primary target concepts and .74 for secondary target concepts.

### 3.5 Data Analysis

These experiments were meant to examine possible differences in *D* score and explicit results that could be caused by gender, age, or level of education. As such, the participant group was divided into subgroups by gender and their median of age and education level. The *D* score was calculated using the scoring algorithm from Greenwald, Nosek, and Banaji (2003) with some exceptions. First, the Greenwald, Nosek, and Banaji scoring algorithm truncated results so that any latency below 300 ms was re-coded as 300 ms and any latency above 3,000 ms was re-coded as 3,000 ms. Observations were eliminated where latency was greater than 10,000 ms. Participants whose latency was out of the 300-3,000 ms boundaries for more than 10% of the trials were eliminated. The present study eliminated the extreme responses rather than truncating. Second, the present study removed the first trial of blocks three and five. Responses for these trials were consistently above 3,000 ms. As the IATs were conducted online, it is difficult to say what caused the extreme latency, but it is possible that participants did not fully read the instructions before beginning or were confused during the first trial. The psychological indices were calculated by averaging the results from each semantic differential scale with respect to the target concept. Relative preference is the mean of participants' responses to the relative preference question. Correlations were calculated using Pearson's correlation coefficient, and statistical significance was calculated using a two-tailed *t*-test.

# CHAPTER FOUR: RESULTS

Observations that were outliers of the limits of 300 ms to 3,000 ms were eliminated. The outlier data eliminated from each IAT experiment was low and is reported in Table 4 below.

*Table 4: The Summary of Total Number and Percentage of Observations*

| | Qualified | | Eliminated | |
|---|---|---|---|---|
| | *n* | % | *n* | % |
| Prefer US Names to Arabic, Written | 19,545 | 96.26 | 759 | 3.74 |
| Prefer US Fonts to Arabic | 17,797 | 97.09 | 533 | 2.91 |
| Prefer US Names to Arabic, Spoken | 15,298 | 95.91 | 652 | 4.09 |
| Prefer US Accents to Arabic | 12,588 | 95.65 | 572 | 4.35 |
| Prefer Female Voice to Male | 17,080 | 95.63 | 780 | 4.37 |
| Prefer Human Voice to Machine | 15,582 | 94.72 | 868 | 5.28 |

The results are reported in three main sections. Section 4.1 discusses the implicit results of the IAT experiments. Section 4.2 discusses the relative preference from the participants' self-reporting measures conducted with each IAT session. Section 4.3 discusses the correlations among the implicit, relative preference, and the psychological indices.

4.1 Implicit results

*Table 5: The Summary of IAT D Scores, Explicit Preferences, Cohen's d of Warmth, Effect Size of Warmth, Cohen's d of Pleasure, and Effect Size of Pleasure.*

| | IAT *D* | | Relative Preference | | Warmth | | Pleasure | |
|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *d* | $r^2$ | *d* | $r^2$ |
| Prefer US Names to Arabic, Written (*n*=216) | .23 | 0.38 | 0.89 | 1.28 | 0.28 | .14 | 0.32 | .16 |
| Prefer US Fonts to Arabic (*n* =195) | .33 | 0.31 | 0.58 | 1.41 | 0.29 | .14 | 0.18 | .09 |
| Prefer US Names to Arabic, Spoken (*n* =144) | .28 | 0.30 | 0.91 | 1.30 | 0.30 | .15 | 0.24 | .12 |
| Prefer US Accents to Arabic (*n* =139) | .18 | 0.36 | 0.80 | 1.50 | 0.19 | .10 | 0.29 | .14 |
| Prefer Female Voice to Male (*n* =190) | .30 | 0.39 | 0.36 | 1.48 | 0.48 | .23 | 0.39 | .19 |
| Prefer Human Voice to Machine (*n* =175) | .32 | 0.30 | 2.38 | 0.85 | 2.25 | .75 | 2.01 | .71 |

Table 5 shows agreement between the *D* score and the relative preference item. It also includes the warmth and pleasure indices converted to Cohen's *d* for comparison. Positive *D* scores indicate an implicit association between the first target concept (e.g., human voices, US names) and the pleasant attribute dimension, whereas positive relative preference scores indicate a preference for the first target concept. A one-sample *t*-test was used to ensure that all *D* scores and relative preference results were significant ($p$= .001 to .000).

Results from the written *US names–pleasant, US fonts–pleasant*, spoken *US names–pleasant*, and *US accents–pleasant* IATs indicate little self-presentational bias. *D* scores, relative preference, and Cohen's *d* and effect sizes from the warmth and pleasure indices all seem to correspond. The low effect size ($r^2$=.09) in the pleasure index in the *US fonts—pleasant* IAT could be explained by the playfulness of the Arabic fonts. Results in the *female voice–pleasant* IAT indicate both a stronger implicit association between female voice (the first target concept) and *pleasant* and a relative preference for female voices. The low relative preference (*M*= 0.36, *SD*= 1.48) compared to the relatively strong *D* score (*M*= .30, *SD*= 0.39) and the comparatively high results for the warmth (*d*= 0.48, $r^2$= .23) and pleasure (*d*= 0.39, $r^2$= .19) indices indicate self-presentational bias. The results indicate that while participants had a stronger implicit association with female voices and pleasant than with male voices and pleasant and also found the female voices warmer and more pleasant, participants did not express this preference in the relative preference item. This bias may indicate the social desirability of gender equality.

The results of the *human voice–pleasant* IAT also indicate self-presentational bias. The *D* score ($M$= .32, $SD$= 0.30) indicates a stronger implicit association between human voices and pleasant than machine voices and pleasant, but compared to the other IAT experiments, the *D* score is not the largest. The relative preference item indicated a strong preference for human voices ($M$= 2.38, $SD$= 0.85) as did the results from the warmth ($d$= 2.25, $r^2$= .75) and pleasure ($d$=2.01, $r^2$= .71) indices. The strong explicit results compared to the average *D* score indicate a high rate of self-presentational bias toward human voices.

