

**Multimodal User Interfaces for Newspaper Distribution Transcription on
Mobile Devices with Accelerometer Sensors**

Tayyab Zaheer

University of Tampere
Department of Computer Sciences
Interactive Technology
M.Sc. thesis
Supervisor: Roope Raisamo
May 2009

University of Tampere

Department of Computer Sciences

Interactive Technology

Tayyab Zaheer: Multimodal User Interfaces for Newspaper Distribution Transcription on Mobile Devices with Accelerometer Sensors

M.Sc. thesis, 46 pages, 5 index pages

May 2009

Mobile devices are being used to benefit a variety of personal and working environments. One potential working environment is newspaper distribution, where mental tiredness during night-time working hours can result in loss of efficiency and motivation. Reading the addresses manually from the address book whilst simultaneously keeping track of the subscriber's newspaper preferences and delivering them to the right door increases the cognitive load on the distributor and can lead to early exhaustion. Distributors could benefit from the automation of their manual home address book by having its electronic version on their mobile devices. However, merely automating the manual book keeping does not decrease the required visual attention during the work. Incorporating multimodality decreases the cognitive load on the distributors and makes the system more usable, allowing them to complete their job efficiently and accurately. A prototype system of the presented solution was developed and tested in a real working environment to prove its feasibility. A new interaction technique was introduced combining simple gesture recognition for accelerometer equipped mobile devices, as well as providing auditory and haptic feedback for the distributor. An intermediate comparative study between manual distribution and presented solution was made in order to measure the usefulness of the system in the context of time and efficiency. The feedback reveals that the system is more easily operable if the presented solution allows the distributors to add customizable gesture recognition.

Keywords and Terms: cognitive load, multimodality, interaction technique, auditory feedback, haptic feedback, distribution transcription, usability, gesture recognition

Acknowledgements:

I would like to thank my supervisor Professor Dr. Roope Raisamo for his valuable guidance and feedback. He guided me in every theoretical and technical aspects of my thesis. His feedback comments enabled the timely completion of my thesis.

I would also like to thank Professor Dr. Erkki Mäkinen for reviewing the final draft of my thesis and giving valuable feedback.

Finally, I would say many thanks to my family and best friend for their moral support and prayers for the successful completion of my thesis.

Contents

List of Acronyms and Abbreviations.....	v
1 Introduction	1
2 Key concepts and related work for mobile devices	3
2.1 Accelerometer based gestural interfaces.....	3
2.2 Speech Synthesis.....	7
2.3 Haptic user interfaces.....	7
2.4 Attention and Cognitive load.....	8
2.5 Multimodal user interfaces	9
3 Manual newspaper distribution transcription	10
3.1 Procedure for manual newspaper distribution	10
3.2 Shortcomings of manual distribution transcription	10
4 Proposed system	12
4.1 Input/Output Modalities.....	12
4.2 Platforms and technologies.....	12
4.2.1 Symbian operating system	12
4.2.2 Nokia’s Series 60 platform for mobile devices	13
4.2.3 Symbian C++ programming environment	14
4.3 User Interface Design Issues.....	16
4.3.1 Technique for gesture recognition	16
4.3.2 Haptic user interface	16
4.3.3 Text to speech synthesis	17
4.4 Detailed description	17
5 Technical Software Design	25
5.1 Class Diagram and Class Descriptions	25
5.2 Sequence Diagrams.....	27
6 Evaluation	35
6.1 Goals	35
6.2 Apparatus	35
6.3 Procedure	36
6.4 Evaluation Design.....	37
6.5 Results.....	37
6.5.1 Data Summary	37

6.5.2	Physical Comfort	37
6.5.3	Newspaper distribution speed	39
6.5.4	Newspaper distribution complaint rate	39
7	Discussion	41
8	Conclusions and future work.....	43
	References:	44

List of Acronyms and Abbreviations

3D	3 Dimensional
API	Application Programming Interface
DPM	Delivered per minute
EPOC	Electronic piece of cheese, a family of graphical operating systems developed by Psion for portable devices
Hz	Hertz, frequency unit
ITU-T	International Telecommunications Union (ITU-T) standard
LPC	Linear Predictive Coding
PSOLA	Pitch Synchronous Overlap Add Method
PDA	Personal Digital Assistant
MFT	Multi-Function Transducers
MUNDRE	Multimodal User Interface for Newspaper Distribution Transcription on Accelerometer equipped Mobile Devices
OCSP	Online Certificate Status Protocol
OS	Operating system
RMV	Rotary Mass Vibrator
S60	Series 60
TTS	Text to Speech
UI	User Interface
XML	Extensible Markup Language

1 Introduction

In the recent years mobile devices have become more sophisticated in their use. These devices are being used in everyday life, ranging from a simple calling device to a more personal information management and retrieval system. The rapidly growing market for mobile devices such as personal information managers, cellular phones, pagers, watches, and wearable computers offers a tremendous opportunity to introduce interface design innovations [Hinckley et al., 2000]. Most people interact with multiple portable devices: a personal music player, mobile phone and digital camera. Although other device specific features are added onto mobile phones; however, these are not integrated with the connectivity that the mobile phone represents [Tanaka et al., 2007]. Many mobile manufactures now call them minicomputers. Their use in everyday life is obvious due to their portability and ease of use. These portable devices are also becoming a part of a variety of working environments.

Use of mobile devices in working environments includes special features of mobile interaction, such as changing orientation and position, changing venues, the use of computing as auxiliary to ongoing technology and the general use of mobile devices [Hinckley et al., 2000]. One potential working environment is newspaper distribution, where people can become mentally tired and exhausted during their working hours. One of the important factors leading to early exhaustion during the work is the mental attention required to complete the work on time and with accuracy. This required attention can be eased by reducing the cognitive load on the distributor. This requires the analysis of the actual manual distribution work and finding other ways to ease it. Incorporating multimodality can reduce the cognitive load on the user of the device in this particular environment. A rich multimodal display can be used without requiring any visual attention in order to present information in a natural and non-irritating manner [Williamson et al., 2007]. However, the interactions with the device need to be minimally disruptive and should demand minimal cognitive and visual attention [Hinckley et al., 2000].

Manual newspaper distribution transcription can be automated and transformed into an electronic version of distribution transcription for these mobile devices. However, merely automating and viewing the electronic version of the distribution transcription instead of the paper (manual) one will not reduce a significant amount of cognitive load on the distributor. However, this can be achieved by incorporating multimodal user interface in the system, such as gesture recognition for input and speech synthesis as well as graphical user interface to get feedback from the device. Systems which support multimodal user interfaces are rarely designed and developed as many mobile applications currently only support the key pad for user input. The International Telecommunications Union (ITU-T) standard keyboard present on mobile phones imposes restrictions on the user interface of mobile phone application development [Coulton et al. 2007]. However, several interesting interaction techniques have been proposed through the incorporation of the low-cost tilt sensor within mobile devices [Widgor & Balakrishnan, 2003]. For example, gesture input is one form of continuous interaction that has been underused in interaction with current systems [Crossan & Smith, 2004].

However, gestures remain as an alternative or complementary modality for application control with mobile devices [Kela et al., 2005]. More appropriately, gestural interfaces could reduce the cognitive load by adding a gesture control, as on-screen and keypad buttons are generally grouped together closely making interactions slow and error prone. This is particularly the case in a mobile context, where the user's visual attention may be required elsewhere [Crossan & Smith, 2004]. In addition to gestures playing an important role in building eyes-free interfaces [Kela et al., 2005], realistic synthesis of vibrotactile sensations and audio sensations also play a part in these interfaces

[Williamson et al., 2007]. Moreover, speech user interfaces are added in order to overcome the problem of small screen displays in these mobile devices [Turunen & Hakulinen, 2007]. Moreover, context aware mobile applications enable the rich user interface (UI) with user environment. However, the UI designers may not be familiar with the concepts of context awareness and do not necessarily know the design guidelines for underlying system functionality and the user interaction [Häkikilä & Mäntyljärvi, 2006].

This thesis presents the multimodal design and implementation of the electronic version of newspaper distribution transcription on accelerometer equipped mobile devices that overcome the deficiencies of manual newspaper distribution transcription. Gesture commands are used to control the device's internal functions, such as its own internal states. The presented system introduces a new interaction technique which uses the tilt of the mobile device for simple gesture recognition. The accelerative forces are measured for tilt detection with respect to the orientation of the mobile phone, such that whenever the distributor tilts the phone, the change in values will result in the change of state for the system application.

Speech synthesis technology is augmented in the system, which serves as auditory feedback for the newspaper distributor using the built-in text-to-speech (TTS) synthesizer of the mobile device. The application programming interfaces (APIs) provided by the mobile device manufacturer to access the built-in synthesizer functionality are used to output the speech from the given text. In addition, the vibrotactile feedback is provided to the distributor through a built-in Rotary Mass Vibrator (RMV) embedded in the mobile device, this feature serves as a key to building eyes-free interfaces. The presented system combines gesture input with auditory and vibrotactile feedback for the newspaper distributor in order to reduce the visual attention and cognitive load required during the working hours. This makes interaction more convenient, efficient, useful and informative for the distributor.

This thesis combines graphical, speech, haptic and gestural interaction modalities to gain the advantages of multimodal user interfaces. An intermediate comparative study in between manual and presented system is also done in order to measure the usefulness of the system in the context of time and efficiency for newspaper distribution.

This thesis has been divided into eight chapters. Chapter two contains the key concepts and related work for mobile devices such as accelerometer based gestural interfaces, text-to-speech synthesis, haptic user interfaces and their use in multimodal user interfaces. Chapter three is focused on the Manual Newspaper Distribution transcription and its disadvantages. Chapter four offers a solution to the manual distribution transcription, by incorporating the multimodal user interfaces in the electronic version of newspaper distribution transcription system, whereas chapter five addresses the design issues involved in the proposed system. Chapter six introduces the evaluation of the system, after which chapter seven discusses the comparative study done on the presented system. Finally, chapter eight contains the conclusions and suggestions for future work.

2 Key concepts and related work for mobile devices

This chapter explains the key concepts and related work in accordance with my thesis. It is divided into five sections. The first section will explain the concepts and related work on accelerometer based gestural interfaces and in depth information about 3D accelerometer based sensor data when used in mobile devices; the second section will address concepts related to speech synthesis, the third, fourth and fifth sections will also describe concepts and related work for haptic user interfaces, attention and cognitive load and multimodal user interfaces on mobile devices accordingly.

2.1 Accelerometer based gestural interfaces

Gestures are a natural modality for certain tasks and can augment other modalities, for example, the auditory and vibrotactile feedback can be augmented with the gesture input. For some tasks the gestural control can be natural and quick; however, many targets remain for future work [Kela et al., 2005]. Tsukada and Yasummura [2002] developed the Ubi-Finger and realized the sensuous operations for PDA and information appliances by gesture of fingers. They found that most of the gestural input systems are either very expensive or large and have not been used in mobile environment. Tanaka [2004] studied a network based interactive music engine and a portable rendering player, it used the force sensing resistors to capture grip pressure, and accelerometers were used to sense gesture and motion in three dimensional space in order to create and mix new rhythms and melodies by proactively recognizing the gestures made by the user. However, the hardware used was large. Hughes et al. [2004] developed many prototype sample applications in order to demonstrate the usefulness of gestures for interacting with the mobile device using accelerometers and other sensors; however, the hardware used for this purpose was a backpack to iPaq mobile device, which makes it large and difficult to handle.

Many interactions with mobile devices can be demanding of cognitive load and visual attention. Designing appropriate interaction technique to recognise the correct user gesture in order to minimize the cognitive load and visual attention is a challenging work. Hinkley et al. [2001] proposes an intuitive set of gestures which can be used to rapidly change the user profile on the mobile phones in order to minimize the cognitive load and visual attention on the user.

Gestures provide the possibility to control the device on the move without getting access to keys of the mobile device [Linjama & Kaaresoja, 2004]. Gesture commands can be used to control two types of applications: device internal functions and external devices. The sensor-based user interface type can be categorized into three distinct categories with respect to the operating principle, analogy with speech interface, personalization and complexity and required computing load for those interface types [Kela et al., 2005].

Sensor-based user interface type (movement)	Operating principle	Analogy with speech interface	Personalisation	Complexity and computing load
1. Measure and control	Direct measurement of tilting, rotation or amplitude	E.g. control based on volume level	-	Very low
2. Gesture command	Gesture Command recognition	Speech command recognition	Machine learning, freely personalisable	High
3. Continuous gesture command	Continuous gesture recognition	Continuous speech recognition	Machine learning, freely personalisable	Very high

Table 1: Categorization and properties of movement sensor-based user interfaces [Kela et al., 2005]

Our system falls into the category of measure and control (see Table 1), where the computing complexity is very low and we directly measure the tilting amplitude of the mobile device.

