VOICE BASED FOR BANKING SYSTEM

by

FAIZAH BT MOHAMAD @ CHE OMAR

Dissertation submitted in partial fulfillment of
the requirement for the Bachelor of Technology (Hons)
(BUSINESS INFORMATION SYSTEM)

DECEMBER 2005

UNIVERSITY TECHNOLOGY OF PETRONAS
BANDAR SRI ISKANDAR
31750 TRONOH, PERAK
December 2005

CERTIFICATION OF APPROVAL

VOICE BASED FOR BANKING SYSTEM

by

Faizah bt Mohamad @ Che Omar

A project dissertation submitted to the
Information Technology Programme
University Technology of PETRONAS
in partial fulfillment of the requirement for the
BACHELOR OF TECHNOLOGY (Hons)
(BUSINESS INFORMATION SYSTEM)

Approved by,

(Mr Shuib b Basri)

UNIVERSITY TECHNOLOGY OF PETRONAS
BANDAR SRI ISKANDAR
31750 TRONOH, PERAK
December 2005

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgment, and that the original work contain herein have not been undertaken or done by unspecified sources of persons.

FAIZAH BT MOHAMAD @ CHE OMAR

ABSTRACT

The trouble with traditional banking system service resulted difficulties, latency and low quality of service, not suitable for disable people and require extra manpower to perform simple bank activities. The goal of this project is to build a voice recognition based system which specifies on the banking activities element and specializes in using voice as a medium to run bank activities via telephony network system. Three fundamental objectives were addressed in the study. First, to develop two-way interactive program of banking system, which use voice as important mechanism to receive instruction and response to user. Second, it support to first objective which to develop such a user friendly and high security voice banking system which requires the user first logs on to the system by furnishing the assigned customer identification number and personal identification number before user proceed for further actions. And therefore, there must have a strong database structure development of the application in the voice banking system that purposely to maintain the integrity of the data stored and responds to authorized user only. For third objective, is to determine the best programming in order to implement in telephony network system. There is a study and architecture on how voice can be accepted, manipulated and generated by using combination two types of programming which are Cold Fusion and VoiceXML, which is goes to the third objective. The functions of this system is proved and demanded by user as it provides such convenience and easy services with just use voice to transmit the instruction. Hence, this strategy will grab large number of customers and simultaneously will generate huge profit too to the bank institution that applies this system. It is hoping that, by developing this system it will be a platform for next developer to host the system and can be use a large number of customers simultaneously and efficiently.

Keyword: Voice based, telephony, combination of programming, architecture

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

VBBS
Voice Based for Banking System

VXML
Voice Extension Markup Language

XML
Extension Markup Language

DTMF
Dial Tone Multi-Frequency

MADK 3.0
Mobile Application Development Kit 3.0

CF
Cold Fusion

TTS
Text-to-Speech

UTP
University Technology PETRONAS

W3C
World Wide Web Consortium's

IVR
Interactive Voice Response

IDE
Interactive Development Environment

CHAPTER 1

INTRODUCTION

As widespread the e-business paradigm has become, it is limited to users of computers equipped with a monitor, keyboard, and pointing device. Because of this limitation, a huge potential customer base has been ignored - people who, due to time, location, and/or cost constraints, do not have access to a computer. Many of these people do, however, have access to a telephone. Therefore, providing conversational access (that is, voice input and audio output) to services over the telephone permits companies to reach this as yet untapped market.

Hence, the telephone is only one device through which businesses can provide service via voice access. Customers will expect to use mobile and pervasive devices (such as smart phones) to access and communicate as well. By providing access to information and services through the telephone, businesses have taken customer service to new heights. Since almost everyone has at least one telephone and people can use a phone from virtually.

Apart from that, this research will also include theories, rules and boundaries employed for the development of the application. Those who really interested to perform banks activities have a slight advantage to explore and understand further content of the dissertation.

1.1 Background of Study

This project research is about to develop a Voice-Based Banking System (VBBS), that combining the technologies of VoiceXML and Cold Fusion. Well, this study will explore whether the voice recognition product developed could be used to support the delivery of telephony network system and enabled to create telephony application in "naturalness" way. The system will enable consumer banks to deliver telephone banking services such as ***Transfer Money and Check Balance Statement (both for saving and current account).*** The system also will be managed to give command or list of option to user for further actions of banking activity.

The objective of the VBBS is to be able to provide bank services to users at anywhere, using a telephone to access the service, which is in natural language, as extension manner to the e-business model. It seems that this project research have to take tough challenge on what Chris Schmandt from Speech Research Group Media Laboratory with his point of view as "Speech recognition and touch tones are the obvious feedback channels but both are limited. Speech recognition does not work well over telephones because of variable noise levels and line characteristics".

Users are much more likely to interact with a system they feel comfortable with and that responds in a human like way. With that, the basic plan general architecture of the voice banking application that seeks to develop which contains two parts-the front end and the back end. The front end takes input in the form of two choices whether using touch-tone system or voice recognition system. Whereas the back end will response the input from the user and all the necessary data input will be stored in the database.

Although there are many types of sound, which different people have different intonation of speech, this project research will able to explore the methodology of voice detection. For example word "start", some people will speak up as start in a high tone while some people might have such low tone of voice speak. Therefore, the recognizer speech will processes callers' spoken command and ascribe the meaning to the system database.

There will be a study exploration towards the speech production which, it guides and explain on how the voices from user can response the system. In this VBBS, the user first logs on to the system by furnishing the assigned customer identification number and personal identification number. Based on the information provided, the system first authenticates the user, and if the user's credentials are verified successfully, the application starts by prompting the first four options to the user and waits for the user to respond. The user may select any one of the options in order to proceed further, which are contained; Inquiry section, which are: check account balance (for both saving and current account) and Transfer money; and Exit function.

It is hypothesized that, at the end of the product research it will be able to response in both two methods of communication which are by voice recognition and touch tone or Dual Tone Multi-Frequency (DTMF) system.

1.2. Problem Statement

Many businesses are betting that consumers will embrace any technology that provides real-time access to information piped through their regular telephone, wireless phone or voice-connected handheld device. VBBS is standalone testing configuration and it will become as the platform to host it in the future. Thus, this project research can overcome the weaknesses of having usual banking system and also with other conventional banking techniques. Below are the problems statements that author have found from using those conventional methods:

1.2.1 Lack of Mobility

One still has to be tied to the physical network to access information and perform banking task. From using usual banking system, it is only applicable to be use to normal people. Then, it is not consider to old folks and also handicapped people (e.g. blind and paralyze) that they also need to perform banking transaction. They can't perform those activities, as the system does not have special treatment such as telephony service in order to fulfill their needs.

1.2.2. Require manpower to perform simple banking activities

It is not worth for bank institution needs manpower to perform simple banking system despite use machine that have programmed and can give such efficiency task to perform. Besides, it can't increased the speed of warehousing (database record) and receiving operations as customer required to queue and waiting for their turn. In fact, speech offloads routine calls from the agent's queue. Or, if the speech system can't completely solve the caller's issue, it can classify the call or collect necessary information before passing the call to the agent. That means agents focus on complex queries, increasing their job satisfaction. That, in turn, reduces attrition levels and their associated hiring and training costs. With fewer calls reaching agents, call centers can accommodate future growth in call volumes with existing staff levels.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction of Voice

In general, voice and speech has different meaning of definition but it works together in order to produce and can convey such useful message. System needs voice transformation (voice quality and intonation), which it is the process of taking the speech of a source speaker and transforming the characteristic of the signal, such that a human listener would believe the speech was uttered by a target speaker.

Speech involves the intentional production of sound patterns, which are believed to carry a certain meaning; when both speaker and listener attach the same meaning to an utterance, speech is successfully recognized and communication occurs. Thus, speech recognition requires not only the ability to determine that a sound pattern is actually speech but also the ability to attach to that pattern the meaning intended by the speaker. This complex task begins with the identification of acoustic signals, or sounds. Then, some meaning is inferred from the sound pattern. Meaning is arbitrarily associated with sounds in a language, and the physical properties of a signal corresponding to a particular meaning vary about some ideal norm. Therefore, it is not surprising that we have always wanted to communicate with and command various technical devices by voice (Nancy B. Lerner, March 1989).

Every speech sound is produced by articulation or the movement of one or more vocal organ along the vocal tract. When articulation comes into count, it is inevitably to discuss on place of articulation and manner of articulation. Table1 3.0 and figure 3.1 illustrate the place of articulation and manner of individual consonant and vowel, the vocal tract is divided into different segments or regions. The locations of these organs and regions are as shown in the following table:

Vocal organs and regions of articulator	Adjectives
Nose	Nasal
Mouth	Oral
Lips	Labial
Teeth	Dental
Alveoli (or alveola ridge or gum ridge)	Alveolar
(hard) Palate	Palatal
Velum (soft palate)	Velar
Pharynx	Pharyngeal
Uvula	Uvular
Larynx	Laryngeal
Glottis	Glottal

Table 2.1: Name of Vocal Organ and Region or Articulator

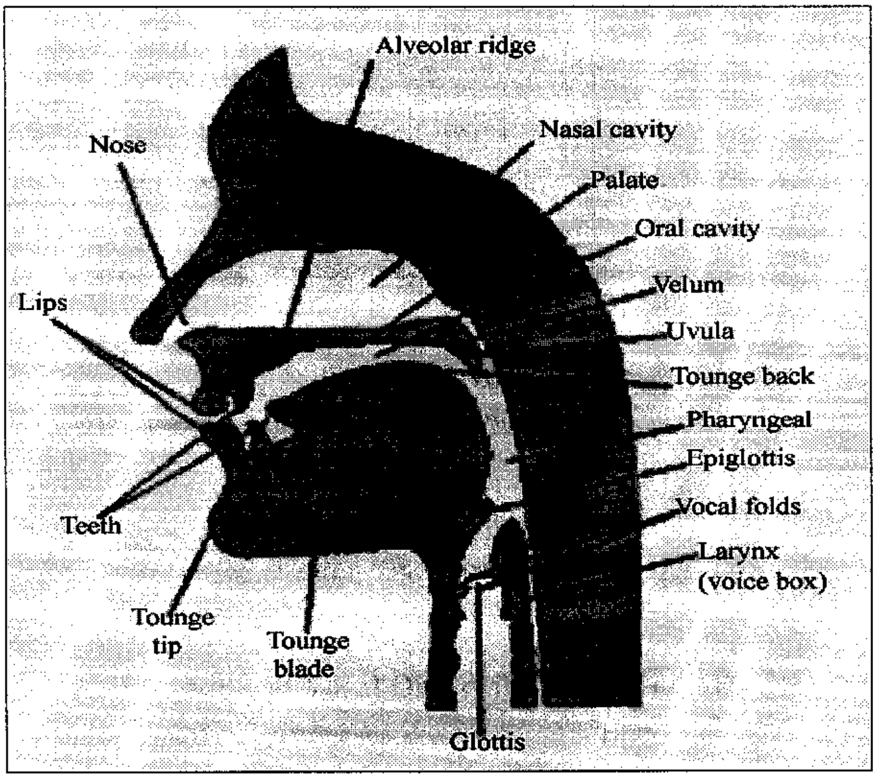


Figure 2.1 Diagram of Vocal Organs and Articulacy Regions

2.3 Mechanism of Voice

Current technology addresses the speech recognition task in two phases. First, analog waveforms produced by a speaker are transformed into digital patterns, usually by a common analog-to-digital converter. Then, the digitized speech is matched to the vocabulary of the machine. The techniques currently used include simple pattern matching and linguistic analysis involving the extraction of certain phonological features (physical properties of the speech waveforms). These techniques will eventually be supplemented by methods of analyzing syntax and prosody (the timing changes in pitch and loudness). Significant advances in these areas will have to be made prior to the development of a continuous natural speech recognizer that can accurately mimic the human recognition process (Blackwell.G, 2003).

The important feature in developing voice based we need to improve the generation of synthetic intonation for speech synthesis. However, we need to meet two goals: The first of them is to understand how linguistic theory can be appropriately exploited to improve the generation of synthetic intonation. The second is to provide a flexible and robust intonation model. This model must be able to generate different intonation patterns for the same input text to match different intended meanings of the text. These goals will become more special and aid towards development of speech synthesis (Robert A. J. Clark, April 2003).

Intonation modeling for speech synthesis is now one of the big issues facing speech synthesis systems. The quality of synthesized phonetic material has progressed sufficiently that "what is being said" is sufficiently clear that "how is it being said" is a question, which is raised in the mind of the listener. From the perspective of speech synthesis we are addressing the question: "How do we best model intonation?" This leads us to first ask: "What is intonation?" As we shall see definitions of intonation vary. Ladd (1996, p. 6) say: "Intonation, as I will use the term, refers to the use of *suprasegmental* phonetic features to convey `post-lexical' or *sentence-level* pragmatic meanings in a *linguistically structured* way". Whereas Cruttenden (1997, p. 7) says: "Intonation involves the occurrence of recurring pitch patterns, each of which is used with a set of relatively consistent meanings, either on single words or on groups of words of varying length". The problem of intonation modeling for speech synthesis is summed up by the following quote regarding segmental effects on pitch: "However, we believe that our understanding of perception of pitch influent, meaningful speech is currently not sufficient to make strong claims about the *im*perceptibility of any aspect of speech, so currently we have no other option but to model any effect on any acoustic feature that can be clearly demonstrated in natural speech" (Van Santen & Hirschberg 1994).

We have some basic intuitive ideas about what natural pitch should sound like, but we just don't understand enough to know how the pitch associated with a specific segment, in a specific syllable with a specific accent, in a specific wording a specific phrase with a specific phrase type, in a specific context, spoken by a specific speaker, should behave (Robert A. J. Clark, April 2003).

We implicitly assumed that an accent is some sort of obvious `bump' in pitch. What these bumps should actually look like turns out to be quite a complex problem, and is one of the questions that this thesis attempts to answer. There are actually two things we need to model here. We need to model how the pitch range changes over the utterance as a whole and we then need to be able to overlay pitch events successfully onto this range to produce a resulting contour (Robert A. J. Clark, April 2003).

Speech synthesis is somewhat easier to pin down than intonation. In the most general sense speech synthesis is the production of `synthetic' speech using a personal computer or other computing device. By this we mean producing an electronic signal which when played through a speaker or similar transducing device resembles human speech enough for the human brain to interpret it as such. More technically, this means that the signal must contain a reasonable representation of the voicing and the different harmonic resonances associated with the underlying formants in the vocal tract (Robert A. J. Clark, April 2003).

An investigation by Van Lacker et al (1985) tested the ability of listeners to recognize voices when played normally as well as backwards. He found that for some speakers the listeners were able to recognize the speaker nearly as well, whereas for others they performed poorly. On the basis of this and other research, Van Lanchker concluded that the critical cues for recognition are not the same for all speakers.

In a paper by Zetterholm (2000) it was shown that quality, pitch register, intonation and other prosodic aspects of the voice and speech style are important features to capture in order to succeed in imitating another voice. Thus was demonstrated through a series of perceptual tests. Very little work has been done on the transformation of linguistic features. It is particularly hard problem since in order to do so, one must recognize the words spoken, identify those words which should be mapped, and then synthesize the appropriate word with the voice quality of the target speaker (Halpern, 2001; Eisenzopf, 2002; Sharma & Kunins, 2002).

2.4 Quality of Voice Transformation

In order to carry out voice transformation there are a number of different parameter to be mapped, including spectral dynamics, fundamental frequency and timing. These characteristics can broadly be decomposed into two parts; firstly voice quality, and secondly characteristic of the fundamental frequency and timing (William H Edmonson and Li Zhang, 2002).

There has been a considerable amount of research directed at the problem of voice quality transformation (Arslan 1999, Arslan & Talkin 1997, Stylianou et al. 1995). The general approach has been to begin with a training phase in which material from source and target speakers is aligned and used to define a transformation which maps the acoustic space of the source speaker to that of the target. There are two key questions to be addressed; how should the speech signal be represented and how should the mapping be achieved? (Joram Meron, 2001).

Intonation plays an important role in speaker identity. The only approach which has so far been proposed to the problem of transforming the F0 contour of one speaker to another simply consists of modifying the source F0 contour such that it has the mean and standard deviation of that of the target speaker (Arslan 1999). However, two contours may have the same mean and standard deviation, but differ greatly in how they are perceived as was noted by Ladd and Terken (1995). Clearly, a more sophisticated approach would benefit voice based transformation system. The Tone and Break Indices (ToBI) system proposed by Silverman et al (1992), offers a methods for describing intonation contours in term of a series of intonation events. These comprise tones, which describe pitch accents and the nature of the contours at the end of a phrase, and brake indices which describe the nature of pauses.

VoiceXML is a mark-up language for developing speech user interfaces. The development of the VoiceXML standard by AT & T, IBM, Lucent Technologies and Motorola has meant that freed developers from the necessity of having to learn about speech recognition algorithms or proprietary Application Programming Interfaces (API) for speech recognition engines (Mc Glashan *et al*, 2001). With the development of VoiceXML 2.0, a range of supporting standards has emerged for describing TTS, recognition grammars and call control. These standards have been grouped by the W3C into a suite called the W3C Speech Interface Framework and will likely form the basis for future voice enabled web applications (Larson, 2003).

Intonation analysis generally involves three basic tasks: detection, identification, and placement. Detection of intonation events involves determining where, in the speech signal, accent and boundary events are located. Identification of intonation events consists of giving names to the event. In the Tilt model, for example, identification involves determining whether an event is an accent, a boundary, or perhaps a combination of both. Using the ToBI model, the process involves not only determining whether the event is an accent or boundary, but what the tones are that make up the event. The third task, placement, is the act of linking an event with a portion of linguistic text (e.g. syllable nucleus, demi-syllable, syllable, word, phrase). The task being undertaken is a combination of detection and identification (Kurt Dusterhoff, 1999).

Pitch is one of the most important prosodic information for automatic speech recognition and understanding. It can be used to break up speech signals into some prosodic phrases such as accent phrases (W.A.Lea, M.F.Medress and T.E.Skinner, 1975), (H.Shimodaira and M.Kimura, 1992) and also used to estimate the structure of sentence (A.Komatsu, E.Oohira and A.Ichikawa, 1988).