*Table 6: The Summary of IAT D Scores and Explicit Preferences*

| | Overall *D* | | Target of *D* | | Attribute of *D* | | Relative Preference | |
|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Visual IAT | 0.23 | 0.38 | 0.22 | 0.41 | 0.24 | 0.44 | 0.89 | 1.28 |
| Audio IAT | 0.28 | 0.30 | 0.21 | 0.34 | 0.39 | 0.43 | 0.91 | 1.30 |
| Visual IAT Secondary Feature | 0.33 | 0.31 | 0.27 | 0.37 | 0.40 | 0.39 | 0.58 | 1.41 |
| Audio IAT Secondary Feature | 0.27 | 0.36 | 0.21 | 0.44 | 0.36 | 0.42 | 1.19 | 1.58 |

Table 6 shows the *D* scores of the overall IAT, the target concept, and the attribute. The table is arranged by IAT type. The visual IAT type (where both the target concept and the attribute were presented as text) shows only a small difference between the *D* of the target concept and the *D* of the attribute. The audio, visual with secondary features, and audio with secondary features IAT types show a significant difference between the target concept *D* and the attribute *D*.

*Table 7: IAT D Scores by Demographics*

| | Prefer US Names to Arabic, Written | Prefer Western Fonts to Arabic | Prefer US Names to Arabic, Spoken | Prefer US Accents to Arabic | Prefer Female Voice to Male | Prefer Human Voice to Machine |
|---|---|---|---|---|---|---|
| Gender | | | | | | |
| Female | .21 | .32 | .29 | .19 | .40 *** | .32 |
| Male | .27 | .34 | .26 | .16 | .02 | .31 |
| Age | | | | | | |
| Under 22 | .24 | .34 | .24 | .20 | .28 | .31 |
| 22 and Above | .22 | .32 | .33 | .16 | .31 | .32 |
| Education | | | | | | |
| Under 15 | .23 | .34 | .26 | .19 | .22 * | .30 |
| 15 and Above | .24 | .32 | .31 | .17 | .36 | .33 |
| View Order | | | | | | |
| First | .22 | .36 | .25 | -.03 *** | .12 *** | .29 |
| Second | .25 | .30 | .32 | .34 | .50 | .35 |

* $p<.05$ ** $p<.01$ *** $p<.001$

Note: In View Order *first* represents the first IAT target concept paired with pleasant in block 3 (e.g., *female voice–pleasant* or *human voice–pleasant*). *Second* represents the second IAT target concept paired with pleasant in block 5.

Table 7 shows three interesting results. Gender difference only appeared statistically in the *female voice–pleasant* IAT with female participants having a stronger association with female voices and pleasant (*M*=.40, *SD*= 0.37) than male participants (*M*= .02, *SD*= 30, $F(1,189)= 44.61$, $p= .000$). In the same IAT experiment, view order caused a significant difference in the *D* scores when participants heard *male voices* paired with *pleasant* (*M*= .50, *SD*= 0.35) as compared with *female voices* paired with *pleasant* in block 3 (*M*=.12, *SD*= 0.33, $F(1,189)=59.50$, $p= .000$). In the *US accents–pleasant* IAT, view order also caused a significant difference in the *D* scores when participants heard *US accents* paired with *pleasant* (*M*= -.03, *SD*= 0.33) as compared with *Arabic accents* paired with *pleasant* in block 3 (*M*= .34, *SD*= 0.29, $F(1,137)= 49.24$, $p= .000$).

*Table 8: Relative Preference by Demographics*

| | Prefer US Names to Arabic, Written | Prefer US Fonts to Arabic | Prefer US Names to Arabic, Spoken | Prefer US Accents to Arabic | Prefer Female Voice to Male | Prefer Human Voice to Machine |
|---|---|---|---|---|---|---|
| Gender | | | | | | |
| Female | 0.68 ** | 0.39 ** | 0.77 | 0.75 | 0.19 ** | 2.39 |
| Male | 1.28 | 1.02 | 1.14 | 0.88 | 0.83 | 2.36 |
| Age | | | | | | |
| Under 21 | 0.99 | 0.49 | 0.70 | 0.95 | 0.02 | 2.33 |
| 21 and Older | 0.78 | 0.66 | 1.15 | 0.68 | 0.52 | 2.41 |
| Education | | | | | | |
| Under 15 | 0.98 | 0.67 | 0.83 | 0.89 | 0.31 | 2.40 |
| 15 and Older | 0.82 | 0.49 | 0.99 | 0.73 | 0.40 | 2.37 |
| View Order | | | | | | |
| First Item First | 0.95 | 0.60 | 0.81 | 0.72 | 0.40 | 2.32 |
| Second Item First | 0.81 | 0.56 | 1.02 | 0.86 | 0.32 | 2.46 |

∆ $p$<.1 * $p$<.05 ** $p$<.01 *** $p$<.001

Table 8 indicates a significant gender difference in relative preference in the written *US Names–pleasant* ($M_{female}$= 0.68, $SD_{female}$= 1.25, $M_{male}$= 1.28, $SD_{male}$= 1.27, $F$(1,182)= 9.56, $p$= .002), *US fonts–pleasant* ($M_{female}$= 0.39, $SD_{female}$= 1.38, $M_{male}$= 1.02, $SD_{male}$= 1.38, $F$(1,172)= 7.60, $p$= .006), and *female voice–pleasant* IATs ($M_{female}$= 0.19, $SD_{female}$= 1.44, $M_{male}$= 0.83, $SD_{male}$= 1.50, $F$(1,189)= 7.23, $p$= .008). In all demographic categories and by view order, the relative preference for the *human voice–pleasant* IAT was consistently the highest.

*Table 9: D Score by Relative Preference*

| | Prefer US Names to Arabic, Written | Prefer US Fonts to Arabic | Prefer US Names to Arabic, Spoken | Prefer US Accents to Arabic | Prefer Female Voice to Male | Prefer Human Voice to Machine |
|---|---|---|---|---|---|---|
| Relative Preference | | | | | | |
| Low | .14 * | .24 *** | .19 ** | .07 *** | .28 | .30 |
| High | .27 | .43 | .33 | .28 | .31 | .33 |

* $p<.05$ ** $p<.01$ *** $p<.001$

Table 9 shows the *D* score of participants separated by either low or high relative preference by mean. The significant differences between the *D* scores for participants with low and high relative preference indicate a lack of self-presentational bias. However, the lack of significance in the difference between *D* scores in the *female voice–pleasant* and the *human voice–pleasant* IATs indicate self-presentational bias in these IAT experiments. If there were no self-presentational bias in the explicit measures, the *D* score should be lower if the relative preference is lower or higher if the relative preference is higher.