Sensors need to be incorporated within the mobile devices in order to recognize the gestural input. Accelerometers allow the user to input data and commands by tilting the device [Crossan & Smith, 2004]. In gestural interfaces, the tilt gesture means that a user may tilt a device, for example, left to right or backward/forward, to initiate some UI (user interface) action. The resulting control and navigation of the user interface can be seen to enhance the user experience [Mäntyjärvi et al., 2006]. Harrison et al. [1998] proposed new interaction technique for navigating through the sequential lists by measuring the extent or degree of tilt. In general, the tilt interface follows a simple to learn pattern [Mäntyjärvi et al., 2006]. More complex gestures require machine learning based algorithms [Mäntyjärvi et al., 2007]. Hinkley et al. [2000] studied a tilt-based gesture system for scrolling and automatic screen orientation of a Personal Digital Assistant (PDA) and demonstrated the potential for tilt-based gestures to provide a fast and natural method for interaction.

Earlier, the possibility of using tilt input to navigate menus and scroll large documents and maps had been explored by Rekimoto [1996]. Later, Rekimoto [2001] introduced two input devices for wearable computers, called GestureWrist and GesturePad, which allowed the users to interact with wearable or nearby computers by gesture based commands. These devices were designed to be as much unobtrusive as possible. Bartlett [2000] used the two-axis single chip accelerometer to scroll through the picture slides on an Itsy device (Compaq Research’s experimental platform for “off the desktop” computing) by tilting the device in horizontal and vertical directions.

However, the potential for using 3D accelerometers on phones appears to be completely unexplored. This is principally due to the fact that the first 3D accelerometers have only recently been integrated into mobile phones and these phones have yet to become widespread [Vajk et al., 2007]. Accelerometer equipped mobile devices are often suggested as being useful for continuously monitoring background acceleration and providing context information for the current task. The advantage of having the accelerometer on a mobile device over stylus based gesture systems are that they offer the possibility of one handed, screen free gesture control [Crossan & Smith, 2004]. Linjama et al. [2005] used the simplest gesture, tapping the mobile device to control the movement of the graphical objects for their bouncing ball game. They also incorporated haptic sensing and vibrotactile feedback to offer a new dimension to multimodal interactions. Williamson et al. [2007] developed “The Shoogle System”, which was a metaphor for physical bouncing balls sensing (vibrotactile feedback) on the accelerometer based mobile phones. The system sensed the shaking, picking-up and placing of the mobile phone in the pocket. It can be stated that accelerometer based gesture control is used as a complementary modality for handheld devices. Predetermined gesture

commands or freely trainable by the user can also be used for controlling functions in other devices [Mäntyjärvi et al., 2004].

Gesture recognition requires careful study as the variation in how the user holds the phone could produce anomalous output and there is no method of obtaining the phone's physical position within the actual space. The rotational and translational accelerations affect the accelerometer's output, as both of these can produce the same internal forces on the accelerometer [Vajk et al., 2007]. Nokia has produced a number of phones incorporating built-in motion sensors, including the N95 and N96 as well as Nokia 5500 sports. This was one of the first such equipped phones and was basically targeted at sports users, utilizing the built-in motion sensors as a pedometer and speed/distance tracker for various exercising purposes.

It utilized a 6g accelerometer, which means it could detect acceleration forces with a magnitude of up to six times that of the earth's gravity. The accelerometer outputs three 12-bit signed data values at a frequency around 37Hz. These outputs correspond to the three phone axes (x, y, z) as shown in Figure 1 [Vajk et al., 2007].



Figure 1: Nokia 5500 accelerometer axes [Vajk et al., 2007]

Figures 2 and 3 show the recorded output from the sensors on Nokia 5500 in three different scenarios. The output has been expressed as acceleration forces in g for clarity [Vajk et al., 2007].

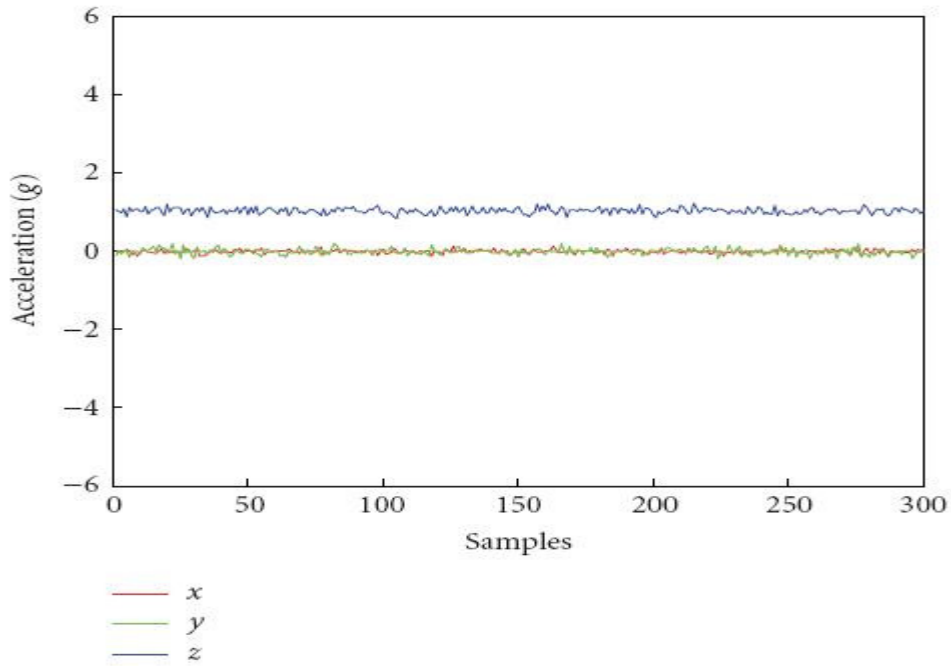


Figure 2: Nokia 5500 accelerometer data (phone at rest) [Vajk et al., 2007]

Figure 2 shows the raw data values for the x, y and z axes when the phone is at rest. It is obvious that x and y values are approximately zero (although with some noise). However, output z shows the positive 1g force, which is the effect of gravity on the device. The gravity provides a means of deducing the orientation of the phone, which can be utilized to provide a “tilt-based” controller [Vajk et al., 2007].

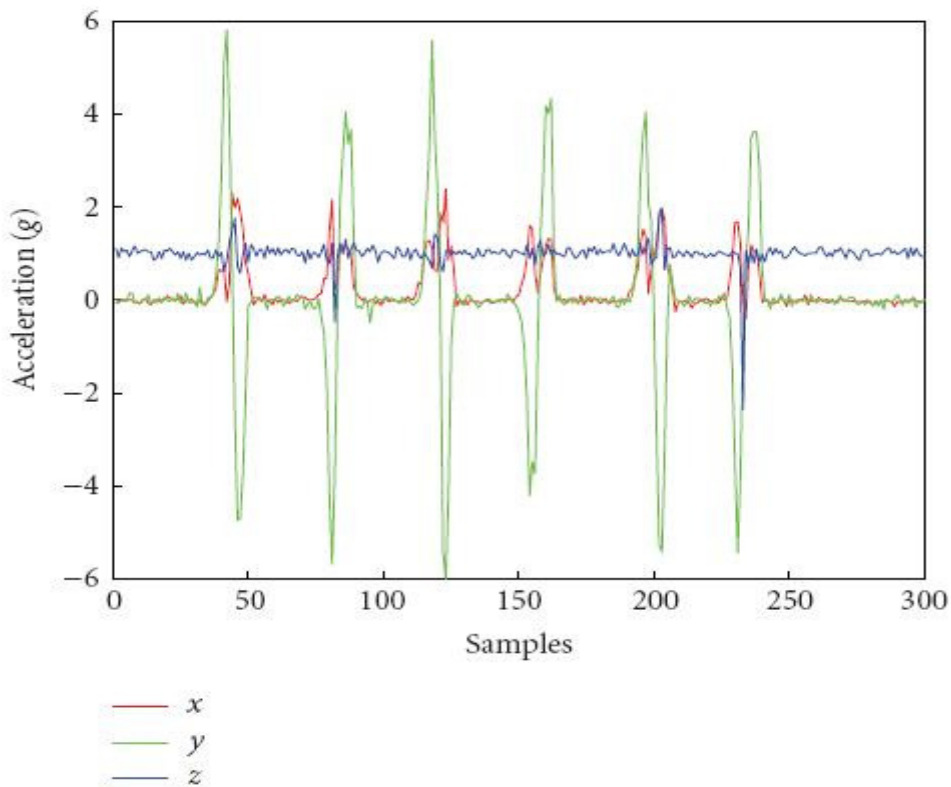


Figure 3: Nokia 5500 accelerometer data (lateral movement of phone across table) [Vajk et al., 2007]

Figure 3 depicts the forces produced by very large accelerative forces along the y axis. The values alternate between the two positive and negative levels, which is due to the acceleration and deceleration of the device and vice versa [Vajk et al., 2007].

The acceleration data provides a basis for understanding the underlying functionality of the 3D accelerometer in these mobile devices. This also helps in design and development of applications for recognising the gestural commands accurately.

2.2 Speech Synthesis

Speech synthesis is a mechanism to artificially produce human-like speech. A system used for this purpose is called as a speech synthesizer. Speech synthesizers can be implemented either in the software or the hardware. A text to speech system converts the normal language text into speech, called a text-to-speech synthesizer [Wikipedia-Speech Synthesis, 2008]. The main advantage of producing synthetic speech over natural speech is that any text can be converted to speech, and thus, any information which is written and stored in some form can be spoken [Neerinx et al., 2008]. In addition, speech can be a natural and efficient interaction modality in mobile environments [Turunen & Hakulinen, 2007].

There are three basic methods for speech synthesis, which are formant, articulatory and concatenative synthesis. Formant synthesis is based on the modelling of resonances in the vocal tract. Articulatory synthesis, on the other hand, tries to model the human vocal organs as perfectly as possible and, therefore, requires much more computational power. The level of success is lower as compared to other synthesis methods. Concatenative synthesis is more popular because it sounds more natural as it uses pieces of pre-recorded speech. One form of concatenative synthesis is unit selection synthesis [Neerinx et al., 2008]. Unit selection synthesis uses diphones, half-phones, syllables, morphemes, words, phrases and sentences [Wikipedia-Speech Synthesis, 2008]. One form of unit selection synthesis, diphone synthesis, uses diphones, which are formed by concatenating the small fixed-sized speech units or the transition between two subsequent sounds [Neerinx et al., 2008]. It sounds more natural than formant synthesis but suffers from the sonic glitches of concatenative synthesis. Diphone synthesis requires the minimal speech database of diphones. At runtime, the prosody of the sentence is superimposed on these smaller diphone units, by using the digital signal processing techniques, such as Linear Predictive Coding (LPC) and Pitch Synchronous Overlap Add Method (PSOLA) [Wikipedia-Speech Synthesis, 2008]. This method provides great naturalness as only a small amount of speech processing is required [Neerinx et al., 2008]. This is of benefit for mobile devices, where low processing is required in order to reduce battery power consumption.

2.3 Haptic user interfaces

Feedback can be tightly coupled to the input we make on the device. The user can just shake, tilt or wobble the device to stimulate the auditory and/or vibrotactile feedback. The presentation of timely haptic responses greatly improves the sensation. The vibrotactile waveforms are similar to the enveloped sine waves, with center frequency at 250Hz (both the resonant frequency of the transducer, and the peak sensitivity of the skin receptors involved in vibrotactile perception) [Williamson et al., 2007].

Currently, there are three common solutions for presenting vibrations in mobile devices, the Rotary Mass Vibrator (RMV), piezoelectric actuators and Multi-Function Transducers (MFT). RMVs are the most commonly used vibrotactile display available for mobile devices, as these are

inexpensive and can be available in a much smaller size. However, the piezoelectric actuators require a large amount of electric power to actuate them, which is not intended for mobile devices. MFT is a speaker capable of producing both audio and vibration [Chang & O'Sullivan, 2005].

2.4 Attention and Cognitive load

Attention is required in everyday tasks we perform to achieve our goals. We get attention from our visual, auditory or any other sensory stimuli. Focused attention or selective attention, is studied by presenting people with two or more stimulus inputs simultaneously, and instructing them to respond to only the single stimulus. On the other hand, divided attention is studied by presenting at least two stimulus inputs at the same time, but with instructions that participants must attend and respond to all stimulus inputs. What we attend to in the real world is largely determined by our current goals and these goals are followed by our actions or tasks [Eysenck & Kaene, 2005]. We discuss divided attention here, as the newspaper distribution requires that the distributor must respond to all the stimulus inputs, such as handling the newspapers in his arm and hand, while simultaneously looking at the address book, and getting an audio and haptic feedback when the newspaper is dropped in the post box.