When studying intonation alone, finding and describing an accent of a given type in a given place is often the final goal, but in the context of speech synthesis knowing that an accent is found in a certain place under certain conditions is not enough. We need to be able to recreate that accent that is we need to know how to place such an accent with respect to its prosodic context in a given speaker's pitch range. To do this we need some understanding of the underlying prosodic structure and some understanding of how this relates to pitch range. Prosodic structure is used to account for high-level patterns in intonation, such as the difference in pitch range between two phrases of the same utterance. Different theories take different views to how prosodic structure should be represented (Robert A. J. Clark, April 2003).

The usefulness of prosodic information for speech recognition has been known for a rather long time and emphasized in numerous papers (Lea, Wayne A., 1980). Nevertheless, only a very few speech recognition system did actually make use of prosodic knowledge. The role of prosody in speech recognition is that of supplying side information. In principle, a speech recognition system can do its main task without requiring or processing prosodic information. However, Vaissiere pointed out, prosodic information can (and does) support automatic speech recognition on all levels (Vaissiere, Jacqueline (2002).

A synthesizer voice may be constructed fully automatically from waveform file and phone label files (although the process may take several hours). Some of the database have phone label automatically assigned from word labels using an aligner, thus, making the process the database construction require less skilled work. Because it use acoustic measures and search for appropriate join points during selection, accurate phoneme boundaries are not very important (Alan W Black and Nick Campbell, 2000).

Human speech is produced by the vocal tract, which starts at the glottis (vocal folds) and ends at the lips. The lungs contract to force air through the trachea and pharynx and out through the nasal and oral cavities. In English there are four different types of sounds that can be created; aspiration noise, friction noise, plosive (release of breath) and voicing. Voicing is a quasi-periodic vibration of the vocal folds- for example the syllable in 'lay'. The frequency of the vibration is called the fundamental frequency or F0 and is perceived as pitch.

Macmillan English Dictionary (2002) defines each term as "the study of the pattern of speech sounds used in a particular language" and "the system of spelling that a language uses". It can be concluded that phonology is study of the sounds of the speaker's language and orthography can be concluded as a system that uses written symbols to uniquely identify each other through the representation of alphabet and graphemes.

In fact, the sound were produced at the glottis and modified by the vocal tract. One useful way of describing speech production is the source-filter model. In this view, a source (excitation) waveform is modified by a filter. This model is able to represent most speech phenomena. A simple form of this model works as follows; during unvoiced speech the excitation may be modeled as noise, and during voicing as a series of impulses at the appropriate fundamental frequency. The filter simulates the effect of the vocal tract resonance, to create the resulting speech. The locations of these organs and regions are as show in the following table, taken from Yousif A. El-Imam and Zuraida Mohammed Don (2000). We have determined that speech is a collection of words shaped by voice. Figure 3.0 shows the model of source filter, and the words are called the source. Since the words are modified by voice, we say the source passes through a filter. Another essential to create this model, it can visualize the idea to develop a detail system process flow later.

This knowledge will brings to the source filter model of speech as beneath:

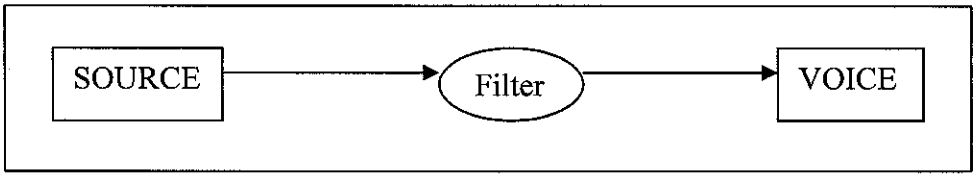


Figure 2.2 a Simple Model of Speech

2.5 VoiceXML

VoiceXML provides a high-level programming interface to speech and telephony resources for application developers, service providers and equipment manufacturers. As such, the language follows all of the syntactic rules of XML with semantics that support the creation of interactive speech applications. Standardization of VoiceXML will simplify creation and delivery of Web-based, personalized interactive voice response services; enable phone and voice access to integrated call center databases, information and services on Web sites, and company intranets; and help enable new voice-capable devices and appliances. VoiceXML is expected to expand access to the internet through telephones and other devices using both speech and ordinary touch-tone user interface.

Every now and then, there comes an agreed-upon, widely adopted standard or enhancement in a technology that starts a tidal wave in the industry by enhancing performance, by reducing costs at least a factor of 10, or by allowing for application or services that were not easily attainable before. XML, VoiceXML and speech recognition are such new standards and technologies for the voice industry (p.6 Chetan Sharma and Jeff Kunins, 2002)

VoiceXML is a new flavor of XML that defines structures for playing prerecords voice prompts as well as text-to-speech generation for presentation to the user over the telephone. The integrated response from the user is handled by either DTMF (touch tone) or speech recognition. The World Wide Web Consortium's (W3C) working draft on "Voice Browser" activity defines the standard for VoiceXML. W3C is diligently working to expand access to the Web by allowing people to interact with Web sites via spoken commands. This technology allows any telephone to access Web-based services and is especially helpful to people with disabilities. (Les Hamilton, 2001)

2.6 Interactive Voice Response (Voice-based Technologies)

A modern IVR system uses a simple touch-tone or Dual Tone Multiple Frequency (DTMF) user interface. In this type of interface, a user simply dials a telephone number and a computer picks up the call and starts the automated dialog process. An automated speech interface guides the user to enter choices using keypad of the telephone device. Once the session is over, the IVR system disconnects the call. By the 1980s, the DTMF-based user interface became a de facto standard for IVR systems. However, this interface was soon to be replaced by a technique capable of recognizing the human voice with better accuracy. In the late 1980s, the Advance Research Project Agency (ARPA) funded some of the organizations working on the development of computer-based speech recognition systems. However, modern IVR systems handle a large numbers of users effectively because they are able to recognize voice inputs supplied by different callers, by interpreting them using highly advanced techniques like AI. Despite remarkable advances in speech recognition technology, DTMF-based IVR systems still survive, as they less expensive and faster to develop and implement compared to natural language recognition-based or directed dialog-based IVR systems. An estimated comparison graph of development expenses is show below:

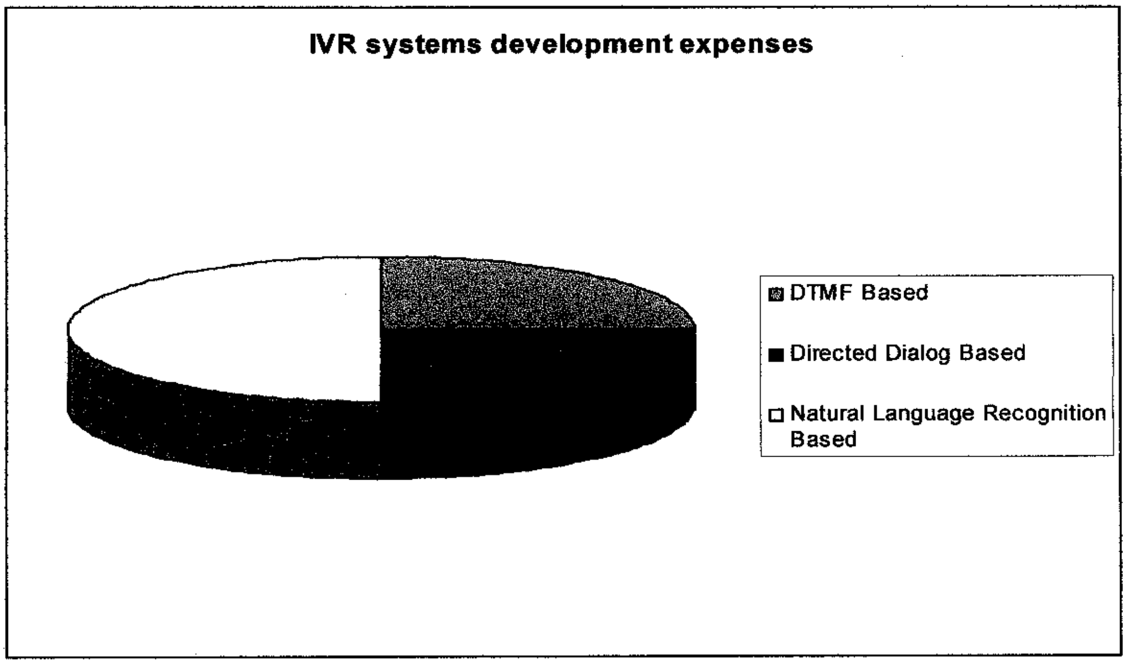


Figure 2.3 IVR System Development Expenses

The speech recognition facility has added a new dimension to IVR development and implementation. Many factors contributed to the success of IVR systems and their popularity in the corporate sector. Some of these factors are:

- The most significant benefit offered by IVR systems is the diminishing cost of operations with increased customer participation.
- With fast, increasing processing power, speech recognition systems are becoming more and more accurate.
- From the customers' point of view, IVR system have proved their worth by providing 24/7 services.

(Chetan Sharma, Jeff Kunins; 2002)

CHAPTER 3

METHODOLOGY / PROJECT WORK

Throughout the project completion, it implies two main activities. Firstly, it involved information mining and grammar construction on voice technology with its architecture integration. And the second activity is by debugging of prototype with the use of programming tools and toolkit identified.

In developing the application, VBBS used four essential phases in order to guide the project flow throughout its development-planning schedule. To view big picture of project methodology is as attached in APPENDIX F. Hence, the basic phases are shown as beneath:

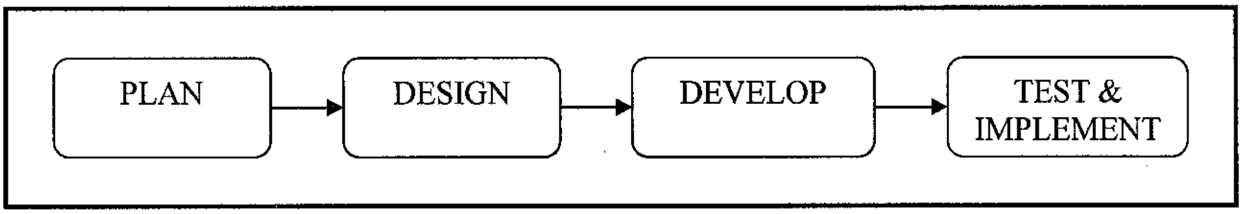


Figure 3.1: Project Methodology

- Project Planning and Functional Analysis
- Design System Sequence
- Development
- Testing and Implementation

3.1 Project Planning and Functional Analysis

The initial stage of project methodology is ultimately the most critical for determining whether a project will succeed. At this stage, there are several planning items that need to take into consideration such as gather information through distribute survey questionnaires to end-users, what tool to be used, when to be done and how to apply those information to develop the system. The planning come to more focus and narrow by making decision on what research area to be choose and all the information description able to support in this planning and functional analysis phase.

3.1.1 Research Area Analysis: Speech Production

Speech production selected to be essential part in the research area of VBBS development. It describes the execution of system for users' voice match to the voice in the dictionary and grammar.

In this stage, the output derived from grammar and vocabulary mapping management. It is consider, as an intuitive presentation of retrieval results and one of the most important aspects of any speech application is its grammar and vocabulary. A properly designed grammar is absolutely critical to the voice user interface of any speech application. A grammar is the union of words and phrases comprising the expected range of input and output for the application. It basically defines what an application needs to listen for from the user at every step. For example, for this VBBS the following grammar use to direct the user:

```
<grammer type = "x-ref">
        checking [account] | savings [account] | transfer [funds]
</grammar>
```

In this example, the user has to select from three available options, each of which can be communicated two different ways (to have a saving account, the user could say "account saving" or just "account"; both inputs would be acceptable). Similarly, if the system prompts, "Please say the checking or transfer," the user needs to pick from the list of checking and/or transfer that the application is designed for. The system should be able to account for variations in the utterances. And if the user picks something that is not defined in the grammar, then the application should help guide the user to the correct choices (by reprompting) or enable the user to connect to a human operator or exit the application. Thus, a vocabulary is the list of words used in a given application. Another term under this grammar and vocabulary mapping management is *phonetic mapping*. Phonetic mapping is another major part in speech production process, which during the process the spectral analysis information (set of numbers) is rendered into the closest and the best possible phoneme results. This is the most vital step as it assists grammar and vocabulary management to directly influence the accuracy of the speech recognition capability of the system. Phonetic mapping results in the generation of a qualifier set of phoneme strings. Every string has its own score that indicates its proximity to the spoken word or text. Then, these sets of phonemes are matched against the provided grammar library to select the closest option available based on the grammar. Consider the following example to illustrate phonetic mapping:

One	Score	75%
Won	Score	30%
Wan	Score	45%

Then, the system will look at the grammar that defines a set of expressions for an element named choice:

$choice = (right | wrong | none);

When the system performs matching, it picks the best option based on the grammar and produces the result shown as:

One Score 75%

Therefore, the system will recognize the word as *One* when processing input.

After all, the roles of grammar pave the way for users' interactions with such applications. The author will ensure that the user isn't forced to interact with a system that full of computer-directed dialogs and will give instruction to users about to speak naturally with using 'Standard English'.

3.2 Design System Sequence

Design system sequence will portray the detail design of the project content arrangement and also its flow. It architecture the system based on user requirement and it gives the idea to develop such a user friendly system.

3.2.1 VBBS Content Organization

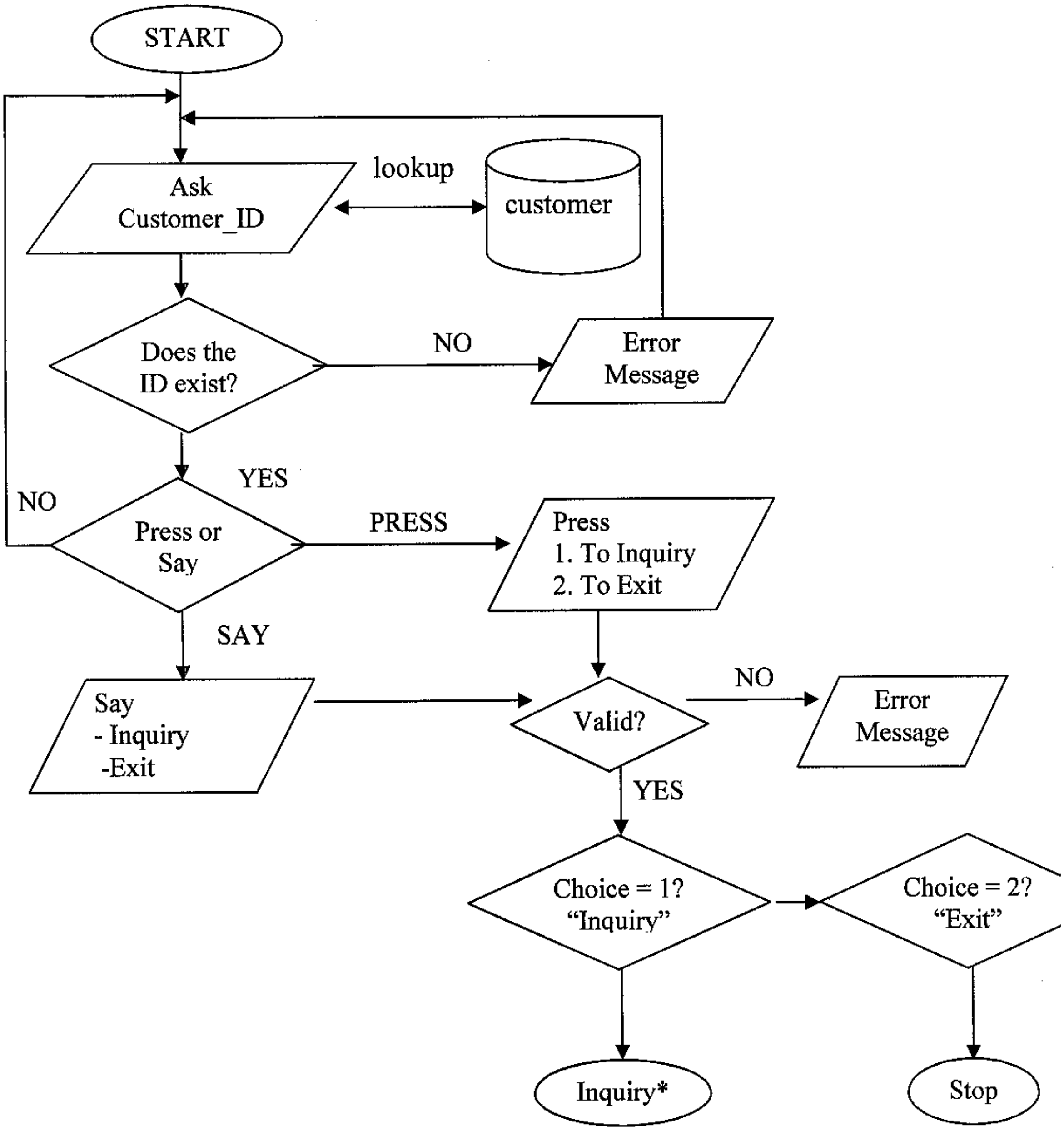
Based on the information in functional work analysis, it recognized the system credibility and responsibility to perform in the VBBS functionality. Thus, the idea of the analysis is to identify the structure of **content organization.**

The purpose of content organization is to make decision on provide content to the users through the voice application. Content organization able to makes an impression, good or bad, on users' right in the maiden interaction in the application. It is based on this impression that the user will appraise through the system. The VBBS goes in the hierarchical structure of content organization. This organization developed as it offers more than one service to choose from VBBS. The system dialog structure designed will generate a service as below statement:

