A Model of Persuasion for Speaking Rate Adaptation

by

Ke Deng

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2010

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Proper speaking rate is a key attribute of effective communication. Emerging persuasive technologies use computers as a tool to induce human behavioural and attitude changes. This thesis established a computational framework which can persuade people to slow down their speech and communicate more effectively.

We defined a conceptual model and implemented a computer software system, both serving as the cornerstones of our persuasion framework. The computer system is designed to persuade people to be aware of their speaking rate and to slow down their speech. The combination of computer technology and persuasive technologies and theories are embedded in the system. In order to conduct effective persuasion, a number of computer-based survey questions were asked and a short tailored letter was generated for each participant. A virtual coach system monitored and reminded the participant to slow down. A few adaptive cues were used to enhance the effects of the persuasion.

We evaluated the feasibility and effectiveness of the overall system. At the same time, we evaluated the feasibility of individual elements. A total of 22 participants was selected to make up the sample. The experiments were conducted under controlled conditions. The results indicated that our system is effective in persuading people to speak more slowly. The feedback from users indicated that our system raised their awareness about speaking rate.

Acknowledgements

First and foremost, I would like to thank my wonderful supervisor, Professor Chrysanne DiMarco, without whom, this thesis would never have been completed. I thank her for her constant support and guidance.

I would like to thank Professor Michael Terry and Professor Neil Randall for spending their precious time and effort in giving feedback, which was critical in helping me improve the thesis.

My thanks also go to my course instructors, fellow students, and other faculty members at the University of Waterloo. Your help has been kind and generous.

Finally, my thanks go to my family and my friends. Thanks for your support and encouragement.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Effective communication is an essential skill in both personal and professional life. Communication involves messages, message senders, and message receivers. Effective communication happens when the message is clearly delivered from the sender and correctly interpreted by the receiver. Effective communication is one of the determinant factors for successful teamwork, which has become a widely adopted practice for any non-trivial sized projects to deal with the ever increasing real-world competition.

1.1 A Communication Scenario

Consider this scenario:

> Mr. Slow, Mr. Right, and Mr. Quick are all experts in their own fields. Before they met, each of them worked on small-sized projects which needed only a small amount of communication with other people. One day, Mr. Manager at a giant company started a huge project that required the expertise of all three. After working together for a while, Mr. Slow complained to Mr. Manager that he often misunderstood some parts because Mr. Quick spoke too quickly; Mr. Quick also complained that he often found it difficult to concentrate because Mr. Slow spoke too slowly. No one had any complaints about Mr. Right, but Mr. Right appeared the most upset because he found both Mr. Slow and Mr. Quick to be problems: one too slow and the other too quick!

Obviously, Mr. Slow, Mr. Right, and Mr. Quick failed to coordinate effectively, which definitely negatively affected the quality of their teamwork, and could have led to project failure. Mr. Manager knew the consequences that this communication problem could cause. Mr. Manager again told Mr. Quick that he should speak more slowly, and Mr. Slow that he should speak more quickly. A few weeks later, Mr. Manager heard the same complaint again from all three: Mr. Quick still spoke much too quickly, and Mr. Slow very slowly. Mr. Manager was at a loss how to solve the problem.

Although the above scenario seems simple, it is not uncommon in our everyday life. The only difference is that most of us just leave the problem unresolved. Some people simply do not care about the serious consequences poor communications may cause; others just do not know how to resolve the problem. After all, persuading people to change any behaviour is a challenging task. An intervention based solely on intuition is generally not an effective startegy. We believe that approaches grounded in formal psychological theories of persuasion and behaviour modification are needed.

1.2 Persuasion Is Challenging

Persuading people to change their rate of speaking can be challenging for the persuader and frustrating for the speaker. The speaking rate of a speaker is a habitual behaviour [81]. A habitual behaviour is supported by an individual's belief system. Without a proper strategy for modifying someone's speech behaviour, a persuader may end up with an ineffective outcome, and the speaker may experience frustration because of his failure to change. A number of external factors (such as different situations and emotions [49]) can also affect the rate of speaking, making persuasion even more complex. For example, people tend to speak quickly when they are excited. Without practice or training, people often speak too quickly in public speaking. Although speech coaches will advise speakers to slow down in public speaking, the persuasion effect is greatly offset by the high cost of speech coaches and the limited areas where human coaches are feasible. For example, a human coach may tell the speaker to slow down before a presentation, but the speaker may still speak too quickly during the presentation.

1.3 The Role of Computers in Facilitating Persuasion

Computer technologies provide a promising solution to behaviour modification in the form of automated virtual coaches. The emerging persuasive technologies use computers as a tool to induce human behavioural and attitude changes. For example, Soler et al. [85] designed and implemented a mobile persuasive technology that aims at convincing adolescents to be aware of oral and dental health. The authors used a game design approach which embedded persuasive strategies. Their target audience was teenagers and young adults. As another example of a persuasive health technology intended for children, The Playful Toothbrush [19] targets kindergarten children, teaching them proper tooth-brushing techniques.

Soler et al. follow Fogg's [33] persuasive technology design process, which focuses on a simple target behaviour to be modified, a carefully chosen target audience, well-considered possible barriers for persuasion, and an easily accessed technology channel. In Soler et al.'s context, the target behaviour was dental hygiene, the target audience was adolescents, who were considered to be easily motivated by the perceived benefit of an improved appearance. Soler et al. also believed that lack of awareness of potential illnesses caused by poor dental habits would be key barrier to behaviour change. Mobile phones were chosen as the technology channel because adolescents were considered to be "married" to their mobile phones.

In the dental game, a mouth space is metamorphized as a city in which different types of bacteria (the inhabitants) live. These bacteria survive by exploiting resources in the city (i.e., the mouth). Over time, exploitation of resources can lead to nature's revenge in the form of natural disasters. Dental hygiene activities, such as tooth-brushing are metaphorized as good natural disasters which destroy the areas inhabited by the bacteria. Under serious conditions, a catastrophe (e.g., fluorogenesis) can happen, leaving only a small number of survivors. These survivors will easily adapt to the new oral world, and continue to exploit resources.

A player takes the role of one type of bacteria, represented as an antagonist role (i.e., it is assumed the bacteria are good inhabitants). During the game, the player (in the role of bacteria) tries to learn techniques for exploiting more resources without considering the consequences of natural disasters. By learning how bacteria exploit resources, the player will acquire a better understanding of how bad dental habits lead to dental illnesses. The game is designed as a scrolling puzzle-platform game, which is well-suited to the small

screens of mobile phones. The characters were designed with the aesthetics of adolescents in mind. Overall, the game reflects three persuasion strategies: cause-and-effect simulation, suggestion, and attractiveness.

The game was evaluated on a group of 17 adolescents. The results showed that the game was effective in raising adolescents' awareness of dental health, but the long-term impact on dental health habits remains unknown. These examples demonstrated that computer-related technologies can serve as a powerful tool for modifiying the habitual behaviour of people.

1.4 Pervasive Persuasion

Computer technologies, especially mobile devices, make pervasive persuasion possible. Mobile devices can not only be connected to the Internet, but also be equipped with sensors, which can collect real-time information about the users. Moreover, the portable size of mobile devices makes them ubiquitous, i.e., accessible any time anywhere. For example, in the context of speaking rate persuasion, a mobile device might be embedded with a voice-capturing sensor and the related software modules that can analyze speech and give proper reminders.

Pervasive Persuasion also means the overall social cost of persuasion will be reduced. For example, in the context of speaking rate persuasion, general and tailored persuasion strategies can be implemented by means of automated virtual coaches. Speakers would have access to more convenient and less expensive coaching than personal vocal coaches. With this benefit, more and more people might become motivated to improve their speech skills, which would potentially lead to more effective communication in a variety of business and life situations. And, as the volume of persuasion technology applications increases, improvements in persuasion strategies will result from careful analysis of the outcomes of these applications, which will in turn increase the spread of persuasive systems.

1.5 Problem Statement

1.5.1 Overview

The benefits of effective communication have been discussed in previous sections. For example, effective communication can improve the quality of teamwork. In specialized domains such as business and politics, communication skills are highly important, and often taught by communication experts. In the rest of the world, improving one's communication skills is still is a challenge task because of the high cost and limited availability of expert human voice coaches. In many job situations, most especially in technical industries, lack of communication skills is a common problem [82], partly caused by the overwhelming emphasis on technical skills.

As the size and complexity of projects in organizations escalates, coordination among team members becomes a critical factor. The basis for better team coordination is effective communication. However, different people have their own styles of communication, which are formulated according to their own habits and often rooted in their belief systems [93]. For example, a person who habitually speaks very quickly may hold a well-established set of beliefs about good speech habits. For example:

- "Speaking quickly is not a problem".
- "Speaking quickly gives me an advantage in social situations".
- "Speaking quickly helps listeners understand me better".
- "I do not think I speak too quickly".
- "My friends value me for my quick speech".
- "I know I speak too quickly but I can't change this."

Persuading people to change their rate of speaking is challenging because speech is a habitual behaviour. However, even habitual behaviours can be changed if appropriate methods are chosen for the particular individual. For example, smoking is known to be extremely harmful, but many people still smoke because smoking has become a habit. Tang et al. [55] conducted an experiment in which a doctor met with each smoker once

for less than 5 minutes to try to persuade the person to quit smoking. The results of the experiment showed that only 2% of smokers quit smoking. Although this cessation rate is quite low, the experiment is still a successful example of habitual behaviour modification. Furthermore, as Reiter et al. [76] commented, if computers could be used to persuade people to change their behaviours, even with such a high failure (98%), the overall benefits would still be significant, given the low costs involved.

A person's rate of speaking is an essential attribute of their overall speech style. Studies [44, 95] have shown that speaking rate can affect listener perception, especially for certain groups of listeners such as non-native speakers and those with hearing impairments. Janse [45] investigated word perception for both naturally produced rapid speech and artificially time-compressed speech. The author found that even if the natural rapid speech was completely intelligible, listeners still found it difficult to process. One reason might be that rapid speech causes listener perception issues (e.g., degraded interaction between high-level concepts and low-level lexical information). On the other hand, speaking too slowly may delay the delivery of information and cause an audience to lose interest. For a detailed discussion about speaking rate, see Section 2.7.

Rate of speaking is affected to some extent by a speaker's personal belief system, which is often interwoven with the speaker's emotions. For example, when a speaker becomes excited or angry, he will tend to speak quickly. It is not uncommon that poor speakers avoid speech therapy, although they are aware of their problems with speaking too rapidly or too slowly. People may be reluctant to seek help due to embarrassment. In such situations, a non-judgemental virtual coach may serve as an effective replacement for a human therapist, and may encourage more people to act on improving their speech skills.

Since emotions can affect a speaker's communication skills, it is reasonable to assume that controlling affect can play an important role in speaking rate persuasion. For example, vocal messages delivered by a speaker can convey different emotional effects: gentle, upset, angry, etc. In a persuasive technology interface, emotionally related elements, such as icons indicating various emotional cues, could be displayed in order to affect the user's emotional response. For a detailed discussion about how emotions are related to persuasion, see Section 2.5.

1.5.2 Research Goals and Questions

The ultimate goal of this research is to establish a computational framework which can persuade people to communicate in the most effective manner. Specifically, we focus on how to persuade people to adjust their speaking rate to a "just right" point for the most effective communication. To the best of our knowledge, little research to date has been done on this problem. The fundamental research problem we are addressing is to define this new area and identify open research questions.

We believe that changes in a speaker's rate of speaking can affect a listener's comprehension, and that a properly designed model of speech model can be used to control rate of speaking. Our current objective is to investigate the effectiveness of persuasive technologies in controlling speaking rate. More specifically, our hypothesis for this thesis is that a persuasion model can be used to increase awareness of speaking rate among speakers, and slow down their speaking rate. The following research questions will be studied in this research:

1. Which components should be included in such a persuasion model?
2. Which persuasive strategies are effective for controlling speaking rate?
3. Can such a model raise awareness of speaking rate among speakers?
4. How should emotional elements be applied in such a computational model?
5. How should one implement such a computational model?
6. How should one evaluate such a computational model?

This research is aimed at helping to develop a new research area in persuasive technologies for speech communication. We believe that the following long-term research challenges should be considered and investigated:

Ontologies:

As discussed in Chapter 2, the study of persuasion involves research across multiple areas, including philosophy, psychology, physiology, rhetoric, and sociology. Our research investigates how to use computer technologies to relate knowledge from these

various areas for conducting effective persuasion to modify speaking rate. To systematically study the interrationships among these areas, it is necessary to construct an ontology, which can represent multidisciplinary knowledge in terms of concepts and the formal relations between these concepts. This ontology should be constructed and evaluated incrementally as our research progresses.

Affect:

Human speaking rate depends on a number of known and unknown factors. One of the essential factors is the emotions of a speaker. A detailed discussion about emotions can be found in Section 2.5. However, studies about emotions are still limited, especially for speaking rate adaptation. Thus, an important research task is to further investigate the relationship between emotions and speaking rate, aiming at building models of affect that can reflect these relations. More importantly, we need to investigate how to realize these affect models in a computational framework.

Evaluation:

Due to a number of uncertain factors, such as human factors and our current limited understanding about speech behaviour modification, evaluating speech persuasive systems is difficult. For example, how do we determine whether the speaker is using the most effective speaking rate? As research progresses, a standard set of evaluation methods need to be defined and implemented.

Ethics:

Ethics plays an important role in persuasive systems as it directly affects human behaviour and attitudes. Typically, persuasiveness is considered as the primary objective in designing persuasive technologies. However, how to ensure ethical practice is addressed in persuasive system design should be addressed as this field progresses.

1.5.3 Results

Chapter 5 will discuss the results of this research, which reflect the feasibility and effectiveness of our perusasion model. The results comprised of two parts: computer-recorded data analysis and human-interpreted feedback results. The effectiveness of the model refers to

1. Whether the model can raise the awareness of speakers on speaking rate.

2. Whether the model can actually slow down the speaking rate of speakers.

The results are also used as a source of finding possible directions for future research in this new field.

Chapter 2

Background

Speaking rate is an essential property of human speech. Different people speak using different rates, and even the same person often uses different rates in different situations. Modern brain research has proved that human speech is mainly controlled by the left hemisphere of the brain [94], indicating that speaking rate, as part of the overall speech activity, has a physiological basis in the brain. Vocal organs of a speaker move successively to produce speech events, which result in acoustical signals perceived by listeners. The rate of these movements determines the speaking rate of a speaker.

Our literature search reveals that little work has been done on persuading people to change their speaking rate, especially using a computer-based persuasion approach. Speaking rate persuasion involves a wide range of research fields. This section will review these fields and discuss the why these fields are important for developing our persuasion model. Section 2.1 reviews the philosophical aspect of persuasion. Section 2.2 reviews the general persuasion models. Section 2.3 reviews the relevant work in the field of persuasive technologies. Section 2.4 discusses persuasion models based on rhetorical techniques. Section 2.5 reviews the affect aspect of persuasion. Section 2.6 discusses the relevance of tailored techniques for persuasion. Section 2.7 reviews the literature of speaking rate. Section 2.8 reviews two relevant behavioural theories which serve as the foundation of our persuasion model. Although these sections review different aspects for developing a speaking rate persuasion model and discuss the relevant background to help establish this new field, how to effectively combine this wide range of knowledge into a computer platform is a question *per se*.

2.1 Philosophical Persuasion

The philosophical aspect of persuasion should be considered as one of the cornerstones of our proposed persuasion model, which aims at inducing speech behaviour among human speakers. This section reviews a number of different philosophical views on persuasion to help establish the foundation and coverage of this new research field. Our research need to consider how to apply these views to our computer-based persuasion model.

The Latin root for the word persuasion is *persuadere*, which can be divided into two sub-words: *per* (meaning thoroughly) and *suadere* (meaning to urge). Socrates, one of the most inscrutable Greek philosophers, believed that both true and false ideas already exist in people's minds, and that asking people questions is the best way to pursue the truth. He travelled around Greece and taught philosophy for no money. Because he frequently questioned established ideas and rules, Socrates was sentenced as a capital criminal by the city of Athens (for dishonouring the gods and corrupting youth) and was offered two options: 1) being executed and 2) being exiled but stopping questioning. Socrates chose the second option and finished his life with a cup of hemlock [26, 12].

In contrast to Socrates, Sophists were a group of Greek teachers who travelled through Greek-speaking countries and taught people rhetoric by charging fees. They believed that pursuing truth was not their top priority. Instead they tried to prove their positions using arguments. Arguments were constructed to make people believe a proposition, which might not be true. Although Sophists' argumentative approach could be immoral, they did represent an early form of persuasion.

The philosophy of Socrates was not directly written by Socrates. Plato wrote down his ideas through literary dialogues. Plato believed that rhetorical strategies and argumentation could be completely separated, and that Sophists were taking advantage of this separation by using rhetoric without content [40]. Thus, Plato considered all rhetorical strategies should not be used for arguments, but instead arguments should be proved using only evidence.

Rhetoric focuses on the linguistic aspects of persuasion. Plato's student, Aristotle, believed that the rhetorical style was as important as the argument itself, establishing a foundation for modern rhetoric. Aristotle infused argument and presentation skills into formal and informal reasoning systems. Aristotles rhetoric considered audience as the

central part, and believed that the overall rhetoric was about persuading audience using rhetorical proofs: pathos (emotion aspect), ethos (credibility of the speaker), logos (logical aspect), and topoi (argument discovery) [50].

Coercion is a concept that has been linked with persuasion since the beginning of persuasion. Coercion implies forcing people to do what they are not willing to do. Modern philosophy differentiates persuasion from coercion. Powers [72] categorized and discussed the differences from six aspects: 1) intention, 2) willingness to harm, 3) message, 4) result, 5) choice available, and 6) choice constrained.

Persuasion and coercion both carry an intention, Persuasion carries a non-harmful intention, but coercion carries a harmful one. The harmful intention involved for coercion often results from a willingness to harm on the part of the message sender. Both persuasion and coercion require a message to realize the sender's intention. The message in persuasion often involves positive, non-threatening information, but the message in coercion contains threats. Persuasion often has a positive result on the audience, but coercion often produces a bad result. With persuasion, an audience often have various choices available to them, but with coercion, choice hardly exists. In the case of a persuasive message, the audience is not constrained, but with coercion, the intended behaviour of the audience is often constrained by certain conditions [72].

2.2 General Persuasion Models

Guerini [42] summarized the definitions of persuasion into five categories, each considering a different aspect of persuasion: 1) how to effectively structure logical arguments 2) why audiences can be persuaded 3) how to conduct persuasion using cognitive devices 4) what are the goals of persuasion 5) what is the relationship between the message sender's goals and the audience's goals Guerini also discussed computational systems that belong to each of these five categories. Only a few representative examples are selected and briefly discussed below. For more examples, please refer to [42]. Guerini's model represents a general persuasion model, and reveals a direction of how to use computer to conduct persuasion. However, Guerini's model does not consider the specific problems for speech behaviour persuasion.