Treisman and Davies [Treisman & Davies, 1973] concluded that two monitoring tasks interfered with each other in the participant's mind, when the stimuli on both tasks is in the same sensual modality (visual or auditory) rather than different modalities [Eysenck & Kaene, 2005]. In our case, not any of the tasks for newspaper distribution are in same sensual modality, which means the tasks don't interfere with each other.

The ability to perform two tasks together depends on each of their difficulty level separately, and it is hard to define the task difficulty level precisely. The demands of resources of two tasks performed together might be thought to equal the sums of the demands of the two tasks when performed separately. Additionally, Segal and Fussella experiments suggest that auditory image task was more demanding than visual image task [Segal & Fusella, 1970]. However, the auditory image task was less disruptive than the visual image task when each task was combined with a task requiring detection of visual signals, suggesting that visual image task was more demanding than the auditory one [Eysenck & Kaene, 2005]. In our case, we see that required tasks for newspaper distribution require more visual tasks, as compared to auditory tasks, because the distributor needs to repeatedly look at the address book and the newspapers during one's working hours. We can make use of the vacant auditory channel in order to make the tasks less disruptive and less demanding for the distributor.

Second most important factor which builds-up the cognitive load on the distributor is one's working memory. Three types of memory stores have been proposed [Eysenck & Kaene, 2005]:

- Sensory stores, each of them holding information very briefly and are modality specific.
- A short-term store of very limited capacity (only about seven digits can be remembered), which is a fragile storage, as any distraction usually causes forgetting.
- A long-term store of essentially unlimited capacity holding information over extremely long periods of time.

Memory stores are used in everyday life. Short-term memory store is mostly used to carryout our routine tasks. Long-term memory store is used to recall some event from the past or any information which has been repeatedly taken place in short-term memory store. Short-term memory store constraints build an important factor in the design of user interfaces for computer systems.

In our case, it is already understandable that the distributor can't remember all the names and addresses and the newspapers required by each person, and one is dependent on the address book. However, as already argued, visual tasks are more disruptive than auditory ones, resultantly we can make use of auditory modality to let the distributor know about the names and addresses of the persons and their newspaper preferences on the go. This helps the distributor overcome the short-term memory constraint as well as avoid making any mistakes while distributing the newspapers.

2.5 Multimodal user interfaces

Multimodal user interfaces are a breakthrough for building better user interfaces. Multimodal interfaces allow the simultaneous use of input modalities and present the information with the synergistic representation of many output modalities [Raisamo, 1999]. Many desktop solutions for multimodal user interfaces have been proposed. However, little attention is paid to develop mobile device applications, which could make use of all the modalities and technological advancement available in these devices. Multiple modalities can be used in several ways in mobile devices and its applications. Multimodal user interfaces help the user in preventing and correcting the errors and bring robustness to the interface and add alternative communication methods to different situations and environments [Cohen et al., 1994]. Moreover, speech interfaces can be added in addition to other commonly used modalities in order to overcome the problem of small screen displays in the mobile devices [Turunen & Hakulinen, 2007].

Gestures play an important role in the design of multimodal interfaces. Gestures can be used in navigation and in providing control commands. As a result, a user interface supporting gestural interaction must combine at least the graphical and gestural modalities [Mäntyjärvi et al., 2007]. We start from adding the speech interface to the system in order to increase its usability, we can further add other modalities, such as graphics, haptics, gesture and other information sources (e.g. positioning information) [Turunen & Hakulinen, 2007]. Several modalities can be used in a complete interaction for a mobile device, such as graphical for prompt, gestures for providing input, and vocal, tactile and/or graphical feedback to the user [Mäntyjärvi et al., 2007]. Moreover, design considerations for multimodal user interfaces suggest that any multimodal system should not require the user to operate on all input/output modalities simultaneously [Wikipedia-Multimodal Interaction, 2008].

3 Manual newspaper distribution transcription

This chapter describes the manual newspaper distribution procedure in detail and its shortcomings. First section will describe the procedure for manual newspaper distribution transcription and the steps involved to perform it. Shortcomings of manual distribution transcription are described in the next section.

3.1 Procedure for manual newspaper distribution

The steps involved in manual newspaper distribution are as follows:

- The distributor already possesses the address book, in which the Street Name, Building Number, Surnames of the persons, their flat number and the newspaper/s required by them are written in the form of a symbolic table. It also shows the days on which address that particular newspaper is to be delivered.
- The distributor receives the bundle of newspapers and the list of changes to be made in the address book.
- After the address book has been updated by the distributor, he reviews the book in front of each building in order to know how many newspapers are needed to take it with oneself from the newspaper bundle.
- When the distributor is inside the building with the required number of newspapers for this particular building, one starts distributing them. One has to check the address book each time after each and every delivery of the newspaper for the next person in the address book.

The manual distribution transcription procedure requires much attention from the distributor, as he needs to correctly maintain the address book entries everyday and requires to repeatedly look for flat numbers and their corresponding deliverable newspapers.

3.2 Shortcomings of manual distribution transcription

This section describes the shortcomings of manual distribution transcription in accordance with the manual distribution transcription procedure described in the previous section.

- The distribution address book is in the form of symbolic table, marked with X, HS, or sometimes with actual name of the newspaper, which is quite difficult to read and this imposes a huge amount of cognitive load on the distributor (Figure 4).

Last Name	Mon-Fri	Sat - Sun
Lehtonen	X	
Turunen	HS	⊗
Library	5 X, 5 HS, 5 6-Degree	
Technopolis	2 HS, 3 X	

Figure 4: Snapshot of the Manual Newspaper Distribution Address Book

In Figure 4, symbol X corresponds to Aamulehti (the local newspaper) and HS corresponds to Helsingin Sanomat newspaper. The numbers appearing before these symbols determine the number of newspapers to be delivered at that particular address. Special symbols like ⊗ determine that particular newspaper needs to be delivered on Saturday and Sunday only. The distributor should be habitual to these symbols in order to deliver the newspapers efficiently and accurately.

- Everyday the distributor needs to update the address book for the required number of newspaper changes. This consumes a lot of time before distribution, and might cause mistakes.
- When the distributor reviews the address book pages in front of each building in order to know how many newspapers one needs to deliver in that particular building, this causes a huge wastage of time.
- The distributor needs to check the address book again and again after each and every delivery of the newspaper. This makes more cognitive load on the distributor and reduces the efficiency, when one has delivered through many addresses after some time.

The shortcomings described above directly emphasize the need for a better solution for newspaper distribution transcription, which can reduce the cognitive load on the distributor and make his job easier.

4 Proposed system

This chapter describes the proposed system in detail. First, the input/output modalities used by the system are discussed and afterwards the platforms and technologies used to develop the system are described in the following sections.

4.1 Input/Output Modalities

This section describes the input/output modalities used by our proposed system. Graphical, speech and haptic user interfaces are used as the output modalities for our proposed system; whereas gesture user interface is used as an input modality.

- **Graphical user interface**
The graphical user interface, which is the display of the mobile device, shows the step by step information to the distributor. A simple text editor shows all the information regarding each and every address. Additionally, the list of newspaper names and their respective quantities, required to be distributed in each building are shown in a summary format, just before the distributor starts distributing them in each and every new building.
- **Gestural user interface**
Tilt of the mobile device is used as an input modality for gesture recognition from the user. The built-in embedded accelerometer is used to measure the tilt of the mobile device.
- **Speech user interface**
Text to speech synthesis is used as an output modality, for step by step address information guidance from the mobile device, in speech format.
- **Haptic user interface**
The haptic user interface in this context is based on vibration: it is one of the output modality for the user, which is enabled through the built-in embedded vibrator of the mobile device.

The above mentioned user interfaces are combined to make a better multimodal user interface for our system. We minimized the chance of getting the simultaneous use of two or more input modalities by selecting gestural user interface as the sole input modality for our system. Moreover, we made the synergistic use of output modalities by synchronizing the timely activation of auditory and haptic feedback for the distributor.

4.2 Platforms and technologies

This section describes the underlying platforms and technologies for the proposed system in detail.

4.2.1 Symbian operating system

Symbian is a mobile operating system (OS) targeted for mobile devices that offers high level of integration with communication and personal information management functionality. As mobile devices need to run reliably for days, months and years, mobile users do not want either of their communication disrupted or their contacts, emails, messages being lost. As a result, the basic mobile phone functionality, personal information management services and other network based

services need to run reliably and efficiently at all time. All of these requirements can be met if the operating system is designed for reliability and efficiency. However, the most important requirements are multi-tasking (with multi-threading), real-time operation of the cellular software, effective power management, small size of operating system itself, as well as applications built on it, ease of developing new functionality, reusability, modularity, connectivity (which includes interoperability with other devices and external data storage), and robustness. Now a days sensor technology is also becoming an important part of these devices. In this regard, modularity plays an important role, as new software and hardware components need to be added to the device. Symbian OS was designed and implemented, while keeping in view all the above requirements.

As Symbian OS fulfils most basic requirements for operating mobile devices, it became the most common choice for extending the mobile phone functionality. The world's top mobile phone manufacturers with the largest market share (Nokia, Motorola, Samsung, Sony Ericsson and Siemens holding almost 80% of the market), are using Symbian as the preferred operating system for their mobile devices. Symbian was formed from Psion software by Nokia, Motorola, Psion and Ericsson in June 1998. After the release of EPOC (Electronic piece of cheese, a family of graphical operating systems developed by psion for portable devices) release 5 (last version used mainly in Psion 5) of Psion operating system for embedded devices, the operating system renamed 'Symbian OS'.

Symbian develops and licenses Symbian OS, containing the base (microkernel and device drivers), middleware (system servers, such as the window server), a large set of communication protocols, and a test user interface for the application engines of the operating system. Licensees develop the user interfaces to suit their purposes, and they also have the ability to license their user interface and application set on top of the Symbian OS to other Symbian licensees – as Nokia did this with Series 60 platform.

4.2.2 Nokia's Series 60 platform for mobile devices

Nokia's Series 60 (S60) platform runs on top of Symbian OS. S60 platform and Symbian OS provide a user experience and application set unlike any other smartphone platform in a very competitive package, as it takes a great effort from a mobile phone manufacturer to create a comparable user interface and application set – even if they start building this on top of Symbian OS.

Nokia wants to license its smartphone platform to all manufacturers in order to make the developer base larger and to create more content. Additionally, it is easier to enter the Smartphone market with Series 60 platform as there is already an established developer base. Moreover, Series 60 terminals are binary-compatible with each other, which means that the application developed for Nokia's Series 60 terminals are interoperable with other Series 60 terminals (with the exception of unsupported features, such as trying to make use of sensor's application programming interfaces in a device without that sensor hardware).

The Symbian OS provides the text to speech engine, which can be accessed through the utility classes and the functions. This text to speech engine is provided in various languages to cover a broad range of users. For example, Nokia 5500 provides the high quality text-to-speech synthesizer. Other Nokia mobile phones that also provide this feature are Nokia E50, E51, E61i, E65, E90 Communicator, N75, N78, N81, N96, 5700 XpressMusic and 6110 Navigator phones. High quality text-to-speech synthesizer provides the ability to correctly pronounce the large words in a natural and understandable manner. The voice packages are provided for many common natural languages,

such as English, French, German, Greek, Italian, Norwegian, Polish, Russian, Portuguese, Spanish, Swedish, Japanese, and Thai etc. Each language is supported by at least one person's voice. The output quality for these languages also depends on the language accent and the speech output quality provided by the manufacturer of the mobile device. The Symbian OS Platform SDK provides the Test-to-speech (TTS) utility API to access the TTS functionality on S60 devices, including the high-quality TTS engine supported in some devices.

S60 Platform SDK C++ Plug-ins, for Symbian OS, are software packages for developers creating C++ applications for S60 devices. The plug-ins extend the features of a S60 Platform SDK, improving device emulation, adding functionality that is not included in shipped devices or adding functionality that is specific to a certain device. The plug-ins allow a greater range of applications to be built and tested using the S60 Platform SDK emulator. Sensor plug-in for S60 3rd edition SDK for Symbian OS is used to access the functionality and motion sensing values of the embedded accelerometer of the mobile device. This plug-in adds Sensor Server API support to S60 public SDKs and it only supports compiling for terminals.

The Symbian OS Platform SDK for C++ provides the vibration API to control the vibrator of the S60 3rd edition devices. The Vibration API provides the ability to control the device vibration feature. This API allows setting a custom intensity for the vibration. The intensity is defined as percentage of the maximum rotation speed of the vibra motor. Negative values are also allowed, in which case the vibra motor rotates in the opposite direction. According to SDK documentation, the allowed value range for the intensity is -100 to 100. The S60 platform provides the ability to do programming in Symbian C++ environment. The SDK provides all the API's for the required functionality of our proposed system.