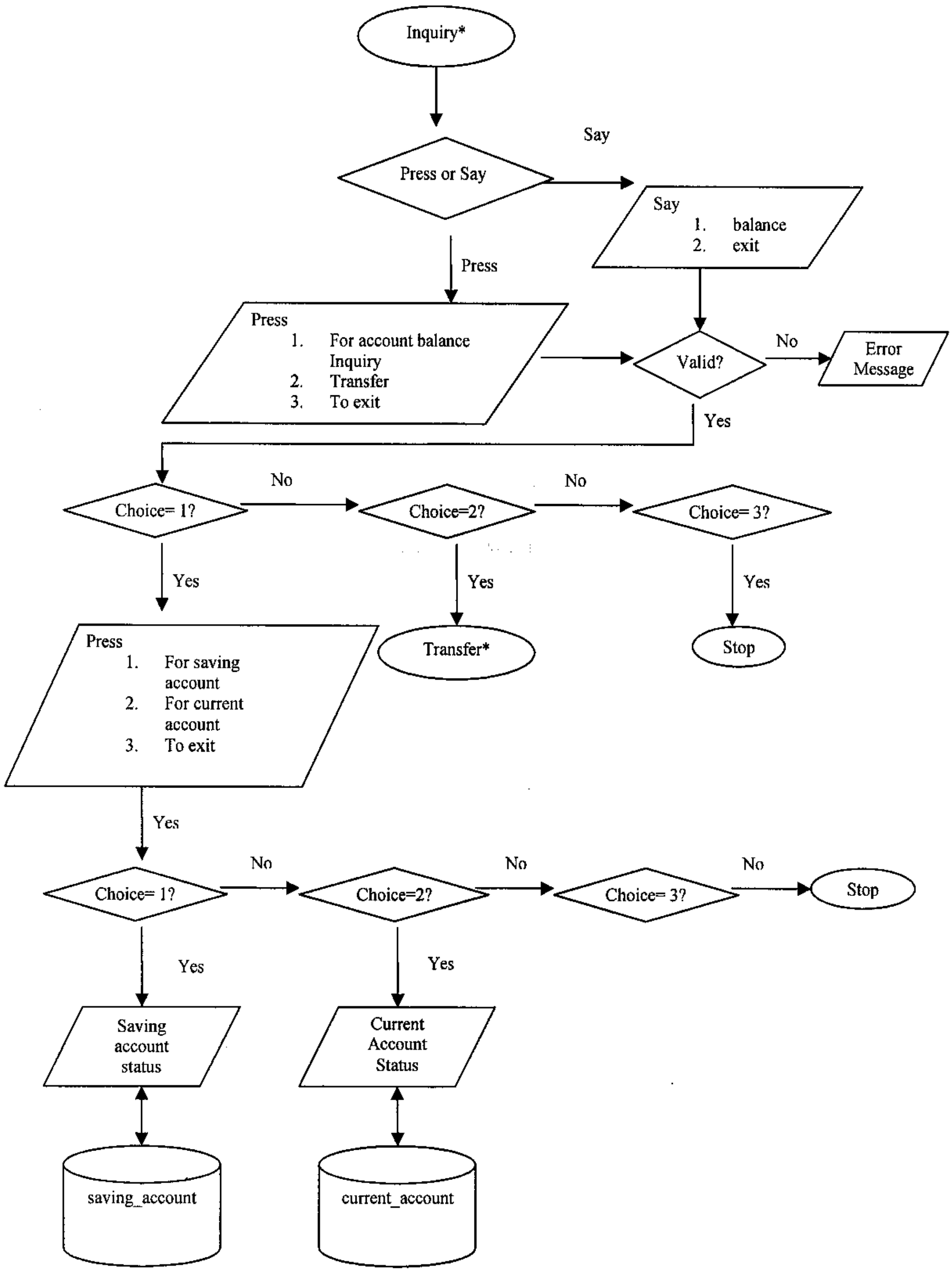
- **System:** "Please choose on of the following option: Press 1 or say 'balance' for checking account balance. Press 2 or say 'statement' for getting details of your current account balance. Press 3 or say 'exit' to close the application.
- **User:** "balance".
- **System:** "Press 1 or say 'current' for current account. Press 2 or say 'saving' for saving account. Press 3 or say 'exit' to close the application.
- **User:** "exit".
- **System will stop.**

3.2.2 System Flow

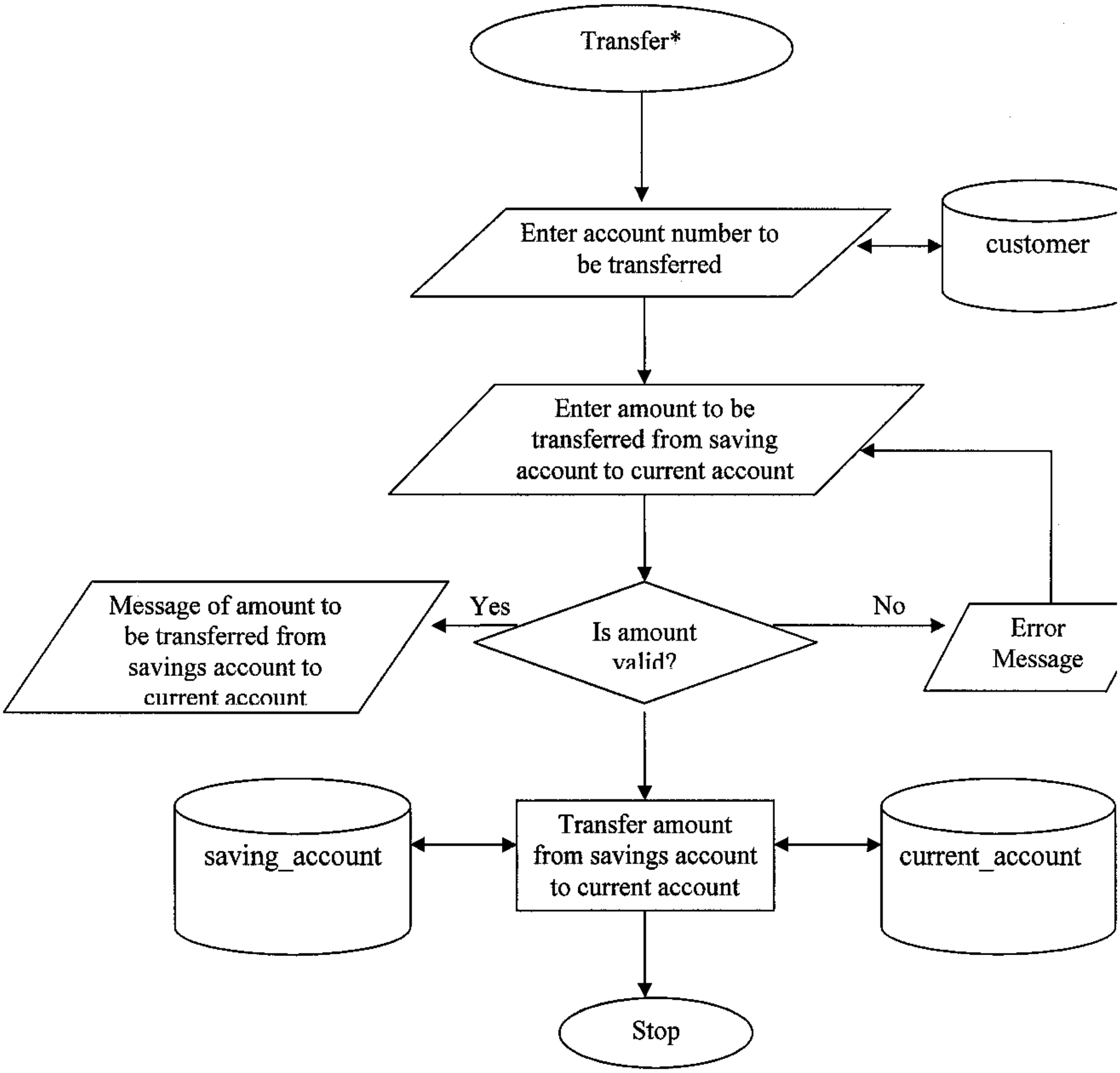
3.2.2.1 Flowchart of Main Menu



3.2.2.2 Flowchart of Inquiry Section



3.2.2.3 Flowchart of transfer section



3.2.3 Database Information

Another essential part of system development is database structure that requires including all the entity involves throughout the system. The various types have been identified and all data required to build to the application and stored in five tables. The author chooses Microsoft Access 2000 to generate database and might assist in data retrieval system. Below table is a brief description of each of the tables along with a representation of its schema.

TABLE NAME	TABLE DESCRIPTION
customer	Stores the personal information of customers
account_statement	Stores the account statement information of all users
current_account	Stores information related to current accounts
saving_account	Stores information regarding savings accounts
transfer	Stores all the information regarding cash transfer conducted by users

Table 3.1 Banking Services Application Tables

3.2.3.1 Customer Table

This table stores all personal information regarding customers that the application needs as follows:

- Customer identification number
- Telephone identification number
- Name of the customer
- Address of the customer

3.2.3.2 Account_statement Table

This table stores all the information used to generate the account statement for the user such as debit and credit balance of the user. Specifically, the account_statement table contains the following:

- Savings account number
- Customer identification number
- Debit balance
- Credit balance
- Transaction date
- Closing balance

3.2.3.3 Current_account table

This table stores the information related to the current accounts, such as the balance of account. In particular, the current_account table contains the following:

- Current account number
- Customer identification number
- Balance
- Modified date

3.2.3.4 Saving_account table

The saving_account table stores data regarding savings accounts. The following fields are used to store the data:

- Savings account number
- Customer identification number
- Balance
- Modified date

3.2.3.5 Transfer Table

The transfer table stores information regarding money transfer transaction conducted by users in the following fields:

- Amount transfer
- Customer identification number
- Savings account number
- Current account number
- Transfer date

3.3 Develop Methods and Approaches

The function of develop methods and approaches is to establish relationship between all elements into one general idea. There is a view on components of speech recognition in VBBS used and it tells about the sequence of how voice is capture and response to users. And the most important part is the VBBS process flow and it has focuses to more detail by having the combination of speech component and VBBS process flow

3.3.1 Components of Speech Recognition in VBBS

Basically, speech recognition systems have two main functions: to understand the words being spoken and then convert them into text for further use, and to convert text to speech for the purposes of information access. It allows a convenient and natural way to use voice as an input mechanism (and is thus interactive). By the way, VBBS have three basic components in a way to generate speech:

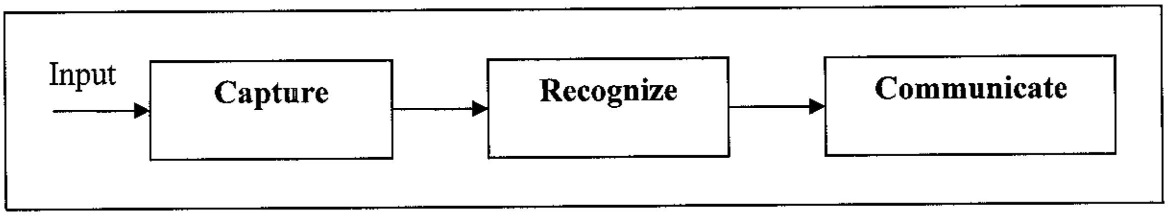


Figure 3.2 Component of Speech Recognition in VBBS

1. Capture and preprocessing. Speech is an analog signal and it needs to be captured and converted into digital format (spectral representation and segmentation). Once the speech is captured, computer algorithms analyze acoustic signals and recognize common sound patterns called phonemes. Any language can be broken down into phonemes including the sounds of the individual letters and their pairings, such as *ao* and *sh*. Words, phrases and sentences can be represented digitally as sequences of these phonemes.

2. Recognition and feature extraction. Once the speech signal is segmented, phoneme probability (i.e., the probability that a particular phoneme represents what was spoken) is calculated. Based on the statistical analysis, words are reconstructed by matching the phonetic sounds to the lexical database. Complex neural network programs are being used to accurately predict the sequence of the words that make up a conversion or a sentence.

3. Communication with other application software and hardware. After speech input has been identified, it is communicated to the application software or speech-aware applications for further processing.

3.3.2 Primary VBBS Approach Flow

Based on the research, author has identified the whole process works in the system.

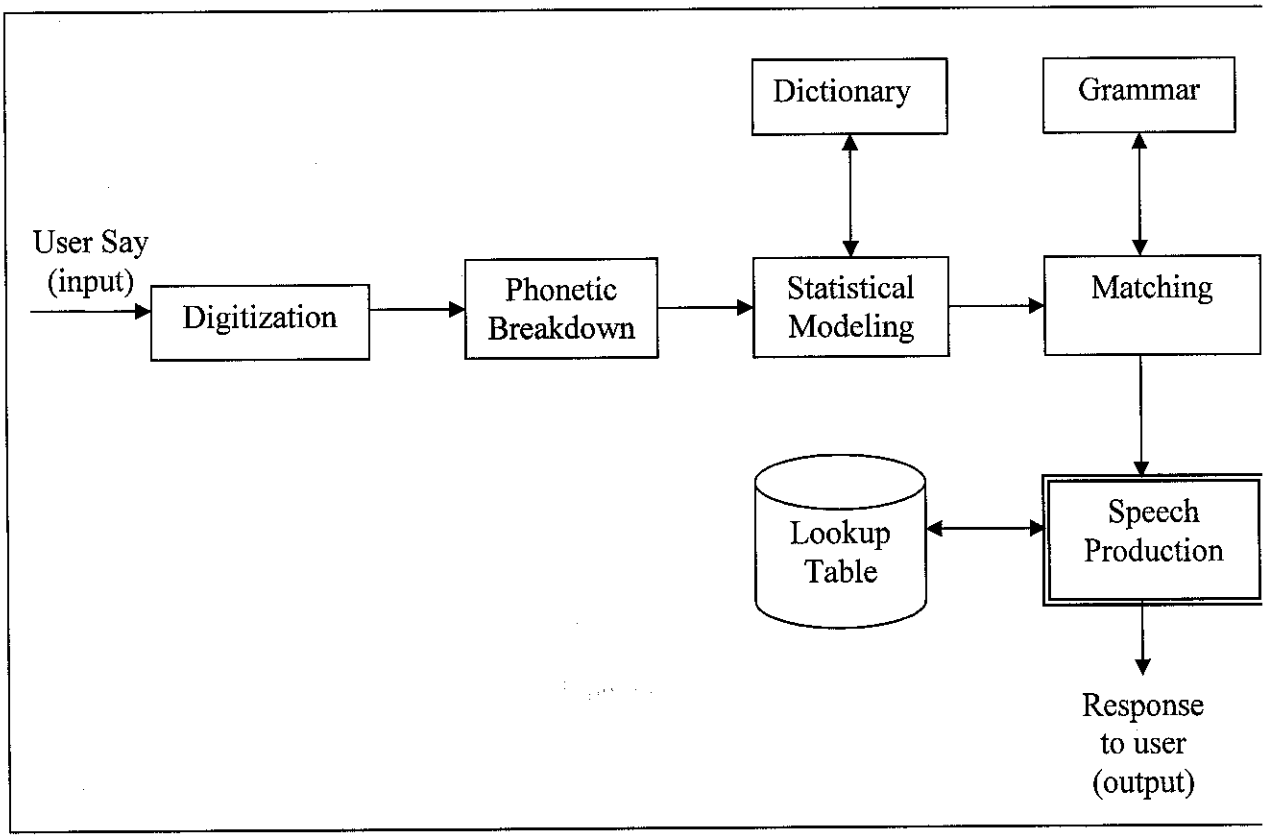


Figure 3.3 VBBS Process Flow

Step1. User input. The user says a word or phrase (or a combination of words and phrases) using an input device such as a telephone or wireless phone. The system captures the speech (in its analog form) in the form of an acoustic signal.

Step2. Digitization. The analog speech signal is converted into a digital signal so the computer can further process it. The conversation process takes into account the acoustic properties of the human ear.

Step3. Phonetic Breakdown. The speech recognition software breaks down the digital signal into basic components of speech-consonant or vowel sounds. *Phonemes* are the single distinctive speech sound of a language-its primary unit. Phonemes combine to form syllables and the syllables combine to form words. Each phoneme is distinguished by its own unique pattern in the spectrogram. For voiced phonemes, the signature involves large concentrations of energy called *formants*; within each formant, and typically across all active formants, there is a characteristic waxing and waning of energy in all frequencies, which is the most salient feature of human voice. But unfortunately the author was not being able get spectrogram to analyze phonemes in a waveform. However, based on the research this spectrogram is not possible to distinguish phonemes in a waveform, if the waveform is chopped into frequency components, it can be viewed as the spectrogram.

Step4. Statistical Modeling. The system then tries to match these sounds to their phonetic representations. A dictionary is the phonetic representation of words that are used by the recognition engine to recognize speech.

Step5. Matching. The system application then tries to match the possible phonetic representations to words or phrases defined in the grammar of that application. For example the user might say "balance" to check balance in the account statement. But if the word balance hasn't been defined as acceptable synonym for balance grammar detector, the application will not be able to accept the utterance.

Step 6. Speech Production. The speech production is the main area in producing speech, in a way to give information and response to users' inquiry. This essential step is the research area that will explain detail in the section 3.4.4.

3.3.3 Combination of Speech Component and VBBS Approach Flow

Below is the architecture of the overall system execution and procession. This combination of speech component and VBBS process flow can give a better view and ease to identify the flow of the system development idea. The relationship between three phases differs for each and the system will eliminate if there is incomplete requirement performed. It contains three main portion of speech component and, speech production is the main topic research area.