## 4.2 Psychological Indices

*Table 10a: Reliability (Cronbach's α) of Psychological Indices for First Item Viewed by IAT*

| | Reassurance | Warmth | Competence | Pleasure | Arousal |
|---|---|---|---|---|---|
| Prefer US Names to Arabic, Written | .80 | .92 | .91 | .90 | .83 |
| Prefer US Fonts to Arabic | .85 | .91 | .90 | .91 | .77 |
| Prefer US Names to Arabic, Spoken | .76 | .93 | .93 | .92 | .87 |
| Prefer US Accents to Arabic | .78 | .93 | .91 | .87 | .78 |
| Prefer Female Voice to Male | .78 | .91 | .91 | .89 | .75 |
| Prefer Human Voice to Machine | .72 | .89 | .90 | .85 | .63 |

*Table 10b: Reliability (Cronbach's α) of Psychological Indices for Second Item Viewed by IAT*

| | Reassurance | Warmth | Competence | Pleasure | Arousal |
|---|---|---|---|---|---|
| Prefer US Names to Arab, Written | .78 | .91 | .91 | .90 | .65 |
| Prefer US Fonts to Arab | .74 | .90 | .86 | .91 | .69 |
| Prefer US Names to Arab, Spoken | .80 | .90 | .92 | .91 | .78 |
| Prefer US Accents to Arab | .67 | .91 | .86 | .85 | .74 |
| Prefer Female Voice to Male | .70 | .87 | .88 | .82 | .70 |
| Prefer Human Voice to Machine | .30 | .72 | .82 | .78 | .65 |

Tables 10a and 10b show the strongest internal reliability in the warmth (Cronbach's α=.72 to .93), competence (Cronbach's α=.82 to .93), and pleasure indices (Cronbach's α=.78 to .92). Table 7b shows very low internal reliability for the reassurance index in the *human voice–pleasant* IAT (Cronbach's α=.30), which indicates that participants had inconsistent criteria when judging the machine voices.

Factor analysis showed that for both the first and second item viewed the items in the psychological indices loaded on three factors (see Appendix A). For both the primary and secondary target concepts the items loaded on three factors. For the primary target concept the first factor explained 53.45% of the variance and for the secondary target concept the first factor explained 50.35% of the variance.

*Table 11a: Psychological Indices for First Target by IAT*

| | Reassurance | | Warmth | | Competence | | Pleasure | | Arousal | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Prefer US Names to Arabic, Written | 0.57 | 0.86 | 0.41 | 0.88 | 0.46 | 0.91 | 0.44 | 0.86 | 0.19 | 0.78 |
| Prefer US Fonts to Arabic | 0.60 | 0.93 | 0.52 | 0.92 | 0.72 | 0.97 | 0.46 | 0.94 | 0.16 | 0.78 |
| Prefer US Names to Arabic, Spoken | 0.85 | 0.94 | 0.60 | 1.02 | 0.57 | 0.96 | 0.52 | 0.98 | 0.28 | 0.96 |
| Prefer US Accents to Arabic | 0.82 | 0.93 | 0.61 | 0.96 | 0.70 | 1.00 | 0.51 | 0.91 | 0.03 | 0.84 |
| Prefer Female Voice to Male | 0.71 | 0.90 | 0.86 | 1.09 | 0.71 | 0.96 | 0.78 | 1.04 | 0.67 | 0.83 |
| Prefer Human Voice to Machine | 1.15 | 0.83 | 1.26 | 0.91 | 1.07 | 0.96 | 1.23 | 0.91 | 0.69 | 0.76 |

*Table 11b: Psychological Indices for Second Target by IAT*

| | Reassurance | | Warmth | | Competence | | Pleasure | | Arousal | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Prefer US Names to Arabic, Written | 0.15 | 0.72 | 0.18 | 0.79 | 0.32 | 0.86 | 0.19 | 0.77 | 0.21 | 0.60 |
| Prefer US Fonts to Arabic | 0.31 | 0.79 | 0.25 | 0.91 | 0.42 | 0.93 | 0.28 | 0.97 | 0.35 | 0.76 |
| Prefer US Names to Arabic, Spoken | 0.31 | 0.88 | 0.31 | 0.89 | 0.54 | 0.91 | 0.31 | 0.84 | 0.31 | 0.73 |
| Prefer US Accents to Arabic | 0.45 | 0.79 | 0.43 | 0.94 | 0.43 | 0.87 | 0.25 | 0.92 | 0.21 | 0.82 |
| Prefer Female Voice to Male | 0.75 | 0.81 | 0.38 | 0.88 | 0.83 | 0.91 | 0.41 | 0.84 | -0.07 | 0.78 |
| Prefer Human Voice to Machine | 0.03 | 0.71 | -0.63 | 0.75 | -0.01 | 0.96 | -0.55 | 0.86 | -0.47 | 0.90 |

Tables 11a and 11b show strongest favor for human voices (*M*= .0.69 to 1.26, *SD*= 0.76 to 1.07) and lowest favor for machine voices (*M*= -0.63 to 0.03, *SD*= 0.71 to 0.96) in all five indices. When comparing responses for the spoken and written *US names–pleasant* IAT within tables, the spoken *US names–pleasant* IAT resulted in higher responses. Information stored in short-term memory has a higher acoustic. It is possible that audio target concepts are more efficient in the IAT and require less processing to associate (Shulman, 1970; Tversky, 1969).

4.3 Correlations

*Table 12a: Correlations between D, Relative Preferences, and Psychological Indices for Primary Target Concept*

| | Prefer US Names to Arabic, Written | Prefer US Fonts to Arabic | Prefer US Names to Arabic, Spoken | Prefer US Accents to Arabic | Prefer Female Voice to Male | Prefer Human Voice to Machine |
|---|---|---|---|---|---|---|
| | *D* | *D* | *D* | *D* | *D* | *D* |
| Relative Preference | .21 ** | .34 ** | .26 ** | .34 ** | .07 | .08 |
| Reassurance | .13 | .22 ** | .14 | .31 ** | .13 | .17 * |
| Warmth | .06 | .25 ** | .14 | .30 ** | .04 | .17 * |
| Competence | .01 | .15 | .09 | .19 * | .11 | .17 * |
| Pleasure | .03 | .19 * | .11 | .31 ** | .08 | .13 |
| Arousal | .07 | .23 ** | .09 | .01 | .08 | .25 ** |