STOP [76] is a persuasive system with a clear goal (i.e., persuade users to quit smok-

ing). STOP is based on Natural Language Generation (NLG) techniques, and conducts computer-based persuasion. STOP generates tailored smoking cessation letters for patients by analyzing their questionnaires. A user (smoker) fills in a 4-page paper questionnaire or a Web-based HTML form. The user-entered data is stored in a database. The core system of STOP uses document-planning schemas to produce a document tree, which is then used to generate a four-page tailored letter for each recipient. STOP represents a possible cost-effective way of conduct text-based persuasion. Although the results of STOP did not confirm its effectiveness, we believe that this tailored approach can still represent a valuable approach for conducting speaking rate persuasion if a tailored letter can be properly designed.

ARGUER [77] is a typical example of argumentative persuasion. ARGUER is an interactive argumentation system through which a user can have a dialogue with the system. The user enters an utterance into the system. The system detects and interprets the user's utterance with the help of argument schemata, which represent how an argument may be structured. After reasoning based on the user's belief models, the system can generate responses with a proper argument for rebuttal.

Promoter [42] is an intelligent user interface for inducing actions from users. Promoter aims at defining a general purpose persuasion model with consideration of user cognitive state, social relations, emotional state, and the context. Persuasive strategies are represented as rules in the system. The reasoning model creates a taxonomy of persuasive strategies, and selects the proper strategies. The selected strategies can be modified when necessary, and then tagged with rhetorical relations for generating text messages. The messages are translated into a intermediate language that can be realized in different modes. Promoter focuses on general persuasion modeling and multimodal realization.

ARGUER and Promoter are relevant to our research because they both employ interactive interfaces. However, for speaking rate persuasion, a different set of persuasive interfaces and strategies must be developed and tested.

2.3 Captology

Recently, Fogg [30] at the Stanford Persuasive Technology Lab established a new research field called "captology" to study how to use computers to conduct persuasion, especially

mobile persuasion. Captology covers a wide range of fields including computer technologies (e.g., websites, mobile phones, video games) and persuasive technologies (e.g., motivation, attitude, behaviour change). Captology can be used for inducing both attitude or behaviour changes. Because it is easier to track behaviour changes than attitude changes, most captology research currently only aims for inducing behaviour changes.

Captology focuses not only on persuasion modelling, but also on system design. Fogg recommended a set of principles when developing a persuasive system. Development starts with finding a target behaviour to change. The next step is to find an audience that is persuadable. After finding the target behaviour and proper audience, an analysis must be performed to find the key factors that prevent audiences from changing their target behaviour. These factors can be discovered by answering three questions:

1. Does the audience lack motivation?
2. Does the audience lack ability?
3. Does the persuasion happen at the proper time?

With these factors in mind, possible persuasive strategies can be constructed and integrated into a system. Coming up with effective strategies requires knowledge from multiple research areas such as human-computer interaction, information system, affective computing, philosophy, rhetoric, psychology, etc. [30, 23].

Fogg also emphasized the importance of choosing a proper persuasion channel such as websites, video games, mobile applications, social networks, etc. Some persuasive technologies may benefit from using multiple channels. Some forms of persuasion may work better when implemented as a multimodal system, i.e., a combination of text, audio, and video messages.

2.3.1 Persuasive System Design Examples

This section discusses persuasive systems from three aspects that relate to speaking rate persuasion: 1) Interface Design-Based Approach describes an example of how to persuade drivers not to overspeed; 2) User Interaction-Based Approach describes an example of how to persuade users to take breaks during work; 3) Mobile Device-Based Approach describes

an example of how to persuade users to achieve their exercise goals by monitoring their heart rate. Although many other persuasive systems, such as story-based game playing [85], are equally interesting systems, this section only discusses related examples involving speed changes (e.g., driving speed, frequent breaks during work, heart rate changes) because these may suggest some aspects for speaking rate persuasion, the focus of this thesis.

Interface Design-Based Approach

Speeding is a serious and common problem related to road safety in many countries. Transport Research Laboratory reported that every one mile per hour average speed increase leads to 5% more accidents [87, 53]. Kumar et al. [53] conducted a survey and identified as one reason for speeding that drivers are often not aware of the current road speed limit as well as their current speed. They adopted Fogg's design principles [30] and brainstormed 20 different possible design methods [1]. Through ideation, they designed a Dynamic Speedometer, which can constantly display the current speed limit and driver's speed.

Dynamic Speedometer increases drivers' awareness of the current speed limit by visualizing it so that drivers do not have to search for this information. For example, if the current speed limit is 40 mph, the speedometer will highlight all the speed readings above 40 mph. When the driver exceeds the speed limit, the speedometer will change its appearance (e.g., changing its background colour) to remind the driver. In addition to the visual reminders, the speedometer can also play audio beeps with different frequencies representing different levels of speeding. Kumar et al. believe that automobiles will be able to automatically retrieve road speed limit information using technologies such as GPS digital map databases. By conducting experiments in a driving simulator, Kumar et al. reported that their dynamic speedometer could effectively persuade drivers to slow down.

User Interaction-Based Approach

Dynamic speedometer persuades drivers using visual and audio-based reminders. Although reminder-based persuasion seems to be an effective approach, frequent interrupting with reminders may reduce the persuasive effect. This situation is especially obvious when using break-reminder software programs, which try to increase a computer user's productivity by persuading the user to take frequent breaks. For example, many computer users believe

that taking frequent breaks may decrease their productivity, and thus either turn off their break-reminder program or ignore break notifications. However, Barredo et al. [9] reported that frequent, long-term, non-stop use of keyboard and mouse can cause repetitive strain injury (RSI).

SuperBreak [63] is designed to persuade a computer user to take frequent breaks while continuing to do activities based on the user's preference. SuperBreak replaces the solely-wait breaks with interesting activities that do not involve finger uses. For example, a user can play games, read documents, or watch videos using arms (e.g., stretching arms). The movements of arms are captured by webcams. SuperBreak supports three types of activities: 1) vision-based games, 2) vision-based document reading, and 3) passive video presentation.

During a vision-based game, a user tries to catch the red and green spots in the screen. If the user catches them in time, she will get a score, and the spots will appear in other positions. The whole purpose is to encourage the user to stretch her arms instead of typing. Vision-based document reading enables a user to navigate a document using arm movements. The passive video presentation plays video clips found on the Internet (e.g., youtube.com) so that the user can have a rest. During the breaks, the keyboard can also be locked based on the user's preference. SuperBreak reported that the activities during breaks did encourage people to take more breaks.

Time Aura [57] is a computer system designed to minimize users' cognitive demands for pacing while performing tasks. The system consists of a set of interfaces which contain visual cues. These interfaces assist presenters to pace their progress during presentation. The results indicated that the visual cues in Time Aura is effecive in reducing users attention needed for task pacing. Time Aura is relevant to our work because pacing monitoring and alternation are also important aspects of speaking rate adaptation. Presentation can be considered as a specific realization of the more general communication style.

Mobile Device-Based Approach

The ubiquitous and portable features of mobile devices provide an excellent platform for implementing persuasion. A detailed discussion of mobile persuasion can be found in [34]. This section focuses on an aspect of design principles for mobile persuasion (i.e., glanceable interface). A glanceable interface conveys information in a concise and simplified way,

which is an essential feature for conducting real-time persuasion where intervention may cause intolerable effects. For example, when persuading a speaker to change their speaking rate, a complex interface may significantly interrupt the speaker's thought processes, and negatively affect the speaker's performance. Thus, a glanceable interface embedded with aesthetic elements can realize the principle of simplicity for users.

TripleBeat [25] aims at persuading people to do more physical exercises and achieve their specific research goals. TripleBeat monitors a user's heart rate, which reflects the current exercise performance level of the user. If the user's heart rate falls below the training zone, a glanceable interface with appropriate recommendation will be displayed (e.g., a pink heart image with +6 means the user needs to increase his heart rate by 6 beats per minute). The simple design of the interface reduces a user's cognitive effort for interpreting it so that the user can concentrate on his physical exercise. TripleBeat also use musical feedback to encourage users to speech up or slow down their exercises. That is, based on current heart rate, the system can calculate a piece of music with a particular feature and play it for the user.

TripleBeat also uses elements of social pressure, i.e., having users compete with each other. However, TripleBeat defines its own standard for the competition: users define their own exercise goals and different users have different goals. How well a user performs is not based on how fast he/she can run, but how well he/she can reach his/her predefined goals. TripleBeat also implemented its own score functions for measuring a user's performance against his/her goal. This strategy enables users to enjoy the environment of competition, but keeps expectations realistic and best-suited to the individual users.

Current Research Persuasive Systems

From the example systems discussed above, a common pattern can be observed: they all deliver messages (or conduct interventions) through multiple channels (i.e., text, images, sounds, or video). Depending on the context, persuasion often requires a carefully designed interface that is both simple and enjoyable. The persuasive strategies chosen are often gentle and encouraging. The three examples all involve some form of persuasion related to a rate of change, and the results were all positive, indicating a possible solution may be found for implementing speaking rate persuasion. These systems also focus only on behavioural changes rather than attitude changes, which agrees with Torning's survey [88]:

84.4% of current research papers address only behavioural change.

Captology is relevant to our research because our model is designed to conduct computer-based persuasion. Furthermore, our speaking rate persuasion systems should be considered as a branch of the captology field. We address a specific set of problems within the area of persuasive technologies.

2.4 Rhetorical Persuasion

Rhetorical persuasion is relevant to our research because our model also uses text-based persuasion techniques, which are the focus of rhetoric. Rhetorical models represent a systematic way of organizing text so that they can be more persuasive. As research progresses, applying rhetorical models into the persuasion model will become necessary.

The origin of modern rhetoric can be traced back to Ancient Greece when Aristotle, a student of Plato, first infused argument and presentation skills into formal and informal reasoning systems. Aristotle's rhetoric considers audience as the central part, and the overall rhetoric is about persuading audience using rhetorical proofs including logos (logical aspect), pathos (emotion aspect), ethos (credibility of the speaker), and topoi (argument discovery). Aristotle's perception established the foundation of traditional rhetoric [50].

Kenneth Burke [15], one of the most influential contemporary rhetoricians, defined rhetoric in a different way. Burke considered rhetoric as "the use of words by human agents to form attitudes or to induce actions in other human agents." He also contended that rhetoric was "rooted in an essential function of language itself" and "the use of language as a symbolic means of inducing cooperation in beings that by nature respond to symbols."

Aristotle's rhetoric focused only on what the rhetor did to the audience. In contrast, Burkean rhetoric emphasizes "consubstantiality", in which the audience can actively participate in the persuasion rather than only passively receive messages from the rhetor. That is, the audience receives messages from the rhetor, and identifies with the received messages. This identification process is based on Burke's assumption that each symbol used by the rhetor is associated with an attitude. Identification can be used in three different forms: 1) the rhetor establishes commonality with the audience, 2) the rhetor creates a common enemy (antithesis), and 3) the rhetor includes the audience as part of

the persuasion. Burke believed that human nature can encourage separated individuals to establish consubstantiality through which an identification process can occur [15].

Burke's rhetoric also provides a method, called "Pentad", for understanding the motivations embedded in symbols. Pentad consists of five different types: act, scene, agent, agency, and purpose. Act represents what events happened; scene describes the context of the act; agent refers to the person who conducts the act; agency is the method used; purpose is the reason for the act. The symbols in a discourse can be classified into the five types. Each type will take a certain ratio of the discourse. Understanding the distribution of these ratio values is important for interpreting the motivations of the discourse [16].

Burke's rhetoric augmented traditional rhetoric by introducing the concept of consubstantiality. Persuasion conducted by the rhetor involves the audience as a central part. The audience identifies with the symbols from the rhetor. These symbols are assumed to carry attitudes or motivation. To interpret the motivation, an identification process and Pentad are employed. For example, according to Burke's Pentad, the motives of a speaker can be analyzed based on what was done (act), where was it done (scene), who did it (agent), how was it done (agency), and why was it done (purpose).

2.4.1 Rhetorical Models

Although current research in persuasive technologies does not directly use rhetorical models, we will consider such models in relation to speech persuasion as part of the "seeding" nature of this thesis, i.e., we are trying to establish a new research topic in persuasion for speaking rate adaptation. Rhetorical models will serve as an important component in our future research.

Natural Language Processing (NLP) methods provide a platform on which tailoring strategies can be realized. Statistical NLP models use machine learning and data mining techniques to process and analyze corpus through a quantitative approach. This quantitative approach often requires a huge amount of corpus, which is not always available for certain fields such as health documents of a specific type.

Kelly et al. [51] took an alternative approach which analyzes a tailored medical corpus using descriptive rhetorical models. The authors believed their rhetorical models could generate low-cost, high-quality tailorable medical documents. These models are based

on theories from Rhetoric of Science, Rhetorical Genre Theory, and Rhetoric of Health. Persuasive strategies from these theories are selected and embedded into the models.

The initial setup is to generate persuasive weight-loss documents for diabetes, involving a health educator and patients. The health educator communicates with the patients by posting tailored documents for each patient through an online course system. Each patient communicates with the educator by logging on to the course management system and answering health-related questions. A rhetorical model-based NLP system analyzes the answers from the patients and generates tailored persuasive documents for individual patients.

The overall model of Kelly et al. consists of two sub-models: Generic Structure Model (G_eSM) and Rhetorical Appraisal Model (R_hAM). G_eSM and R_hAM together represent a top-down approach analyzing the responses of patients from three different levels: thematic, rhetoric, and linguistic. G_eSM is designed to analyze the contextual situation and typology of documents at a generic level. G_eSM classifies the documents into three different genres: introduction, responses, and special greetings. G_eSM conducts thematic level analysis only. The more detailed semantic analysis and rhetorical analysis of each document is left to the R_hAM model.

R_hAM encompasses a wide range of rhetorical strategies covering eight dimensions: Social/Situational, Agency, Accountability/Responsibility, Risk, Reward, Emotional Tone, Rhetorical Question, and Chronos. Each dimension can be associated with a selectable value. Most of these dimensions are interrelated. Each value of a dimension represents a possible rhetorical strategy which can be used by the health educator. R_hAM analyzes each documents and interprets the value of the dimensions based on the communications between the health educator and the patients.

2.5 Emotions and Persuasion

Contemporary research in the psychology of persuasion believes that there is a relationship between affect and persuasion [27]. That is, it is possible to persuade people by changing their mental states. For example, commercial advertisements are usually embedded with persuasive multimedia content along with the essential product information. This embedded multimedia content is aimed at inducing the audience to switch to a different mental

state so that they can be persuaded to buy the advertised products. Thus, effective affect models can contribute to a promising approach to successful persuasion.

An affect model is a conceptualization of affective experiences, aiming at constructing a dimensional space that can capture affective experience similarities and differences [13]. One important criterion for evaluating an affect model is how many dimensions of factors are considered. For example, one simple but reliable single-dimensional model, called bipolar valence model, describes affect as a continuum whose ends are opposite emotion pairs such as happy-sad and positive-negative. Multi-dimensional models such as discrete emotion models view affect from several discrete aspects, attempting to interact with the environment in a more flexible way. However, this flexibility introduces extra complexity to the model, making its benefits questionable [27].

Two-dimensional affect models [78, 80] have been considered as an effective and economic approach over the past 50 years, despite disagreement from some researchers [38]. Two-dimensional models are derived through data reduction algorithms, which can analyze and reduce a large set of affective states into a small set which contains only two essential factors. That is, after a series of similarity sorting on a large set of emotion words, only two types of emotion words (i.e., two factors) are observed to be stable.

The exact selection of the two factors depends on their context, e.g., if some cultural context is considered, the two stable factors could be valence and arousal. When other different contexts are considered, the two factors could be one of the pairs: pleasantness-arousal, pleasantness-dominance, calm-alert, etc. This inconsistency is often considered as a weakness of two-dimensional affect models. Compared to the simple bipolar valence model, two-dimensional models conceive affect in a more precise way without adding much overhead. Compared to multi-dimensional models, two-dimensional models effectively capture the two necessary stable factors while lowering the complexity cost.

Affect models provide an essential platform on which persuasion can be conducted. Different affect models focus on different scenarios, while all of them aim at providing an accurate but economic dimensional space for categorizing emotions. Two-dimensional affect models are currently the most common approach because of their capability and economy features. Our persuasion model is embedded with emotional elements to increase its persuasiveness. Thus, an understanding of basic affect models will help develop the affective side of a persuasion model, especially for speaking rate persuasion.

2.6 Tailored Persuasion

Fogg [31] referred to tailoring as "taking the information from the system's databases and providing it to people during the normal routines of life". Tailoring in context is an open problem about how to tailor information retrieved from databases and use the information to change users' attitudes or behaviours through persuasion.

According to Fogg [32], a tailoring technology persuades users to change their attitudes or behaviours by presenting them with user-relevant information, so the users can avoid wading through voluminous information. Tailoring information to individuals has been studied, but tailoring information to context, the next big step, is still an open problem. For example, if a persuasive system is created to help a user lose weight, tailoring information to individuals may refer to customizing the persuasive information to the user's age, health records, education level, etc. To be more persuasive, the system may need to consider the user's current context which might include their current location (e.g., in a fastfood restaurant, at home, or in the workplace), current mood (e.g., happy, sad, or excited), current health condition (e.g., sick or well), people nearby (e.g., with friends or with family), etc. A message containing tailored information can then be transmitted to the mobile device carried by the user to be persuaded. For example, if the user is in a restaurant, the message may contain healthy menu choices; if the user is at home, the message may be a reminder of medical appointments; if the user is at workplace, the message may give tips on stress reduction or suggest exercises at the user's desk.

Mobile devices or health devices such as pedometers can further tailor the information received so that its presentation can be more persuasive. The content of the information can be summarized according to the individual user's needs and the capacity of health devices. For example, a text-based application reminder may better suit a cell phone while a simple voice warning message may better fit on a resource-constrained pedometer. If the user is a doctor, a detailed summary written in medical language may contain precise details. However, a summary written in everyday language may better serve for a patient with limited medical knowledge.

In the context of speaking rate persuasion, different persuasion strategies may serve better for different speakers, e.g., if a speaker speaks too quickly, the selected strategy should induce slower speech, but if the speaker speaks too slowly, the strategy should induce a quicker rate of speech. Furthermore, if a speaker does not realize that speaking

too quickly is a problem, an appropriate strategy would first raise the speaker's awareness about their speaking rate.

2.7 Literature on Speaking Rate

This section reviews a number of research areas that are closely related to conducting speaking rate persuasion. Section 2.7.1 reviews a number of factors that are related to speaking rate; Section 2.7.2 describes how speaking rate affects both speakers and listeners; Section 2.7.3 reviews speech perception models and how they are related to speaking rate.