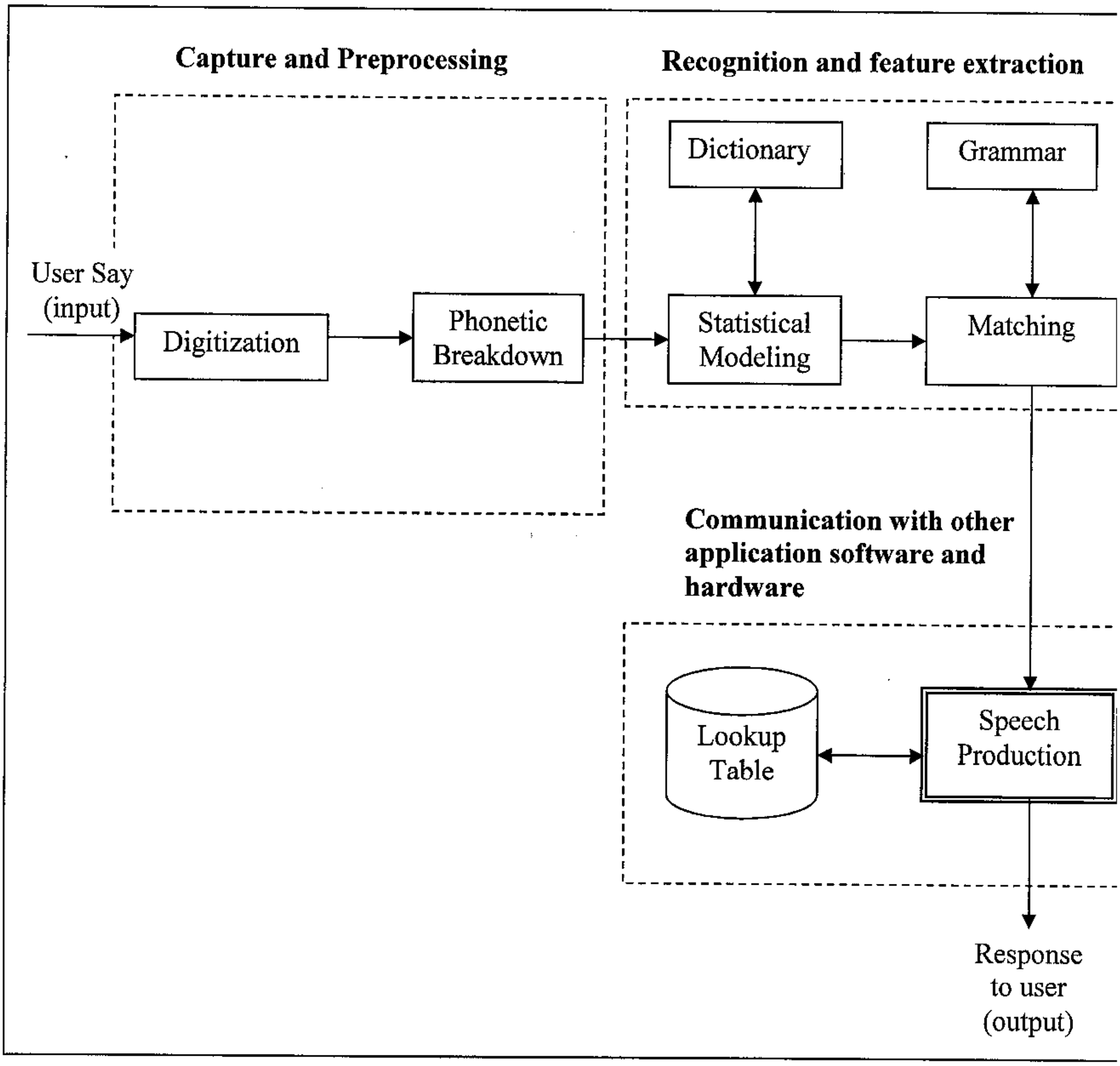


Figure 3.4 Combinations of Speech Component and VBBS Approach Flow

3.3.4 Model Architecture of Speech Production

The architecture model assumed by the specification which has the following components:

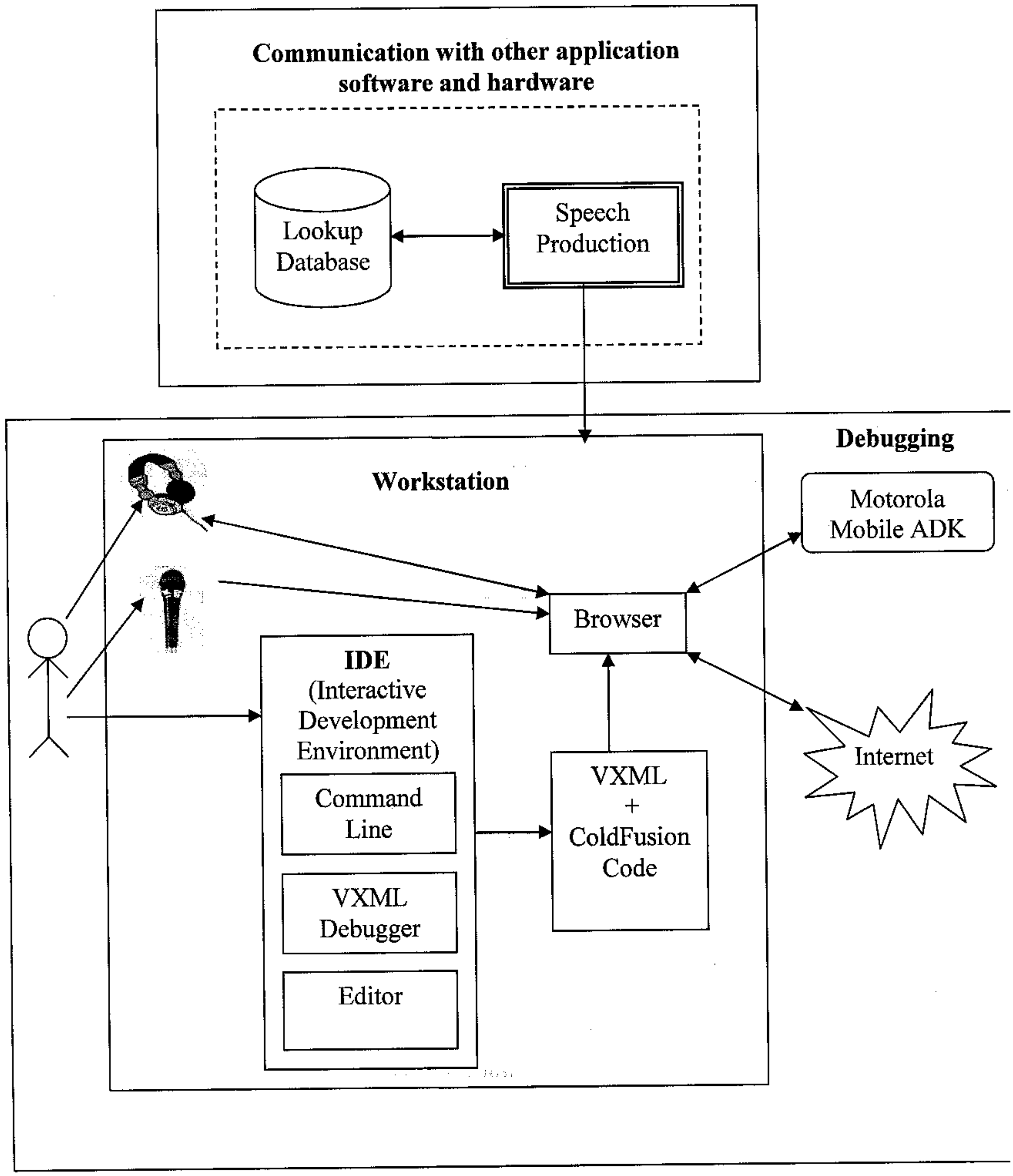


Figure 3.5 VBBS Model Architecture of Speech Production

As VBBS run in standalone configuration, the "gateway" runs on the developer's own workstation. In this testing development, the users will uses a headset and microphone to listen and speak, and all speech synthesis and recognition is done on the workstation. Interactive Development Environment consists of command line, VXML debugger and editor. Those three elements are essential in a mean of developing GUI. Then another spectrum of speech production architecture is VXML and Cold Fusion. On this part, it is the main requirement that can execute and process features from IDE. The acceptance of structure from code execution will enter to the browser segment, which require to install specific software (Motorola Mobile ADK) locally and launch the browser from the command line or from within an IDE. Debugging session will occur in the Motorola toolkit. It has the capability to execute the program and act as a gateway to process input and output of the speech configuration. To experience how the toolkit works, it shows as in APPENDIX B. Therefore, these are where all the important part combining and integrate together in order to create or to support speech production phase.

3.3.5 System Programming

The decision made on using VXML and Cold Fusion is based on the analysis during functional analysis phase. There are several ideas and several tools (programming) to use, but then it should be match with the main purpose of this project area which is developing voice-based application and user will apply it with using telephone or mobile phone devices. Thus, VoiceXML has several advantages that might help to implement in the phone application. As the researchers mentioned, different people have different key of voice tones, accents, and frequency and VoiceXML has the ability to fix the grammar and it uses the <grammar> element to define grammars for form items and other elements.

Grammar defines the keywords and a set of acceptable inputs that a user can use to interact with the application. They pave the way for users' interactions with such application. While writing grammar of VBBS, there is an important element that should take into consideration which is to ensure that user isn't forced to interact with a system full of computer-directed dialogs. Hence, this grammar used to extract the desire phrases from the given input and it is the key role that plays in enabling the user to interact with a voice application through natural speech.

For the Cold Fusion (CF), it comes with all the hooks necessary to link to almost any database application and any other external system. Apart from that, Cold Fusion is fast, scalable, multithreaded, service-based architecture and easy to maintain because no compilation or linking step is required

3.4 Testing and Implementation

Programming used are VoiceXML and Cold Fusion and this combination able to illustrate the flexibility of VoiceXML technology by presenting a server-side voice application that integrate the technologies of both language. There will be automatically link through the grammar defined by the system and able to do checking automatically.

As this system is standalone mode, Motorola is providing the best environment that found for this testing or debugging stage. It was the Mobile ADK from Motorola, Beta 3, which provides a simulated Voice Gateway which can run off system development machine. It is great for debugging, since don't have to use your phone, nor upload files to servers and so it saves a lot of time and it had to add the type= "application/x-gsl" to the <grammar>'s.

3.4.1 Tool Required Supporting Programming and Testing Phase

In performing this project, several tools and equipment are needed. Following are minimum requirement needed to perform this project and to use the application.

3.4.1.1 Hardware Requirements

1. Personal Computer/Desktop
   - Window XP
   - Pentium 4 processor 1.4 GHz or higher
   - 64 MB RAM memory or higher
   - 1 GB hard disk space or higher

2. Audio Tools
   - Speaker or Headphone
   - Microphone
   - Sound Card

3.4.1.2 Software Requirements

1. Software and Programming Tool
    - VoiceXML 1.0 and 2.0
    - Cold Fusion MX7
    - Motorola Mobile Application Development Kit (MADK) version 3.0 (Mobile ADK for voice)
    - Microsoft Sound Recorder 5.1 (Audio Processing Tool)

CHAPTER 4

RESULT AND DISCUSSION

4.1 Research Findings

Upon development of this project, the biggest findings were the adoption of two technologies that can work together in developing VBBS. In the model architecture of speech production is the most fundamental platform for next developer to host it and implement it in bank institution.

Journal from previous researcher helped the extraction of the methodology of the project and though the language or programming is still new and should be testing further in order to be involved from many in the future. Researchers also said that there is one criterion that can caused failure to any voice recognition system which is background noise. Hence, it is one of the other studies to develop a new method to eliminate unimportant noise from the voice recognition engine and it can be part of the research area of project in the future.

Based on the interview done with Mr. Ikhwan Essendi bin Nazaruddin, Software Engineer, Motorola (Cyberjaya), he mentioned about the security system should implement in the voice-based telephony system as it unable to unauthorized user to intrude the system. There are issues on either to use voice or other method as password to enter the system and Mr. Ikhwan suggested that the best choice to protect personal account for voice-based system is by using DTMF for user to first logon their password and ID and thus can proceed to their activities safely.

Additionally, there are four phases of system methodology that developer should perform and for the testing phase Mobile Application Development Kit 3.0 from Motorola has chosen to be the best product to test the application as the VBBS is a standalone configuration.

Generally, VBBS is a two way communication system, but then, there are still other factors need to consider which is the technique of user say it. User should be trained and know how to say word. The concept inspired from the grammar mapping used in the VXML and the system rely on this essential factor. It is not what user say, but how users say words.

From the survey questionnaires distributed, it shows that 100% of people are having mobile phone and 80% are preferred and agree to use voice based as their alternative to perform bank activities with their mobile phone. 70% of them are not preferred to use internet banking with a reason of issues on Internet hackers, theft, and so forth. Then they also comment that they need to be tied to the physical network to access information and perform banking task and this is not suitable for busy people and do not have extra time to go to bank.

With that, it is proved that Malaysia's bank institution should accept and adapt voice-based system as one of their new services provider offer to customer. And survey done resulted that consumers or users have such awareness and positively affect if this product implement to bank institution in Malaysia.

4.2 Result Discussion

In meeting with the project objectives, the end product prototype should be able to produce a result triggered from word said by user. This prototype will process a detail of account statement from each user's account.

This system fulfilled with security system which requires the user first logs on to the system by furnishing the assigned customer identification number and personal identification number before user proceed for further actions. The system will evaluate the customer ID by number keyed to the system. This is to ensure that people surround would not be able to get and use the same ID to perform bank activities without permission. All the detail description of the system functional testing is in section 4.2.1.

4.2.1 Functional Testing

All the code snippets are based on the VoiceXML and Cold Fusion and fully tested in a desktop PC using Motorola Application Development Kit version 3.0. Thus, it needs to run the server (CF server) at first, and execute the system in the MADK. Based on the system tester feedback, user is advised to have well trained with the system so that system will familiar with voice input. That is one of the system limitation and further research need to enhance this element and have the capability to execute voice as the user instructed. All the system interface views are in the Appendix B.

As stated earlier in the functional analysis, developer provide two types of communication with the system which via DTMF and voice to process information. Hence, below of the system it provides the keypad view for user to enter the number or the instruction of the system. It will facilitate user to input their password and able to secure their account instead of using voice to transfer the password which might allowed people surround hear the secret identification number.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

This project simultaneously contributes the automatic of voice recognition system, Text-to-Speech (TTS), which VBBS encourage user to use DTMF element, grammar checking and manipulation of both VoiceXML and Cold Fusion language. The examined on flexibility of use both language are proved and can be apply in host configuration (voice web browsing), which contains PSTN, VoiceXML Gateway, Web and Application Server and Internet.

Also, illustrated in this project is the general view of voice production architecture and how it works. This essential architecture also discuss successfulness in portraying the combination and ability several element work together.

Finally, from the research and development done, the objectives that are stated earlier are reached successfully. And hope the effort performed will later become a stepping stone for an advance level of VBBS that uses or involved many users and perhaps able to promote this product to all bank in Malaysia. And hence, Bank Institution might treat this new marketing strategy as to be different than other rival and will attract many customers to use it.

5.2 Recommendation

Further research is needed to get good in analyzing accent, intonation and noise from customer. It is possible can produce wrong output from the system if all those criteria not taken seriously in developing the system.

And apart from fulfilling the goals set at the beginning, several other interesting possibilities for extensions have opened up during the course of this project. The database collection process could be modified to support scalable condition. Scalable here means the ability of system to support many customers in a time and for each customer activities should produce such accurately and stable of result.

For the security features, two locations that would secure the system either to put the security feature at the client site or server site. And for VBBS, it concentrates only at client site, which is use DTMF. Hence, for future developer it is encouraged to put it on two sites which can shield system more accurate and confident.

Lastly, it would be much more efficient to have a second version of VBBS that can test it with using real phone and can only accept if customer speaks a Standard English.

REFERENCES

[1] Alan W Black and Paul Taylor, 1994, *A Generic Speech Synthesis System*, ATR Interpreting Telecommunications Laboratories, University of Edinburgh.

[2] Alan W Black, 1995, *Predicting the Intonation of Discourse Segments from Example in Dialogue Speech*, ATR Interpreting Telecommunications Laboratories, University of Edinburgh.

[3] Paul Taylor, Simon King, Stephen Isard and Helen Wright, 1998, *Intonation and Dialogue Context as Constraints for Speech Recognition*, Centre for Speech Technology Research, University of Edinburgh.

[4] Ben Gillett, 2003, *Transforming Voice Quality and Intonation*, University of Edinburgh.

[5] Simon King, Todd Stephenson, Stephen Isard, Paul Taylor and Alex Strachan, 1998, *Speech Recognition Via Phonetically Featured Syllables.*

[6] Yoshinori Shiga and Simon King, *Accurate Spectral Envelope Estimation for Articulation-to-Speech Synthesis*, Centre for Speech Technology Research, University of Edinburgh, U.K.

[7] Y. Shiga and S. King, Sept. 2003, *Estimation of Voice Vource and Vocal Tract Characteristics based on Multi-Frame Analysis*, in *Proc.Eurospeech2003*.

[8] Bryan Duggan and Mark Deegan, *Considerations In The Usage Of Text To Speech (TTS) In The Creation Of Natural Sounding Voice Enabled Web Systems*, National College of Ireland.

[9] Sharma, C. Kunnins, 2002, *VoiceXML: Strategies and Techniques for Effective Voice Application Development with VoiceXML 2.0,* John Wiley & Sons; 1st edition.

[10] Alan W Black, Joke Van de Plassche, Briony Williams, 1991, *Analysis of unknown words through morphological decomposition*, University of Edinburgh and University of Nijimegan.

[11] Yousif A. El-Imam, Zuraida Mohammed Don, Feb 2000, *International Journal of Speech Technolgy.*

[12] Mark Miller, Kenneth R. Abbott, 2002, VoiceXML 10 Projects to Voice Enable Your Web Site.

[13] Stephen Breitenbach, 2002, Early Adopter VoiceXML.

[14] James A Larson, Larson technical service, 2002, VoiceXML: Introduction to Developing Speech Applications.