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

*Table 12b: Correlations between D, Relative Preferences, and Psychological Indices for Secondary Target Concept*

| | Prefer US Names to Arabic, Written | Prefer US Fonts to Arabic | Prefer US Names to Arabic, Spoken | Prefer US Accents to Arabic | Prefer Female Voice to Male | Prefer Human Voice to Machine |
|---|---|---|---|---|---|---|
| | *D* | *D* | *D* | *D* | *D* | *D* |
| Relative Preference | .21 ** | .34 ** | .26 ** | .34 ** | .07 | .08 |
| Reassurance | -.10 | -.24 ** | -.06 | -.10 | -.01 | -.07 |
| Warmth | -.14 | -.23 ** | -.04 | -.16 | .09 | -.09 |
| Competence | -.04 | -.20 ** | .05 | -.12 | .08 | .03 |
| Pleasure | -.12 | -.31 ** | -.06 | -.17 | .08 | -.08 |
| Arousal | -.08 | -.07 | .03 | -.01 | .03 | -.06 |

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table 12a shows four patterns: The *D* score significantly correlated with the relative preference and the psychological indices in *US fonts-pleasant* ($r_{\text{Relative Preference}}$=.34, *p*=.000; $r_{\text{Psychological indices}}$=.15 to .25, *p*=.052 to .000) and *US-accented speech-pleasant* ($r_{\text{Relative Preference}}$=.34, *p*=.000; $r_{\text{Psychological indices}}$=.01 to .31, *p*=.915 to .000), the *D* score significantly correlated only with the relative preference in written *US names-pleasant* ($r_{\text{Relative Preference}}$=.21, *p*=.004) and spoken *US names-pleasant* ($r_{\text{Relative Preference}}$=.26, *p*=.002), the *D* score significantly correlated only with the psychological indices in *human voice-pleasant* ($r_{\text{Psychological indices}}$=.13 to

.25, $p$=.093 to .001), and no significant correlation between the $D$ score, the relative preference, or the psychological indices in *female voice-pleasant*.

Table 12b shows only three of the patterns: The $D$ score significantly correlated with the relative preference and the psychological indices in *US fonts-pleasant* ($r_{\text{Relative Preference}}$=.34, $p$=.000; $r_{\text{Psychological indices}}$=-.07 to -.31, $p$=.355 to .000), the $D$ score significantly correlated only with the relative preference in written *US names-pleasant* ($r_{\text{Relative Preference}}$=.21, $p$=.004), spoken *US names-pleasant* ($r_{\text{Relative Preference}}$=.26, $p$=.002), and *US-accented speech – pleasant* ($r_{\text{Relative Preference}}$=.26, $p$=.002), and no significant correlation between the $D$ score, the relative preference, or the psychological indices in *female voice-pleasant*.

.

# CHAPTER 5: DISCUSSION

In the written *United States names–pleasant* IAT, as predicted the *D* score shows stronger correlation with *US names* and *pleasant* than with *Arabic names* and *pleasant*. The relative preference and psychological indices express preference for written *US names*. There is strong correlation between the *D* score and the relative preference. These results were consistent with other IAT experiments concerning associations with Arabic target concepts (Nosek et al., 2006).

In the *US fonts–pleasant IAT*, the *D* score indicates stronger associations with *US fonts* and *pleasant* than with *Arabic fonts* and *pleasant* ($M$=.33, $SD$= 0.31, $F$(1,189), $p$= .000). The relative preference also indicates a preference for *US fonts* and is significantly correlated with the *D* score. The psychological indices are also significantly correlated with the *D* score for both the first and second targets. This is as expected; results were predicted to follow the written *popular US names–pleasant* IAT. These results indicate that the IAT is capable of measuring association strength with a secondary feature of a target concept and an attribute dimension. With the results from the first and second IAT, the study could move forward with an IAT using audio target concepts. H1 predicted that the IAT can be used to measure the strength of association between a secondary feature of a target concept and an attribute dimension. The results of the *US fonts–pleasant* IAT and the *US accents–pleasant* IAT support this hypothesis.

H2 predicted that the IAT can be used to measure the strength of association between target concepts that are presented as audio stimuli and an attribute dimension. If H2 were supported, the spoken *US names–pleasant* IAT would demonstrate a degree of sensitivity at least approaching that of the written *US name–pleasant* IAT. The mean *D* score of the spoken *US names–pleasant* IAT was .28 ($SD$= 0.30), and the mean relative preference was .91 ($SD$= 1.30, $F$(1,131), $p$= .11) . As in the written version, there was no significant correlation between the *D* score and the psychological indices, which may be interpreted as indicating self-presentational bias with respect to the socially sensitive topic of race. Considering the similarities with the results of the more traditional written version of this IAT, it is reasonable to believe that the IAT is capable of measuring the association strength between the target concepts and attribute dimensions using audio stimuli. The results from the spoken *US names–pleasant* IAT are also similar to previous visual IAT experiments concerning Arabic target concepts, either in text or images (Nosek et al., 2006). In an IAT experiment using *Muslim Arab names* versus *other names,* Nosek et al. found a *D* score of 0.14 (*SD*= 0.42) and an explicit result of 0.45 (*SD*= 0.77). The explicit results in the Nosek et al. experiment are also based on a single question: "Which statement best describes you?" The possible responses to this question ranged from -2 (I strongly prefer other people to Arab Muslims) to 2 (I strongly prefer Arab Muslims to other people).

The results from the *US accents–pleasant* IAT also indicate evidence of successful use of audio stimuli in an IAT. This experiment was designed to test the ability of the IAT to measure the associative strength of a secondary feature of audio target concepts. The *D* score indicates a stronger association between *US accents* and *pleasant* than between *Arabic accents* and *pleasant*. It is interesting to note that participants who had *US accents* paired with *pleasant* in block 3 of the IAT experiment showed a slightly stronger association between *Arabic accents* and *pleasant* than with *US accents* and *pleasant* (-0.03). It is also interesting to note that the strongest association between *US accents* and *pleasant* was with participants whose view order was the opposite (*Arabic accents* paired with *pleasant* in block 3). The *D* score was significantly correlated with the relative preference item as well as with the psychological indices of Reassurance, Warmth, Competence, and Pleasure for *US accents*. The psychological indices show that collectively participants felt the *US accents* were more reassuring, warmer, more competent, more pleasurable, and more arousing than *Arabic accents*.

The significant correlation between the *D* score and the relative preference in both the spoken *US names–pleasant* and *US accents–pleasant* IAT experiments and the similarities of these results to those of the written *US names–pleasant* IAT provide evidence that implicit associative strength between a target concept presented as an audio stimulus and an attribute dimension can be measured using an IAT.