2.7.1 Factors Related to Speaking Rate

In addition to brain research, researchers from other fields (linguistics, physiology, psychology, computer science, engineering, etc.) have also discovered a number of factors (e.g., emotion, age, demographic, personality, linguistic structure, etc.) that are related to speaking rate. For example, emotional states such as excitement can affect speaking rate, and different speech rates can elicit different emotional responses from listeners [49]; older adults tend to speak slower [7]; many people slow down their speech when using a second language because of increased cognitive demands [43]; utterance length and content can significantly affect speaking rate [73]; relationships may affect speaking rate [11].

Yuan et al. [96] studied these factors in an integrated way using conversations in both English and Chinese. Verhoeven et al. [91] investigated how linguistic background and extralinguistic variables can affect speaking rate. Verhoeven et al. explained linguistic background as variations of a pluricentric language (Dutch) used in different geographical regions. The extralinguistic variables were age and gender. The authors observed that linguistic background significantly affected speaking rate, that young people spoke more quickly than older people, and that males spoke more quickly than females. These findings are consistent with the traditional observations as discussed above.

2.7.2 The Effects of Speaking Rate

Speaking rate affects the brain activities of both speakers and listeners. For example, rapid speech often requires speakers to compress some syllables of their speech, leading to decreased acoustic quality [45]. A positive relationship between speaking rate and the number of speech errors has been observed [92]. Decreased acoustic quality and increased number of speech errors lowers comprehension for listeners, especially for elderly listeners [18, 46] and non-native listeners [48]. On the other hand, too-slow speech may significantly limit the information to be conveyed, and thus decrease the effectiveness of communication [43]. Using a proper speaking rate is essential for improving both the performance of speakers and the comprehension of listeners.

2.7.3 Speaking Rate and Perception Theories

Speech perception refers to the process by which humans interpret and understand the expression of ideas, thoughts, and emotions using voice. Vocal utterances carry acoustical energy produced from the movements of the speaker's lungs and vocal folds. Listeners perceive the acoustical energy and extract acoustic cues from it. The extracted cues are then mapped to linguistic information stored in human memory. Researchers have proposed many models (e.g., motor theory [56], exemplar theories [41, 47]) of speech perception.

Motor theory argues that humans do not perceive speech by identifying the incoming acoustic patterns. Instead, speech perception is done by identifying the vocal tract gestures (i.e., vocal contractions) of the speaker through a specific physiological module. Exemplar theories postulate that every word sound perceived by a listener leaves a unique trace in the memory. The speech perception process is performed by matching the new word sounds with the unique traces existing in the memory through similarity comparison.

Theoretical modeling of speech perception helps investigate how well a listener comprehends a speech, and what are the important factors for an effective communication. However, modeling speech perception is a difficult task because: 1) acoustic signals are physically continuous (segmentation problem) and 2) the relationship between acoustic signals and speech sounds are affected by many factors such as a phoneme's context, coarticulation, and the variety of speech styles. A phoneme's context refers to the context in which a phoneme is realized acoustically, e.g., "b" is realized differently in the words

(i.e., different contexts) “bike” and “climb” [8]. Coarticulation refers to the phenomenon in which the articulation of adjacent phonemes are overlapped, e.g., in the word “sweet”, the articulation of ‘s’ and ‘w’ are overlapped simultaneously [29].

Speaking rate is an essential attribute of a individual’s speech style. Studies show that speaking rate can affect speech perception, especially for certain groups of listeners, including foreign language learners and people with hearing impairments. Janse [45] studied word perception for both naturally produced rapid speech and artificially time-compressed speech. The author found that even if natural rapid speech is completely intelligible, listeners still find it difficult to process. One reason might be that rapid speech causes listener perception issues (i.e., degraded interaction between high-level constructs and low-level lexical access).

Yang et al. [95] developed a digital speech repeater based on the Pitch Synchronous Over-Lap Add (PSOLA) [20] technique. This repeater can change the speaking rate of recorded materials with a high voice quality. The authors used this repeater to investigate how speaking rate affects the listening comprehension of English language learners. They found that generally the comprehension level increased as the speaking rate decreased, but if the speaking rate was over-slow, the comprehension level was also reduced. Iwasaki et al. [44] studied the relationship between speaking rate and speech perception of cochlea implant users. They found slowing speech was an effective approach to improving perception. Findings from [68, 69, 70, 89] clearly indicated that speaking rate was an important factor for clear speech, and that slow speech improved the perception of hearing-impaired people.

2.7.4 Speech Rate Persuasion

The literature search described above confirms that speaking rate does matter in many situations. However, how to persuade people to change their speaking rate is still an open area to be investigated, especially through the newly created persuasive technologies such as Captology. The literature indicates that neither too-slow nor too-rapid speech is an ideal behaviour. However, a “just-right” rate is very difficult (if not impossible) to achieve without the ability to persuade a too-quick speaker to slow down and persuade a too-slow speaker to speed up. Although both slowing down and speeding up speech involve changing speaking rate behaviour, they each have a different set of problems to solve. We believe

that persuading speakers to slow down their speech should be investigated first because it may represent a sub-problem of persuading speakers to speed up. For example, a speaker can often more easily slow down his/her speech than speeding it up.

2.8 Behaviour Theories

Our persuasion model is based on two behaviour theories to be discussed in the following sections 2.8.1 and 2.8.2. This section starts with a brief introduction of behaviour theories. Behaviour theories explain the underlying reasons why humans follow a certain behaviour. An understanding of these theories is fundamental for developing a persuasion model, which aims at changing a specific human behaviour. Over two thousand years ago, the physician Hippocrates assumed that human behaviour follows certain patterns. Hippocrates believed that the human body contains four types of liquids (blood, black bile, yellow bile, phlegm). The behaviours of each individual were considered to be influenced by the mixture of these four liquids inside the person's body. When the four liquids are balanced, the individual behaves normally; when they are not balanced, different behaviour will occur. Although Hippocrates' theory may today seem too simple or lacking evidential support, it did represent a rejection of superstitions and an active attempt at explaining human behaviour [58]. Shakespeare also thought that human behaviour is an overt expression of the inside nature of individuals [62].

In 1928, Dr. William Moulton Marston [59], who developed polygraph (i.e., lie detector), proposed his famous DISC theory, which categorized human behaviours into patterns. He posited that human behaviours can be categorized into four types: Dominance, Influence, Steadiness, and Conscientiousness. Dominance focuses on how to deal with problems; Influence focuses on how to influence others; Steadiness focuses on how to react to dynamic environment; Conscientiousness focuses on how to comply with rules. The individual uses a combination of the above types of behaviour to achieve a certain goal. Certain types of behaviours work better for specific goals than other types. The DISC model only categorized human behaviours; it did not state one type was more advantageous than others.

2.8.1 Theory of Planned Behaviour

In 1985, Ajzen [4, 5] proposed a now well-known behaviour theory called Theory of Planned Behavior (TPB), which was extended from a previous theory called Theory of Reasoned Action (TRA). TPB and TRA took the view that human behaviour is mostly goal-oriented. A goal is realized by carrying out an intention and unfolding an action plan. When a human conducts a behaviour, a plan is either explicitly or implicitly constructed. For example, people make an explicit plan before they go to a meeting, but people often drink water without conscious thought, although a plan has been implicitly constructed.

Intention is an antecedent of action. TRA postulated that intentions are determined by two factors: attitude toward the behaviour and a subjective norm. Both factors are mostly effective for behaviour that requires only volitional control. For example, a person can sit down and drink a cup of water easily; the sitting and drinking behaviour is under volitional control of the person. However, if a behaviour is complex or complicated, TRA may become ineffective because pure volitional control seems insufficient. TPB augmented TRA with a new factor, perceived behavioural control, to consider behaviours that are not under volitional control.

TPB postulated that intention is determined by three factors: 1) attitude toward the behaviour, 2) subjective norm, and 3) perceived behavioural control. These three factors are based on Expectancy-Value Models (EVM) [4], which consider salient beliefs and evaluation of the expected outcomes. Attitude toward the behaviour is determined by behavioural beliefs, which are in turn determined by two factors: 1) salient beliefs about how likely the behaviour will bring about a positive or negative outcome; and 2) the degree to which the outcome will be evaluated as positive or negative. Subjective norm is determined by normative beliefs, which are in turn determined by two factors: 1) salient beliefs about social pressure from significant others on performing the behaviour; and 2) the motivation to comply with other people's opinions. Perceived behavioural control is determined by control beliefs, which are in turn determined by two factors: 1) salient beliefs about how often inhibiting factors will be encountered, and 2) how likely these inhibiting factors will be overcomed.

Ajzen also stated that intentions may change with the passing of time and information newly received. This indicates that it is possible to change the salient beliefs of a person with new information, resulting in a change in the person's intention. Ajzen also observed

that time may affect the stability of an intention, thus transforming intention to action should be done in a timely manner. Strength of salient beliefs is also an important factor when relating to intention. For example, if a person has a very strong belief in a certain behaviour, it may be difficult to change the person's intention within a short period of time.

2.8.2 Fogg Behaviour Model

In 2009, Fogg [36] presented a behaviour model, Fogg Behaviour Model (FBM), which directly guides the persuasive design process. A key purpose is to discover the underlying factors that drive human behaviour, assisting designers to understand why persuasive technologies work. As Fogg stated, "many attempts at persuasive design fail because people don't understand what factors lead to behavior change." FBM focuses on influencing the behaviours rather than the attitudes of people.

FBM postulates that the realization of a target behaviour is determined by three factors: 1) motivation, 2) ability, and 3) triggers. That is, if these three factors are satisfied at the same time, the target behaviour can take place. To persuade people to perform a behaviour, a designer must know which of these factors need to be enhanced. The combination of these three factors represents the probability of the occurrence of the target behaviour. FBM also implies that motivation and ability have a trade-off relation. That is, if the ability is extremely high but the motivation is low, the target behaviour may still occur. For example, if a person is not motivated to buy a house, but if the house costs only one dollar, this person may still buy it if this person has a high ability to pay the one dollar. Similarly, if the motivation is extremely high but the ability is low, the target behaviour may also occur. That is because the high motivation may drive a person to do extra work that increases the person's ability. In both cases, triggering must be involved.

FBM considers triggering as an essential condition for a target behaviour to happen. Even if both motivation and ability are high, without proper triggering, the behaviour may still not occur. For example, even if a person is willing to play basketball and has spare time and sufficient skills, the person may still stay at home without proper triggering such as a phone call from friends. FBM also posits that an improper type of triggering may be annoying. Thus, choosing and designing a proper type of triggering is a key element for successful persuasive design.

Based on his knowledge and past design experience, Fogg recommended several elements for each persuasive factor. That is, to enhance the motivation factor, manipulating elements such as pleasure/pain, hope/fear, and acceptance/rejection can be considered. To enhance the ability factor, manipulating elements such as time/money/efforts may help. Triggers can be realized through a spark/facilitator/signal. Spark triggers are connected with the factors to be enhanced. Facilitator triggers simplify the behaviour while triggering it. Signal triggers are simple reminders.

2.8.3 Behaviour and Attitude

TPB explains human behaviour from the aspect of attitudes, intentions, and beliefs. FBM focuses on the direct manipulation of human behaviours. A distinctive element of FBM is the triggering system. A proper combination of TPB's belief evaluation and FBM's triggering system may lead to an effective real-time persuasion system.

2.8.4 Speech and Persuasion

Understanding which factors affect the persuasiveness of messages is important for developing a persuasion model. Thus, this section discusses the relevant voice characteristics which can be manipulated for having an effective speaking rate persuasion model. Persuasion messages can be realized in a number of ways, including text, sound, video, and so forth. Sound can in turn include a number of formats such as music and speech. Speech characteristics can significantly influence the effect of persuasion [84]. Vaart et al. [90] describe voice characteristics from five acoustic aspects: 1) pitch, 2) intonation, 3) speaking rate, 4) fluency, and 5) loudness. Pitch measures the mean frequency of acoustic signal; intonation measures the changing of frequency; speaking rate measures the acoustic signals produced in a fixed time duration; fluency measures the fractions between voice intervals; loudness measures the energy carried by voice.

A low-pitched voice is often perceived to be a strong and friendly attribute of the speaker, and thus generates more persuasive power. Intonation involves switching from one tone to another, e.g., from a high tone to a lower one. In English speech, switching to a lower tone often indicates that the statement is assertive, and switching to a higher tone often indicates that the statement is a question. Tone switching may also be accompanied

with emotional state changes, which in turn affect the listener's cognitive process [66]. Laplante et al. [54] also discussed voice tone, voice intensity, and a positive coorelation between changes in voice tone and changes in perception intensity. Laplante also observed that tone changes may indicate changes in perceived politeness. For example, switching to a high tone may be perceived as being more polite.

Changing one's rate of speaking is a well-known technique that can affect the persuasive effect of a speech. For example, increasing speech rate may be perceived as mastery of a topic at a high level of competency. Decreasing speech rate may be perceived as emphasizing the material. However, changing to a too-high rate of speaking may also damage the persuasive power of the speech by adding too much cognitive burden to the listeners [83]. Vaart et al. briefly discussed the relationships among persuasive power and vocal characteristics, e.g., voice loudness may positively affect persuasive power, but speaking rate and intonation may be negatively related to persuasion.

Vaart et al. cite Cialdini's six compliance principles [22], which are directly related to the persuasion process. Cialdini posited that people decide whether to comply with persuasion based on six principles: 1) reciprocation, 2) social validation, 3) consistency, 4) liking, 5) authority, and 6) scarcity. For example, the liking principle states that people are more likely to comply with people they like; the authority principle states that people are more likely to accept opinions from an authority source. For a detailed explanation of these principles, please refer to Cialdini's original work. Vaart et al. investigated whether the vocal characteristics of interviewers can affect the cooperation rate of participants. The work of Vaart is based on the liking and authority principles mentioned above.

Voice persuasion can be realized either by human or computer speech. However, the research literature seems to agree that human-recorded voice has greater persuasive effect than computer-synthesized voice. Stern et al. [86] studied this issue by differentiating the voice source as either human or computer. Stern et al. found that if listeners were told the speaker was a human, they preferred human voice to computer-synthesized voice. However, if listeners were told the speaker was a computer, the listeners did not show much bias. Stern et al. suggested this difference may be caused by the different expectations of listeners. Stern et al. also found using videotape as the source produced a greater persuasion effect than using only tape-recorded voices. Stern et al. believed that using videotape can promote experimental realism, which has been discussed in [79].

Linguistic styles can also affect the persuasion power of a speech. For example, hedges, hesitations, disclaimers can directly decrease persuasive effect. When the language contains too many hedging expressions such as "sort of", "maybe", and so on, a speech is often perceived as weak and lacking persuasive power. These commonsense usages of linguistic elements are consistent with the aforementioned Cialdini's six compliance principles, where authority and credibility of a message source can affect its power of persuasion.

2.8.5 Voice and Images

When producing persuasion through voices generated by computers (either pre-recorded human voice or synthesized voice using text-to-speech systems), a properly matched "talking head" can enhance the interaction between the computer and the user [39]. This may in turn increase the persuasive effect of the speech. Talking heads are a commonly used technique for creating effective virtual characters such as Embodied Conversational Agents (ECAs) [17]. Research in both human-computer interaction and virtual reality has investigated how to use talking heads to express complex feelings and emotions [6, 67].

A simple talking head may only involve animations being played during speech. Complex talking heads may include detailed emotional expression that matches the mood of the voice and text. For example, Poggi et al. [71] investigated the effects of using facial expressions in ECAs. Gong et al. [65] posited that using a matched voice and talking head delivers better results than using a high quality voice but a head that does not match. This indicates visual cues of a message can enhance the persuasive power of a voice message if these cues can be properly combined. Talking heads can be considered a persuasion technique inspired by human-to-human interaction. For example, when one person speaks to another, facial expressions and body language often carry emotions and emphasized messages, which are usually expected by the listeners due to the non-verbal cues. These types of cues, while are easily expressed by a human speaker, are often missing when a computer is the message source. Foster [39] discusses how representing an interactive system as a human face or body can enhance the interactions between a human and a computer. Kuhnel et al. [52] evaluated three talking-head systems together with two speech synthesis systems, and investigated how both heads and voices can influence smart home systems.

The human voice typically carries emotions. For example, when people are reading positive content, their rate of speaking tends to be rapid, their vocal pitch high. In con-

trast, when people are reading negative content, their rate and pitch tend to be slow and low, respectively. These changes can also be observed when reading happy or sad contents [64]. Nass et al. [64] observed that a happy voice can make the content being read sound happier, and that extroverts prefer a happy voice to a sad voice. A sad voice can even cause extroverts to lose interest in the content. Nass et al. also reported that when the emotion carried by the content matches the emotion of the voice, the appeal of the content is heightened. When the emotions do not match, the credibility of the content can be increased. The underlying theory in Nass et al.'s argument is consistency theory. Consistency theory [28] states that people tend to be consistent, and if inconsistency occurs, people tend to change inconsistency to be consistent by adapting their perceptions.

Some researchers argue that recorded speech differs from computer synthesized speech when emotion of speech is considered [64]. Even if synthesized speech can convey emotions, because a computer is known to be the source, people tend to interpret them as entities without emotion. As a result, synthesized speech may not influence people through the emotions it carries. However, other researchers found that computer-synthesized speech may produce similar effects to human interaction when emotion is considered [75].

Chapter 3

Methodology

3.1 Overview

The purpose of this project is to create a computer software system that can persuade people to be aware of their speaking rate and slow down their speech. The combination of computational design and persuasive technologies and theories are embedded in the system. In order to conduct effective persuasion, a number of computer-based survey questions are asked and a short tailored letter is generated for each participant. A virtual coach system monitors and reminds the participant to slow down. A few adaptive cues are used to enhance persuasion effects. To the best of our knowledge, very little work has been done so far in this area. Thus, one of our primary tasks is to lay out a framework which can serve as a foundation for our future research.

To evaluate the effectiveness of the instruments (i.e., computer-based survey, tailored letter, virtual coach, adaptive cues), a total of 22 respondents was selected to make up the sample. Each participant was asked to go through the following five-step experiment:

1. Answer a prepared set of computer-based survey questions in Likert format;
2. Read silently the computer-generated letter;
3. Read aloud one article without the intervention of the persuasive system;
4. Read aloud another article with the intervention of the persuasive system;

5. Discuss a set of feedback questions with the investigator.

Data collected from each instrument for each participant was computed and interpreted; feedback discussion was analyzed and interpreted. Primary data from the participants was mostly used. The design of survey questions involved the use of secondary resources such as published articles and advice from speech experts and user interface design experts. While investigating the effectiveness of the present system, discovering other possible approaches and theories was also an important goal during the evaluation experiments.

3.2 Research Design

A combination of descriptive and experimental methods was used for this research. Descriptive methods focus on the present conditions of the subject or events, with the main task being to observe the behaviour rather than change it. The focus of descriptive methods is on collecting information to describe and formulate hypotheses based on the existing conditions. We are in the initial stage of this new type of research (i.e., computer-based persuasion of speaking rate). Thus, descriptive methods are considered as flexible and cost-effective when constructing a foundation framework.