[15] http://www.roadnews.com/html/Articles/a37.htm

[16] http://www.voicexml.org/

[17] http://www.w3.org/MarkUp/Forms/

[18] http://dmoz.org/Computers/Speech_Technology/Telephony/

[19] www.mot.com.my

[20] http://www.oreillynet.com/pub/a/oreilly/web/news/coldfusion2_0901.html

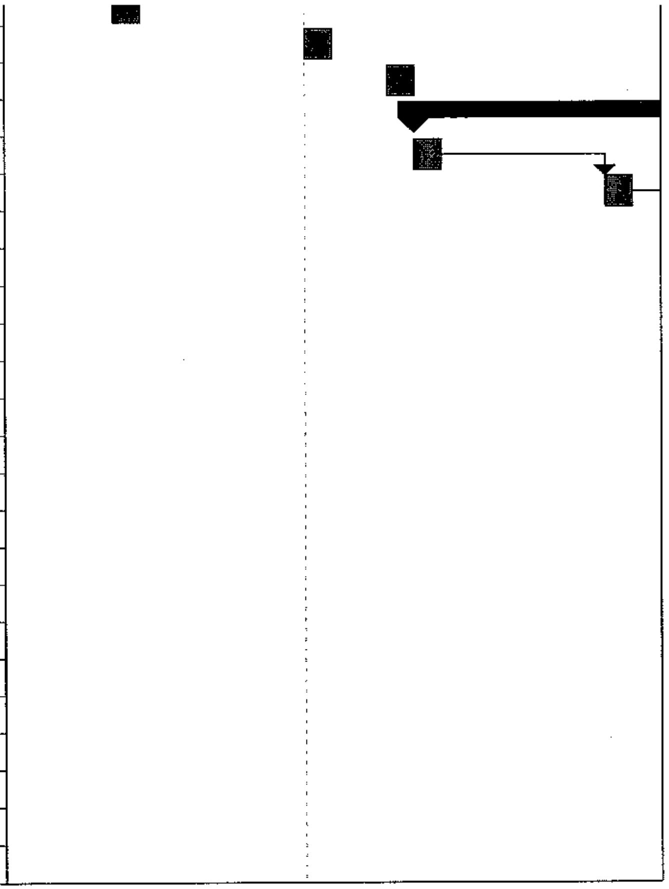
[21] http://www.jan.wvu.edu/media/speechrec.html

[22] www.w3.org/Voice/

APPENDICES

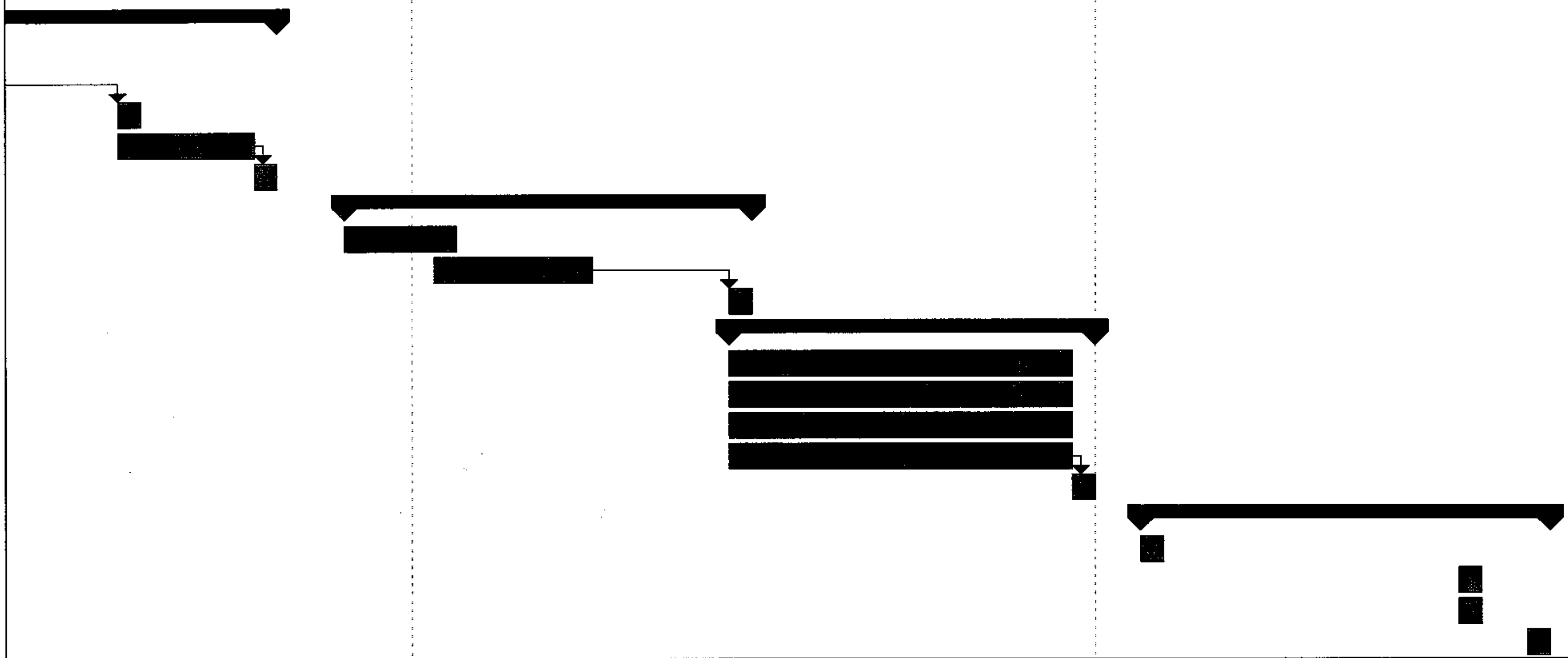
APPENDIX A: PROJECT TIMELINE/GANTT CHART

2		1.1 Proposal Submission	1 day	Mon [illegible]	Mon [illegible]
3		1.2 Preliminary Report Submission	1 day	Mon 8/1/05	Mon 8/1/05
4		1.3 Project Scheduling	1 day	Thu 8/4/05	Thu 8/4/05
5		**2.0 Project Planning and Functional Analysis**	**15 days?**	**Fri 8/5/05**	**Thu 8/25/05**
6		2.1 Data Gathering	1 day?	Fri 8/5/05	Fri 8/5/05
7		2.2 Data Analysis	1 day	Fri 8/12/05	Fri 8/12/05
8		2.3 Method Generation	1 day	Fri 8/19/05	Fri 8/19/05
9		2.4 Requirement Specification	4 days	Fri 8/19/05	Wed 8/24/05
10		2.5 Progress Report Submission	1 day	Thu 8/25/05	Thu 8/25/05
11		**3.0 System Design**	**14 days**	**Mon 8/29/05**	**Thu 9/15/05**
12		3.1 Interface Design	5 days	Mon 8/29/05	Fri 9/2/05
13		3.2 System Design Modelling	5 days	Fri 9/2/05	Thu 9/8/05
14		3.3 Data Structure Design	1 day	Thu 9/15/05	Thu 9/15/05
15		**4.0 Programming**	**12 days**	**Thu 9/15/05**	**Fri 9/30/05**
16		4.1 Interface Development	11 days	Thu 9/15/05	Thu 9/29/05
17		4.2 Algorithm Generation	11 days	Thu 9/15/05	Thu 9/29/05
18		4.3 System Programming	11 days	Thu 9/15/05	Thu 9/29/05
19		4.4 System Intergration	11 days	Thu 9/15/05	Thu 9/29/05
20		4.5 SV Final Draft Submission	1 day	Fri 9/30/05	Fri 9/30/05
21		**5.0 Implementation Phase**	**14 days**	**Mon 10/3/05**	**Thu 10/20/05**
22		5.1 System Implementation	1 day	Mon 10/3/05	Mon 10/3/05
23		5.2 Product Demonstration	1 day	Mon 10/17/05	Mon 10/17/05
24		5.3 System Delivery	1 day	Mon 10/17/05	Mon 10/17/05
25		5.4 Final Delivery	1 day	Thu 10/20/05	Thu 10/20/05



Project: proj planning
Date: Tue 12/20/05

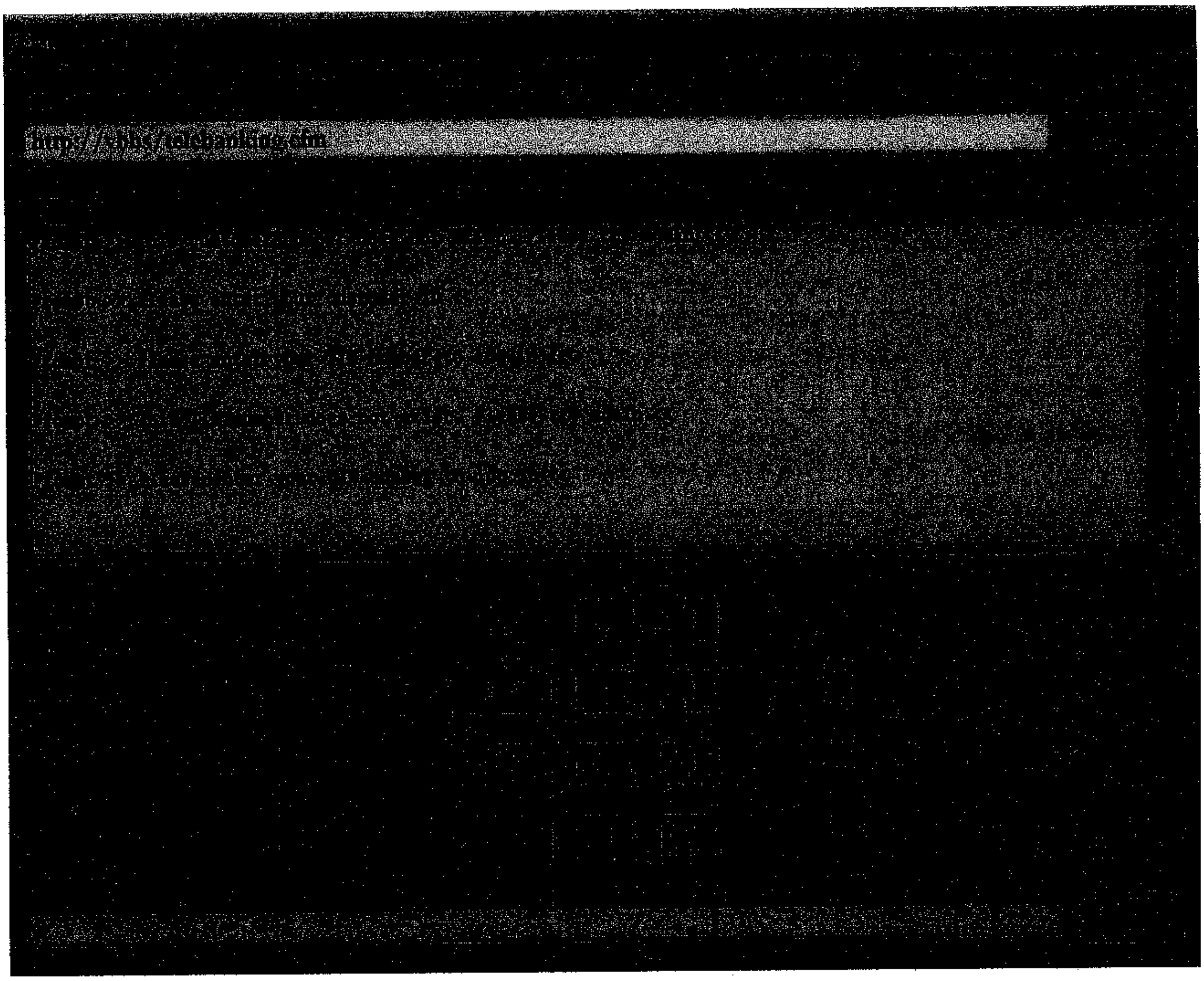
Task
Split
Progress
Milestone
Summary
Project Summary
External Tasks
External Milestone
Deadline



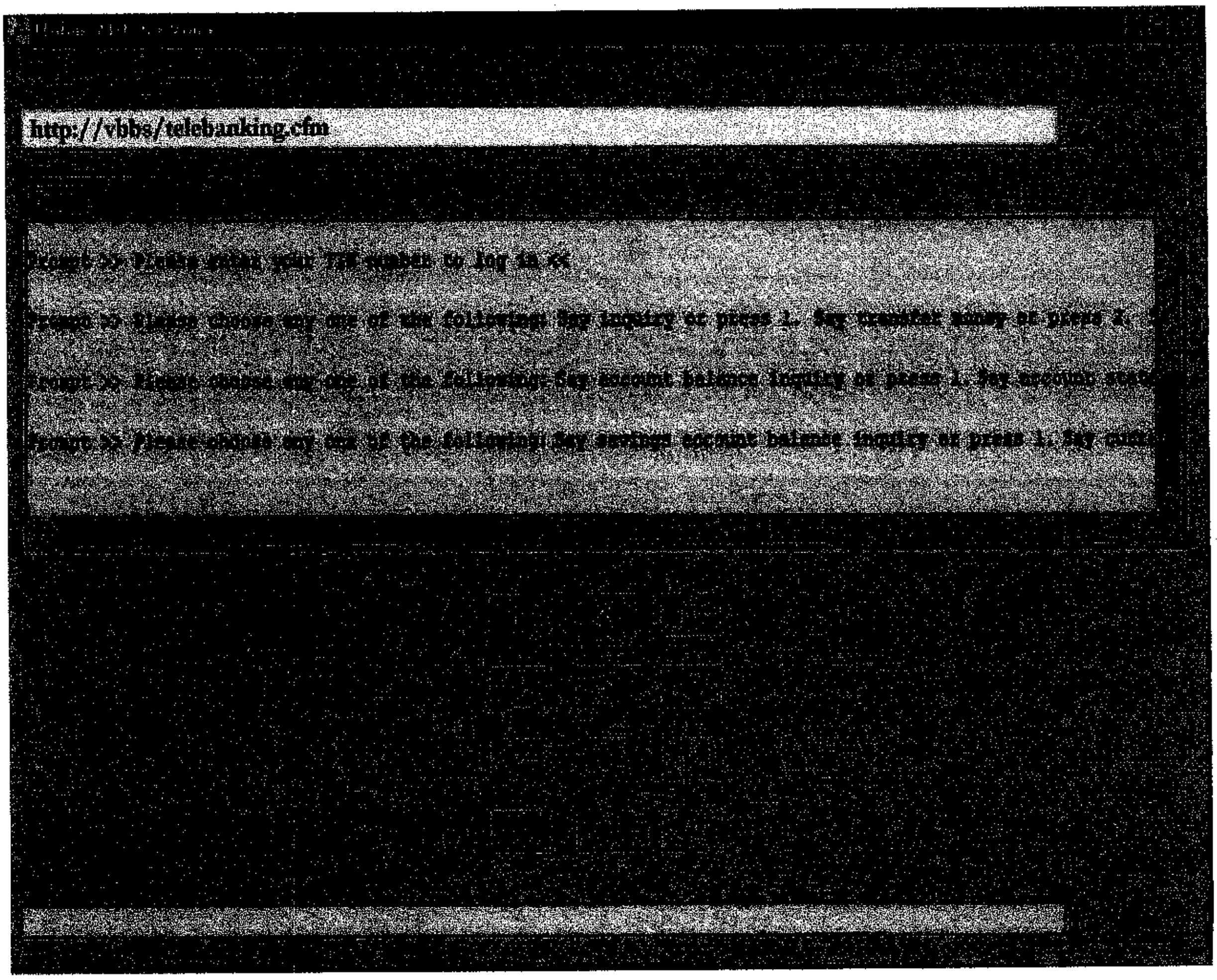
Task		Milestone		External Tasks	
Split		Summary		External Milestone	
Progress		Project Summary		Deadline	