4.2.3 Symbian C++ programming environment

S60 is built on top of open Symbian operating system and it allows third party applications to be developed using c++ application environment. Such applications can benefit from the best possible performance and have access to comprehensive device functionality. Figures 5 and 6 show the S60 platform 3rd edition and the S60 3rd edition application development framework respectively.

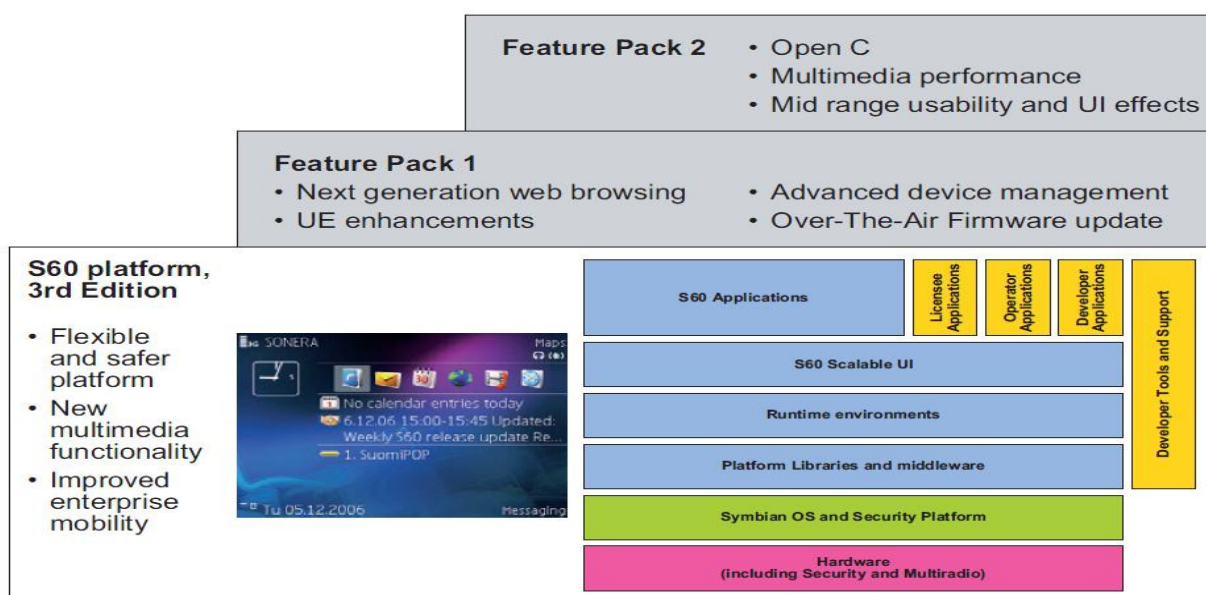


Figure 5: S60 Platform 3rd edition [Symbian, 2009]

Examples of APIs available for Symbian S60 C++ applications include:

- User interface, including support for the SVG-T graphics to provide flexible support of various screen resolutions,
- Telephony,
- Communications,
- Multimedia,
- Integrated messaging,
- Built-in application engines,
- Built-in sensors and vibrators,
- Data handling, and
- Security.

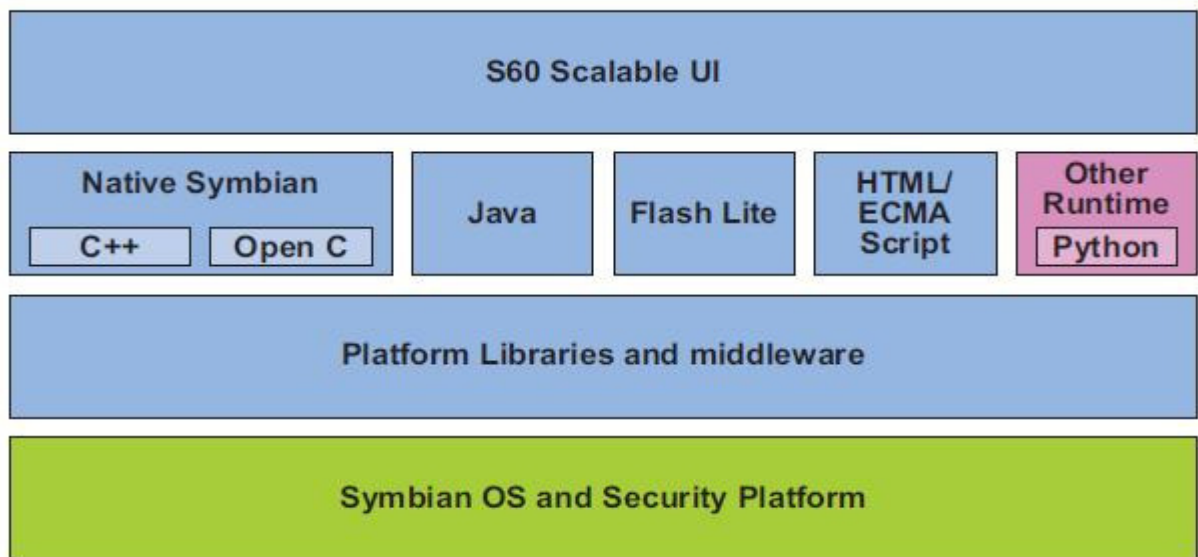


Figure 6: S60 3rd edition application development framework [Symbian, 2009]

The security architecture of S60 3rd edition / Symbian OS C++ is based on the principle of controlled access to the protected system resources. These resources can be other processes, application programming interfaces (APIs), or user/system files stored in the file system. Controlled access to these resources means that only trusted applications or processes are allowed to access them. For example, applications can store information in their private directories that are protected from other applications.

In order to create different levels of trust, all C++ applications are categorized into different domains. How much the application is trusted, depends on the application and application domain. Trusted applications must include a digital signature with the application file. Applications can be granted a signature, for example after passing the Symbian signed test criteria. When the application is installed on S60 3rd edition based device, the device can use the digital signature to verify:

- Whether the application is signed by a trusted party and therefore can be given access to the protected system resources it requires.
- That the application has not changed since signing the certificate. For example, the application signature can be used to detect Trojans. This verification is an automated task to protect the user's device.

The S60 3rd edition based devices can also use the Online Certificate Status Protocol (OCSP) in order to check if the certificate for the application can still be trusted or not. It is also recommended that the device is allowed to contact the OCSP server on the internet to check for the certificate's validity, before the application is actually installed on the device. OCSP servers can be maintained by all the concerned trusted application signers.

4.3 User Interface Design Issues

The proposed system was designed while keeping in view the user interface and the usability issues discussed and found in related systems. This section will describe the user interface design issues related to our proposed system.

4.3.1 Technique for gesture recognition

As much of the related research has been done in the development of gesture based systems, which require the user to tilt the device to initiate some action, we use the same technique for our proposed solution. Whilst designing the proposed system, the hand gesture interaction was made very simple, limited in number and easy to remember. As in this particular working environment, the distributor already had too many other hand related tasks, and one should not be given a complex hand gesture for interacting with the mobile device. Additionally, for a positive user experience, the system ensured that it is correctly able to recognise the intended hand gesture, which was holding the device in a normal horizontal way and then tilting the mobile device toward its right to navigate to the next address in address list and vice versa. The measured tilt of the mobile is directly mapped onto interaction events of the system.

The gesture recognition is implemented, by measuring the tilt of the device for positive and negative y and z-axis. A certain range of values is associated with the condition, which determines the success or the failure of the gesture recognition. The distributor tilts the device for changing the state of the system. Once the system has successfully changed its state, it starts observing for the next tilt, while simultaneously speaking out the announcements and displaying the text on the display of the mobile device.

4.3.2 Haptic user interface

Currently, mobile devices have only simple tactile feedback as a haptic user interface. Rotary Mass Vibrator (RMV) is the most commonly available haptic interface for the mobile devices, which can either be in 'On' or 'Off' state with a specified frequency for a certain time interval. However, Nokia will introduce more expressive feedback in its future products; quite recently it has licensed Immersion's VibeTonz mobile player for its mobile handsets to provide more multisensory experience for its users.

Haptic feedback is one of the output modalities used in our proposed system. Whilst activating and deactivating the RMV of the mobile device, the developer must specify the required intensity for the rotation of the mass, to let the user sense about comfortably. Timely activation and deactivating the RMV also played a major role in the design and development of the proposed system and made it a better experience for the distributor. Two different intensities and time durations were used for the RMV, such that the distributor was able distinguish between them, while holding the device with one's hand.

4.3.3 Text to speech synthesis

Text to speech synthesis is used for building eyes free interfaces. Text to speech synthesizer available on the mobile device is not much accurate in pronouncing all dictionary words. As a result, the words provided to the synthesizer should be simple and concise. Each word or phrase or sentence provided to the text to speech synthesizer must be completed before the next word or phrase is pronounced. The user interactions with other features of the system must be kept pending until the current announcement of the text is completed.

The announcements made for the distributor are kept small enough such that synthesizer is able to pronounce the words with more clarity and speak out each person's name, newspaper name and the required newspaper quantity separately. Only the English text to speech synthesizer engine was tested as a proof of concept. Finnish text to speech synthesizer engine is not yet provided by Nokia for its mobile devices.

4.4 Detailed description

Following is the detailed description of the proposed system along with its working procedure:

- 1) The address book file, which is in XML (Extensible Markup Language) format, is copied to the mobile device, when the application is installed on the device. Figure 7 shows a sample address book XML file for the system.

```

<info>
    <Address>
        <street>Finninmaenkatu</street>
        <building>4A</building>
        <person>
            <name>Total</name>
        </person>
        <paper>Aamu</paper>
        <quantity>6</quantity>
        <paper>Helsingin Sanomat</paper>
        <quantity>3</quantity>
    </Address>
    <Address>
        <street>Finninmaenkatu</street>
        <building>4A</building>
        <person>
            <name>George</name>
            <floor>7</floor>
        </person>
        <paper>Aamu</paper>
        <quantity>2</quantity>
        <paper>Helsingin Sanomat</paper>
        <quantity>1</quantity>
    </Address>
</info>

```

Figure 7: Address book XML file

This address book file is divided in a number of address tags contained within a single info tag. In order to make the information easily readable and to categorize it properly, each address tag is further divided in street, building, person, paper and quantity tags. Furthermore, each person tag is divided in name and floor tags, which contain the name of

the person and the floor number. Each paper tag is followed by the quantity tag, which contains the corresponding number of newspapers to be delivered.

- 2) When the application is started, it parses the address book XML file, and fills the data structure with all the information. When the distributor starts distributing the newspapers, he/she can start the application by using the Options menu and selecting the 'Start' option (Figure 8). The application makes an announcement about the Start of distribution procedure (Figure 9).



Figure 8: The user starts the application and selects the Start option from the Options menu to start the distribution procedure of the application



Figure 9: The application announces the start of distribution procedure

- 3) The “New Address Summary” is displayed and the corresponding announcement is made for the distributor (Figure 10). The summary information contains the number of newspapers to be delivered at that particular street address along with the building number. This announcement is made for every new building and street address encountered in the address book file.

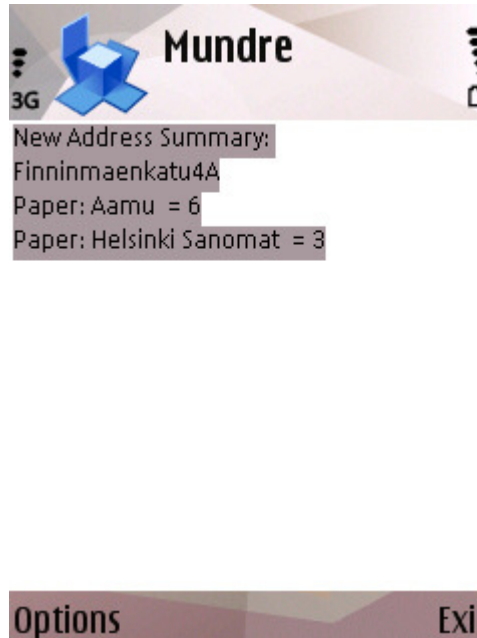


Figure 10: The New Address Summary is displayed and the corresponding announcement is made for the distributor

- 4) While traversing through the simple linked list of addresses and newspaper information in the data structure, the application announces the street address along with the building and activates the vibrator of the mobile device whenever a new street address is encountered.
- 5) After the street address has been announced correctly, the building and floor information are also announced. The application announces each and every person's name and newspaper preferences living on the same floor after a certain time interval. This time interval can be large or small, depending on the newspaper delivery speed, which can be adjusted through the Speed option in the Options menu (Figure 11).