Experimental methods focus on intervention rather than only observing. Measurements are taken on subjects before and after the intervention. Experimental methods are used to evaluate the effectiveness of the framework for each subject. Compared to descriptive methods, experimental methods are more expensive and less flexible, but give more direct and measurable results about the framework.

In this research, we used both descriptive and experimental methods for different parts of the research. When constructing and evaluating the persuasiveness of the computer-based survey and tailored letter, descriptive methods were adopted to obtain first-hand data from the users. The flexibility of descriptive methods enables the rational recommendations from the experts and users to be reflected in the framework. When evaluating the effectiveness of the virtual coach and adaptive cues, experimental methods were adopted because they provide direct and objective indication about system performance. When analyzing the individual elements of the framework, descriptive methods were again used to enhance the performance of individual elements.

A qualitative approach was used in the present study. A qualitative approach focuses on analyzing the data collected in depth and richness, and aims at providing a holistic view of the study. It is not based on statistical analysis, but the interpretation of the researchers. The qualitative approach is considered a necessary method at certain stages in research. For example, when the research is in its early stage and the complete picture is not clear, qualitative methods help determine which parts are important and which parts are not. More importantly, qualitative methods can discover and formulate a rich set of hypotheses, which can be tested later using quantitative methods. The qualitative approach also has the advantage of being flexible. It is suitable for situations in which the research needs to be incrementally changed and refined. This flexibility also allows the application of inductive reasoning, which serves as a cornerstone in constructing new theories.

The quantitative approach, on the other hand, is based on statistical analysis of data. The aim of quantitative methods is to establish reliable relationships among variables by statistically compiling a large amount of data. This approach provides high reliability provided the sample size is large enough, and can construct solid relationships between variables using deductive reasoning. Thus, quantitative methods are often employed at the later stage of a research project when hypotheses are well-understood. If it is adopted in the early stages of the research, the cost involved will be very high.

In this research, only qualitative methods were used because we are still in the early stage of this research. The focus of the present research is to generate a complete picture about how to use computers to induce slower speech. Due to the limited knowledge available, the project is subject to being incrementally changed and refined as the study progresses. Once a solid framework and its related hypotheses have been constructed, quantitative methods will be employed to investigate the reliability of the framework.

The participatory design methodology is used to create the Graphical User Interface (GUI) of the software instruments. Participatory design methods invite participants to help improve the design of user interfaces. For example, at the beginning of the research, participants are invited to give their opinion on how to define the problem and solution; during the development, participants are invited to give advice; at the evaluation stage, participants are invited to evaluate the system and give feedback about the system. Participatory methods are based on community opinions, and are open to change during development. Their flexible nature fits well with the exploratory nature of our present research.

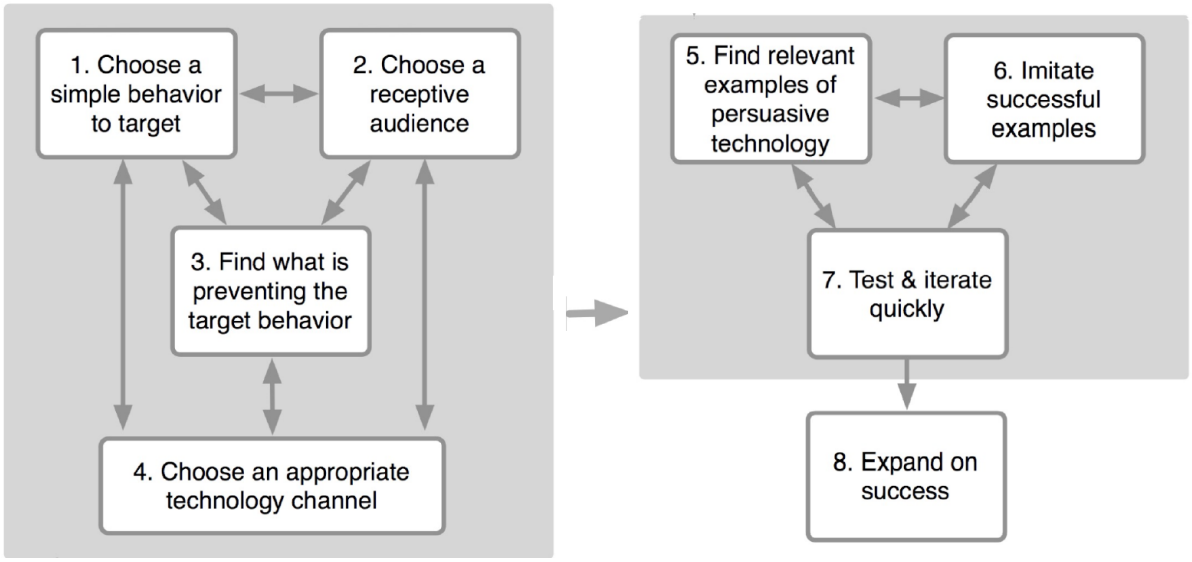


Figure 3.1: BJ Fogg's eight-step persuasive design (adapted from [33]).

Persuasive design is still in an early stage of research. Fogg [33] proposed an eight-step design process for creating effective persuasive systems. The process starts with finding a simple targeting behaviour. As Fogg states, "the design team should select the smallest, simplest behavior that matters." With a simple behaviour to change, the designers should also find a receptive audience rather than the toughest audience. It is also important to target a specific type of audience rather than all types. Next, the designers should figure out what is preventing the target behaviour (i.e., lack of motivation, lack of ability, lack of well-timed trigger). The designers also need to choose the proper technology channel for the audience. For example, Web information may work better for some audiences, but mobile texting may work better for other audiences. Then, the designers must search for relevant example systems, imitate successful examples, test and iterate, and expand the successful system. Figure 3.2 illustrates Fogg's eight-step persuasive design.

We adopted Fogg's eight-step design because we believe it is so far the most practical, cost-efficient, and complete method for designing persuasive systems, especially for academic research projects. Our targeting behaviour is to slow down the speaking rate of speakers. Although we do not have real restrictions on audience, we chose to begin our study with non-native English speakers. We use the theory of planned behaviour and relevant literature to figure out what may prevent people from slowing down their speech. The

technology channels we used include: computer-based survey, tailored letter generation, and virtual speaking rate coach.

Torning et al. [88] reviewed the state-of-the-art of persuasive system design and predicted future directions. A collection (see Table 3.1) of design principles such as tailoring, tunneling, and reduction were summarized in their paper. In our research, we carefully reviewed their summarized persuasion techniques proposed in recent conference papers, and selected the most relevant techniques: tailoring, suggestion, self-monitoring, rewards, reminder, personalization. For example, our system can generate tailored letters for speakers; within the letter, we used rhetoric-based suggestions. Our virtual coach can monitor the speaking rate of a speaker and give rewarding advice such as encouragement messages. If a speaker speaks too quickly, our system can send out voice reminders along with a set of adaptive cues.

Our work is considered as a branch of persuasive technologies, focusing on both attitude and behavioural change of speech behaviour. The above discussed methods are widely adopted by many persuasive systems. We chose these methods because we believe they are suitable and promising for our persuasion model.

3.3 Instruments

A survey, tailored letter, virtual coach, and adaptive cues are the persuasive instruments that will be used to induce slower speech. These instruments are all realized using computer technologies, which enable future large-scale deployment because of lower cost compared with human coaches. Survey and tailored letter are considered as pre-persuasion which targets a speaker's belief system. Survey questions are constructed in the Likert format. A computer-based algorithm is to used to analyze the survey answers and generate a tailored letter. Virtual coach and adaptive cues serve as the real-time intervention, which have immediate force on reducing speaking rate. The construction of survey questions is based on a well-known psychological theory (i.e., the theory of planned behaviour (see Section 2.8.1)). Tailored letters are generated based on an automatic analysis of a user's beliefs about slow speech.

Virtual coach and adaptive cues take advantage of computer technologies which can enable multiple interventions to happen at the same time. The virtual coach is based on the

Persuasion technique	Support dimension	Frequency
Tailoring	Primary task	11
Social comparison	Social	11
Tunneling	Primary task	10
Reduction	Primary task	10
Suggestion	Dialogue	9
Surface credibility	System credibility	8
Normative influence	Social	7
Self-monitoring	Primary task	6
Social learning	Social	6
Praise	Dialogue	5
Liking	Dialogue	5
Simulation	Primary task	4
Reminders	Dialogue	4
Authority	System credibility	4
Recognition	Social	4
Rewards	Dialogue	3
Similarity	Dialogue	3
Trustworthiness	System credibility	3
Cooperation	Social	3
Personalization	Primary task	2
Rehearsal	Primary task	2
Social role	Dialogue	2
Expertise	System credibility	2
Real-world feel	System credibility	2
3rd party endorsements	System credibility	2
Social facilitation	Social	1
Verifiability	System credibility	0
Competition	Social	0

Table 3.1: Design principles used in the persuasive conferences 2006-2008 (Torning et al. [88]).

assumption that computer-synthesized voice may produce better persuasion when coupled with computer-generated graphical facial expression. Voice reminders are intentionally embedded with different levels of emotions. Adaptive cues aim at adding extra persuasive power to the virtual coach.

3.4 Participants

To evaluate the effectiveness of the framework, a total of 22 participants were recruited through in-class recruitment and on-campus postings. Considering the scope of current research, the following inclusion criteria were imposed:

- The participants must be able to read English articles both silently and aloud.
- The participants must include both non-native and native English speakers.
- The participants must be young adults (i.e., between 20 and 35 years old).
- The participants must be familiar with simple computer operations.

Each participant received remuneration of $20 for participation. Withdrawal was without penalty. Participants were randomly divided into three groups (Group 1, Group 2, Group 3). A random selection procedure was used to ensure participants would have an equal chance to become part of a group. Lottery sampling was used for group selection. Each participant in the participation list was assigned a unique number. The numbers were then put into a box and mixed thoroughly. Each group was then blindly selected by the researcher from the box. This simple random sampling method has the following properties:

1. A total number of N participants were included in the population.
2. A sample group contained n participants.
3. All possible groups of n participants were equally likely to occur.

Twenty-two young adults (15 women and 7 men), aged 20–35, participated in this study. All participants are either current university students or university (college) graduates. Seventeen participants are non-native English speakers (first languages are Asian languages); five are native English speakers. All participants had sufficient computer skills to use the system.

Group A involved the testing of all persuasive elements (i.e., the virtual coach, emotional face, flying butterfly, dynamic speedometer). Group 2 involved the virtual coach, emotional face, flying butterfly. Group 3 involved the virtual coach and dynamic speedometer). The division of groups was to evaluate the effectiveness of individual persuasive elements. While one subject was doing experiments, other subjects were not allowed to observe the process.

3.5 Procedure

For each group of Group 1, Group 2, and Group 3, the following procedure was conducted:

1. Select a small group of subjects.
2. Prepare two different articles ($Article_1$ and $Article_2$) with similar difficulty levels.
3. Repeat (a)–(h) for each subject:
   (a) Ask a subject to answer a computer-based multi-choice survey.
   (b) The system will generate a tailored letter for the subject.
   (c) Instruct the subject to read and understand the tailored letter.
   (d) Instruct the subject to read $Article_1$ without the presence of the model.
   (e) Record the average speaking rate of the subject as $Rate_1$.
   (f) Instruct the subject to read $Article_2$ with the presence of the model.
   (g) Record the average speaking rate of the subject as $Rate_2$.
   (h) Interview the subject to collect feedback about the system.
4. Compare $Rate_1$ and $Rate_2$ for the subject.

5. Analyze the results of comparisons (collected from 4), and draw conclusions on the effectiveness of the model in reducing speaking rate.

6. Analyze the feedback (collected from 3h), and study the effectiveness of the model in raising awareness on speaking rate and other related aspects.

To determine the difficulty level of $Article_1$ and $Article_2$, 10 people were invited to read the two articles. The reading times for $Article_1$ and $Article_2$ used by each person were recorded as t_1^i and t_2^i, respectively, where i represents the identification of each person. The time difference is δ_i. The mean value of $\delta_i = |t_1^i - t_2^i|$ for all 10 people are 1 seconds, indicating that $Article_1$ and $Article_2$ are of the similar difficulty level.

The system was installed on a laptop, which was connected with the investigator's laptop through a wireless network. Subjects were not aware of this wireless connection. Subjects were told that the virtual coach system was an advanced intelligent system, which could hear and interpret their speech, in particular, speaking speed. They had to read aloud to the built-in speaker of the laptop so that the system could better catch the voice. The subject was seated inside an empty conference room with the laptop computer. The room door was nearly closed. The research investigator sat outside the conference room where he could hear the voice of the subject.

First, the subject answered 34 computer-based survey questions and read silently the automatically generated letter. Then, the subject read aloud one article shown in the system without any persuasive elements being presented. Next, the subject read aloud another article with the same level of difficulty while the persuasive elements were presented. During the experiment, the subject was told to remain seated until the end of the experiment and to follow the voice instructions of the intelligent system. All of the actions were conducted within the same GUI program. At the end of the experiment, each subject was interviewed by the investigator, and was asked about their experiences and comments about the system. Each subject took approximately 45 minutes to complete the experiment. Finally, after the overall experiments were finished, subjects were told that the intelligence part of the system was still being researched, and that it was the investigator who was remotely instructing the virtual coach to give advice.

3.6 Data Analysis

Data analysis was done through two steps: 1) computer-based automatic analysis and 2) human-based feedback interpretation. A computer-based algorithm analyzed the user's belief system about slow speech and constructed a tailored letter for the user. The core of the automatic analysis is the theory of planned behaviour. The data collected for each user, together with the feedback reports, was manually analyzed to draw conclusions about the effectiveness of the system. The feedback reports were analyzed in-depth (i.e., using qualitative approach) to both investigate current hypotheses and elicit possible hypotheses.

3.7 Ethical Considerations

Because the evaluation of the system involves human subjects, each step of our experiment was carefully designed and approved by the Office of Research Ethics at the University of Waterloo. The participants were informed of the purpose of the study, i.e., to create a computer software system that can persuade people to be aware of their speaking rate and slow down their speech. The participants were also informed about the potential benefits of the experimental results: this research may contribute to the knowledge base of Persuasive System research and lead to the development of improved usability and effectiveness of future persuasion and speech coaching systems. The data collected during interviews will contribute to a better understanding about which persuasive strategies and elements are feasible and effective for speaking rate adaptation. All of the data pertaining to participants are being kept confidential.

Chapter 4

System Architecture

4.1 System Overview

The overall system is composed of two sub-systems: 1) pre-persuasion system (PPS) and 2) real-time persuasion system (RTPS). PPS targets the belief systems of speakers while RTPS imposes real-time interventions on speakers. PPS and RTPS are responsible for different phases of persuasion and are based on different theories. PPS is based mainly on the Theory of Planned Behaviour, which can induce behavioural changes through manipulating the belief systems. RTPS is based on persuasive technologies, which apply real-time interventions through triggering systems.

The psychological root of TPB allows attitude changes to be made through persuasion, but an attitude change often take a longer time to happen even if it can happen. RTPS targets behavioural changes rather than attitude changes since changing attitudes is considered a much more challenging task than changing behaviours. In the psychological field, studying how to change attitudes has involved considerable amount of effort such as using the TPB theory. In the emerging field of persuasive technologies, most research aims at changing only behaviours, leaving changing attitudes as future work.

To take advantage of established research in both the psychological and persuasive technologies fields, our persuasive system integrates and extends the successful TPB theory, persuasive design theories, triggering theory, tailoring theory, natural language generation theory, and user interface design theories. Our system provides a computing framework

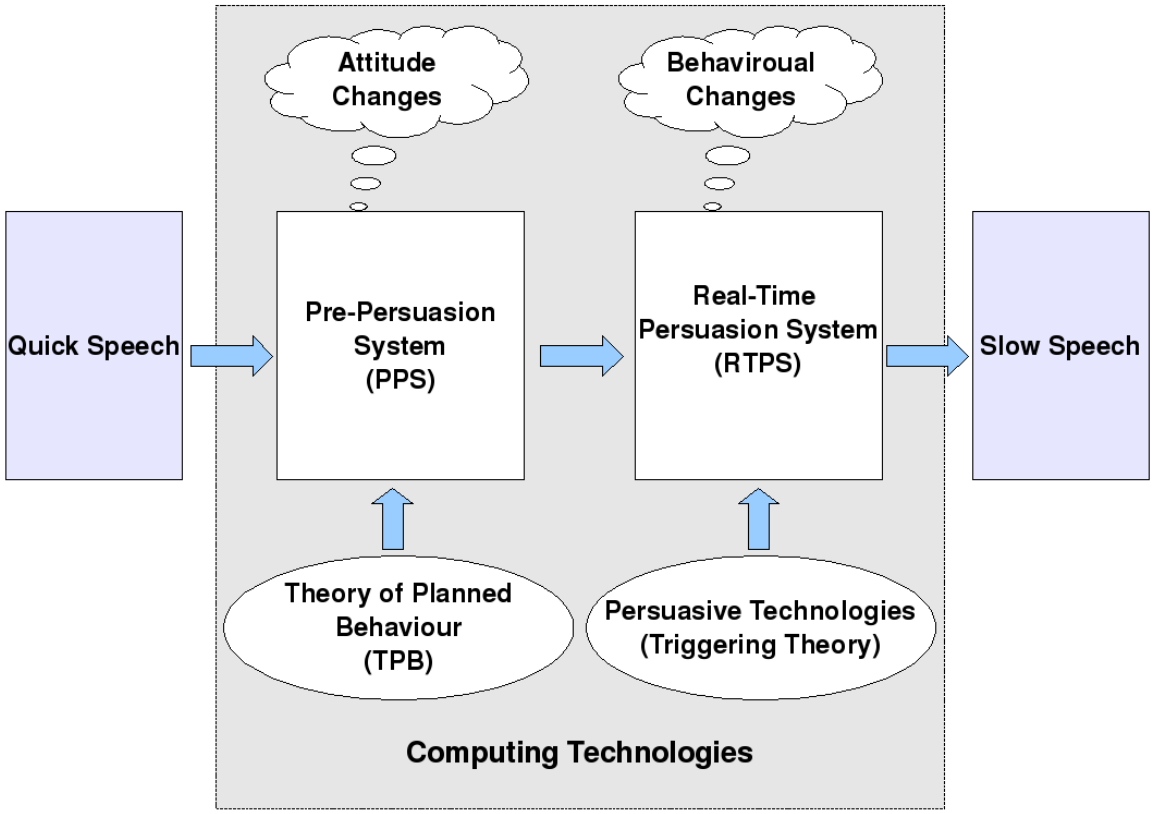


Figure 4.1: Overview of the System Architecture

that can persuade people to be aware of their speaking rate and slow down their speech. Each module of the framework will be the subject of further study and extension in future work. Figure 4.1 illustrates the overview of the system architecture.