Project: proj planning
Date: Tue 12/20/05

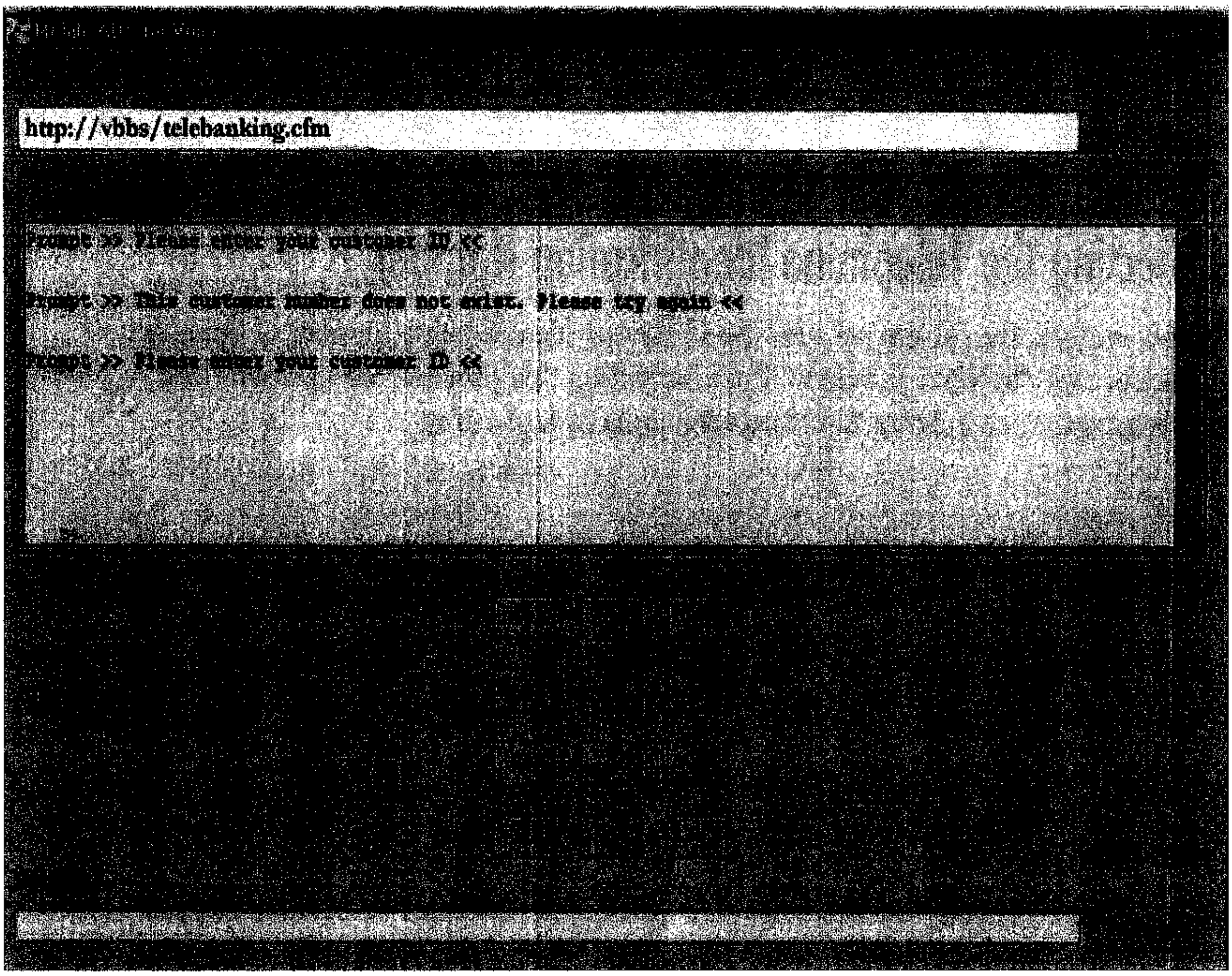
APPENDIX B: SYSTEM INTERFACE



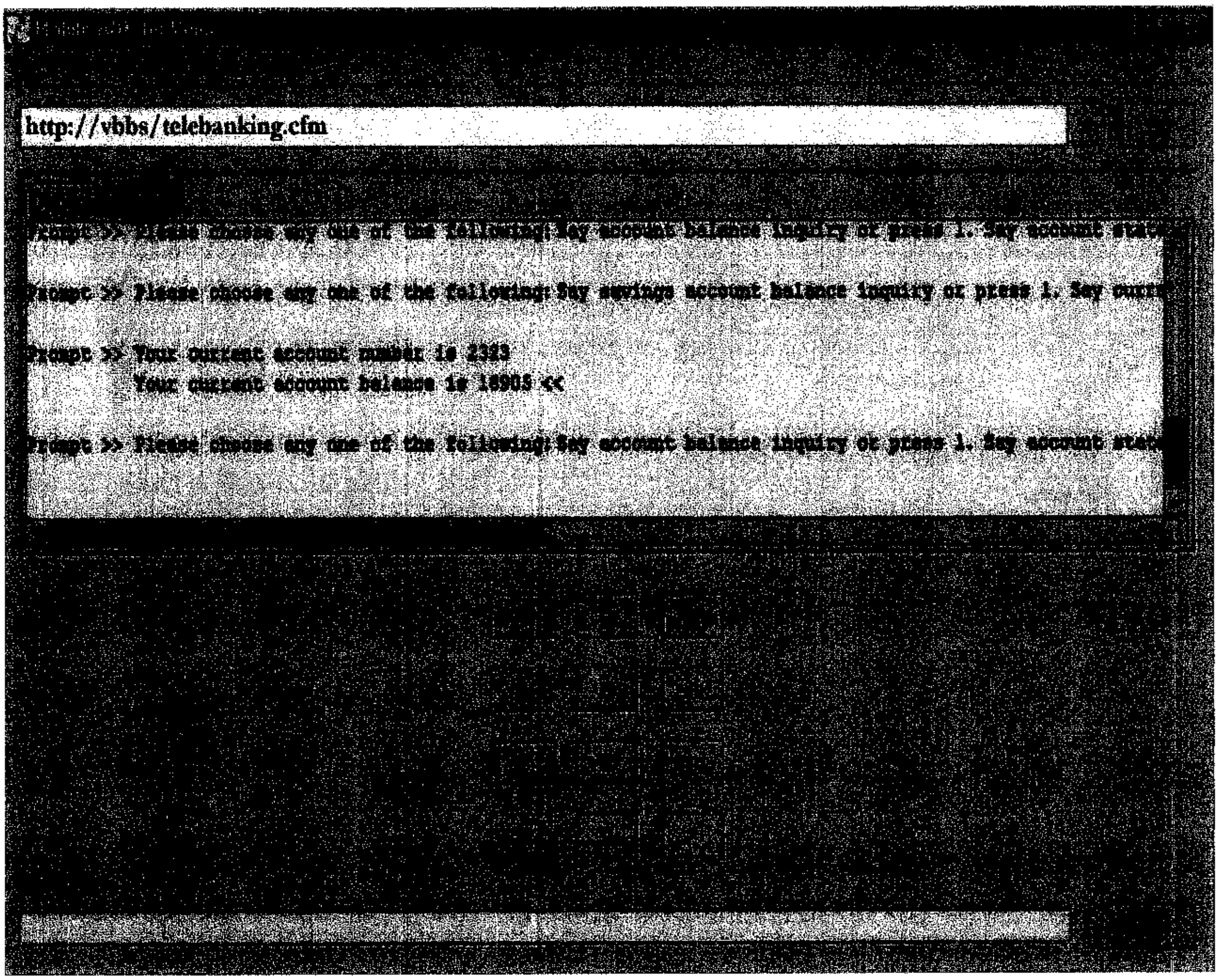
<u>Output if the authentication fails</u>



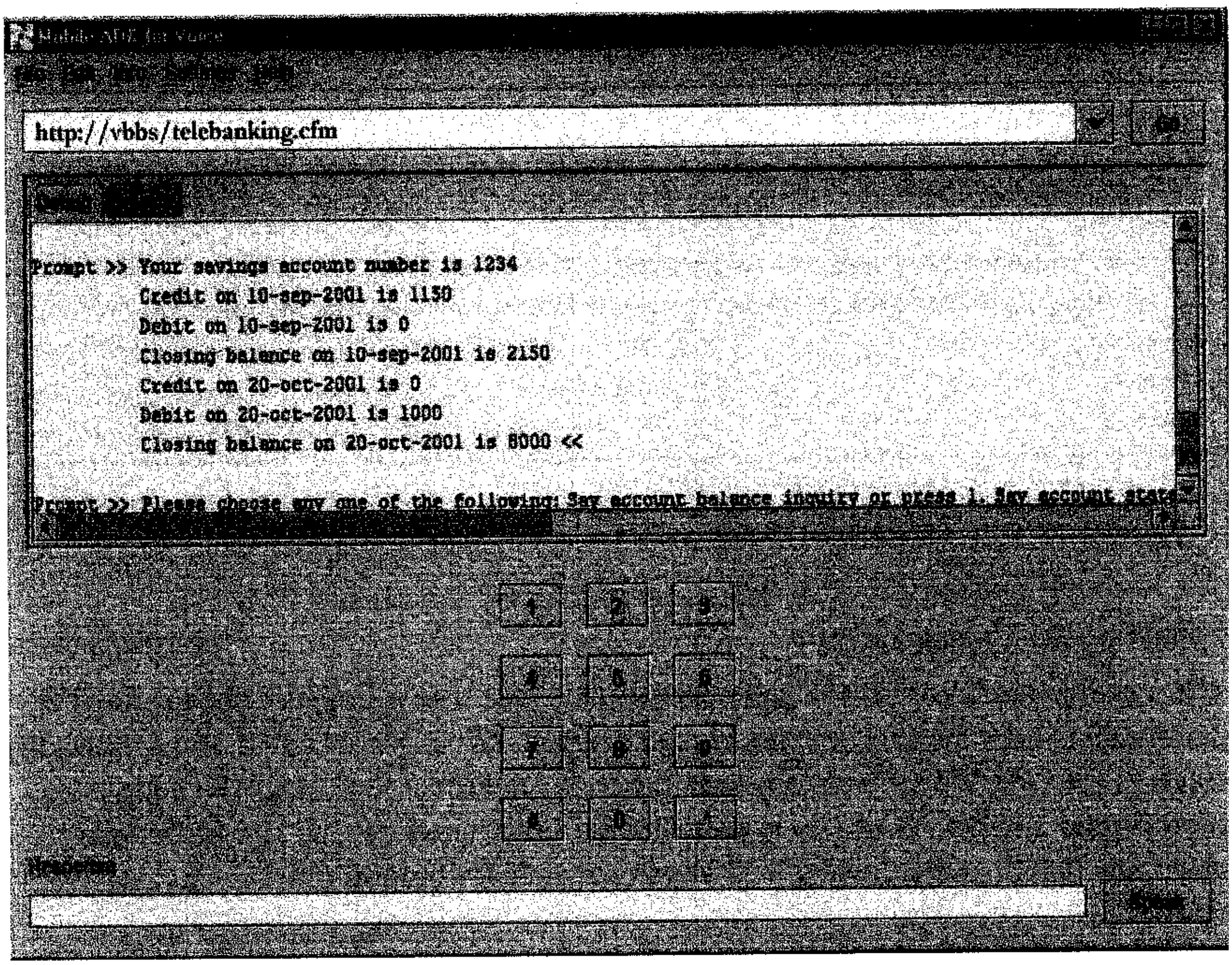
Output of balance enquiry



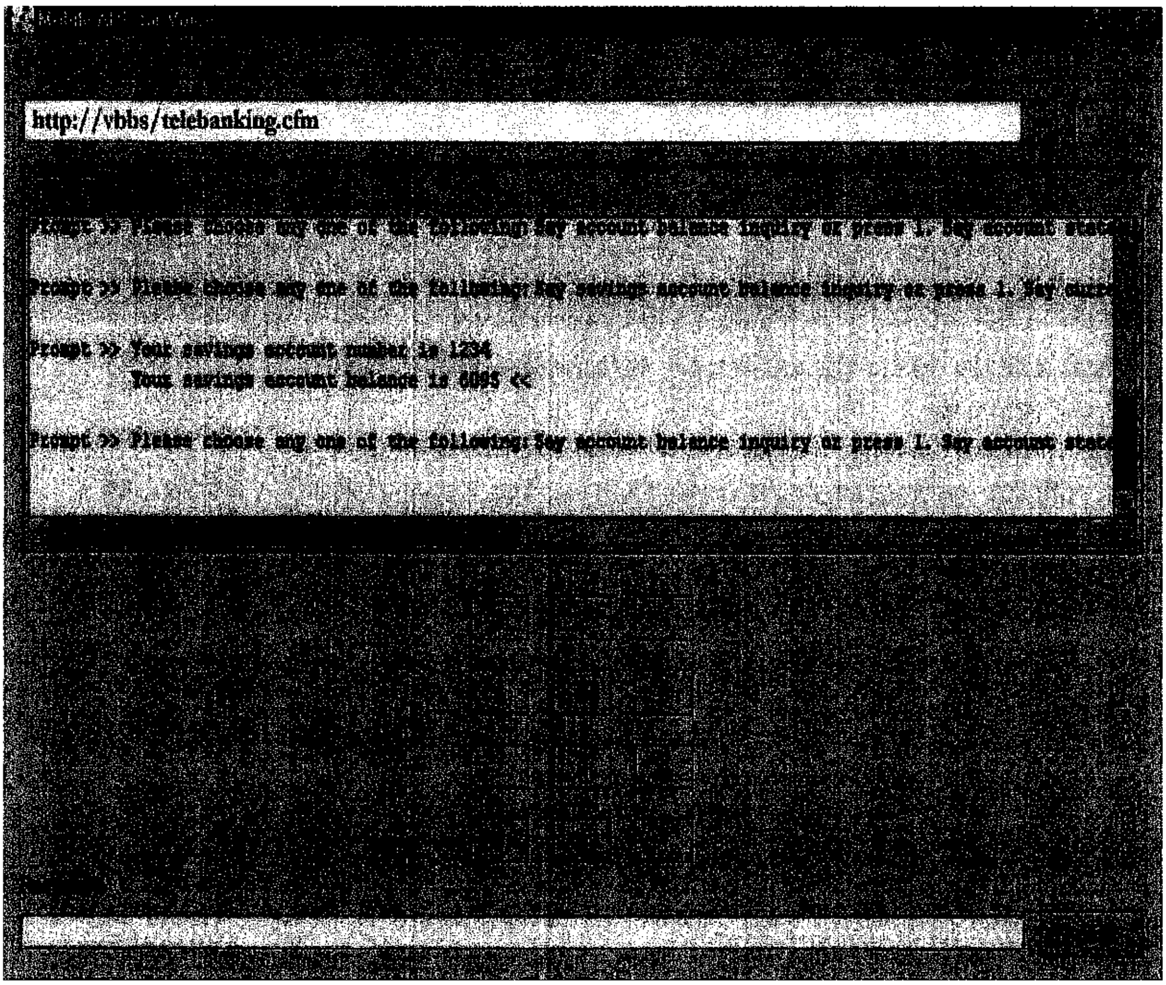
<u>Output of customer login (in case of nonexistent record)</u>



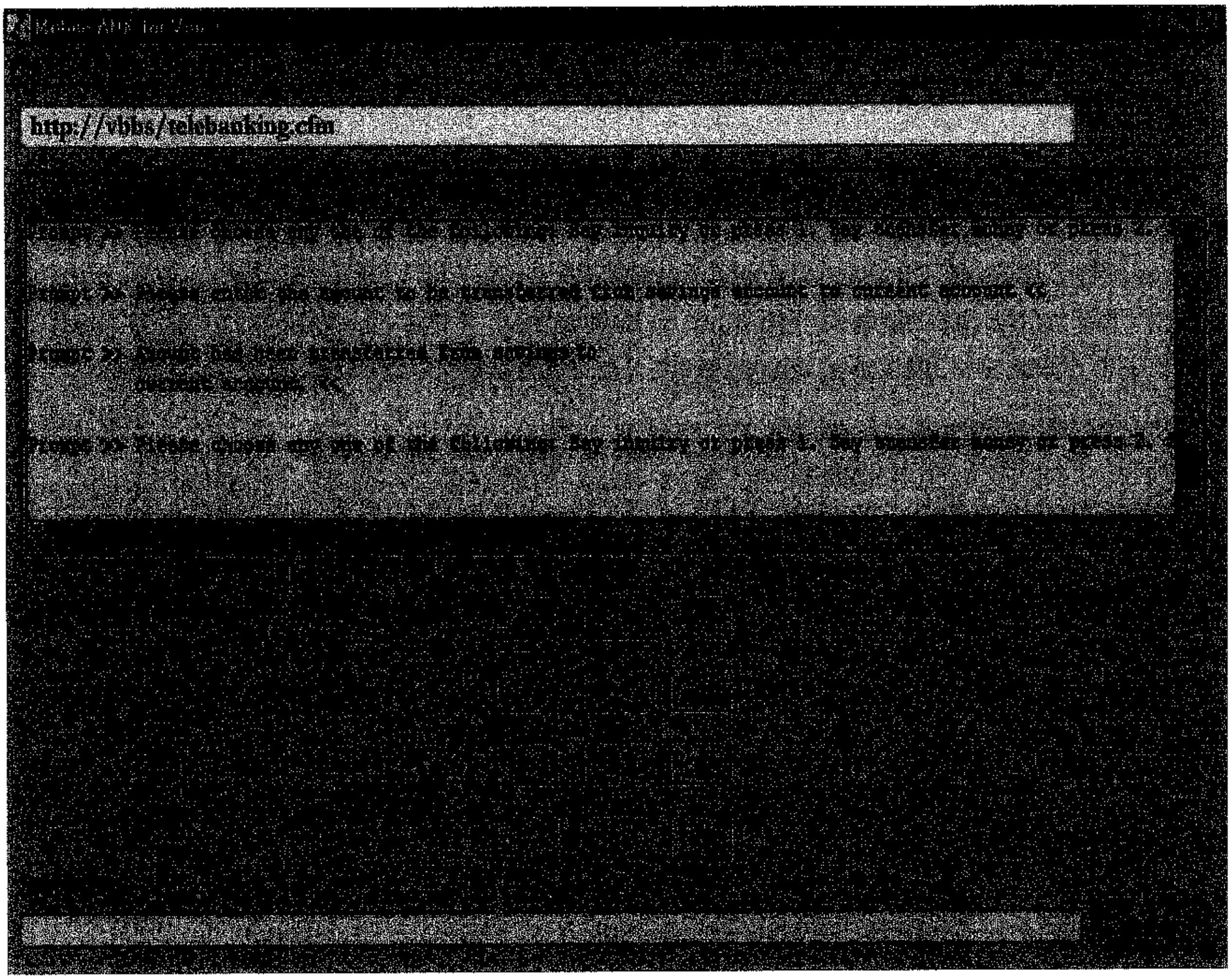
Output of current account



Output of account statement, only for saving account



<u>Output of saving account number</u>



Output of transfer money

APPENDIX C: SYSTEM CODING

1) SOURCE CODE FOR TELEBANK.CFM

```
<cfcontent type = "text/xml">
</cfcontent>
<vxml version="1.0">
<var name="custno"/>
<form id="user">
<field name="custno" type="digits">
<prompt>Please enter your customer ID</prompt>
<filled>
<assign name="document.custno" expr="custno"/>
<goto next="#call_login"/>
</filled>
</field>
</form>
<form id="call_login">
<block>
<var name="custnumber" expr="document.custno"/>
<submit next="custlogin.cfm" method="get" namelist="custnumber"/>
</block>
</form>
</vxml>
```

2) SOURCE CODE FOR CUSTLOGIN.CFM

```
<CFAPPLICATION NAME="TELEBANKING" SESSIONMANAGEMENT="yES">
</CFAPPLICATION>
<CFQUERY NAME="cust" DATASOURCE="teleDSN" DBTYPE="ODBC">
SELECT customer_no FROM CUSTOMER WHERE customer_no =
#url.custnumber#
</CFQUERY>
<CFSET SESSION.custnum = url.custnumber></CFSET>
<CFSET CUSTNO = cust.recordCount></CFSET>
<cfcontent type = "text/xml"></cfcontent>
<vxml version="2.0">
<form>
<block>
<cfif CUSTNO is 0>
<CFOUTPUT>This customer number does not exist.
Please try
again</CFOUTPUT>
<goto next="telebank.cfm"/>
<cfelse>
<CFOUTPUT><goto next="login.cfm"/></CFOUTPUT>
</cfif>
</block>
</form>
</vxml>
```

3) SOURCE CODE FOR LOGIN.CFM

```
<cfcontent type = "text/xml">
</cfcontent>
<vxml version="2.0">
<var name="tinno"/>
<form id="user">
<field name="tinno" type="digits">
<prompt>Please enter your TIN number to log in</prompt>
<filled>
<assign name="document.tinno" expr="tinno"/>
<goto next="#call_login"/>
</filled>
</field>
</form>
<form id="call_login">
<block>
<var name="tinnumber" expr="document.tinno"/>
<submit next="telelogin.cfm" method="get" namelist="tinnumber"/>
</block>
</form>
</vxml>
```

4) SOURCE CODE FOR TELELOGIN.CFM

```
<CFQUERY NAME="tin" DATASOURCE="teleDSN" DBTYPE="ODBC">
SELECT TIN_no FROM CUSTOMER WHERE TIN_no = #url.tinnumber#
</CFQUERY>
<CFSET TINNO = tin.recordCount></CFSET>
<cfcontent type = "text/xml"></cfcontent>
<vxml version="2.0">
<form>
<block>
<cfif TINNO is 0>
<CFOUTPUT>This TIN number does not exist. Please
try
again</CFOUTPUT>
<goto next="login.cfm"/>
<cfelse>
<CFOUTPUT><goto next="menu.cfm"/></CFOUTPUT>
</cfif>
</block>
</form>
</vxml>
```

5) SOURCE CODE FOR MENU.CFM

```
<cfcontent type = "text/xml"></cfcontent>
<vxml version="1.0">
<link next="http://vbbs/telebanking/menu.cfm">
<grammar type="application/x-gsl">
Home
</grammar>
</link>
<menu>
<prompt>Please choose any one of the following:<enumerate/></prompt>
<choice dtmf="1" next="enquiry.cfm">Inquiry</choice>
<choice dtmf="2" next="transfer_money.cfm">Transfer money</choice>
statement</choice>
<choice dtmf="3" next="#exit">Exit</choice>
<help>
Press 1 for inquiry
Press 2 to transfer money
Press 3 to exit from the application
</help>
<noinput>I didn't detect anything, please enter<enumerate/></noinput>
<nomatch> I didn't get that, <enumerate/></nomatch>
<catch event="nomatch noinput" count="4">
<prompt>You have exceeded the number of allowed retries. System will
now stop
the application.</prompt>
<throw event="telephone.disconnect.hangup"/>
</catch>
</menu>
<form id="exit">
<catch event="exit">
<throw event="telephone.disconnect.hangup"/>
</catch>
</form>
</vxml>
```