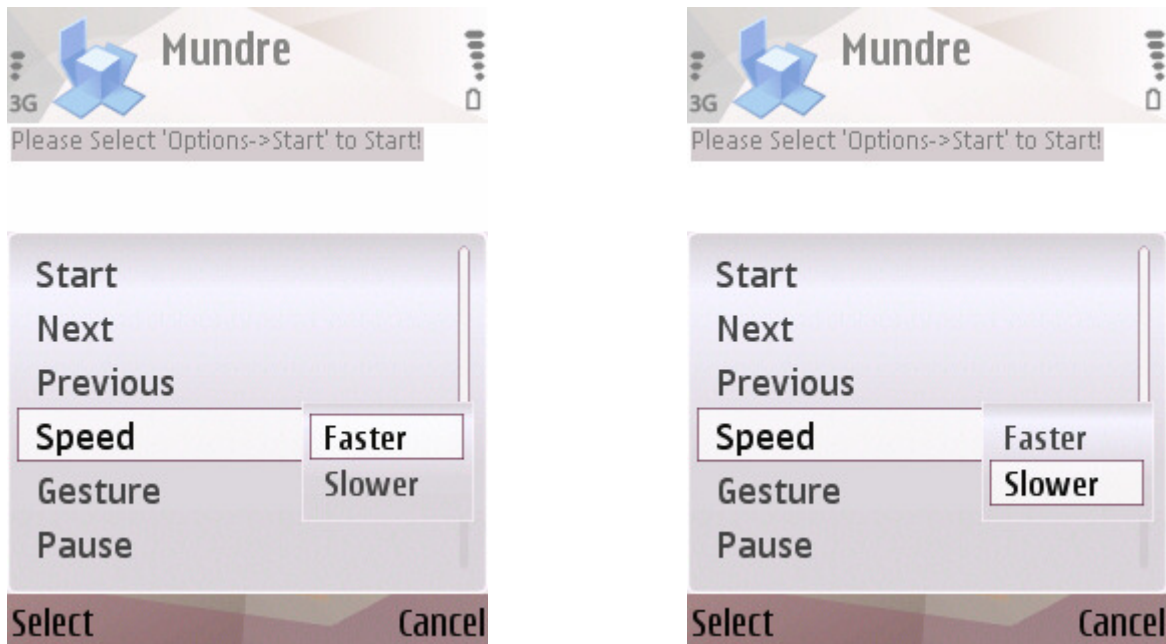


Figure 11: Newspaper delivery speed can be adjusted according to the personal choice of the distributor, it can either Faster or Slower

The address, floor number, name of the person and the names and quantities of the required newspapers are displayed (Figure 12), while only name of the person and required paper names and their corresponding quantities are announced.

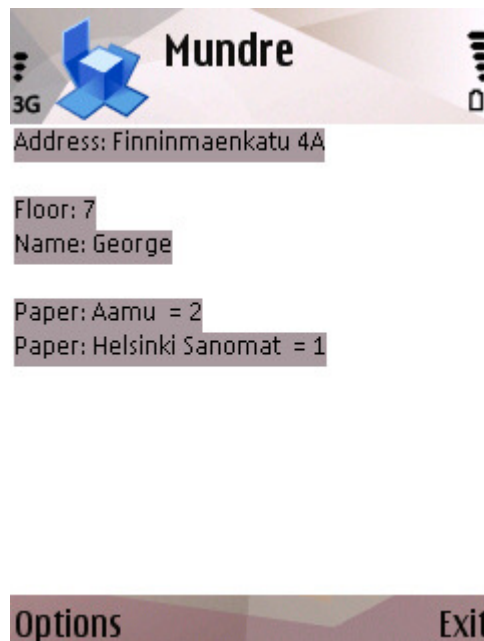


Figure 12: The Address, floor number, name of the person and the names and quantities of the required newspapers are displayed

In addition, the distributor can make a pause to the system, and the system will stop at that particular point of announcement. This can be achieved by selecting the pause from the Options menu (Figure 13). On the other hand in order to restart the announcements, the

system needs to be unpaused. This is done by selecting the 'Unpause or Restart' from the Options menu (Figure 13).

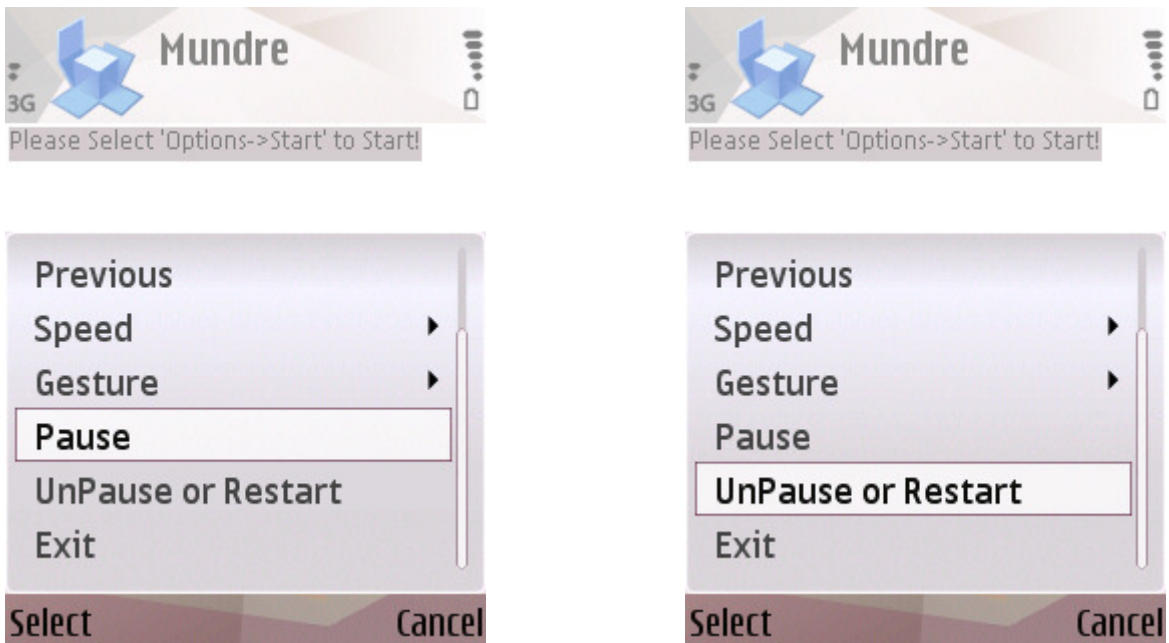


Figure 13: The distributor can make the system pause, and UnPause/Restart it later

- 6) The distributor can navigate between the addresses either through the available Options menu or can use the gestures. These gestures can be enabled or disabled through the Options menu of the application (Figure 14).



Figure 14: The Options menu can be used to navigate between the addresses

The voice, vibration and gestures can be enabled and disabled using the Options menu (Figure 15).

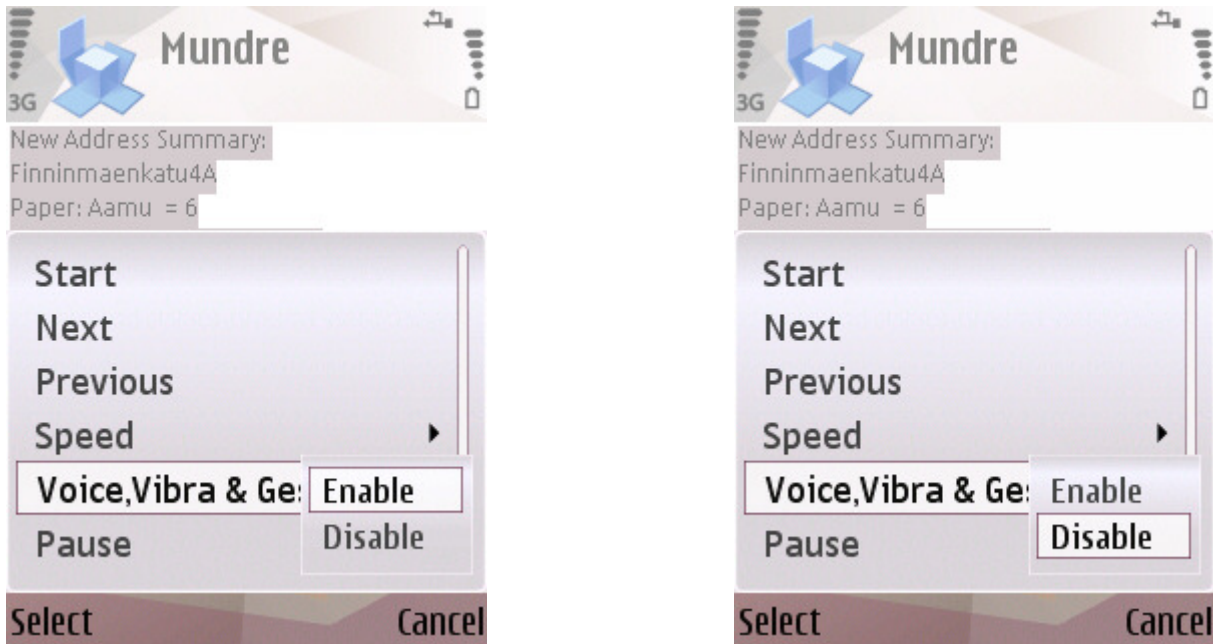


Figure 15: Gestures for the device can be enabled or disabled from the Options->Voice, Vibra & Gesture menu

While the Voice, Vibration & Gestures are enabled, the distributor can hear the announcements made by the system, sense the vibration through one's haptic sense, and use the gestures recognition facility by tilting the device.



Figure 16: Method for tilting the device for "Move Next Gesture" and "Move Previous Gesture"

Keeping the device in hand in a normal horizontal orientation and then tilting the device to the right displays and announces the next address information or new address summary from the address book list (Figure 16). While tilting the device to the left displays and announces the previous address information. An announcement regarding the type of gesture recognised by the system is also made, as a feedback for the distributor, for example when the mobile device is tilted toward the right "Move Next Gesture" is displayed and announced, and when it's tilted on the left "Move Previous Gesture" is displayed and announced (Figure 17).



Figure 17: Gesture recognition is coupled with announcement for audio feedback and the text is displayed for visual feedback for the distributor.

- 7) When the announcements for the whole one street address and building are complete, the application activates the vibrator, and the mobile device vibrates in order to let the distributor know that the distribution for the current address has been completed. The last announcement regarding the end of distribution is made and displayed for the distributor, when the system reaches the end of address book file (Figure 18).



Figure 18: The end of distribution notification displayed and announced for distributor

The presented procedure provides a greater flexibility for the distributors by providing them the option to instantly pause and restart the distribution announcements at any point of distribution.

Moreover, the gesture input provides the ability to instantly move between next and previous address announcements at any point of distribution, which makes the system more flexible and usable for the distributor during the working hours. Increasing/Decreasing the speed of newspaper distribution announcements further provides the distributor the ability to match his/her needs for newspaper distribution.

5 Technical Software Design

This chapter illustrates the technical details involved in the software design of the system. The first section will show the class diagram and will give a short description for all the classes. The second section will show the detailed interaction diagrams of class objects as the user interactions with the system.

5.1 Class Diagram and Class Descriptions

This section shows the Class Diagram for our system (Figure 19). The class diagram illustrates how the classes are organized in a class hierarchy. This section gives a short description about the used classes and their purpose in the system.