4.2 System Conceptual Design

The conceptual model of the system serves as the theoretical foundation of our persuasion model. It is composed of the following components: Context Analyzer, Belief Analyzer, Rhetoric Support Module, Rhetoric Strategy Repository, Affect Support Module, Pre-Persuasion Generator, Speech Rate Detector, Triggering System, Persuasive Strategy Repository, Strategy Selector, and Real-time Intervention Generator. Figure 4.2 illustrates the architecture of the conceptual model.

Context Analyzer: Collects and analyzes the background about the speaker and the context of the speech activity. Context-awareness is a key concept of tailored persuasion. (see Section 2.6). A tailoring technology persuades users to change their attitudes or behaviours by presenting them with user-relevant information. Tailoring information to individuals has been studied, but tailoring information to context is still an open problem [32]. By taking into account the individual context of each speaker, the system is capable of accounting for the individuality of each user, and thus can be more accurate provided the context can be analyzed correctly.

Context analyzing is realized through user modeling. A user model collects and analyzes information about a user. For example, a user model may access a user's profile and preference settings, which form the context of a user. A user model may also keep track of the habitual speech behaviours of a user (e.g., a user may speak quickly under certain situations). A detailed user model may increase the accuracy of the system, but a complex model is often expensive to compute and maintain. At this stage of research, we only investigated preliminary usage of a context analyzer. In-depth investigation on the context analyzer needs to be done in the future.

Belief Analyzer: Collects and analyzes a user's belief system about speaking rate, particularly about slow speech. According to TPB (see Section 2.8.1, these beliefs eventually determine whether or not a speaker will take action to change their speaking rate. The following specific beliefs are collected and analyzed:

- Behavioural beliefs determined by:
  - Salient beliefs about how likely the behaviour (e.g., speaking slowly) will bring a positive or negative outcome;
  - The degree to which the outcome will be evaluated as positive or negative.
- Normative beliefs determined by:
  - Salient beliefs about social pressure from significant others on performing the behaviour (e.g., speaking slowly);
  - The motivation to comply with their opinions.
- Control beliefs determined by:
  - Salient beliefs about how often inhibiting factors will be encountered;

– How likely these inhibiting factors will be overcome.

Rhetoric Support Module: Selects and constructs text messages based on rhetorical models (see Section 2.4.1 for details about rhetorical models). Rhetoric studies how to use language effectively and persuasively. Although the use of rhetoric can be a powerful way to improve the persuasiveness of words, to be truly persuasive, the words must be accompanied by meaningful content. Merely persuasive but meaningless content should be avoided because of ethical issues. For the present work, the Rhetoric Support Module acts as a rhetorical strategy selector from the Strategy Repository. Rhetoric models and related features are proposed for future research.

Rhetoric Strategy Repository: Collects and stores rhetorical strategies, serving as a knowledge base for the Rhetoric Support Module. Commonly used strategies are stored in the repository, which can grow as research progresses, e.g., the following list contains some of the commonly used rhetorical strategies:

- Exemplification: Providing examples or cases in point.
- Description: Describing the sensory details of an event or object.
- Narration: Recounting a sequence of events.
- Process Analysis: Describing the process of how to do something.
- Comparison and Contrast: Comparing the similarities and differences.
- Division and classification: Decomposing the whole concept into parts or categories.
- Cause and Effect Analysis: Analyzing the reasons and consequences of specific events.
- Repetition: Repeating a statement or question to emphasize a point.
- Question: Raising a question without a direct answer to provoke and emphasize a point.

Affect Support Module: Collects affect models and strategies and applies them to the persuasion process (see Section 2.5 for how emotions affect persuasion). With proper strategies, different emotions can be embedded into persuasive messages (e.g., text or voice), inducing the audience to switch to a different mental state so that they can be

persuaded to change their rate of speaking. The Affect Support Module may need to work with the Context Analyzer and Belief Analyzer to produce effective persuasion because the dynamic nature of emotions may be easily affected by the speaker's current environmental conditions and relevant beliefs. One-dimensional models simply represent emotions as opposite-emotion pairs, such as happy-sad. One direct benefit of using such a simple representation is reliability if the actual subject is simple. The most commonly used affect models are two-dimensional models, which can analyze and reduce a large set of affective states into a small set which contains only two essential factors. The selection of a proper affect model will become increasingly important as research progresses depending on the emotional nature of the subjects being persuaded.

Pre-Persuasion Generator: Takes input from Context Analyzer, Belief Analyzer, Rhetoric Support Module, Rhetoric Strategy Repository, Affect Support Module, and performs pre-persuasion actions (i.e., generating tailored letter for each speaker). Specific algorithms need to be developed to facilitate the generation process.

Speech Rate Detector: Detects the speaking rate of the persuadee. The system is designed to persuade speakers to change their rate of speaking (see Section 2.7 for why speech rate matters). Speech Rate can be measured in different ways with two most commonly used methods: syllable rate (SR) and articulation rate (AR). Please refer to [14] for a detailed discussion about measuring speech rate. SR counts pauses as part of speech rate, and AR does not include pauses. As the names of SR and AR indicate, SR refers to the syllables produced per second, and AR refers to the movement rate of the physical articulators. As research progresses, both SR and AR methods should be investigated to obtain a comprehensive understanding of speech rate measurement. Another important future task is to automate the speech rate measuring process. For example, Jong et al. [24] discuss a method to measure speech rate by automatically detecting syllable nuclei.

Triggering System: On detecting speech rate changes, the Triggering System should be capable of making quick decisions on whether the system should invoke an intervention and what should be invoked. The most important property of the triggering system is responsiveness. As research progresses, complex hardware and software logic might be introduced. The triggering system should remain highly responsive

and accurate without being overly affected by complex logic.

Persuasive Strategy Repository: Stores a set of persuasive strategies for speech rate adaptation. These strategies include text messages, multimedia rhetorical materials, etc. Persuasive strategies are considered a cornerstone of the system. In addition to the techniques listed in Table 3.1, Cheng [21] lists a set of additional persuasive strategies. As research progresses, the strategy repository should keep growing. At certain points, the repository should be refined and similar strategies should be combined to improve the efficiency of the system.

Strategy Selector: Selects the proper persuasive strategies that fit the targeting behaviour type. Fogg [35] proposed a tool called the Behavior Grid, which categorizes behaviours into a grid framework. The grid framework considers two general aspects of behaviours: 1) What type of behaviour change? and 2) On what schedule? Fogg's Behaviour Grid provides a underlying framework for categorizing behaviours. However, a more intelligent analyzing module needs to to developed to fit in the Strategy Selector.

Real-time Intervention Generator: Realizes the selected persuasive strategies for a persuadee, and delivers them to the persuadee. This module takes input from the Triggering System, Affect Support Module, and Strategy Selector to produce a real-time intervention for the user. The dynamic nature and time-critical nature of this module require it to be built on top of a highly automated computing framework.

4.3 Software Architecture

To realize and evaluate the persuasive framework, we designed and implemented a software system that contains two core sub-systems: Persuasive Survey System (PSS) and Persuasive Coach System (PCS). PSS and PCS correspond to the pre-persuasion system (PPS) and real-time persuasion system (RTPS) (see Section 4.1), respectively. PSS instructs and collects a user's beliefs about speech rate, analyzes the collected beliefs, and generates a persuasive letter tailored to the user. PCS provides an interactive platform on which a user can read articles while being coached on speech rate adaptation.

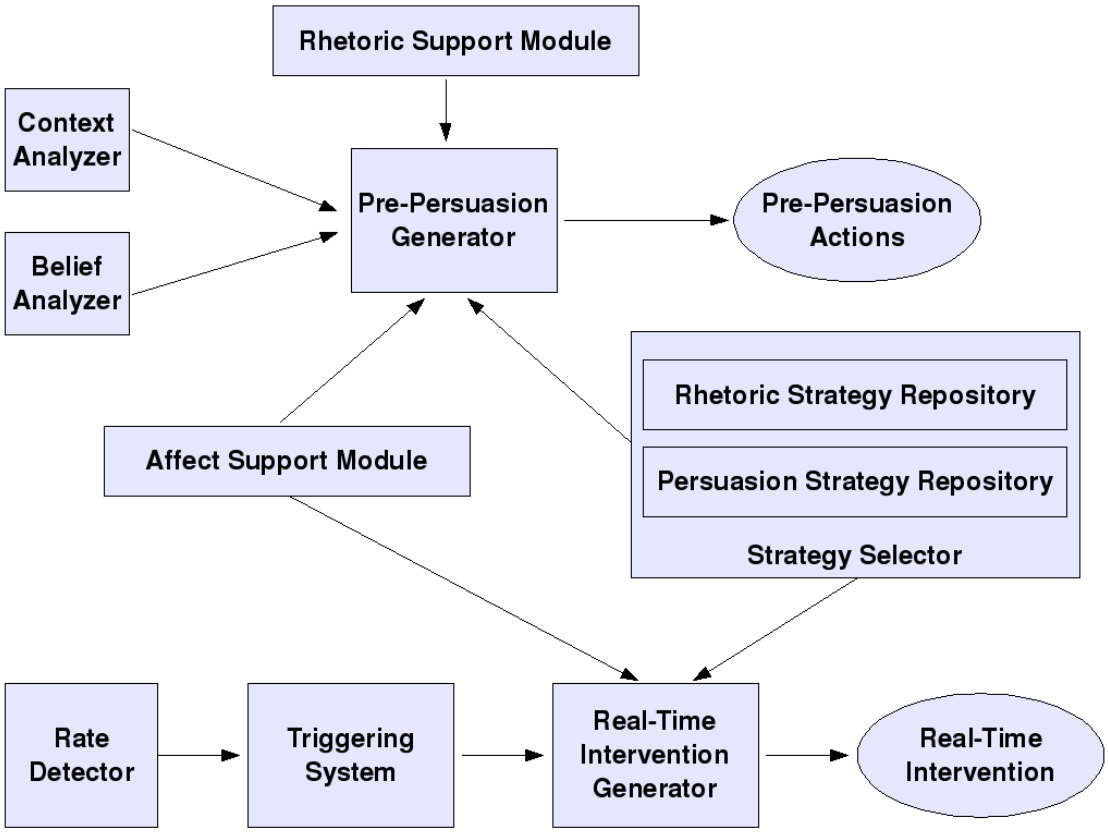


Figure 4.2: Architecture of the Conceptual Model.

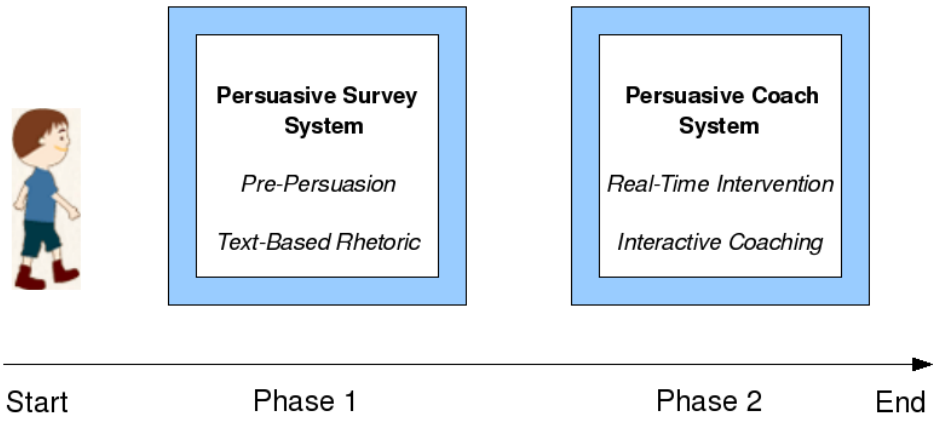


Figure 4.3: Application sequence of PSS and PCS.

PSS focuses on persuasion through text-based rhetoric (i.e., the tailored letter), and PCS utilizes multimedia-based persuasion including text, voice, animation, and adaptive cues. Although PSS is designed to be presented to a user before PCS, PSS and PCS are capable of being used as separate tools when needed. For example, if a user prefers to only read a text, the user may choose PSS; if a user prefers to use interactive software, the user may choose PCS. The present research requires users to use both PSS and PCS in sequence to evaluate their persuasiveness. Figure 4.3 illustrates the application sequence of PSS and PCS. The following sections describes PSS and PCS in detail.

4.3.1 Persuasive Survey System

PSS consists of seven modules: Survey Interface, Survey Engine, XML Source, XML Database, Text Selection Engine, Rhetoric Knowledge Base, and Tailored Letter Generator. Our implementation of Survey Interface and Survey Engine extended the open source project, Java Computer Aided Interviewing Framework (JCaiF) [2], which contains a set of Java interfaces defining commonly used survey elements. We also adopted some features from the open source project Sonar [3], which is a survey administration engine on top of JCaiF. Figure 4.4 illustrates the structure of PSS.

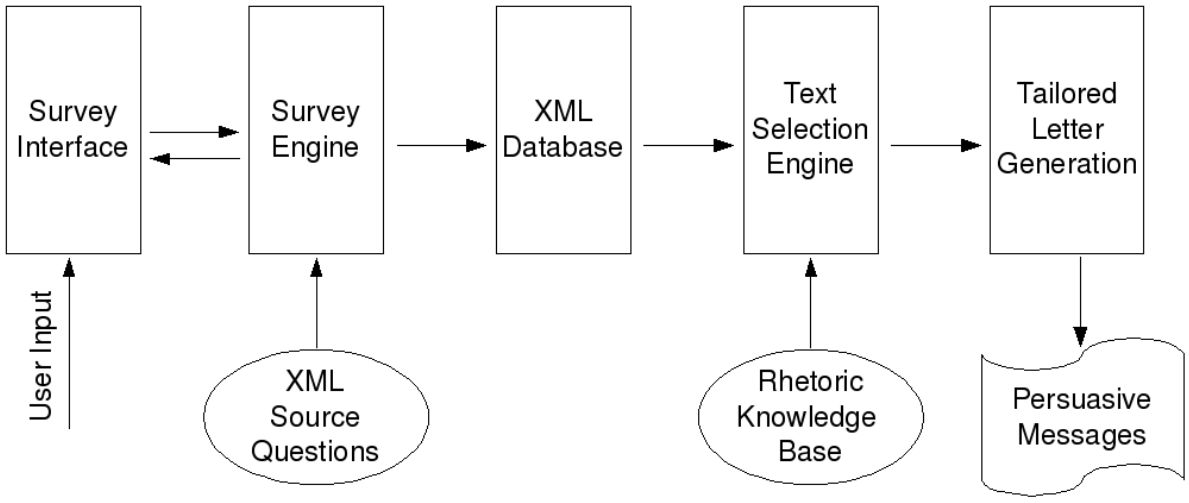


Figure 4.4: The Structure of PSS.

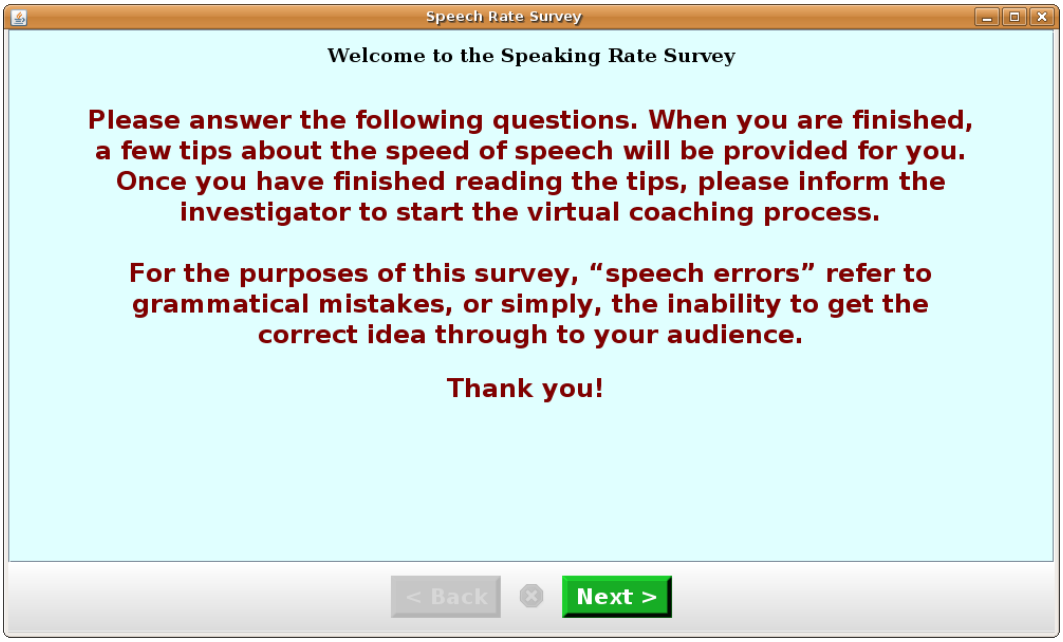
Survey Interface and Survey Engine

The Survey Interface and Survey Engine together provide a computer-based survey interface from which users' beliefs can be collected. The Survey Interface reads survey questions from XML Source, and renders and presents the questions to the user. The Survey Engine is responsible for taking care of the user's answers and store them in a XML database. The Survey Engine also guides the user by presenting proper instructions. Figures 4.5(a) and 4.5(b) show the examples of an instruction and a survey question, respectively.

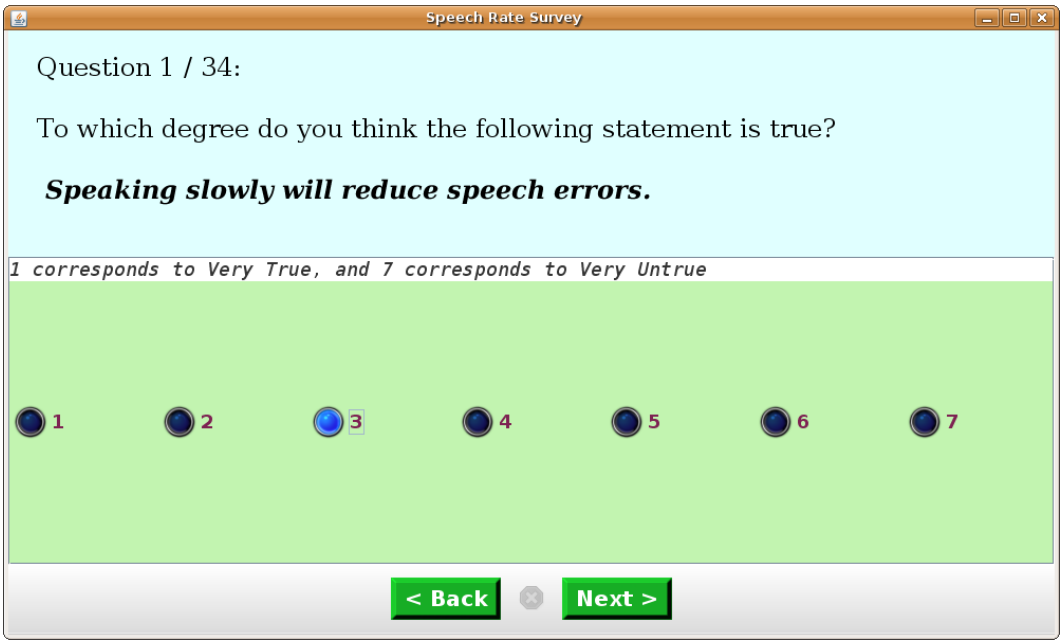
The survey questions and user answers are represented using XML format, which is consistent with the industrial and Sonar's standards. There are a total of 34 questions, designed based on the Theory of Planned Behaviour, which covers a user's behavioural beliefs, normative beliefs, and control beliefs about slow speech. See Appendix A for the complete list of the survey questions.