6) SOURCE CODE FOR ENQUIRY.CFM

```
<cfcontent type = "text/xml"></cfcontent>
<vxml version="2.0">
<link next="http://vbbs/telebanking/login.cfm">
<grammar type="application/x-gsl">
Telebanking
</grammar>
</link>
<menu>
<prompt>Please choose any one of the following:<enumerate/></prompt>
<choice dtmf="1" next="balance_enquiry.cfm">account balance inquiry</choice>
<choice dtmf="2" next="statement.cfm">account statement inquiry</choice>
<choice dtmf="3" next="menu.cfm">return to menu</choice>
<choice dtmf="4" next="#exit">exit</choice>
<help>
Press 1 for account balance inquiry
Press 2 for account statement inquiry
Press 3 to return to the main menu
Press 4 to exit from the application
</help>
<noinput>I didn't detect anything, please choose<enumerate/></noinput>
<nomatch> I didn't get that, <enumerate/></nomatch>
<catch event="nomatch noinput" count="4">
<prompt>You have exceeded the number of allowed retries. System will
now stop the
application.</prompt>
<throw event="telephone.disconnect.hangup"/>
</catch>
</menu>
<form id="exit">
<catch event="exit">
<throw event="telephone.disconnect.hangup"/>
</catch>
</form>
</vxml>
```

7) SOURCE CODE FOR BALANCE_ENQUIRY.CFM

```
<cfcontent type = "text/xml"></cfcontent>
<vxml version="2.0">
<link next="http://vbbs/telebanking/login.cfm">
<grammar type="application/x-gsl">
Telebanking
</grammar>
</link>
<menu>
<prompt>Please choose any one of the following:<enumerate/></prompt>
<choice dtmf="1" next="saving_enquiry.cfm">savings account balance
inquiry</choice>
<choice dtmf="2" next="current_enquiry.cfm">current account balance
inquiry</choice>
<choice dtmf="3" next="#exit">Exit</choice>
<help>
Press 1 for savings account balance inquiry
Press 2 for current account balance inquiry
Press 3 to exit from the application
</help>
<noinput>I didn't detect anything, please choose<enumerate/></noinput>
<nomatch> I didn't get that, <enumerate/></nomatch>
<catch event="nomatch noinput" count="4">
<prompt>You have exceeded the number of allowed retries. System will now stop
the application.</prompt>
<throw event="telephone.disconnect.hangup"/>
</catch>
</menu>
<form id="exit">
<catch event="exit">
<throw event="telephone.disconnect.hangup"/>
</catch>
</form>
</vxml>
```

8) SOURCE CODE FOR SAVING_ENQUIRY.CFM

```
<CFAPPLICATION NAME="TELEBANKING" SESSIONMANAGEMENT="YES">
</CFAPPLICATION>
<CFQUERY NAME="saving" DATASOURCE="teleDSN" DBTYPE="ODBC">
SELECT saving_ac_no,balance FROM saving_account WHERE customer_no =
#SESSION.custnum#
</CFQUERY>
<cfcontent type = "text/xml"></cfcontent>
<vxml version="2.0">
<form>
<block>
<CFOUTPUT QUERY="saving">
Your savings account number is #saving_ac_no#
Your savings account balance is #balance#
</CFOUTPUT>
<goto next="enquiry.cfm"/>
</block>
</form>
</vxml>
```

9) SOURCE CODE FOR CURRENT ENQUIRY.CFM

```
CFAPPLICATION NAME="TELEBANKING" SESSIONMANAGEMENT="YES">
</CFAPPLICATION>
<CFQUERY NAME="saving" DATASOURCE="teleDSN" DBTYPE="ODBC">
SELECT current_ac_no,balance FROM current_account WHERE customer_no =
#SESSION.custnum#
</CFQUERY>
<cfcontent type = "text/xml"></cfcontent>
<vxml version="1.0">
<form>
<block>
<CFOUTPUT QUERY="saving">
Your current account number is #current_ac_no#
Your current account balance is #balance#
</CFOUTPUT>
<goto next="enquiry.cfm"/>
</block>
</form>
</vxml>
```

10) SOURCE CODE FOR STATEMENT.CFM

```
<CFAPPLICATION NAME="TELEBANKING" SESSIONMANAGEMENT="YES">
</CFAPPLICATION>
<CFQUERY NAME="saving" DATASOURCE="teleDSN" DBTYPE="ODBC">
SELECT saving_ac_no,debit,credit,balance,tran_date FROM acc_statement WHERE
customer_no = #SESSION.custnum#
</CFQUERY>
<CFSET ACCNO = saving.saving_ac_no></CFSET>
<cfcontent type = "text/xml"></cfcontent>
<vxml version="2.0">
<form>
<block>
<CFOUTPUT>Your savings account number is #ACCNO#</CFOUTPUT>
<CFLOOP QUERY="saving">
<CFOUTPUT>
Credit on #tran_date# is #credit#
Debit on #tran_date# is #debit#
Closing balance on #tran_date# is #balance#
</CFOUTPUT>
</CFLOOP>
<goto next="enquiry.cfm"/>
</block>
</form>
</vxml>
```

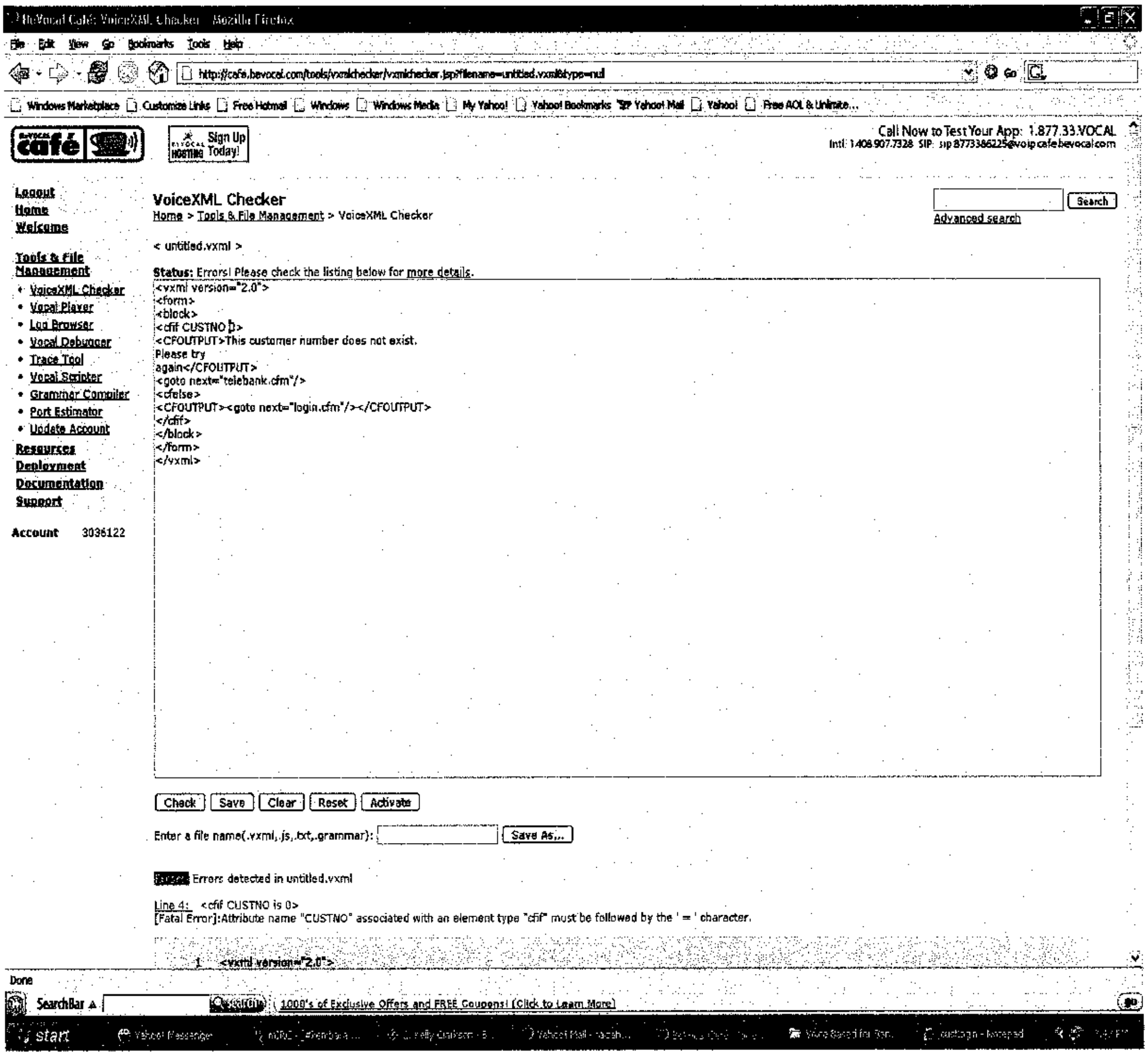
11) SOURCE CODE FOR TRANSFER_MONEY.CFM

```
<cfcontent type = "text/xml">
</cfcontent>
<vxml version="2.0">
<var name="amount"/>
<form id="user">
<field name="amount" type="digits">
<prompt>Please enter the amount to be transferred from savings account to current account</prompt>
<filled>
<assign name="document.amount" expr="amount"/>
<goto next="#call_transfer"/>
</filled>
</field>
</form>
<form id="call_transfer">
<block>
<var name="amounttransfer" expr="document.amount"/>
<submit next="transfer.cfm" method="get" namelist="amounttransfer"/>
</block>
</form>
</vxml>
```

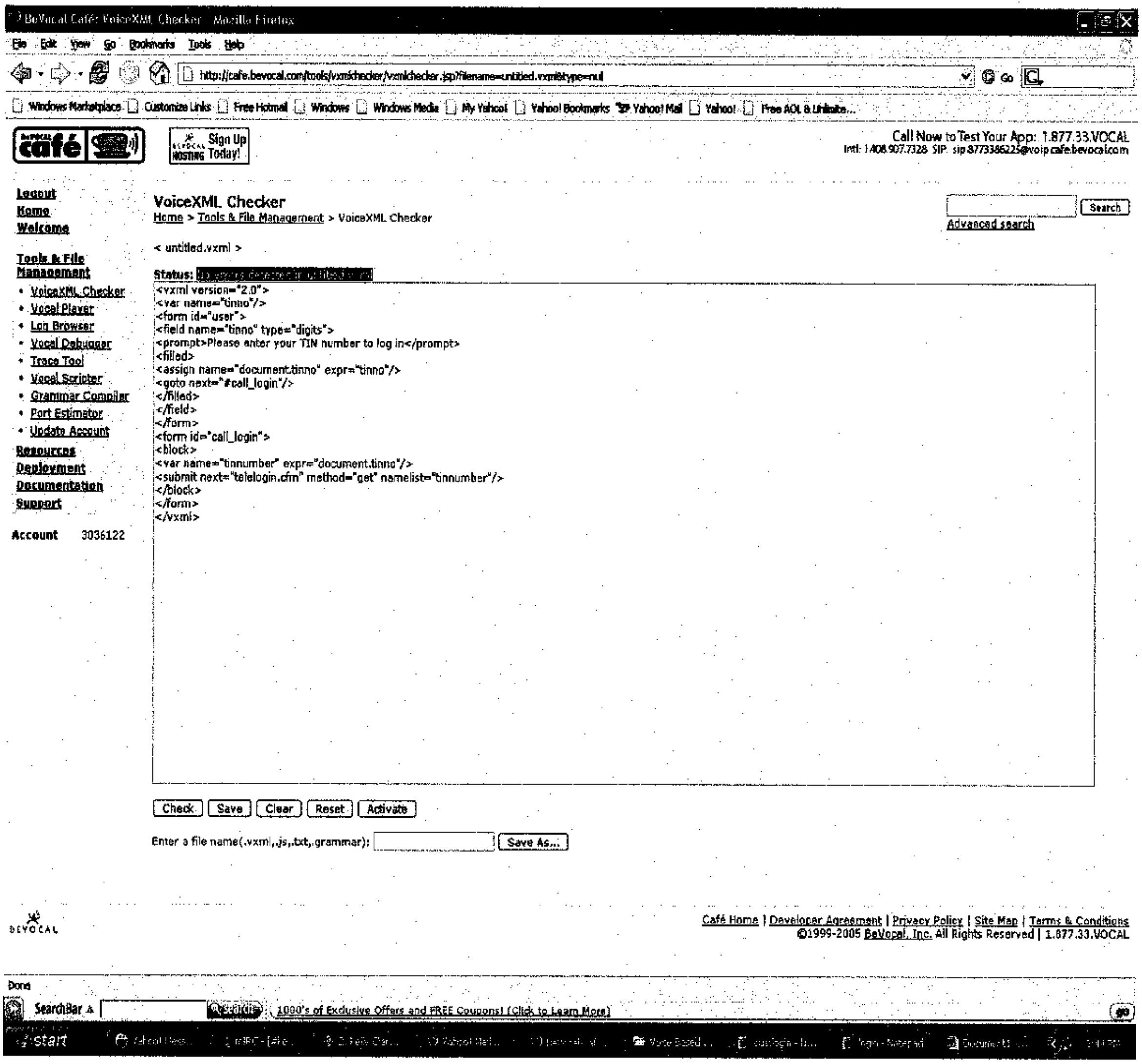
12) SOURCE CODE FOR TRANSFER.CFM

```
<CFAPPLICATION NAME="TELEBANKING" SESSIONMANAGEMENT="YES">
</CFAPPLICATION>
<CFQUERY NAME="saving" DATASOURCE="teleDSN" DBTYPE="ODBC">
SELECT saving_ac_no,balance FROM saving_account WHERE customer_no =
#SESSION.custnum#
</CFQUERY>
<CFQUERY NAME="current" DATASOURCE="teleDSN" DBTYPE="ODBC">
SELECT current_ac_no,balance FROM current_account WHERE customer_no =
#SESSION.custnum#
</CFQUERY>
<CFSET transferamount = url.amounttransfer></CFSET>
<CFSET currentbalance = (current.balance) + (transferamount)></CFSET>
<CFSET validamount = (saving.balance) - (transferamount)></CFSET>
<CFSET today = Now()></CFSET>
<cfcontent type = "text/xml"></cfcontent>
<vxml version="2.0">
<form>
<block><cfif validamount LT 1000>
<CFOUTPUT>You have only #saving.balance# balance in your savings account. Please
try again</CFOUTPUT>
<cfelse>
<CFQUERY NAME="update_current" DATASOURCE="teleDSN"
DBTYPE="ODBC">
UPDATE current_account SET balance =
#currentbalance#, modified_date = #today# WHERE customer_no =
#SESSION.custnum# </CFQUERY>
<CFQUERY NAME="update_saving" DATASOURCE="teleDSN"
DBTYPE="ODBC">
UPDATE saving_account SET balance =
(#saving.balance#) - (#transferamount#), modified_date =
#today# WHERE customer_no = #SESSION.custnum#</CFQUERY>
<CFQUERY NAME="insert_transfer"
DATASOURCE="teleDSN" DBTYPE="ODBC">
INSERT INTO transfer
(customer_no,saving_ac_no,current_ac_no,amount_transfer,
transfer_date) VALUES(#SESSION.custnum#,#saving.saving_ac_no#,
#current.current_ac_no#,#transferamount#,#today#)
</CFQUERY>
<CFOUTPUT>Amount has been transferred from savings to
current account.</CFOUTPUT>
</cfif><goto next="menu.cfm"/>
</block>
</form>
</vxml>
```

APPENDIX D: VoiceXML Checker (café beVocal.com)



<u>The output of the checker if errors in the coding</u>



The output of the checker if the coding is free from errors. Status message will prompt and no errors detected in the coding

APPENDIX E: SURVEY QUESTIONNAIRES

SURVEY QUESTIONNAIRES:
VOICE BASED FOR BANKING SYSTEM (VBBS)

VOICE BASED FOR BANKING SYSTEM description:

Voice based for Banking System (VBBS) is a possible product that can provide voice-based service which is telephone-based banking application in order for user easy to perform chosen banking activities and it is free of charge service. The major advantage of this VBBS is that it provides web content over a simple telephone device, making it possible to access an application even without a computer and an Internet connection

Section1: Personal Details

1. What is your age range?

   A. 8-12 B. 13-17 C. 18-22 D. 23-27 E. 28 and above

2. What is your gender?

   A. Male B. Female

3. What is your race?

   A. Malay B. Chinese C. Indian D. Others

4. Do you have hand phone?

   A. Yes B. No

Section 2: Bank Activities

1. How often are you going to bank to get services?

   A. Often B. Seldom C. Always D. Occasionally E. Never

2. How long usually you queue to wait your turn?

   A. 5-10 minutes B. 15-30 minutes C. 35-50 minutes D. more than 1 hour

3. Which Bank do you think provide the most efficient service in Malaysia?

   A. Maybank B. BCB C. Am Bank D. Others; Name:_________

4. How do you rate Bank Institution services in Malaysia?

Strongly Inefficient	Inefficient	Neither Inefficient nor efficient	Efficient	Strongly Efficient
1	2	3	4	5