Class Diagram:

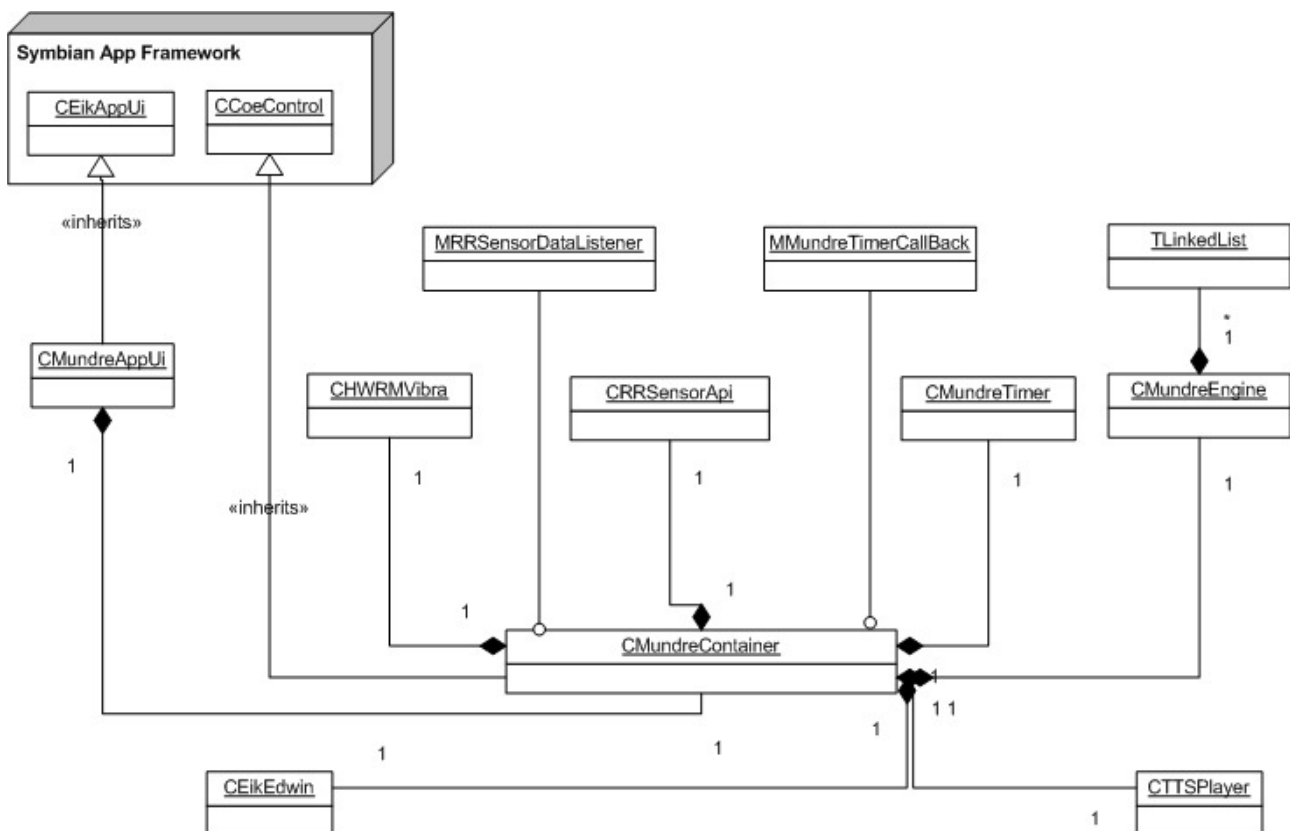


Figure 19: MUNDRE class diagram

Class descriptions:

CMundreAppUi:

Purpose: This class acts as the main class for the construction of the container class. It receives the commands from the “Options” menu and calls on the functionality provided by the container class.

CMundreContainer:

Purpose: This class acts as the main controller class. It is responsible for the creation and deletion of all other classes’ objects. This class controls the behaviour of the whole application. This class

instantiates the CMundreEngine class and retrieves the addresses information from the address file and stores it inside the link list using an array of TLinkedList class. It is also responsible for setting the displaying text using CEikEdwin class and playing it using text-to-speech's CTTSPPlayer class. It instantiates the CRRSensorApi class object is responsible for receiving the accelerometer's sensor data values from the interface class MRRSensorDataListener. It is also responsible for instantiating the CHWRMVibra class in order to activate and deactivate the vibration of the mobile device.

CHWRMVibra:

Purpose: This class is provided by the Symbian framework. It provides the ability to control the activation and deactivation of the device Vibra. The time duration for which the vibration will be activated and the intended intensity of the vibrator must be provided as parameters to the StartVibraL() API.

CRRSensorApi:

Purpose: This class is provided by the Symbian framework. It is used to search for the available list of sensors and provides the ability for the container class to listen the sensor data values in case of accelerometer.

MRRSensorDataListener:

Purpose: This class provided by the Symbian framework. It is an interface class. This class provides callback functionality for listening to the sensor data values by using the HandleDataEventL() function.

CMundreTimer:

Purpose: This class provides the timer functionality for the main container class. This class creates an active object and adds itself to the Symbian's active scheduler, which is used for timer callback functionality of the CMUNDREContainer class.

MMundreTimerCallBack:

Purpose: This class acts as the interface class for the CMUNDREContainer class. It is an abstract base class, which provides the prototype of the callback function TimerFiredL(). This callback function is implemented by the CMUNDREContainer class.

CMundreEngine:

Purpose: This class is responsible for parsing the XML based text file containing all the distribution addresses information. This class parses the XML file and extracts the information and fills the data structure by using TLinkedList class.

TLinkedList:

Purpose: This class is used as a data structure for storing the address information. The CMundreEngine class uses an array of TLinkedList class for storing all the address information.

CTTSPlayer:

Purpose: This class is used to play the text by using the built-in text-to-speech audio player in the mobile device. CMdaAudioPlayerUtility class is used to play the text, which plays the using the active objects and active scheduler of the Symbian framework.

5.2 Sequence Diagrams

This section gives the insight about the dynamic function of the system by illustrating the user interactions through sequence diagrams. The sequence diagrams also illustrate object creation and deletion and interaction between the class objects.

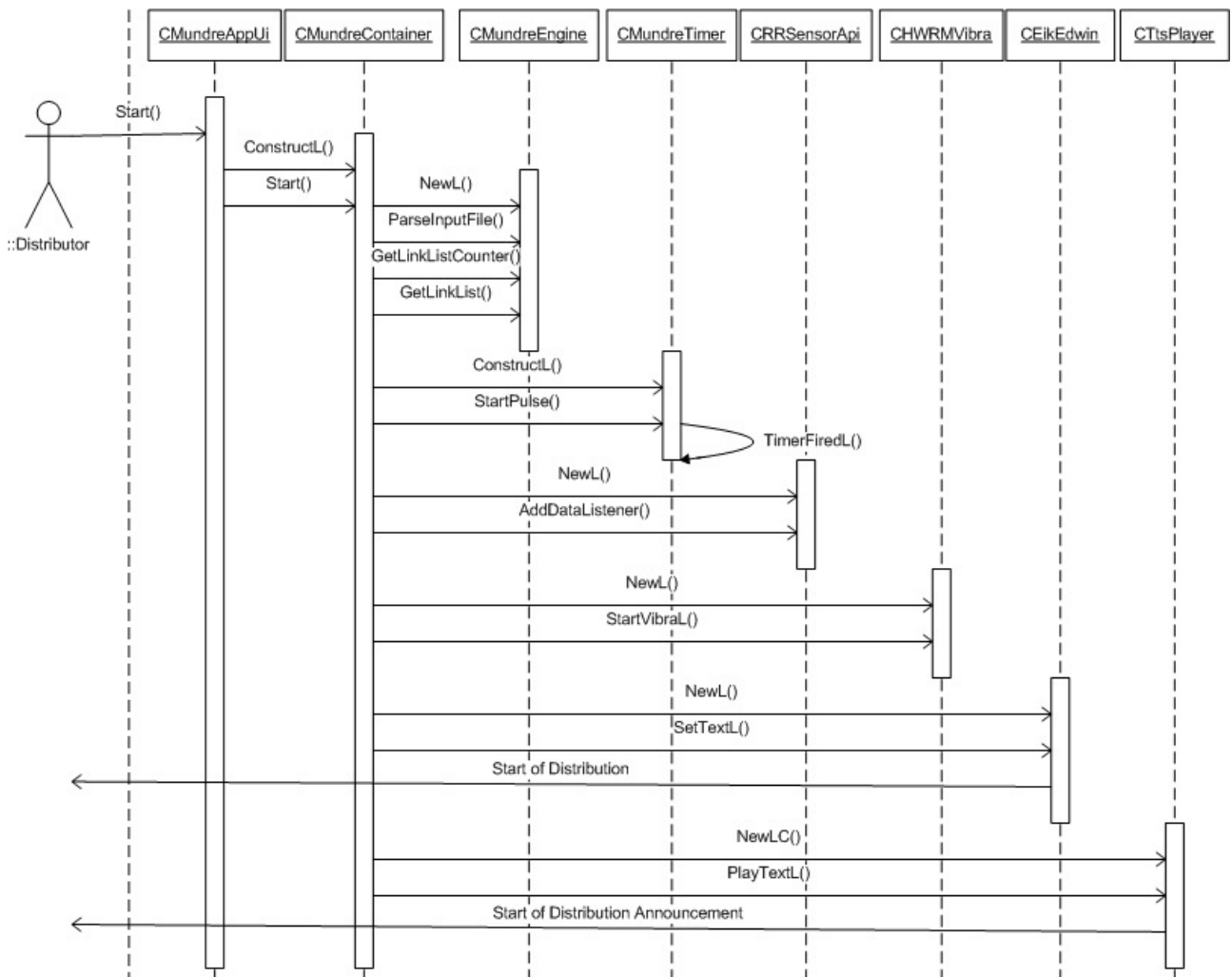
Start of distribution:**Diagram:**

Figure 20: Sequence diagram for Start of the distribution use case

Description: The user starts the application from the mobile device’s “Options->Start” menu. As shown in the sequence diagram the application initializes and constructs the container class object, which further initializes all other dependent class objects. The system response is the announcement coupled with the text displayed, “Start of distribution”.

Enable/Disable multimodalities:

Diagram:

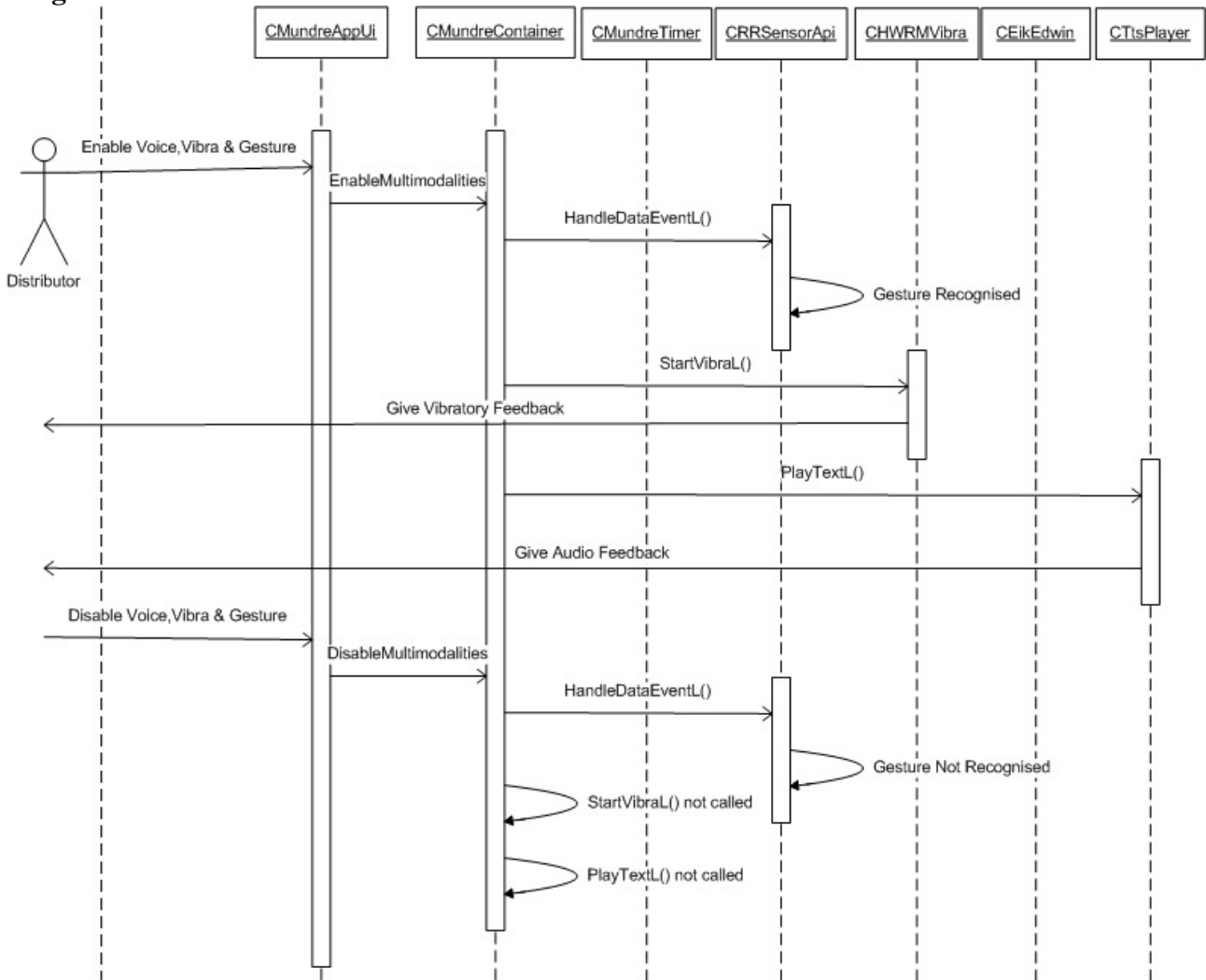


Figure 21: Sequence diagram for Enable/Disable Voice, Vibra & Gesture (multimodalities) use case

Description: The user can use the “Options->Voice,Vibra,Gesture->Enable” menu to enable all the input and output modalities provided by the system. This enables the user to use the multimodal version of distribution transcription. On the other hand when the user disables the voice, vibration and gesture modalities, through the “Options->Voice,Vibra,Gesture->Disable” menu, it enables the use of electronic version of distribution transcription. While using the electronic version of distribution transcription the StartVibraL() for vibration and PlayTextL() for TTS output are not called from the container class, and the HandleDataEventL() do not recognise any gestures.