Text Selection Engine and Rhetoric Knowledge Base

Text Selection Engine reads the answers entered by a user from the XML database, and identifies the belief patterns of the user in order to generate a set of tailored persuasive messages for the user. The Rhetoric Knowledge Base stores a set of rhetorical strategies such as description, repetition, and question (see Section 4.2 for details about rhetorical



(a) A sample survey instruction.



(b) A sample survey question.

Figure 4.5: An example survey.

strategies).

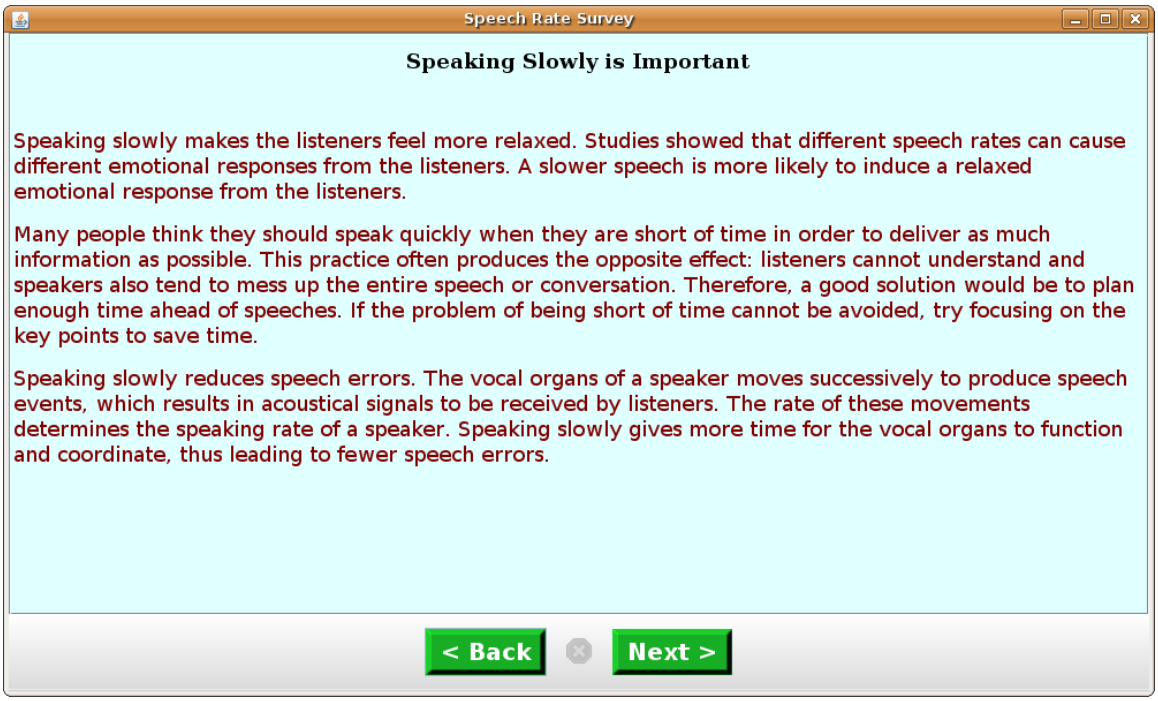
Considering that PSS only serves as the first phase of persuasion, we need to limit the number of messages selected for a user. Otherwise, the user may suffer cognitive demand issues when entering the phase immediately following (i.e., Persuasive Coaching System). The following algorithm is used to select a small number of persuasive messages that can best fit the needs of a user:

Listing 4.1: "Select and rank the best messages."

```
Step 1: Retrieve the answers of a user from the XML
        database. Let v_i represent the value of
        each answer, N represent the total number
        of answers, and M represent the total number
        of messages available.

Step 2: Categorize both the question and answer values
        into groups G_j, where j ∈ {behavioural, normative, control}.

Step 3: Calculate the group score S_j for each group using
              S_j = ( Σ_{v_i ∈ G_j} v_i)/N.

Step 4: Calculate the weight of each group w_j using
              w_j = Sj/(S_behavioural + S_normative + S_control).

Step 5: Select w_j · M messages from each group G_j.

Step 6: Sort the selected messages using its associated v_i.
```

Tailored Letter Generator

Tailored Letter Generator takes affective strategies and embeds them into the messages to be generated. The generated messages are split into several pages with each page containing three or four messages. Figure 4.6 shows an example of a generated tailored letter (the maximum number of messages was set to 6).



(a) Page 1 of a tailored letter.

(b) Page 2 of a tailored letter.

Figure 4.6: An example of a tailored letter.

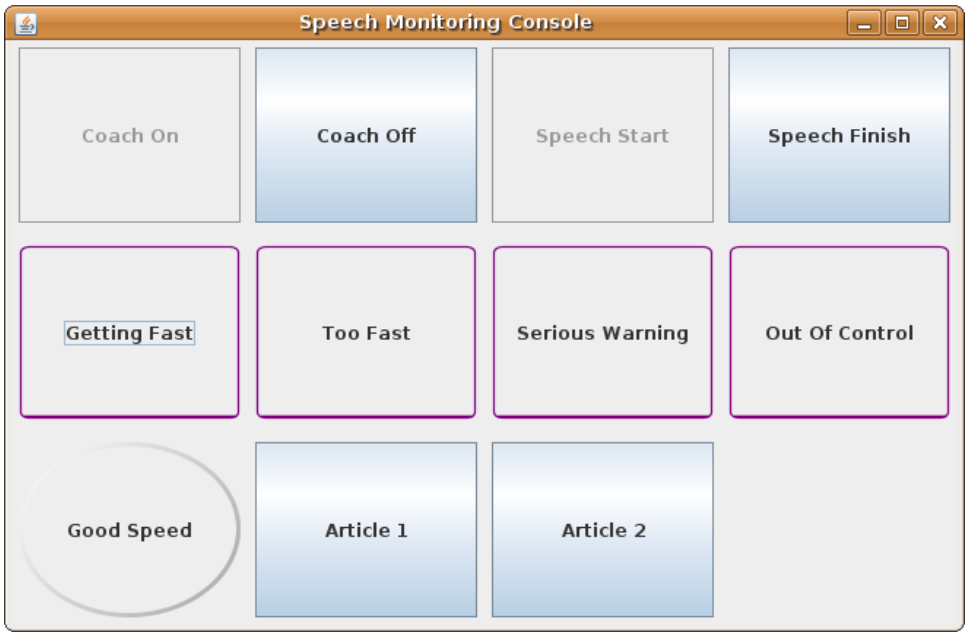


Figure 4.7: A snapshot of the Speech Control Console.

4.3.2 Persuasive Coach System

PCS is designed to realize the real-time intervention feature of the overall persuasive framework. This real-time, direct persuasion approach is complementary to the belief-based persuasion approach used by PSS. The objectives of PSS are to at least raise a speaker's awareness on slow speech and to rhetorically enforce the speaker's attitude towards slow speech, establishing a cognitive basis for the subsequent PCS. PCS is centred around a virtual coach which can provide voice reminders, accompanied with a number of adaptive cues. To be specific, as shown in Figure 4.8, PCS consists of the following elements: a talking head, a speedometer, an emotional icon, a flying butterfly, and a document reader. The present research uses the Wizard-of-Oz design which depends on a human monitor who remotely detects and sends advice to the virtual coach. A human monitor uses a separate computer to operate a control console as shown in Figure 4.7. The control console communicates with the virtual coach through TCP/IP messages. In the future, an independent speech rate detector and an artificial intelligent inferencing system need to be developed.

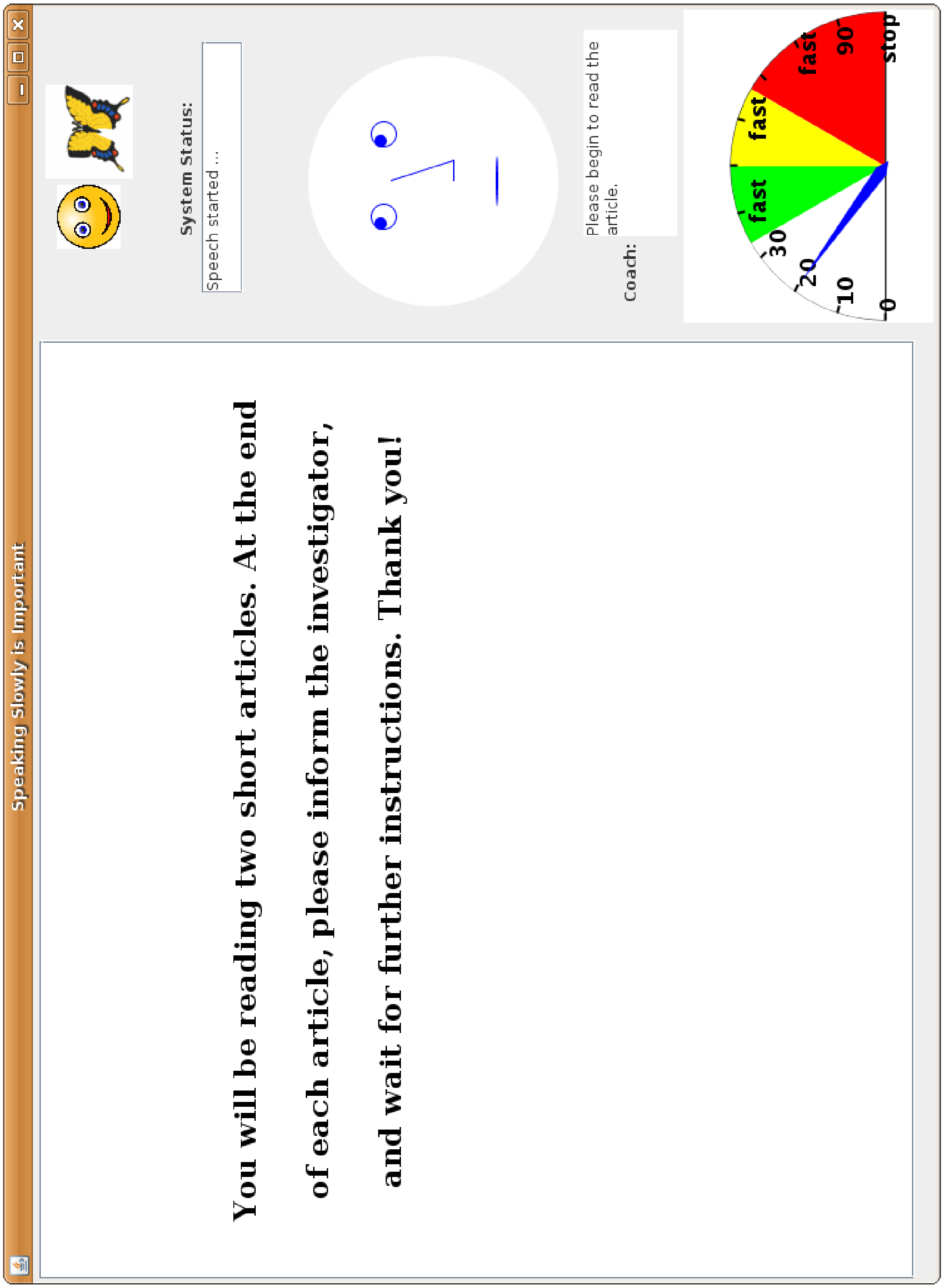


Figure 4.8: A snapshot of PCS.

Virtual Coach

The Virtual Coach is composed of a Voice Generation Component (VGC) and a Talking Face Component (TFC). Both VGC and TFC are solely based on computer synthesis techniques, i.e., VGC synthesizes voice advice using the Text-To-Speech (TTS) technology and TFC produces facial expressions based on pre-defined algorithms that consider the content of the speech. For example, when the virtual coach is saying a word "hello", the VGC component synthesizes the sound based on the phonemes, which are indivisible units of sound in a given language. At the same time, TFC generates the corresponding facial movements based on the same phonemes. For example, the width and height of the mouth can change dynamically according to the current phoneme. We adopted the basic techniques used in the Talking Head field [10].

We chose to design both VGC and TFC using computer synthesis techniques because we plan to test the following hypothesis: Can a consistent virtual voice and virtual face produce enough persuasiveness in the context of speech coaching? The voice is generated using TTS engine. The present research embeds emotions in the voice using rhetorical techniques, i.e., manipulating the advice content. For example, if the speaker speaks a little fast, the voice will be "Please speak slower." If the speaker keeps getting faster, the voice will carry some emotion and say "Please slow down. Your speech is too fast." If the speaker still does not slow down, a voice with a strong emotion (warning) will be produced: "Please slow down immediately, or the system will be locked." In the worst-case scenario, the system will be locked and the voice will be "Your speech is too fast. The system has been locked, and will restart shortly." If the speaker slows down, an encouraging voice will be "Good job! You have excellent speech control." We also plan to investigate the following hypothesis: Can a virtual voice and virtual face achieve emotional effects without annoying the speaker?

In the future, we also plan to change the attributes of the TTS-generated voice so that it can carry more dynamic emotions. For example, by dynamically changing the speech rate, pitch, range, and volume attributes, more complex and dynamic emotions may be generated. Figure 4.9 illustrates the architecture of a complete virtual coach. Scripts are first processed by the TTS engine. The Emotional Enhancement Module (EEM) can adjust voice attributes such as speech rate, pitch, volume, and range while rhetorical techniques can be applied. The output of EEM is then split into VGC and TFC, each of them

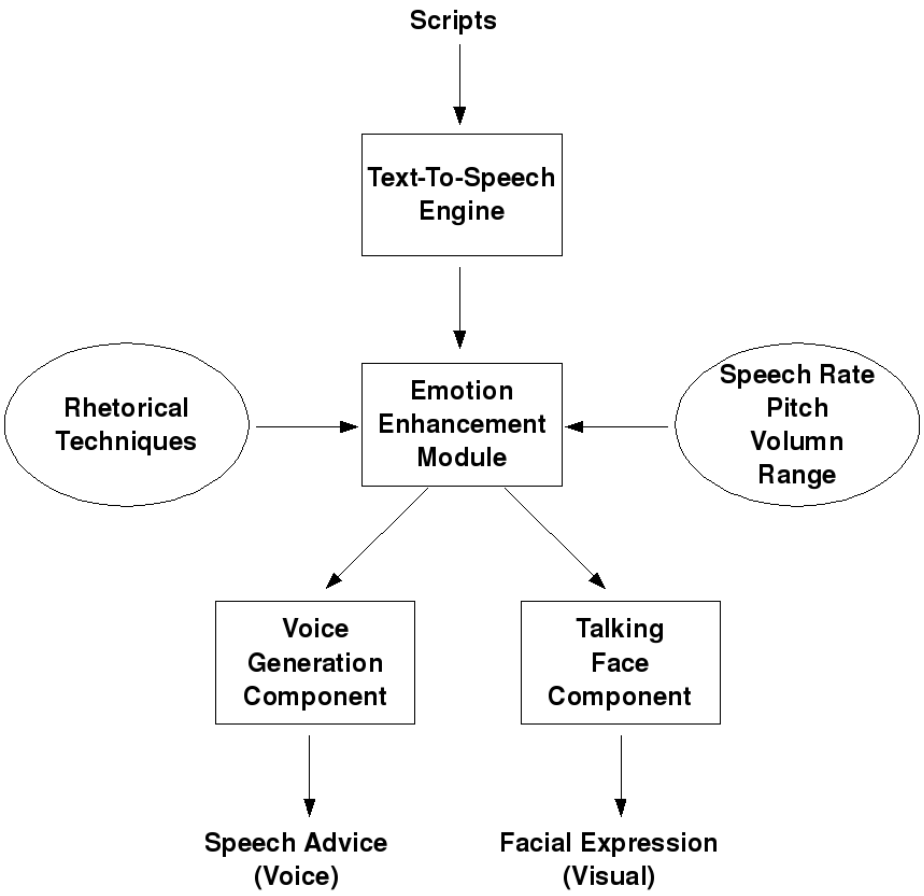


Figure 4.9: The Architecture of virtual coach.

generating coordinating actions (i.e., voice advice from VGC and facial expression from TFC).

Adaptive Cues

In addition to the virtual coach, we also designed visual adaptive cues (i.e., text colour, emotional face, flying butterfly, speedometer, as shown in Figure 4.8) to enhance the persuasiveness of the system. The emotional face can change to be happy, upset, frowning, or angry, based on the current state of the speaker. Similarly, the flying bufferfly changes its flying speed and the speedometer points to different speed zones according to the speech rate of the speaker. The text colour will change from black to red if the speaker is speaking

quickly. Computer technology enables the dynamic changes of these visual cues and the virtual coach at the same pace.

Chapter 5

Experimental Results

Our general methodology has been discussed in Chapter 3. This chapter discusses the experimental results. The primary task is to evaluate the feasibility and effectiveness of the overall system. At the same time, we evaluate the feasibility of individual elements. A total of 22 participants was selected to make up the sample. The experiments were conducted under controlled conditions. The results indicated that our system is effective in persuading people to speak slower. The feedback from users indicated that our system raised their awareness about speaking rate. We also discuss the feedback of users about the system in detail.

This experiment is the first step in investigating how to persuade people to slow down their speech using computer systems. Speech in the context of conversation is a two-way communication, involving voice interaction between two or more speakers. For public speech, often only one speaker is speaking while others are listening. These types of speeches can be affected by many known and unknown factors, which are beyond the scope of our current research. We reduce the form of speech to a simple and well-controlled type: reading articles aloud.

Reading aloud is a basic type of speech. It can be considered a one-way communication. This one-way speech is the basis for more complex forms of speech such as conversation and public speaking. A solid understanding of how to control speech rate during reading will serve as a cornerstone for understanding complex speeches. Thus, our experiments are designed to evaluate how effective the system is at slowing down a user's reading speed. The general evaluation procedure has been discussed in Section 3.5.

5.1 Results

The results shown in Figure 5.1 indicated that all of the 22 subjects slowed down their reading speed when all or parts of the persuasive elements were presented compared to when persuasive elements were completely absent. For example, participants 18 and 19 slowed down from 643 to 382 seconds and from 646 to 426 seconds, respectively. From our experiments, we observed that the absolute reading speed of the 22 subjects varied significantly: for reading article 1, the mean and standard deviation are $\mu_1 = 372$ and $\sigma_1 = 73$, respectively; for reading article 2, the mean and standard deviation are $\mu_2 = 477$ and $\sigma_2 = 87$, respectively.