H3A predicted that participants will have a stronger association between *female voices* and *pleasant* than between *male voices* and *pleasant*, and in the *female voices–pleasant* IAT, the *D* score supports this. When analyzed by demographics and view order, results show female participants have a much stronger association with *female voices* and *pleasant* than male participants. Male participants have a nearly neutral association between *female voices* and *pleasant* and *male voices* and *pleasant*, supporting H3B. Participants who had *male voices* paired with *pleasant* in block 3 of the IAT had a much stronger association with *female voices* and *pleasant* than participants who had *male voices* paired with *pleasant* in block 5.

Relative preference results from the *female voices–pleasant* IAT indicate a preference for *female voices* compared with *male voices*. In the psychological indices, results show *female voices* to be warmer, more pleasurable, and more arousing than *male voices,* while *male voices* scored higher in competence. The results of the psychological indices appear to indicate preference for female voices. The *female voices–pleasant* IAT results support the hypothesis in the implicit results, relative preference, and the psychological indices.

H4 predicted that participants will have a stronger association between *human voices* and *pleasant* than *machine voices* and *pleasant*, and in the *human voices–pleasant* IAT, the *D* score supported this. When analyzed by demographic group and view order, results show no significant differences in *D* scores. Relative preference also indicated a preference for human voices, though it is not significantly correlated with the *D* score. The psychological indices indicated a preference for human voices, and the results for human voice target concepts are significantly correlated with the *D* score in reassurance, warmth, competence, and arousal. The most positive results from a psychological index in all six IAT experiments were the results of the human voice questionnaire, while the most negative results came from the machine voice questionnaire. The results of the *human voices–pleasant* IAT strongly support H4. The results of the *human voices–pleasant* IAT also indicate strong self-presentational bias. This self-presentational bias could be a great hindrance in developing machine voices using only self-reported measures.

The results in Table 6 seem to indicate a higher degree of sensitivity in the D scroe of the attribute than the D score of the target concept. One possible explanation for this is attentional switching. Numerous studies confirm that human memory and cognition are based on the separate coding of imagery and verbal information (Paivio, 1991; Mayer & Sims, 1994; Mayer & Moreno, 1998). Channel correspondence affects the way in which attention is divided between the auditory and visual channels, because attention is limited (Kahneman, 1973). More attentional capacity is required to encode and integrate in memory the two information channels when those two channels convey slightly to moderately different messages. If there is low correspondence between the stimuli in the two information channels, attention to the visual channel will override attention to the auditory channel (Grimes, 1990).

# CHAPTER SIX: CONCLUSION

The audio IAT offers a novel measure of implicit association in audio domains, including features of the audio other than the semantic content. Specifically, this thesis has explored the following vocal dimensions: accent, which indicates national origin, gender, and humanness. These implicit results are less affected by social desirability concerns, which engender a self-presentational bias. By using methods that reduce this self-presentational bias in explicit results and focus on underlying associations, researchers can provide designers with better information on the complexity of the preferences of users. With this information designers can provide machine voices with which users have stronger positive implicit associations. This can improve the perceived reassurance, warmth, competence, pleasure, and arousal of the machine voice. A more favorably perceived machine voice would be beneficial to many facets of human-machine interaction from IVR systems to socially assistive robots. Implicit associations are not the only information needed for audio interaction design, but this research indicates that they should be acknowledged.

## 6.1 Limitations

Stimuli quality is very important. Future research should ensure that the audio stimuli are of the highest possible quality and consistent in volume, tone, speed, and other characteristics across stimuli to avoid possible confounds.

It is possible that the multi-modality of the audio IAT experiments (audio target concepts mixed with text for attribute dimensions) presented issues in change of focus with participants. It would be interesting to develop an IAT experiment with only audio stimuli for comparison to the audio IAT experiments in this study.

Evidence suggests that participants were unprepared for the first trials in blocks 3 and 5 of the audio IAT experiments. Future research should examine participants in a laboratory environment to determine if this is a matter of cognitive overload or simply not reading or understanding the instructions of the IAT experiment.

## 6.2 Future Research

Participants exhibiting self-presentational bias in self-reporting measures due to social desirability, in-group bias, or mere exposure may provide information that helps design a voice that is very politically correct and meets standards of social norms but does not accurately reflect the participant's attitudes and associations with the voice. Several avenues of future study present themselves.

What has not been examined in this study is priming by context. What would the results of the implicit or explicit measures have been if participants were told they were evaluating the voice of an automated customer service system for a travel agency? Would they be different than if they were told the voice would be used in a customer service system for a bank? How would variations indicating different ages in the voice affect the results? Simply asking participants these questions may produce data that is influenced by social desirability concerns. Future audio IAT experiments could prime one group of participants for a specific context and leave a control group unprimed. Comparison of the results could show evidence that implicit associations are context specific and provide a tool for testing machine voices with a specific intended use.

Self-presentational bias may also be tested in experiments concerning adherence to medical advice. A pretest IAT could provide voices that show the greatest positive association for the participants. A second pretest could provide voices with which participants claim the highest explicit measures. The two voices could then be compared in an experiment to determine which voice actually increased patient adherence to medical advice through vocal reminders and encouragement.

A similar study could be conducted using a game of financial consequences. The same two pretests could be used to determine voices: one with the greatest implicit association to positive attributes and one with the greatest positive explicit measures. These voices could then be compared in a game simulation on investing money. The different voices could provide investment advice throughout the game and measures could be taken of which voice's advice is more closely followed.

As machine voices increasingly enter customer service in medical and financial institutions, these experiments may provide insight into both the most effective voices for these institutions as well as continue to validate the need to compare implicit and explicit measures in the study of machine voices.

## REFERENCES

Allen, J.F., Byron, D.K., Dzikovska, M., Ferguson, G., Galescu, L., & Stent, A. (2001). Toward conversational human-computer interaction. *AI Magazine, 22*(4), 27–37.

Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology , 81*, 789–799.

Atkinson, R. K., Mayer, R. E., & Merrill,M. M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology*, *30*, 117–139.

Ashburn-Nardo, L., Knowles, M. L., & Monteith, M. J. (2003). Black Americans' implicit racial associations and their implications for intergroup judgment. *Social Cognition*, *21*, 61–87.