Section 3: Functionality of voice based system

1. Have you experienced any voice recognition system?

A. Yes; Name:____________________ B. No

2. Do you agree if Malaysia's Bank Institution provide voice based service in order to perform banking activities?

A. Agree B. Not Agree; Justify:____________________

3. Do you prefer operator keeps telling all information about your account statement frequently?

A. Yes B. No

4. What language you prefer operator speaks?

A. Standard English B. Bahasa Melayu C. Mandarin D. Tamil

5. What rate do your prefer operator will response from your voice instruction?

A. below 3 seconds B. 4-6 seconds C. 7-10seconds D. 1 minute

6. What types do you prefer system will authenticate your customer ID?

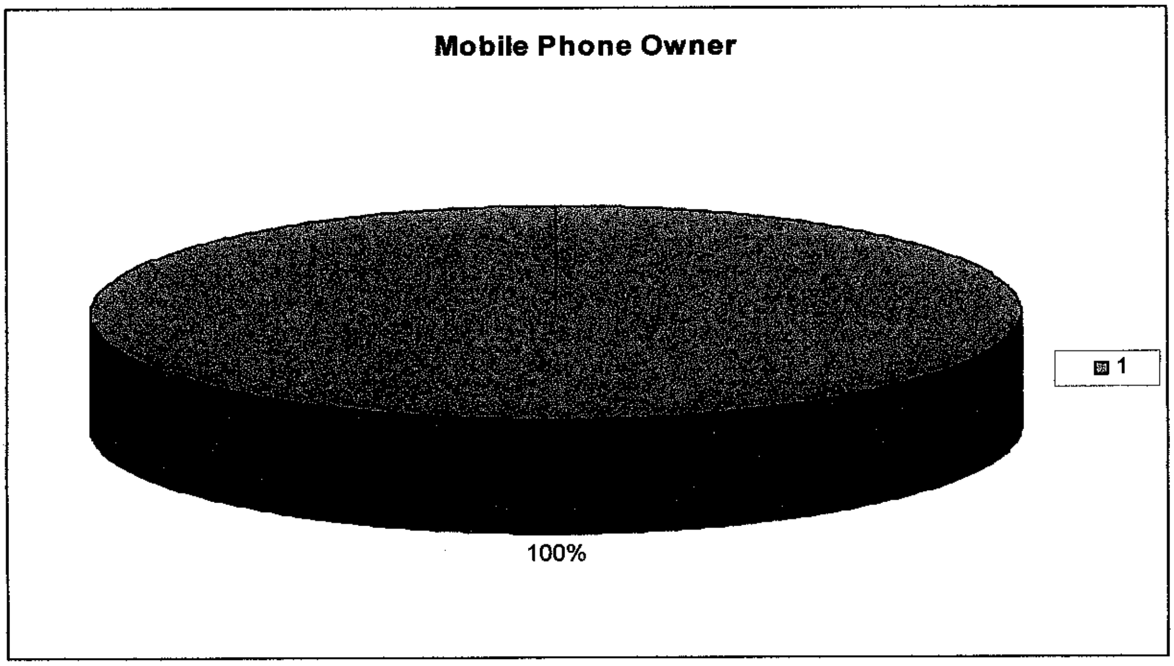
A. Use voice to enter Customer ID B. Dual Tone Multi-Frequency (DTMF) system

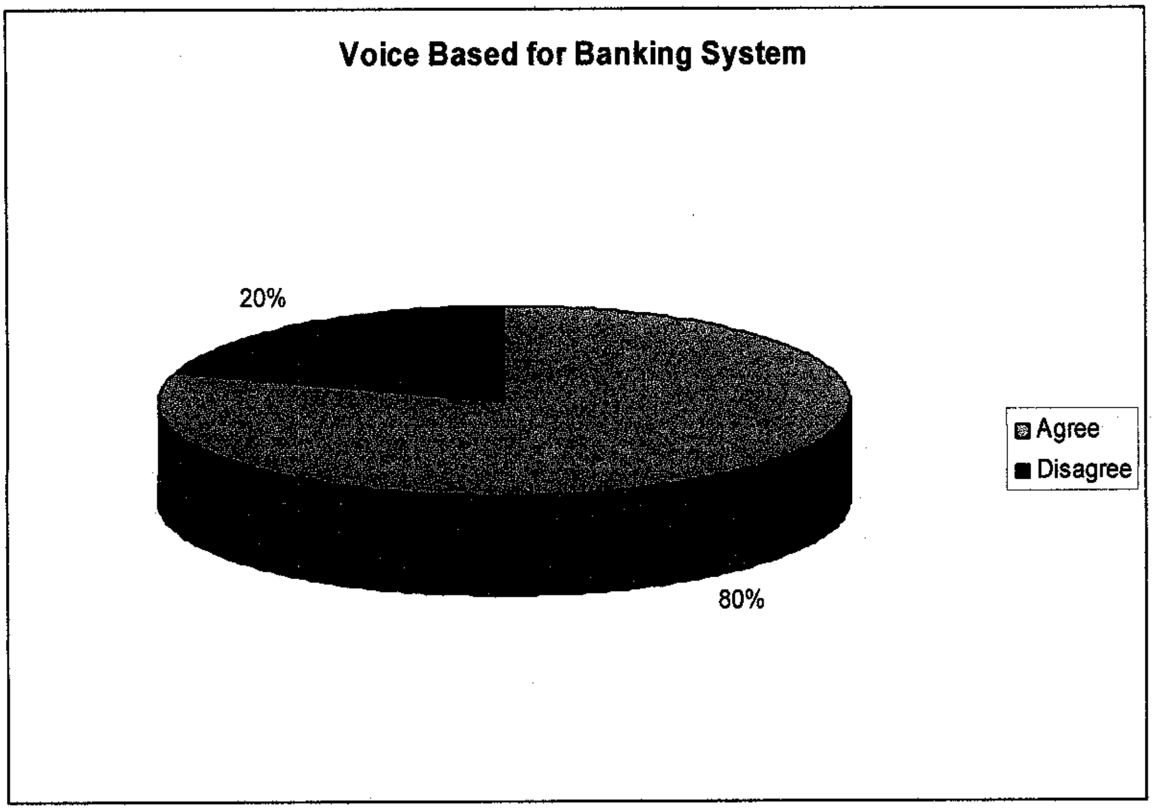
7. Among all the bank activities, please tick the contents of your interest. You can tick more than once.

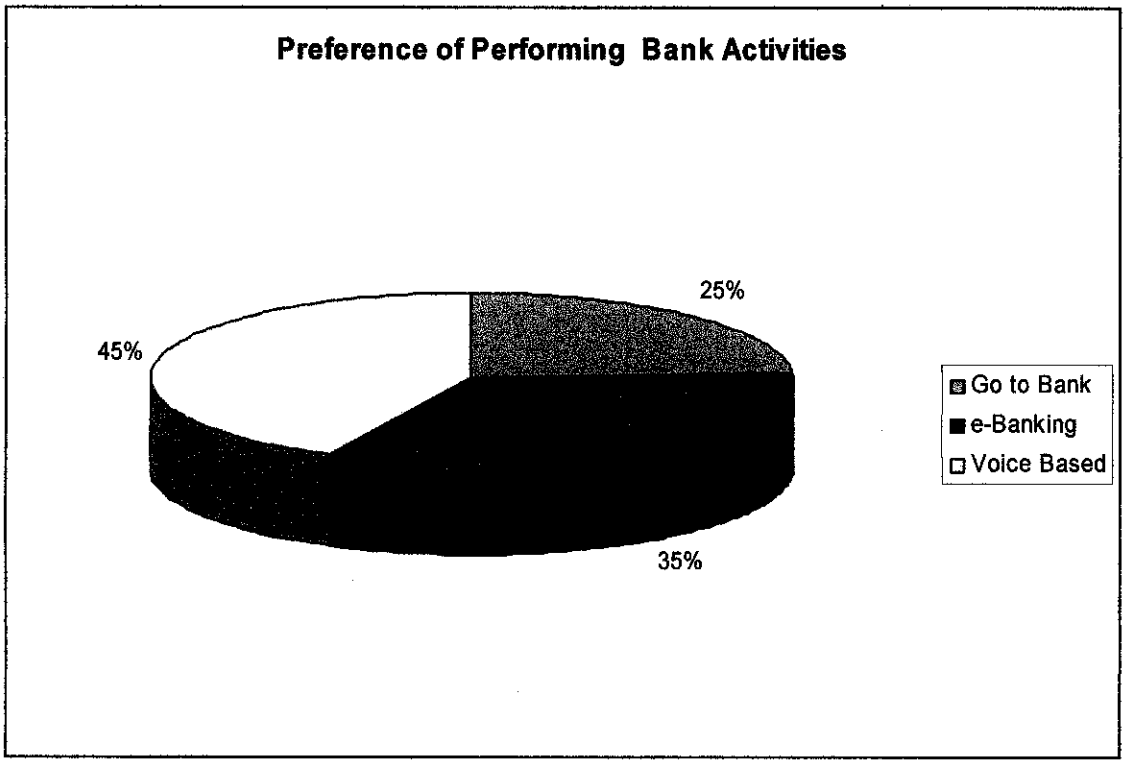
Activities	Tick
A. Transfer Money	
B. Check Balance	
C. Account Statement	
D. Order New Cheque book	
E. Bank Draft	

:: THANK YOU FOR YOUR COOPERATION! ::

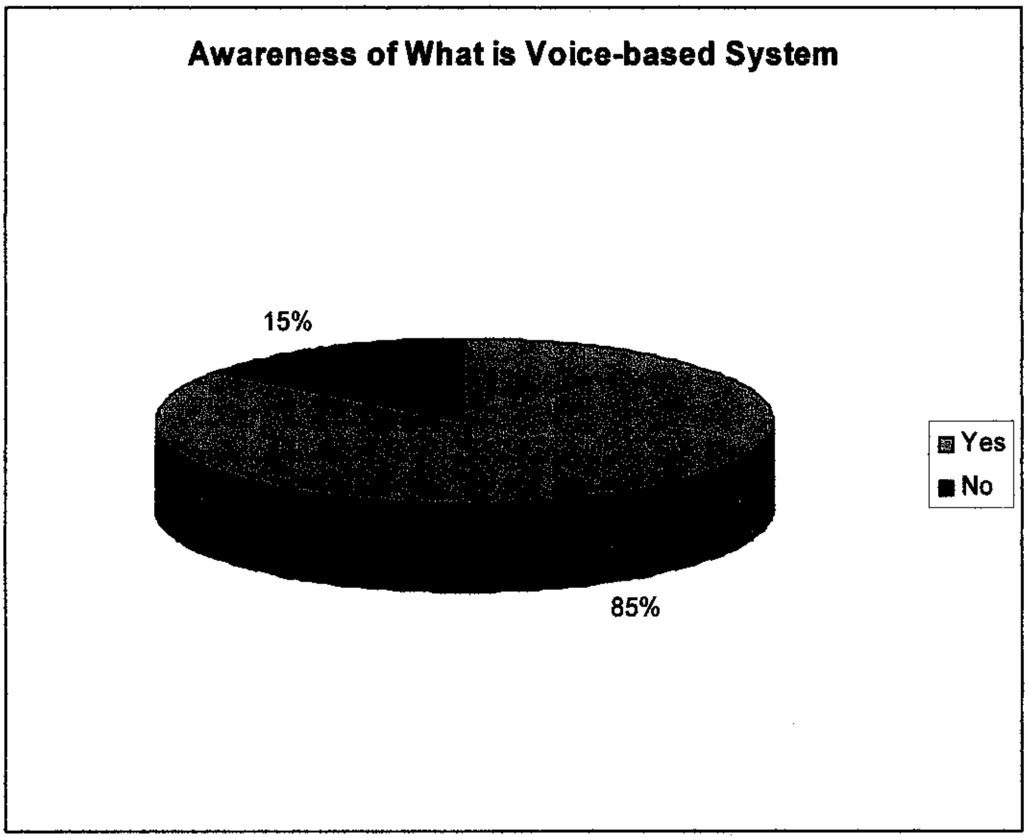
APPENDIX F: RESULT FROM SURVEY QUESTIONNAIRES

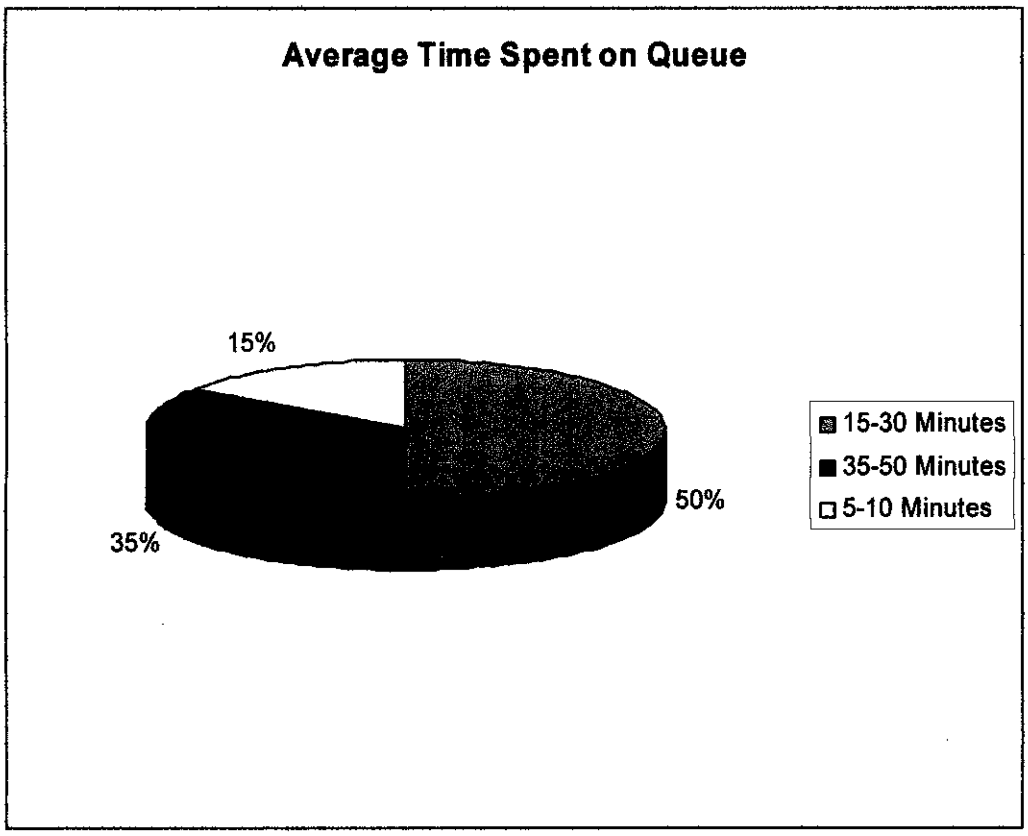

Mobile Phone Owner
1
100%


Voice Based for Banking System
20%
80%
Agree
Disagree

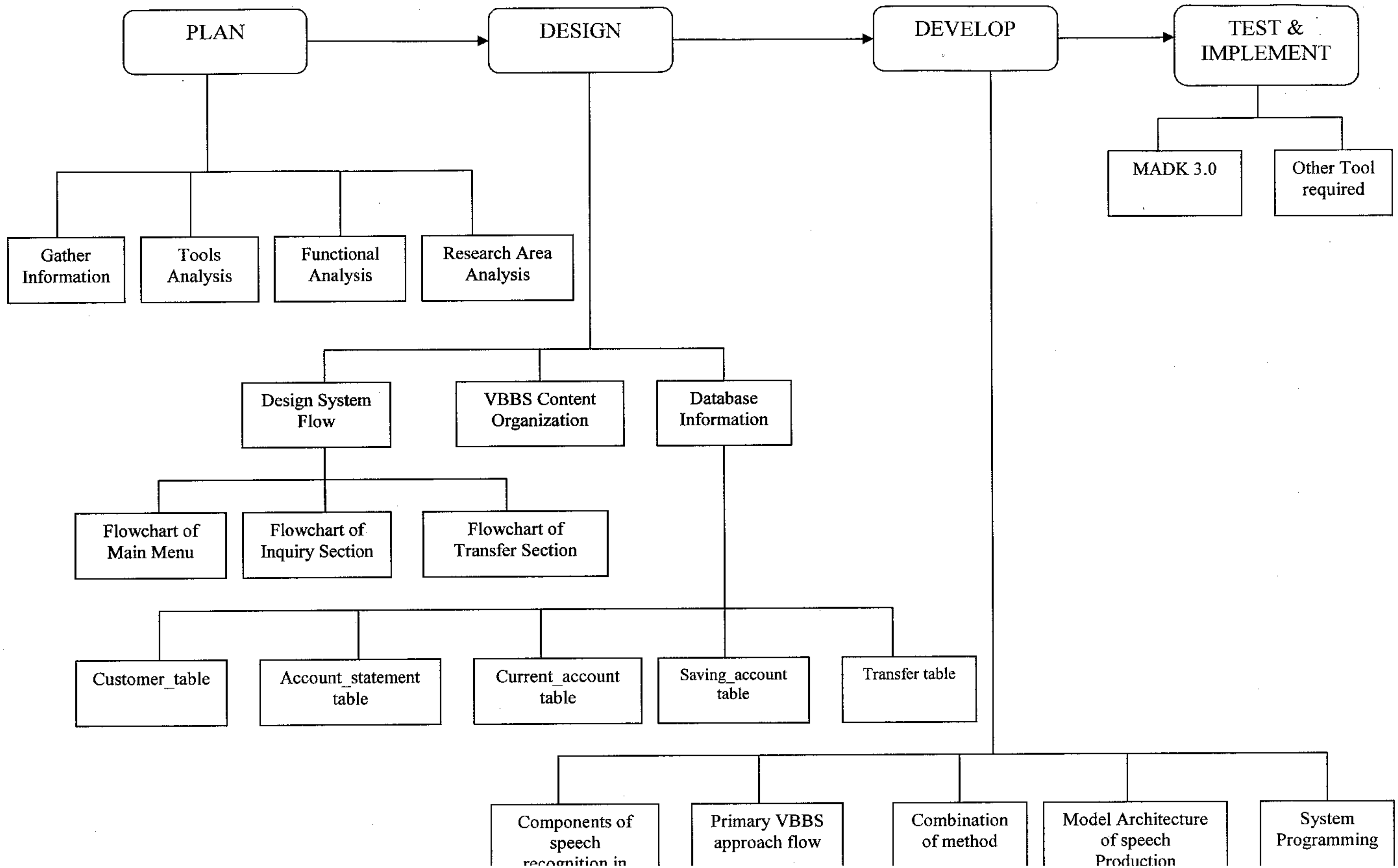


Preference of Performing Bank Activities
25%
45%
35%
Go to Bank
e-Banking
Voice Based


Awareness of What is Voice-based System
15%
85%
Yes
No


Average Time Spent on Queue
15%
50%
35%
15-30 Minutes
35-50 Minutes
5-10 Minutes

APPENDIX G: GENERAL VIEW OF PROJECT METHODOLOGY


PLAN
DESIGN
DEVELOP
TEST & IMPLEMENT
MADK 3.0
Other Tool required
Gather Information
Tools Analysis
Functional Analysis
Research Area Analysis
Design System Flow
VBBS Content Organization
Database Information
Flowchart of Main Menu
Flowchart of Inquiry Section
Flowchart of Transfer Section
Customer_table
Account_statement table
Current_account table
Saving_account table
Transfer table
Components of speech
Primary VBBS approach flow
Combination of method
Model Architecture of speech
System Programming