We believe that the significant standard derivation may be partly caused by varying proficiency levels in English because the 17 non-native English speakers have much larger values in both means and standard deviations for reading both articles than that of the five English speakers. Table 5.1 shows the relevant statistical summary. The present research focused on studying non-native English speakers who are native Asian language speakers. At the same time, we collected preliminary data for native English speakers. Although the overall sample size is small (only 22), the completely positive results from both native and non-native English speakers still serve as a strong indicator for the effectiveness of our system.

Because the absolute reading times (in seconds) varied significantly, we chose the speed percentage changes between two articles for each subject as the metric. Figure 5.1 shows the speed percentage changes for all of the 22 subjects. Table 5.2 shows the statistical summary of the percentage changes for Group 1, Group 2, and Group 3. Group 1 was presented with all of the persuasive elements; Group 2 was presented with the virtual coach, emotional icon, and flying butterfly; Group 3 was presented with the virtual coach and dynamic speedometer (see Section 3.5 for detailed experimental design). Overall subjects slowed down their reading speed by 30%. Surprisingly, the system had less effect on Group 1 than on Group 2 and Group 3. Group 1 was presented with all the persuasive elements (i.e., the sum of the elements presented for Group 2 and Group 3). One possible reason may be that our sample size was too small. However, based on the users' feedback, we conclude that the cognitive demand of dealing with additional persuasive elements may actually reduce the persuasiveness of the system. This hypothesis needs to be further investigated using quantatitive methods in future research.

	Article 1		Article 2	
	mean (μ_1)	SD (σ_1)	mean (μ_2)	SD (σ_2)
All subjects	372	73	477	87
Non-native English speakers	403	50	506	76
Native English speakers	267	25	376	23

Table 5.1: Statistical summary of reading times used by participants. Unit: seconds. SD represents standard deviation.

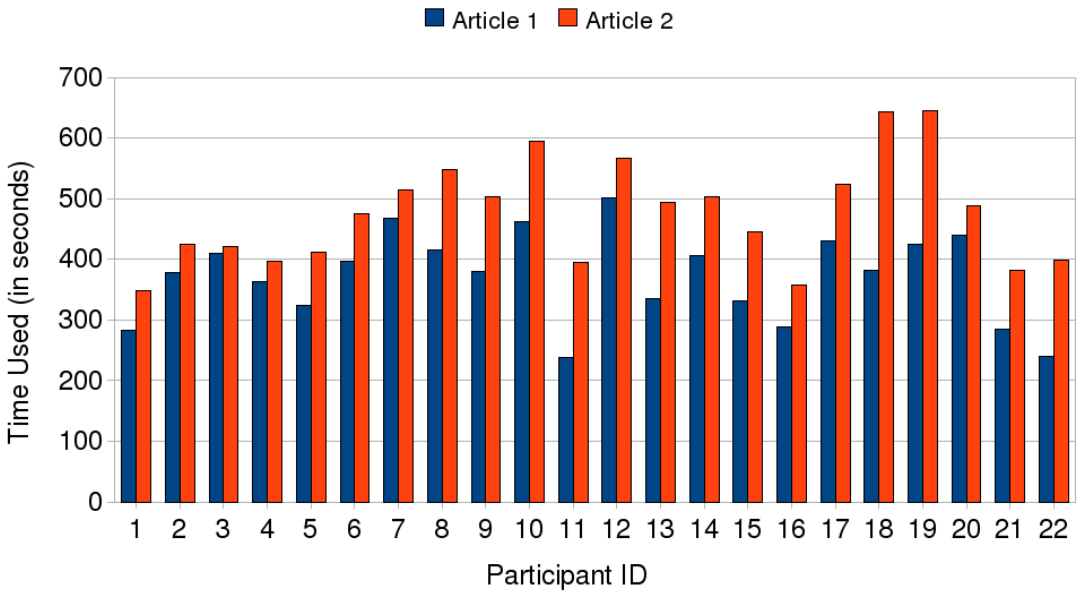


Figure 5.1: Time used for 22 participants (reading two articles).

	Reading Speed Changes	
	mean (μ_1)	SD (σ_1)
All subjects	-30%	-19%
Group 1	-20%	-11%
Group 2	-35%	-19%
Group 3	-42%	-24%

Table 5.2: Statistical summary of reading speed percentage changes.("$-$" means slowed down)

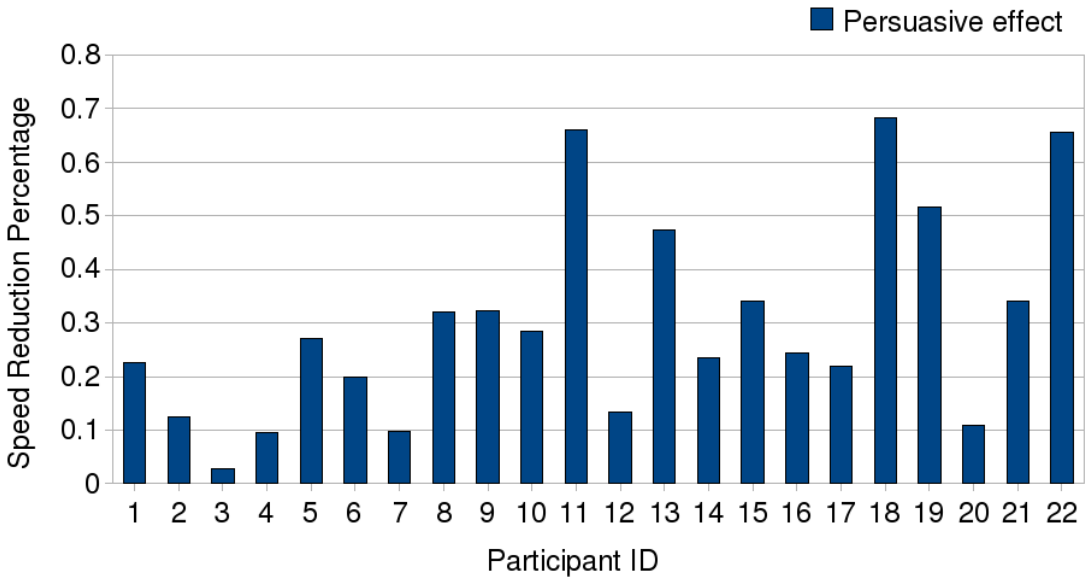


Figure 5.2: Speed reduction effect on 22 participants.

Figure 5.3 shows the box plots [61] for $Article_1$ (control group) and $Article_2$ (experimental group). The results indicated that the Experimental Group slowed down their speech significantly.

5.2 Discussion and Implications

Direct experimental comparisons of reading speed with and without persuasive intervention have indicated that our system is feasible and effective in inducing slower reading speed. This persuasive effect applies to both native and non-native English speakers, although their absolute reading speed may vary significantly. The present experiments used a small sample size and were aimed at testing the feasibility of our approach while discovering possible hypotheses for future studies. Some of the newly discovered hypotheses need to be further investigated using quantitative methods; others may still need to be refined before using costly quantitative methods. In addition to the new hypotheses derived from the experimental results, a number of new hypotheses were also discovered from the feedback and comments from the subjects and the observations of the research investigator.

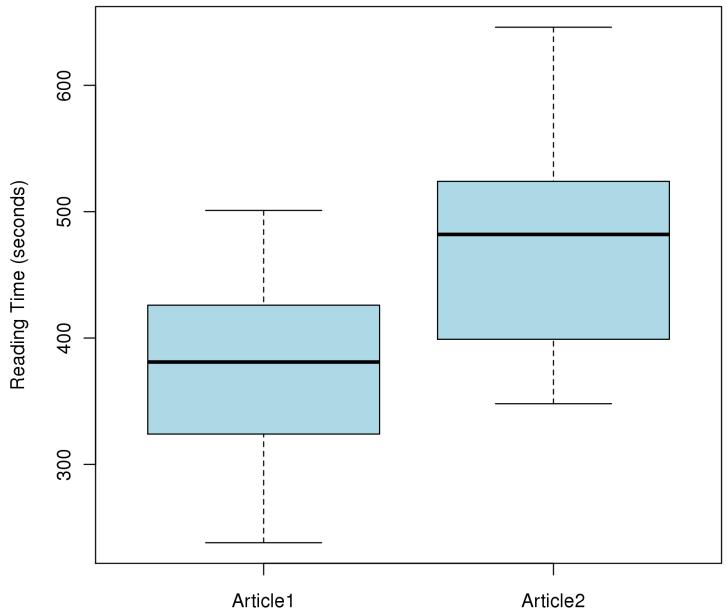


Figure 5.3: Box plots of the control group ($Article_1$) and experimental group ($Article_2$).

The experimental results in Section 5.1 have indicated that our system is effective in persuading people to slow down their reading speed. To have a deeper understanding of the experiences of individual subjects, an interview was conducted for each subject immediately following the experiment. All of the subjects stated that the system was effective in slowing down their speaking rate. The following are the key points summarized from the interviews:

- 22 out of 22 subjects thought the overall system was persuasive and effective.
- 20 out of 22 subjects mentioned that the survey questions raised their awareness about speech rate. Two subjects mentioned that they had some previous knowledge about speech rate from their previous presentation-related training. However, these two subjects agreed that the survey questions helped them to have a much deeper understanding about the effect of slow speech.
- Subjects felt the following survey questions had more persuasive power:
  1. To which extent do the following people want you to speak slowly?
     - Your direct family (excluding spouse).
     - Your spouse or life partner.
     - Your close friends.
  2. To what degree do you think the following statement is true?
     - Speaking slowly will convey less information to the audience.
     - Speaking slowly will allow people to understand me better.
     - Speaking slowly will make me appear more confident.
     - Speaking slowly will reduce chances of people asking me to repeat.
  3. How likely can you keep a slow speaking speed in the following situation?
     - While being short of time.
     - While in an exciting or stressful situation.
     - While people around me are speaking quickly.
     - While I am very familiar with my speech topic.

- Ten out of 22 subjects thought the tailored letter was an effective method to help them to slow down their speech. Five subjects thought the tailored letter might have some effects on their speech control. Seven subjects thought the tailored letter did not have any effect at all.

- All 22 subjects ranked the voice coach as the most effective element that persuaded them to slow down their speech. Eighteen subjects noticed the text font colour changes.

- In Group 1: Nine out of 10 subjects mentioned that they did have time to observe the visual persuasive elements. Three subjects mentioned that they noticed the emotional icon changes. None of the 10 subjects noticed the speed change of the butterfly. Two subjects noticed the speedometer changes.

- In Group 2: Four out of six subjects noticed the emotional icon changes. None of the subjects noticed the speed changes of the flying butterfly.

- In Group 3: Five of the six subjects noticed the speedometer changes.

- Twenty out of 22 subjects mentioned that they would prefer a virtual coach to a human coach because they thought a virtual coach was more subjective, more reliable, less emotional, less expensive, and caused less pressure. Two out of 22 mentioned that they would prefer a human coach because a human is more considerate.

- Fifteen out of 22 subjects mentioned that they noticed the emotion changes of the voice.

- Ten out of 22 subjects mentioned that the encouraging voice was effective. Seven subjects mentioned that warning messages were effective. Three subjects mentioned that they would prefer to hear a real human voice rather than computer synthesized voice. Three subjects mentioned that the repeated encouraging messages were annoying but acceptable. Three subjects felt frustrated because they thought the virtual coach required them to speak too slowly.

Based on our experimental results and the feedback and comments from the subjects, we believe that our current approach is feasible and our current system is effective. From our observation and the feedback of the subjects, we found that 1) emotions appear to have

a very important impact on speech rate persuasion, 2) cognitive demand plays an important role in effective persuasion, 3) the effect of the tailored letter is not that obvious, 4) the Theory of Planned Behaviour is effective in raising speakers' awareness on speaking rate, 5) real-time intervention is a powerful and effective method for slowing down speaking rate, and 6) our computational framework is feasible and effective. Each of the above findings needs to be further investigated in future research. The effect of the tailored letter had been expected to be higher than the actual feedback from the subjects, although subjects all agreed that the letter did raised their awareness on speaking rate. We believe this unexpected result may have two reasons: 1) the tailored letter needs to be redesigned and to consider other unknown design factors, or 2) our sample size was too small.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

Proper speaking rate is a key attribute of effective communication. It is not uncommon to observe people either speaking too quickly or too slowly in both professional and personal situations. The consequences of using improper speaking rate are obvious and often serious. For example, too rapid speech may cause misunderstandings in a teamwork environment, reducing the teamwork quality and efficiency. However, persuading people to change their rate of speech can be challenging for the persuader and frustrating for the speaker. A speaker's rate of talking is a habitual behaviour. Underlying a habitual behaviour is an individual belief system that supports it. Without a proper strategy for modifying behaviour, a persuader may end up with ineffective attempts to persuade. The person being persuaded may experience frustration because of failure to change.

Although persuading someone to alter their speaking rate can be very challenging because this is a habitual behaviour, we believe that this problem can be solved using scientific methods drawn from multiple disciplines: computer science, philosophy, psychology, physiology, rhetoric, and sociology. Computer technologies provide a promising solution to extend traditional human coaching methods. Emerging persuasive technologies use computers as a tool to induce human behavioural and attitude changes. The ultimate goal of this research is to establish a computational framework which can persuade people to communicate the most effectively. Specifically, we focus on how to persuade people to adjust

their speaking rate to a "just right" point at which the most effective communication can occur. To the best of our knowledge, there has been little research work to date on this topic. The fundamental problem we are solving is to define this new area and evaluate the feasibility of our approach.

Our literature search indicated that speaking rate did matter in many situations. However, how to persuade people to change their rate of speaking is still very much an open area to be investigated, especially using the newly created persuasive technologies such as captology. The literature review indicated that neither too-slow nor too-rapid speech are ideal behaviours. However, this "just-right" rate is very difficult (if not impossible) to achieve without the ability to persuade a too-rapid speaker to slow down and persuade a too-slow speaker to speed up.

Behavioural theories explain the underlying reasons for a human to conduct a certain behaviour. A understanding of these theories is fundamental for developing a persuasion model, which aims at changing a human's specific behaviour. We reviewed a number of behaviour theories including the Theory of Planned Behaviour, Fogg Behaviour Model, and multimedia behavioural theories.

On the basis of the literature review, we defined a conceptual model and implemented a computer software system, both serving as the cornerstones of our persuasion framework. The conceptual model served as the theoretical foundation of our persuasion model. It is composed of the following components: Context Analyzer, Belief Analyzer, Rhetorical Support Module, Rhetorical Strategy Repository, Affect Support Module, Pre-Persuasion Generator, Speech Rate Detector, Triggering System, Persuasive Strategy Repository, Strategy Selector, and Real-time Intervention Generator. The computer system is designed to persuade people to be aware of their speaking rate and to slow down their speech. The combination of computer technology and persuasive technologies and theories are embedded in the system. In order to conduct effective persuasion, a number of computer-based survey questions were asked and a short tailored letter was generated for each participant. A virtual coach system monitored and reminded the participant to slow down. A few adaptive cues were used to enhance the effects of the persuasion.

We used both descriptive and experimental methods for different parts of the research. When constructing and evaluating the persuasiveness of the computer-based survey and tailored letter, descriptive methods were adopted to obtain firsthand data from the users.

The flexibility of descriptive methods enables the rational recommendations from the experts and users to be reflected in the framework. When evaluating the effectiveness of the virtual coach and adaptive cues, experimental methods were adopted because they provide direct and objective indication about system performance. When analyzing the individual elements of the framework, descriptive methods were used again to enhance the performance of individual elements.

We evaluated the feasibility and effectiveness of the overall system. At the same time, we evaluated the feasibility of individual elements. A total of 22 participants was selected to make up the sample. The experiments were conducted under controlled conditions. The results indicated that our system is effective in persuading people to speak more slowly. The feedback from users indicated that our system raised their awareness about speaking rate. We discussed the feedback from users about the system in detail. We also discussed a number of directions which need to be further investigated in future research.

From our observation and the feedback of the subjects, we found that 1) emotions appear to have a very important impact on speech rate persuasion, 2) cognitive demand plays an important role in effective persuasion, 3) the effect of the tailored letter is not that obvious, 4) the Theory of Planned Behaviour is effective in raising speakers' awareness on speaking rate, 5) real-time intervention is a powerful and effective method for slowing down speaking rate, and 6) our computational framework is feasible and effective. Each of the above findings needs to be further investigated in future research. The effect of the tailored letter had been expected to be higher than the actual feedback from the subjects, although subjects all agreed that the letter did raised their awareness on speaking rate. We believe this unexpected result may have two reasons: 1) the tailored letter needs to be redesigned and to consider other unknown design factors, or 2) our sample size was too small.

6.2 Future Work

We have defined a general conceptual persuasion model and developed a persuasive system for slowing speaking rate. We also obtained firsthand experimental data which demonstrated the feasibility of our approach. Because we are helping to establish a new research topic in speech persuasion, our current approach was based on qualitative analysis of a

small sample size. In the future, we need to continue extending both the conceptual model and the implemented system while conducting experiments using an quantitative approach.

During our experiments, we observed that the emotions of a speaker can have a huge impact on their speaking rate and the effects of persuasion on the speaker. This finding strengthened our belief that affective models are a necessary component for an effective persuasion model. The affective and rhetorical components were only discussed in the conceptual model and simply prototyped in the software system. For future research, innovative approaches in incorporating affective and rhetorical models need to be investigated. We proposed a basic framework of generating tailored letters for speakers to target their belief systems. In the future, advanced Natural Language Generation techniques need to be adopted to enhance the letter generation component.

From the feedback of subjects, we observed that some subjects continued using a too slow speaking rate once they were induced to speak slowly by the system. Obviously, we need to extend the system to be capable of inducing a speaker to speed up their speech when the current speed is too slow. At the same time, we need to investigate how to determine what is the best rate for a speaker. We believe that the best rate varies among individual speakers and will even be different for the same speaker in different contexts.

Our initial experimental subjects were composed mainly of non-native English speakers and a few native English speakers. Although the results indicated that the system worked well with both types of speakers, we believe that further investigation needs to be done for each type of speaker. In particular, speaking a different language may involve a different set of physiological and psychological factors.

Persuasive technologies are widely considered to play a key role in mobile persuasion. In the future, our persuasion system should be designed to fit on mobile devices so that the virtual coach can become a pervasive technology. The speaking rate detection and analysis mechanism also needs to be realized using software systems in order to become a practical mobile persuasive system. For example, a speaker may use a virtual coach installed on his/her mobile device while giving a public speech.

As a first step towards our ultimate research goal, we have studied how to slow down reading speed. In future, more complex speech patterns such as conversation and public speech need to be investigated. For example, an interesting communication scenario is presentation, which often includes both a public speaking phase (mostly one-way talking)

and answering questions phase (two-way communication). However, the findings from these future studies will all lead to a single outcome—persuasive technologies for speech adaption.