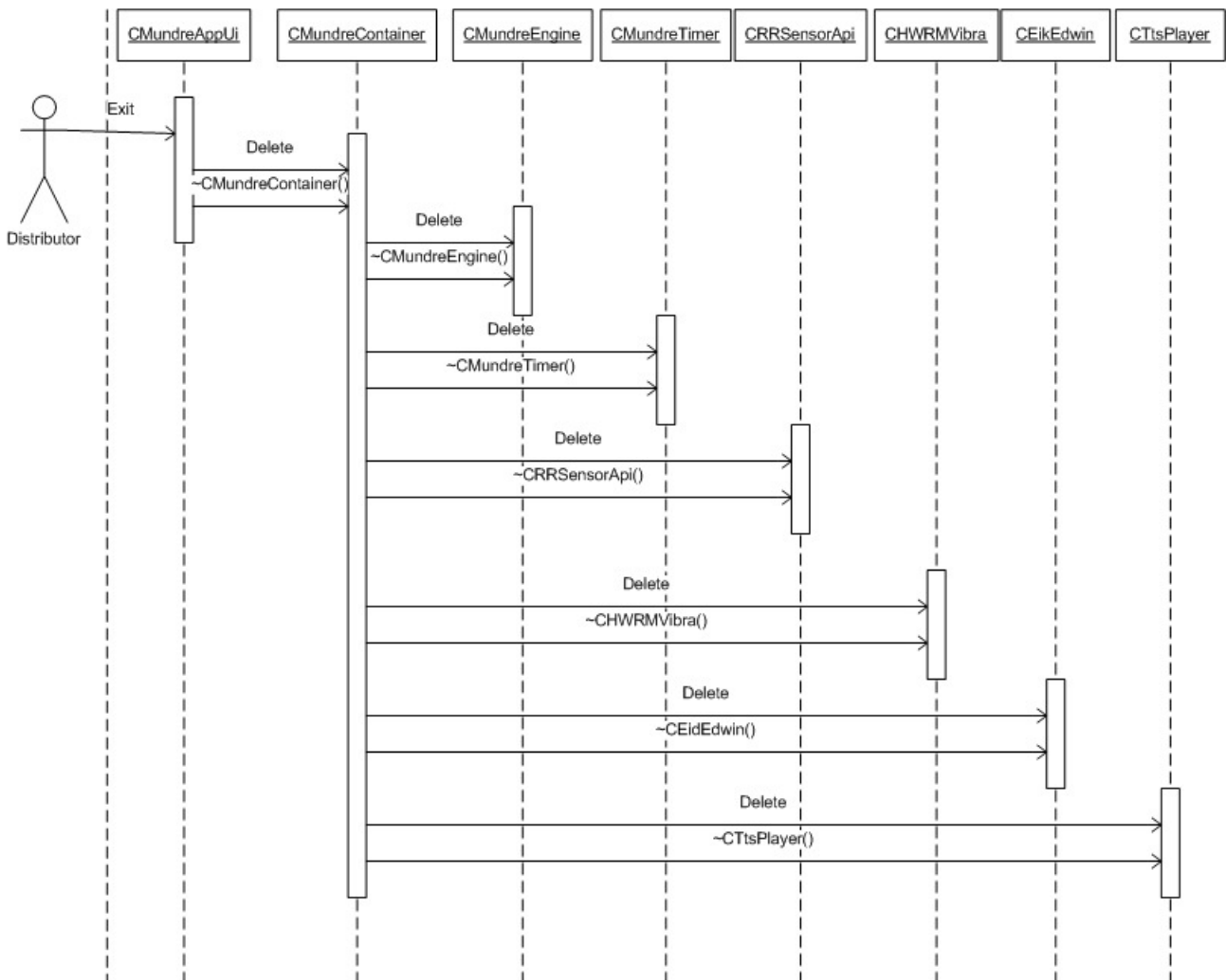
Exit the system:**Diagram:**

Figure 22: Sequence diagram for Exit the system use case

Description: The user can exit the system at any point of distribution by using the “Options -> Exit” menu. The destructor of the container class is called, which further deletes all other created class object and frees the application memory on the mobile device.

Navigate between addresses:

Move next:

Diagram:

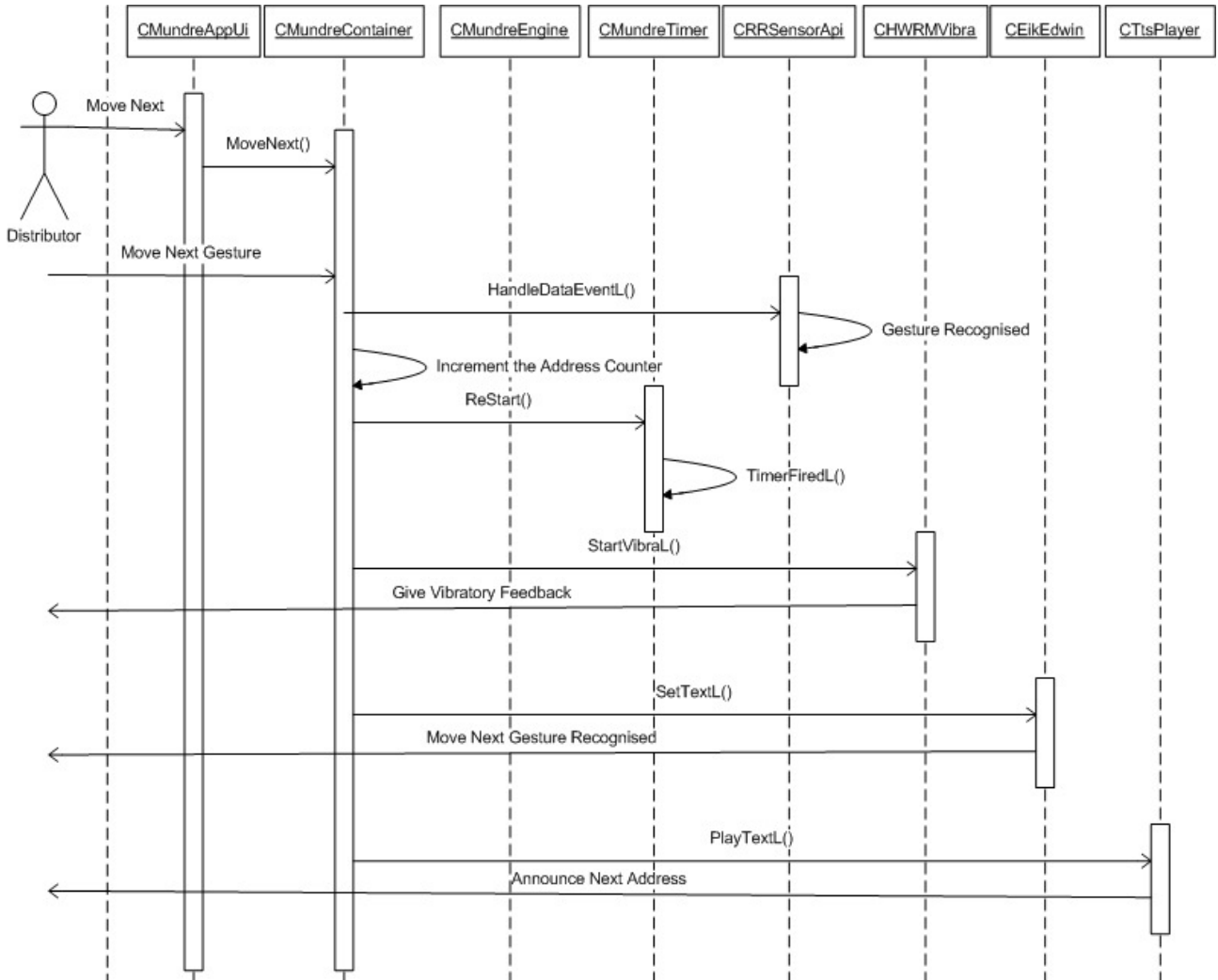


Figure 23: Sequence diagram for move to next address use case

Description: The above sequence diagram shows the interaction for moving to the next address while bypassing the automatic increment of the address counter. The figure illustrates the use of moving to the next address in both electronic and multimodal version of distribution transcription. When the user is using the electronic version only the address counter is incremented, however in case of multimodal version the “Move next gesture” is recognised and the text is displayed on the display of the mobile device as well as announced through the TTS player.

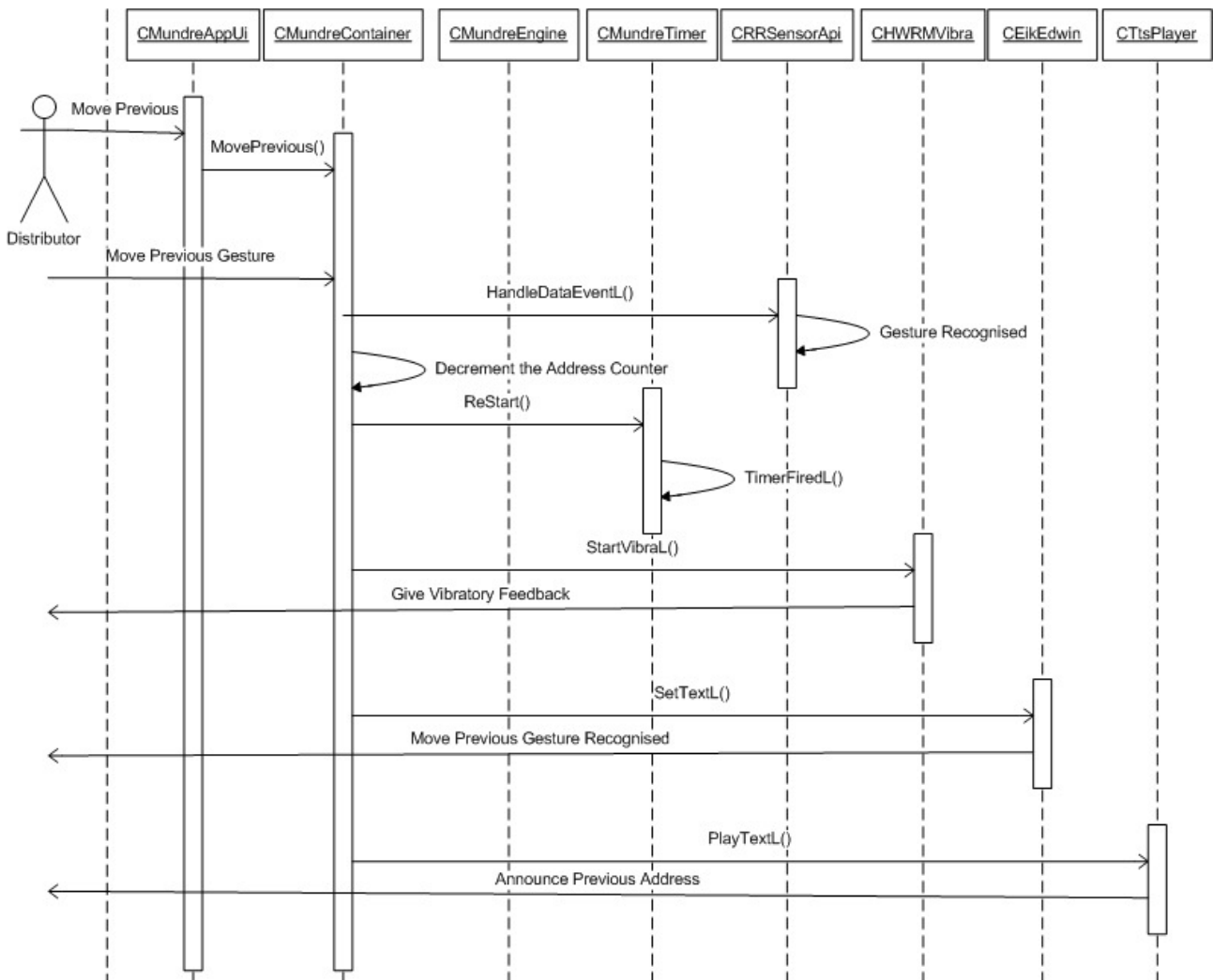
Move previous:**Diagram:**

Figure 24: Sequence diagram for move to previous address use case

Description: The above sequence diagram shows the interaction for moving to the previous address while bypassing the automatic decrement of the address counter. The figure illustrates the use of moving to the previous address in both electronic and multimodal version of distribution transcription. When the user is using the electronic version only the address counter is decremented, however in case of multimodal version the “Move previous gesture” is recognised and the text is displayed on the display of the mobile device as well as announced through the TTS player.