Bellezza, F. S., Greenwald, A. G., & Banaji, M. R. (1986). Words high and low in pleasantness as rated by male and female college students. *Behavior Research Methods, Instruments, and Computers*, *18*, 299–303.

Brebner, J., & Welford, A. (1980). Introduction: an historical background sketch. In A. T. Welford (Ed.), *Reaction Times.* Academic Press, New York.

Card, S.K., Moran, T.P. & Newell, A. (1984). The psychology of human-computer interaction. Hillsdale, NJ: Erlbaum.

Cohen, P. R. & Oviatt, S. L. (1995). The role of voice input for human-machine communication. *Proceedings of the National Academy of Sciences, 92*(*22*), 9921–9927.

Couper, M. P., Singer, E., & Tourangeau, R. (2004). Does voice matter? An interactive voice response (IVR) experiment. *Journal of Official Statistics*, *20*, 551–570.

Crabtree, M., Mirenda, P., & Beukelman, D. (1990). Age and gender preferences for synthetic and natural speech. *Augmentative and Alternative Communication*, *6*, 256–261.

Cuddy, A.J.C., Fiske, S.T., & Glick, P. (2007). The BIAS Map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, *92*, 631–648

Dasgupta, N., McGhee, D.E., Greenwald, A.G., & Banaji, M.R. (2000). Automatic preference for white Americans: eliminating the familiarity explanation. *Journal of Experimental Social Psychology,* 36, 316–328.

De Houwer, J. & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies*. New York: Guilford Press.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 5*(5), 5–18.

Dovidio, J. F., Evans, N. E., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology, 22*, 22–37.

Fazio, R.H. & Olson, M.A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297–-327.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype component: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82,* 878–902.

Gaertner, S. L. & McLaughlin, J. P. ( 1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly, 46,* 23–30.

Gaertner, S. L. & Dovidio, J. F. (1986). The aversive form of racism. In S. L. Gaertner & J. F. Dovidio (Eds.), *Prejudice, discrimination, and racism* (pp. 61–89). New York: Academic Press.

Gellatly, A.W. (1997). The use of speech recognition technology in automotive applications. *Doctoral Dissertation*, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Ghorbel, M., Segarra, M. T., Kerdreux, J., Keryell, R., Thepaut, A., & Mokhtari, M. (2004). Networking and Communication in Smart Home for People with Disabilities. *Proceedings of the 9th International Conference on Computers Helping People* (p. 937). Paris,France.

Gilbert, D. T. & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60,* 509–517.

Graf, P., & Schacter, D. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 501–518.

Greenwald, A. G. (1990). What cognitive representations underlie social attitudes? *Bulletin of the Psychonomic Society, 28,* 254–260.

Greenwald, A. G. & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102,* 4–27.

Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464–1480

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2000). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*, 3–25.

Greenwald, A.G. & Nosek, B.A. (2001). Health of the Implicit Association Test at age 3. *Z. Experimental Psychology, 48*, 85–93

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216.

Grimes, T. (1990). Audio-video correspondence and its role in attention and memory. *Educational Technology Research and Development*, *38*(2), 15–25.

Hewstone, M., Rubin M., & Willis H. (2002). Intergroup bias. *Annual Review of Psychology*, 53, 575–604.

Hick, W. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology , 4,* 11–26.

Houben, K. & Wiers, R. W. (2007). Are drinkers implicitly positive about drinking alcohol? personalizing the alcohol-IAT to reduce negative extrapersonal contamination. *Alcohol and Alcoholism*, *42*(4), 301–307.

Jacoby, L. L. & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340.

Jacoby, L. L. & Witherspoon, D. (1982). Remembering without awareness. *Canadian Journal of Psychology, 36,* 300–324.

Jacoby, L. L., Lindsay, D. S., & Toth, J. P. (1992). Unconscious influences revealed: Attention, awareness, and control. *American Psychologist*, *47.* 802–809.

Jamieson, D. A. & Zanna, M. P. (1989). Need for structure in attitude formation and expression. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 383–406). Hillsdale, NJ: Erlbaum.

Kahneman, D. (1973). *Attention and effort.* New York: Holt, Rinehart & Winston.

Kamm, C. (1995). User interfaces for voice applications. In D. Roe & J. Wilpon (Eds.), *Voice communication between humans and machines.* Washington, D.C.:National Academy Press.

Karpinski, A. & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*, 774–788.

Kawamura, K., Pack, R., Bishay, M., & Iskarous, T. (1996). Design philosophy for service robots. *Robotics and Autonomous Systems, 18*, 109–116.

Kemp, B. J. (1973). Reaction time of young and elderly subjects in relation to perceptual deprivation and signal-on versus signal-off condition. *Developmental Psychology, 8*, 268–272.

Kihlstrom, J. F. (1990). The psychological unconscious. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 445–464). New York: Guilford Press.

Klinger, M. R. & Beall, P. M. (1992, May). *Conscious and unconscious effects of stereotype activation.* Paper presented at the Annual Meeting of the Midwestern Psychological Association, Chicago.

Lane, K.A., Banaji, M.R., Nosek, B.A., & Greenwald, A.G. (2007). Understanding and using the Implicit Association Test: IV. What we know (so far) about the method. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Progress and controversies* (pp. 59–102). New York: Guilford Press

Luce, R. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Mackie D.M. & Smith E.R. (1998). Intergroup relations: insights from a theoretically integrative approach. *Psychological Review*, *105*,499–529.

MacDorman, K. F., Vasudevan, S. K., & Ho, C.-C. (2008). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI and Society, 23*(4), 485–510.

Marshall, W., Talbot, S., & Ades, H. (1943). Cortical response of the anaesthesized cat to gross photic and electrical afferent stimulation. *Journal of Nerophysiology*, *6*, 1–15.

Martin, T. (1976). Practical applications of voice input to machines. *Proceedings of the IEEE,* 64, 487–501.

Mayer, R. & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*(2), 312–320.

Mayer, R. & Sims, V. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, *86*(3), 389–401.

Mayer, R., Sobko, K., & Mautone, P. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology, 95*, 419–425.

Mehrabian, A. (1995). Framework for a comprehensive description and measurement of emotional states. *Genetic, Social, and General Psychology Monographs, 121*, 339–361

Moller, R. (2004). Advanced human-computer interaction methods for the control of technical systems. *Devices, Circuits, and Systems*. Proceedings of the Fifth IEEE International Caracas Conference on Devices, Circuits, and Systems, Dominican Republic, *1*, 130–135.