References

[1] 20 different approaches that could be used to achieve change in driving behavior. http://hci.stanford.edu/research/speedometer.html.

[2] *Java Computer Aided Interviewing Framework*, 2010 (accessed July 1, 2010). http://jcaif.sourceforge.net/jcaif.html.

[3] *Sonar*, 2010 (accessed July 1, 2010). http://jcaif.sourceforge.net/sonar.html.

[4] I. Ajzen. The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2):179–211, 1991.

[5] I. Ajzen and M. Fishbein. *Understanding attitudes and predicting social behavior*. Prentice Hall, 1980.

[6] I. Albrecht, M. Schröder, J. Haber, and H.P. Seidel. Mixed feelings: Expression of non-basic emotions in a muscle-based talking head. *Virtual Reality*, 8(4):201–212, 2005.

[7] JD Amerman and MM Parnell. Speech timing strategies in elderly adults. *Journal of Phonetics*, 20:65–76, 1992.

[8] P. Bamberg and L. Gillick. Phoneme-in-context modeling for dragon's continuous speech recognizer. In *Proceedings of the DARPA Speech and Natural Language Workshop*, 1990.

[9] R.D.V. Barredo and K. Mahon. The Effects of Exercise and Rest Breaks on Musculoskeletal Discomfort during Computer Tasks: An Evidence-Based Perspective. *Journal of Physical Therapy Science*, 19(2):151–163, 2007.

[10] J. Beskow. Talking Heads-Models and Applications for Multi-modal Speech Synthesis. *PhD at KTH Stockholm*, 2003.

[11] H. Bortfeld, S.D. Leon, J.E. Bloom, M.F. Schober, and S.E. Brennan. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, 44(2):123, 2001.

[12] D. Bostock. *Plato's Phaedo*. Oxford University Press, USA, 1986.

[13] M.M. Bradley and P.J. Lang. Measuring emotion: Behavior, feeling, and physiology. *Cognitive neuroscience of emotion*, 25:49–59, 2000.

[14] Angelika Braun and Reiko Oba. Speaking tempo in emotional speech a cross-cultural study using dubbed speech. In *International workshop on Paralinguistic Speech - between models and data*, Saarbrcken, Germany, 2007.

[15] K. Burke. *A rhetoric of motives*. Univ of California Pr, 1969.

[16] K. Burke. Questions and answers about the pentad. *College Composition and Communication*, 29(4):330–335, 1978.

[17] J. Cassell. Embodied conversational agents: representation and intelligence in user interfaces. *AI magazine*, 22(4):67, 2001.

[18] T.C. Cervera, M.J. Soler, C. Dasi, and J.C. Ruiz. Speech recognition and working memory capacity in young-elderly listeners: Effects of hearing sensitivity. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 63:216–226, 2009.

[19] Y.C. Chang, J.L. Lo, C.J. Huang, N.Y. Hsu, H.H. Chu, H.Y. Wang, P.Y. Chi, and Y.L. Hsieh. Playful toothbrush: UbiComp technology for teaching tooth brushing to kindergarten children. 2008.

[20] F. Charpentier and E. Moulines. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. In *First European Conference on Speech Communication and Technology*. ISCA, 1989.

[21] Ran Cheng. Persuasion strategies for computers as persuasive technologies. *Department of Computer Science, University of Saskatchewan*, 2003.

[22] R.B. Cialdini. *Influence: Science and practice*. Scott, Foresman Glenview, IL, 1985.

[23] Sunny Consolvo, Kendra Markle, Kevin Patrick, and Kara Chanasyk. Designing for persuasion: mobile services for health behavior change. In *Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology*, pages 1–1, New York, NY, USA, 2009. ACM.

[24] N.H. de Jong and T. Wempe. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods*, 41(2):385, 2009.

[25] Rodrigo de Oliveira and Nuria Oliver. Triplebeat: enhancing exercise performance with persuasion. In *MobileHCI '08: Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, pages 255–264, New York, NY, USA, 2008. ACM.

[26] E. De Strycker and SR Slings. Plato's Apology of Socrates. *Plato's Euthyphro, Apology, and Crito: critrical essays*, page 72, 2005.

[27] J.P. Dillard and M. Pfau. *The persuasion handbook: Developments in theory and practice*, pages 309–323. Sage Publications, Inc, 2002.

[28] S.T. Fiske and S.E. Taylor. *Social cognition*. 1991.

[29] E. Flemming. Phonetic detail in phonology: Towards a unified account of assimilation and coarticulation. In *Proceedings of the 1995 Southwestern Workshop in Optimality Theory (SWOT), University of Arizona*. Citeseer, 1997.

[30] B. J. Fogg. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):2, 2002.

[31] B. J. Fogg. *Persuasive technology: Using computers to change what we think and do*, page 40. Morgan Kaufmann Publishing, San Francisco, CA, 2003.

[32] B. J. Fogg. *Persuasive technology: Using computers to change what we think and do*, pages 37–41. Morgan Kaufmann Publishing, San Francisco, CA, 2003.

[33] B. J. Fogg. Creating persuasive technologies: an eight-step design process. In *Proceedings of the 4th International Conference on Persuasive Technology*. ACM New York, NY, USA, 2009.

[34] B. J. Fogg and Dean Eckles, editors. *Mobile Persuasion: 20 perspectives of the future of behavior change*. Stanford Captology Media, 2007.

[35] BJ Fogg. The behavior grid: 35 ways behavior can change. In *Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology*, pages 1–5, New York, NY, USA, 2009. ACM.

[36] BJ Fogg. A behavior model for persuasive design. In *Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology*, pages 1–7, New York, NY, USA, 2009. ACM.

[37] BJ Fogg. Three possible futures for persuasive technology. In *Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology*, pages 1–1, New York, NY, USA, 2009. ACM.

[38] JR Fontaine, K.R. Scherer, E.B. Roesch, and P.C. Ellsworth. The world of emotions is not two-dimensional. *PSYCHOLOGICAL SCIENCE-CAMBRIDGE-*, 18(12):1050, 2007.

[39] M. Foster. Enhancing human-computer interaction with embodied conversational agents. *Universal Access in Human-Computer Interaction. Ambient Interaction*, pages 828–837.

[40] D.J. Furley and A. Nehamas. *Aristotle's Rhetoric: philosophical essays*. Princeton Univ Pr, 1994.

[41] S.D. Goldinger. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology-Learning Memory and Cognition*, 22(5):1166–1182, 1996.

[42] Marco Guerini. *Persuasion Models for Multimodal Message Generation*. PhD thesis, University of Trento, 2006.

[43] R. Hincks. Speaking rate and information content in English lingua franca oral presentations. *English for Specific Purposes*, 2009.

[44] S. Iwasaki, S. Ocho, M. Nagura, and T. Hoshino. Contribution of speech rate to speech perception in multichannel cochlear implant users. *The Annals of otology, rhinology, and laryngology*, 111(8):718, 2002.

[45] E. Janse. Word perception in fast speech: artificially time-compressed vs. naturally produced fast speech. *Speech Communication*, 42(2):155–173, 2004.

[46] E. Janse. Processing of fast speech by elderly listeners. *The Journal of the Acoustical Society of America*, 125:2361, 2009.

[47] K. Johnson and J.W. Mullennix. *Talker variability in speech processing*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 1997.

[48] C. Jones, L. Berry, and C. Stevens. Synthesized speech intelligibility and persuasion: Speech rate and non-native listeners. *Computer Speech & Language*, 21(4):641–651, 2007.

[49] K. Kallinen and N. Ravaja. Emotion-related effects of speech rate and rising vs. falling background music melody during audio news: The moderating influence of personality. *Personality and individual differences*, 37(2):275–288, 2004.

[50] W.M. Keith and C.O. Lundberg. *The Essential Guide to Rhetoric*. Bedford/st Martins, 2008.

[51] A.R. Kelly, A. McDougall, and N. Abbott. Rhetorical models for computational systems: an interdisciplinary approach to reusable, tailorable medical information. In *Proceedings of the 27th ACM international conference on Design of communication*, pages 155–162. ACM, 2009.

[52] C. Kühnel, B. Weiss, I. Wechsung, S. Fagel, and S. Möller. Evaluating talking heads for smart home systems. In *Proceedings of the 10th international conference on Multimodal interfaces*, pages 81–84. ACM, 2008.

[53] Manu Kumar and Taemie Kim. Dynamic speedometer: dashboard redesign to discourage drivers from speeding. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*, pages 1573–1576, New York, NY, USA, 2005. ACM.

[54] D. Laplante and N. Ambady. ON HOW THINGS ARE SAID. VOICE TONE, VOICE INTENSITY, VERBAL CONTENT, AND PERCEPTIONS OF POLITENESS. , 5:36, 2010.

[55] M. Law and J.L. Tang. An analysis of the effectiveness of interventions intended to help people stop smoking. *Archives of Internal Medicine*, 155(18):1933, 1995.

[56] A.M. Liberman and I.G. Mattingly. The motor theory of speech perception revised. *Cognition*, 21(1):1–36, 1985.

[57] Lena Mamykina, Elizabeth Mynatt, and Michael A. Terry. Time aura: interfaces for pacing. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 144–151, New York, NY, USA, 2001. ACM.

[58] S.G. Marketos and P. Skiadas. Hippocrates: The father of spine surgery. *Spine*, 24(13):1381, 1999.

[59] W.M. Marston. *Emotions of normal people*. Cooper Press, 2007.

[60] JL McClelland and JL Elman. Interactive processes in speech perception: The TRACE model. In *Parallel distributed processing: explorations in the microstructure of cognition, vol. 2*, page 121. MIT Press, 1986.

[61] R. McGill, J.W. Tukey, and W.A. Larsen. Variations of box plots. *American Statistician*, pages 12–16, 1978.

[62] R. Montemayor. Men and their bodies: The relationship between body type and behavior. *Journal of Social Issues*, 34(1):48–64, 1978.

[63] Dan Morris, A.J. Bernheim Brush, and Brian R. Meyers. Superbreak: using interactivity to enhance ergonomic typing breaks. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1817–1826, New York, NY, USA, 2008. ACM.

[64] C. Nass, U. Foehr, S. Brave, and M. Somoza. The effects of emotion of voice in synthesized and recorded speech. In *Proceedings of the AAAI Symposium Emotional and Intelligent II: The Tangled Knot of Social Cognition, North Falmouth, MA*, 2001.

[65] C. Nass and L. Gong. Maximized modality or constrained consistency? In *AVSP'99-International Conference on Auditory-Visual Speech Processing*. ISCA, 1999.

[66] L.C. Nygaard and E.R. Lunders. Resolution of lexical ambiguity by emotional tone of voice. *Memory & cognition*, 30(4):583, 2002.

[67] J. Ostermann and D. Millen. Talking heads and synthetic speech: An architecture for supporting electronic commerce. In *Proc. ICME*, 2000.

[68] M.A. Picheny, N.I. Durlach, and L.D. Braida. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of speech and hearing research*, 28(1):96, 1985.

[69] MA Picheny, NI Durlach, and LD Braida. Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of speech and hearing research*, 29(4):434, 1986.

[70] MA Picheny, NI Durlach, and LD Braida. Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research*, 32(3):600, 1989.

[71] I. Poggi and C. Pelachaud. Performative facial expressions in animated faces. *Embodied conversational agents*, pages 154–188, 2000.

[72] P. Powers. Persuasion and coercion: a critical review of philosophical and empirical approaches. In *HEC Forum*, volume 19, pages 125–143. Springer, 2007.

[73] H. Quené. Modeling of between-speaker and within-speaker variation in spontaneous speech tempo. In *Ninth European Conference on Speech Communication and Technology*, 2005.

[74] H. Quene. On the just noticeable difference for tempo in speech. *Journal of Phonetics*, 35(3):353–362, 2007.

[75] B. Reeves and C. Nass. *The media equation: how people treat computers, television, and new media like real people and places*. CSLI Publications, 1998.

[76] E. Reiter, R. Robertson, and L.M. Osman. Lessons from a failure: Generating tailored smoking cessation letters. *Artificial Intelligence*, 144(1-2):41–58, 2003.

[77] Angelo C. Restificar, Syed S. Ali, and Susan W. McRoy. Arguer: Using argument schemas for argument detection and rebuttal in dialogs. Banff, Canada, June 1999.

[78] E. Roesch, J. Fontaine, and K. Scherer. The world of emotion is two-dimensional–or is it. *Presentation at the HUMAINE Summer school, Genoa, Italy*, 2006.

[79] R. Rosenthal and RL Rosnow. Essentials of Behavioral Research: Methods and Data Analysis, 1991.

[80] J.A. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161–1178, 1980.

[81] K.R. Scherer and U. Scherer. Speech behavior and personality. *Speech evaluation in psychiatry*, pages 115–135, 1981.

[82] E. Seat and S.M. Lord. Enabling effective engineering teams: a program for teaching interaction skills. *JOURNAL OF ENGINEERING EDUCATION-WASHINGTON-*, 88:385–390, 1999.

[83] K.N. Shahin. Hearing Voices: The Impact of Announcer Speech Characteristics on Consumer Response to Broadcast Advertising. *Journal of Consumer Psychology*, pages 198–204, 2003.

[84] S.M. Smith and D.R. Shaffer. Speed of speech and persuasion: Evidence for multiple effects. *Personality and Social Psychology Bulletin*, 21(10):1051, 1995.

[85] Carmen Soler, Alejandra Zacarías, and Andrés Lucero. Molarcropolis: a mobile persuasive game to raise oral health and dental hygiene awareness. In *ACE '09: Proceedings of the International Conference on Advances in Computer Enterntainment Technology*, pages 388–391, New York, NY, USA, 2009. ACM.

[86] S.E. Stern, J.W. Mullennix, and I. Yaroslavsky. Persuasion and social perception of human vs. synthetic voice across person as source and computer as source conditions. *International Journal of Human-Computer Studies*, 64(1):43–52, 2006.

[87] MC Taylor, DA Lynam, and A. Baruya. *The effects of drivers' speed on the frequency of road accidents*. Transport Research Laboratory, 2000.

[88] Kristian Torning and Harri Oinas-Kukkonen. Persuasive system design: state of the art and future directions. In *Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology*, pages 1–8, New York, NY, USA, 2009. ACM.

[89] R.M. Uchanski, S.S. Choi, L.D. Braida, C.M. Reed, and N.I. Durlach. Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech, Language and Hearing Research*, 39(3):494, 1996.

[90] W. Van der Vaart, Y. Ongena, A. Hoogendoorn, and W. Dijkstra. Do Interviewers' Voice Characteristics Influence Cooperation Rates in Telephone Surveys? *International Journal of Public Opinion Research*, 18(4):488, 2006.

[91] J. Verhoeven, G. De Pauw, and H. Kloots. Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech*, 47(3):297, 2004.

[92] J. Vousden and E.A. Maylor. Speech errors across the lifespan. *Language and Cognitive Processes*, (No. 1-3):48–77, 2006.

[93] P.L. Watkins, C.H. Ward, and D.R. Southard. Empirical support for a Type A belief system. *Journal of Psychopathology and Behavioral Assessment*, 9(2):119–134, 1987.

[94] S.F. Witelson and W. Pallie. Left hemisphere specialization for language in the newborn. *Brain*, 96(3):641–646, 1973.

[95] H. Yang, W. Guo, and Q. Liang. A Speaking Rate Adjustable Digital Speech Repeater for Listening Comprehension in Second-Language Learning. In *Proceedings of the 2008 International Conference on Computer Science and Software Engineering-Volume 05*, pages 893–896. IEEE Computer Society, 2008.

[96] J. Yuan, M. Liberman, and C. Cieri. Towards an integrated understanding of speaking rate in conversation. In *Ninth International Conference on Spoken Language Processing*. Citeseer, 2006.

Appendices

Appendix A

Survey Questions Based on TPB

A.1 Behavioural Beliefs

- Question 1: To which degree do you think the following statement is true?
  - Speaking slowly will reduce speech errors.
- Question 2: To which degree do you value the following outcome?
  - Fewer speech errors.
- Question 3: To which degree do you think the following statement is true?
  - Speaking slowly gives me more time to construct better sentences.
- Question 4: To which degree do you value the following outcome?
  - Be able to construct better sentences.
- Question 5: To which degree do you think the following statement is true?
  - Speaking slowly will make me appear more relaxed.
- Question 6: To which degree do you value the following outcome?

  - Be viewed as a relaxed person while speaking.

- Question 7: To which degree do you think the following statement is true?
  - Speaking slowly will make me appear more confident.
- Question 8: To which degree do you value the following outcome?
  - Be viewed as a confident person.
- Question 9: To which degree do you think the following statement is true?
  - Speaking slowly will allow people to understand me better.
- Question 10: To which degree do you value the following outcome?
  - People understand what I am trying to say.
- Question 11: To which degree do you think the following statement is true?
  - Speaking slowly will reduce chances of people asking me to repeat.
- Question 12: To which degree do you value the following outcome?
  - Fewer people asking me to repeat statements.
- Question 13: To which degree do you think the following statement is true?
  - Speaking slowly will make conversations more enjoyable.
- Question 14: To which degree do you value the following outcome?
  - Have enjoyable conversations.
- Question 15: To which degree do you think the following statement is true?
  - Speaking slowly will cause people to think that I am not as fluent.
- Question 16: To which degree do you value the following outcome?

  - Be viewed as disfluent.

- Question 17: To which degree do you think the following statement is true?
  - Speaking slowly will convey less information to the audience.
- Question 18: To which degree do you value the following outcome?
  - Conveying less information to the audience.

A.2 Normative Beliefs

- Question 19: To which extent do the following people want you to speak slowly?
  - Your direct family (excluding spouse).
- Question 20: To which degree do you value their opinion?
  - Your direct family (excluding spouse).
- Question 21: To which extent do the following people want you to speak slowly?
  - Your spouse or life partner.
- Question 22: To which degree do you value their opinion?
  - Your spouse or life partner.
- Question 23: To which extent do the following people want you to speak slowly?
  - Your close friends.
- Question 24: To which degree do you value their opinion?
  - Your close friends.
- Question 25: To which extent do the following people want you to speak slowly?

  - Your boss, colleagues, or clients.

- Question 26: To which degree do you value their opinion?
  - Your boss, colleagues, or clients.

A.3 Control Beliefs

- Question 27: How often do you encounter the following situation?
  - Being short of time.
- Question 28: How likely can you keep a slow speaking speed in the following situation?
  - While being short of time.
- Question 29: How often do you encounter the following situation?
  - People around me speak quickly.
- Question 30: How likely can you keep a slow speaking speed in the following situation?
  - While people around me are speaking quickly.
- Question 31: How often do you encounter the following situation?
  - I am very familiar with the topic I am speaking about.
- Question 32: How likely can you keep a slow speaking speed in the following situation?
  - While I am very familiar with my speech topic.
- Question 33: How often do you encounter the following situation?
  - An exciting or stressful situation.
- Question 34: How likely can you keep a slow speaking speed in the following situation?
  - While in an exciting or stressful situation.