Möller, S., Krebber, J., Raake, A., Smeele, P., Rajman, M., Melichar, M., Pallotta, V., Tsakou, G., Kladis, B., Vovos, A., Hoonhout, A., Schuchardt, D., Fakotakis, N., Ganchev, T., & Potamitis, I. (2004). INSPIRE: Evaluation of a Smart-Home System for Infotainment Management and Device Control. In *Proceedings of the International Conference on Language Resources and Evaluations* 2004. P-Lisbon.

Möller, S., Krebber, J., & Smeele, P. (2006). Evaluating the speech output component of a smarthome system. *Speech Communication, 48*, 1–27.

Moreno, R., Mayer, R., Spires, H., & Lester, J. (2001). The case for social agency in computer based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, *19*, 177–213.

Mullennix, J., Stern, S., Wilson, S., & Dyson, C. (2003). Social perception of male and female computer synthesized speech. *Computers in Human Behavior, 19*, 407–424.

Nosek, B. & Banaji, M. (2002). The go/no-go association task. *Social Cognition*, *19*, 625–664.

Nosek, B., Greenwald, A., & Banaji, M. (2005). Understanding and using the Implicit Association Test: II. Methodological Issues. *Personality and Social Psychology Bulletin*, *31*(2), 166–180.

Nosek, B., Greenwald, A., & Banaji, M. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J.A. Bargh (Ed.), *Automatic processes in social thinking and behavior*. New York: Psychology Press.

Olson, M. & Fazio, R. (2004). Reducing the influence of extra-personal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*, 653–667.

Paulhus, D., Martin, L., & Murphy, G. (1992). Some effects of arousal on sex stereotyping. *Personality and Social Psychology Bulletin, 18,* 325–330.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*(3), 255–287.

Ramus, F. & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America, 105*(1), 512–521.

Roediger, H., Weldon, M., & Challis, B. (1989). Explaining dissociations between measures of retention: A processing account. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness* (pp. 3–42). Hillsdale, NJ: Erlbaum.

Rudman, L. & Kilianski, S. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin*, *26*, 1315–1328.

Rudman, L. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science, 13,* 79–82.

Rudman, L. & Goodwin, S. (2004). Gender differences in automatic ingroup bias: Why do women like women more than men like men? *Journal of Personality and Social Psychology, 87*, 494–509.

Schacter, D. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 432–444.

Schafer, R. (1995). Scientific bases of human-machine communication by voice. *In Proceedings of the National Academy of Science*, *92*, 9914–-9920.

Sherif, M. (1966). *In common predicament: Social psychology of intergroup conflict and cooperation.* Boston: Houghton Mifflin.

Shulman, H. (1970). Encoding and retention of semantic and phonemic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 499–50

Shriver, S., Toth, A., Zhu, X., Rudnicky, A., & Rosenfeld, R.  (2001). Unified Design for Human-Machine Voice Interaction. In *ACM Conference on Human Factors in Computing Systems*, Seattle.

Stefanov, D., Bien, Z., & Bang, W. (2004). The smart house for older persons and persons with physical disabilities: structure, technology, arrangements, and perspectives. *IEEE Trans Neural System Rehabilitation Engineering, 12*(2), 228–250.

Stern, S., Mullennix, J., Dyson, C., & Wilson, S. (1999). The persuasiveness of synthetic speech versus human speech. *Human Factors, 41*, 588–595.

Tversky, A. (1969). Intrasensitivity of preferences. *Psychological Review*, *76*, 31–48.

Tourangeau, R., Couper, M., & Steiger, D. (2003). Humanizing self-administered surveys: Experiments on social presence in Web and IVR surveys. *Computers in Human Behavior*, *19,* 1–24.

Welford, A. (1980). Choice reaction time: Basic concepts. In A.T. Welford (Ed.), *Reaction Times.* New York: Academic Press, 73–128.

Wilder, D. & Simon, A. (2001). Affect as a cause of intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 153–172). Malden, MA & Oxford, UK: Blackwell.

Woodworth, R. & Schlosberg, H. (1954). *Experimental Psychology*. New York: Henry Holt.

Yerrapragada, C. & Fisher, P. (1993). Voice controlled smart house. In *Proc. IEEE Int. Conf. Consumer Electronics* (pp. 154–155). Rosemont, IL.

Zajonc, R. ( 1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9 (Supp. 2, Pt. 2).

# APPENDICES

## Appendix A: Factor Analysis of Psychological Indices Items

*Table 13a: Total Variance Explained of Psychological Indices Items for Primary Target Concept*

| Component | Extraction Sums of Squared Loadings | | |
|---|---|---|---|
| | Total | % of Variance | Cumulative % |
| 1 | 12.83 | 53.45 | 53.45 |
| 2 | 1.59 | 6.61 | 60.07 |
| 3 | 1.24 | 5.18 | 65.24 |

*Table 13b: Rotated Factor Matrix of Psychological Indices Items for Primary Target Concept*

| | | Factor | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| Warmth | Warm-Cold | **.83** | .17 | .10 |
| Arousal | Upbeat-Depressing | **.76** | .12 | .36 |
| Pleasure | Good-Bad | **.76** | .41 | .12 |
| Warmth | Friendly-Hostile | **.75** | .37 | .06 |
| Pleasure | Wonderful-Terrible | **.74** | .35 | .19 |
| Reassurance | Reassuring-Frightening | **.74** | .35 | .01 |
| Reassurance | Appealing-Disgusting | **.73** | .33 | .13 |
| Warmth | Sincere-Phony | **.73** | .29 | .16 |
| Warmth | Good-Natured-Grumpy | **.73** | .34 | .19 |
| Pleasure | Happy-Sad | **.73** | .27 | .25 |
| Pleasure | Pleased-Annoyed | **.70** | .37 | .16 |
| Arousal | Energetic-Drowsy | **.62** | .09 | .55 |
| Warmth | Well-Intentioned-Spiteful | **.60** | .49 | .12 |
| Warmth | Trustworthy-Unreliable | **.59** | .51 | .15 |
| Reassurance | Indifferent-Shocked | -.03 | **.69** | -.05 |
| Competence | Confident-Timid | .36 | **.66** | .16 |
| Competence | Competent-Incompetent | .45 | **.65** | .22 |
| Reassurance | Calm-Nervous | .44 | **.64** | -.15 |
| Competence | Capable-Helpless | .46 | **.63** | .20 |
| Competence | Efficient-Inefficient | .40 | **.61** | .15 |
| Competence | Intelligent-Stupid | .44 | **.61** | .33 |
| Competence | Skillful-Clumsy | .40 | **.56** | .36 |
| Arousal | Stimulated-Relaxed | -.02 | .12 | **.80** |
| Arousal | Frenzied-Sluggish | .30 | .06 | **.72** |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
[a] Rotation converged in 5 iterations.

*Table 14a: Total Variance Explained of Psychological Indices Items for Secondary Target Concept*

| Component | Extraction Sums of Squared Loadings | | |
|---|---|---|---|
| | Total | % of Variance | Cumulative % |
| 1 | 12.08 | 50.35 | 50.35 |
| 2 | 1.72 | 7.15 | 57.50 |
| 3 | 1.40 | 5.84 | 63.34 |

*Table 14b: Rotated Factor Matrix of Psychological Indices Items for Secondary Target Concept*

| | | Factor | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| Warmth | Warm-Cold | **.83** | .10 | .09 |
| Pleasure | Good-Bad | **.79** | .37 | .14 |
| Arousal | Upbeat-Depressing | **.78** | .07 | .33 |
| Pleasure | Wonderful-Terrible | **.77** | .30 | .17 |
| Eerie | Appealing-Disgusting | **.75** | .24 | .16 |
| Warmth | Sincere-Phony | **.75** | .22 | .15 |
| Eerie | Reassuring-Frightening | **.75** | .28 | .01 |
| Warmth | Friendly-Hostile | **.74** | .36 | .00 |
| Pleasure | Happy-Sad | **.73** | .25 | .22 |
| Warmth | Good-Natured-Grumpy | **.71** | .38 | .14 |
| Pleasure | Pleased-Annoyed | **.69** | .36 | .10 |
| Arousal | Energetic-Drowsy | **.63** | .06 | .55 |
| Warmth | Trustworthy-Unreliable | **.60** | .48 | .12 |
| Warmth | Well-Intentioned-Spiteful | **.59** | .51 | -.01 |
| Eerie | Indifferent-Shocked | -.12 | **.68** | -.09 |
| Eerie | Calm-Nervous | .37 | **.66** | -.21 |
| Competence | Confident-Timid | .30 | **.62** | .22 |
| Competence | Competent-Incompetent | .44 | **.61** | .29 |
| Competence | Intelligent-Stupid | .47 | **.60** | .32 |
| Competence | Efficient-Inefficient | .38 | **.57** | .14 |
| Competence | Capable-Helpless | .44 | **.57** | .27 |
| Competence | Skillful-Clumsy | .40 | **.52** | .38 |
| Arousal | Stimulated-Relaxed | -.08 | .12 | **.76** |
| Arousal | Frenzied-Sluggish | .32 | .01 | **.72** |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
[a] Rotation converged in 6 iterations.

**Appendix B: Stimuli Material**

*Table 15: US Accent, Female Voice, Human voice, and Font IATs stimuli*. There are 16 neutral terms and 4 distinct vocal sources or four distinct font styles.

1. Candle holder
2. Cardboard box
3. Ceiling fan
4. Coffee cup
5. Glass bottle
6. Ironing board
7. Living room
8. Magazine rack
9. Mixing bowl
10. Peanut butter
11. Piano bench
12. Picket fence
13. Plastic cup
14. Remote control
15. Television set
16. Vacuum cleaner

*Table 16: American-Arabic Names IATs Stimuli*. There are 8 popular American and 8 popular Arabic-Muslim names.

1. Michael
2. Christopher
3. Matthew
4. Andrew
5. Justin
6. James
7. Robert
8. John
9. Muzaffar
10. Kashif
11. Jamal
12. Abdul
13. Rashid
14. Naazir
15. Kasim
16. Abbudin

*Table 17: Attribute Words Stimuli*. There are 12 positive terms and 12 negative terms.

1. Wonderful
2. Glorious
3. Happy
4. Love
5. Good
6. Pleasure
7. Success
8. Peace
9. Joy
10. Laughter
11. Affection
12. Ecstasy
13. Horrible
14. Shameful
15. Sad
16. Hate
17. Evil
18. Pain
19. Failure
20. Nasty
21. Awful
22. Hurt
23. War
24. Agony

# CURRICULUM VITA

WADE JOSEPH MITCHELL

## Education

IUPUI
January 2007 – May 2009
*M.S. – Informatics – Human Computer Interaction*

UNIVERSITY OF PHOENIX
September 2004 – October 2006
*M/MIS - Masters of Science in Management of Information Systems*

IUPUI
January 1996 – August 2004
*B.G.S. - Arts and Humanities with Distinction*

## EXPERIENCE

UNIVERSITY INFORMATION TECHNOLOGY SERVICES, IUPUI
*Client Services Analyst II (November 7, 2005 –Present)*
Responsibilities.

Training Coordinator and Supervisor for 16 Indianapolis staff members providing phone, walk-in, email and chat support. Research and solve user issues including all facets of technology. Provide support for over 80,000 users in the IU system with varying degrees of technical knowledge and experience. Update HP Peregrine trouble ticket system as main point of contact and user issue distribution. Distribute contacts to required support groups. Design and implement training activities and coordinate training with other groups for Indianapolis staff.

INDIANA UNIVERSITY SCHOOL OF LAW - INDIANAPOLIS
*Multimedia and Educational Specialist (January 13, 2003 –November 7, 2005)*
Responsibilities.

To research and implement new instructional technologies. Manage installation and integration projects including scheduling, resources, and budgeting. To maintain existing multimedia technology. Record, edit and distribute special events and guest speakers. Produce audio and video media for distance education. Educate the faculty, staff and students on all relevant hardware and software including classroom presentation technology, Distance Education, Microsoft Office, file management, and multimedia design and creation.

Presentations – Computer-Assisted Legal Instruction National Conference
Topic – Cost-effective technology updates for older classroom environments
June,2005

UNIVERSITY INFORMATION TECHNOLOGY SERVICES, IUPUI
*TECHNICAL SPECIALIST (April 15, 1996 –JANUARY 13, 2003)*
Responsibilities.

To supervise the staff and AV/ Computer equipment deliveries. To maintain and troubleshoot all equipment. Special event multimedia set up and control. Audio/Video production.