Increase/Decrease distribution speed:

Diagram:

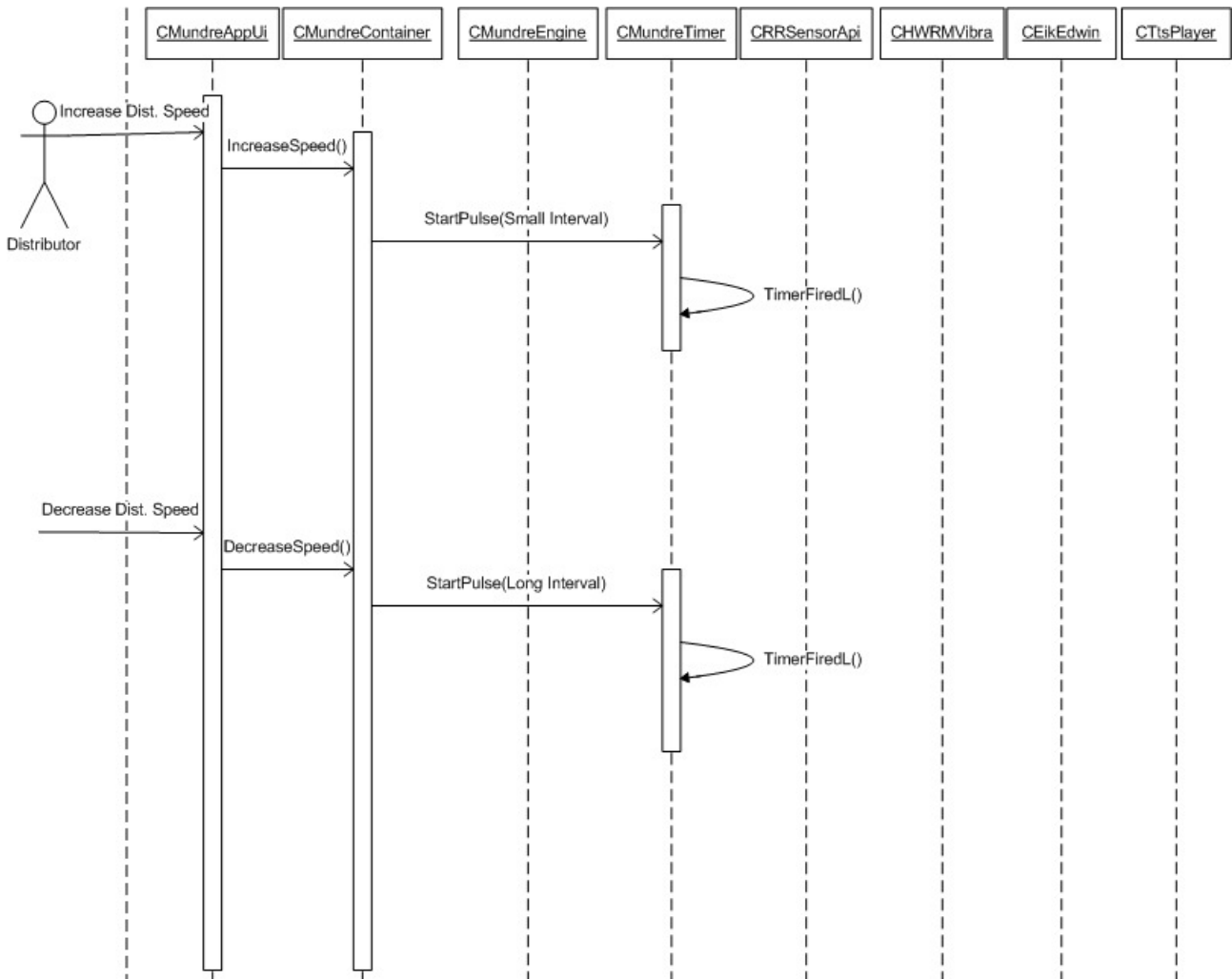


Figure 25: Sequence diagram for Increase/Decrease distribution speed use case

Description: The user can increase or decrease the speed of the system by using the “Options->Speed->Faster” or “Options->Speed->Slower” menu. When the user makes the speed faster, the interval for the firing of the timer is small. On the other hand, when the user makes the speed slower, the interval for firing the timer is long.

Pause and Unpause the system:

Diagram:

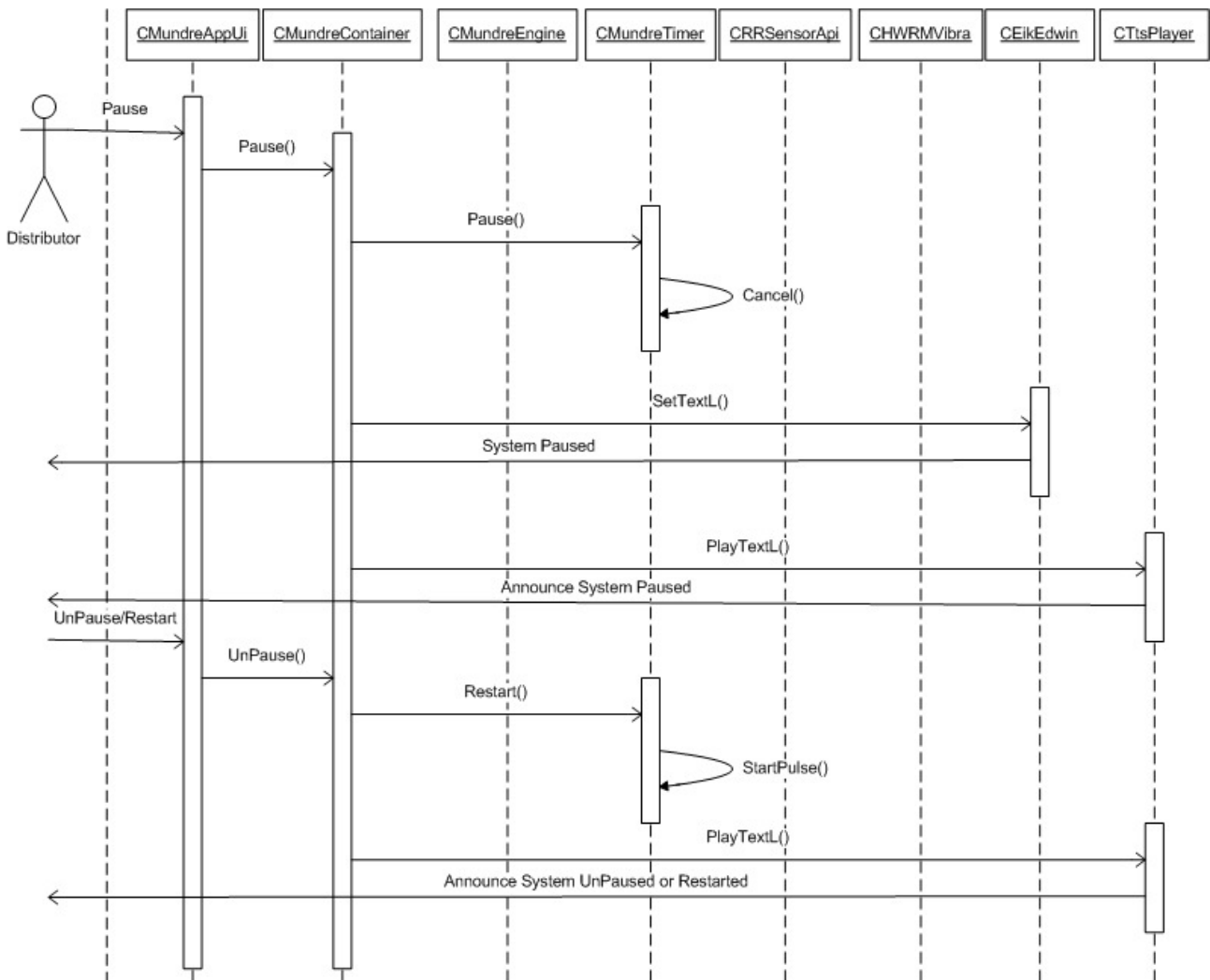


Figure 26: Sequence diagram for Pause and Unpause/Restart the system use case

Description: The user can pause or unpause the system using the “Options->Pause” or “Options->Unpause or Restart” menu. In case of pausing the system the firing of timer is stopped and any other activity of the system is automatically stopped. When the user unpauses or restarts the system, as the address counter is already saved, the system restarts its activities from the same point where it was stopped.

6 Evaluation

The system evaluation was done by practically experiencing the system in a real world scenario and comparing the target goals with the results. The first section will describe the goals of the evaluation in detail; the second section will define the apparatus used for the system, the third section will give details about the procedure followed to carry out the evaluation, the fourth section will describe the evaluation design and finally the fifth section will discuss the evaluation results.

6.1 Goals

Defining the goals of this evaluation will make a base for the comparison between three newspaper distribution techniques. All of the three newspaper distribution techniques are evaluated based on the time and efficiency for precise distribution of newspapers. The distribution technique allowing the distributors to do their job in less time and helping them to minimize errors without creating much cognitive load is considered the best.

The first distribution technique is the manual one and the rest two use mobile device. The manual distribution technique makes use of hand written address book. The second distribution technique is solely based on the electronic version of address book present on the system, where the distributor can only view the text appearing on the mobile device and is not provided with any audio or haptic feedback from the mobile device. However, the third distribution technique is supported by all the modalities provided by system.

6.2 Apparatus

The hardware used to run the system is Nokia N95 and our system application executes as software on the device.

- **Hardware**

The hardware used for carrying out the procedure was Nokia N95. Nokia N95 incorporates the same 3D accelerometer as Nokia 5500, which outputs three 12-bit signed data values at a frequency of around 37Hz. These data values are mapped to positive and negative x, y and z-axis values, ranging from -308 to +308 within the device, and are used for recognizing the gesture by the system. The RMV already embedded in Nokia N95 can rotate in positive and negative directions with the intensity ranging from -100 to 100. Different intensities are used by the system making the distributor easily differentiate between them easily and comfortably.

- **Software**

Our system, the application named 'Multimodal User Interface for Newspaper Distribution Transcription on Accelerometer equipped Mobile Devices' (MUNDRE) is installed on the Nokia N95. The address book XML file is shipped with the MUNDRE installation file, and cannot be modified by the distributor. The experiments for evaluation of distribution techniques were conducted entirely on the mobile device.

The same apparatus was used by all the participants of the experiment for evaluation. The participants were given instructions on how to find and start the installed system application on the device. The instructions for handling the device were also given to the participants before the actual start of the experiment.

6.3 Procedure

10 participants were selected for the experiments, 5 men and 5 women. They were working for one of the newspaper distribution companies in the local city area. The participants did not have any previous experience for using the mobile device to support their newspaper distribution. However, they did have an extensive experience in manual newspaper distribution. The participants took part in this experimental procedure voluntarily and they did not receive any compensation for their participation in the experimental procedure.

The participants were instructed to follow the three distribution techniques separately and report their findings about the usability and feasibility of the system. The time, efficiency and accuracy were considered as important factors evaluating the usability of the system.

Distribution time started when the distributor had just reached outside the first building for distribution and it ended when the distributor had completed distributing all the newspapers for all the addresses. In order to make the distribution even, we made the number of addresses to be reached by the distributor and the number of newspaper deliveries to be equal for every participant.

1. Manual newspaper distribution transcription

The participants followed their daily routine for the newspaper distribution. They noticed and reported the start and the end of the distribution time. These data values were used to calculate the time and efficiency of the distributor. The accuracy for delivering the newspapers correctly for all the addresses was calculated based on the number of complaints made by the people living in that particular newspaper distribution area.

2. Electronic version of newspaper distribution transcription

The second technique for newspaper distribution was taken into account when the participants were instructed to use MUNDRE on their mobile device, while disabling the “Voice, Vibra & Gesture” option from the Options menu. This enabled the participants to use MUNDRE just as an electronic version of their manual address book. The start and the end of the distribution time was noticed and reported by the distributors, while the accuracy for delivering the newspapers was determined by the number of complaints made by the people in that particular newspaper distribution area.

3. Multimodal UI enabled newspaper distribution transcription

The third newspaper distribution technique made complete use of all the modalities provided by the MUNDRE application. The participants were instructed to enable the “Voice, Vibra & Gesture” option, before starting the newspaper distribution. They were allowed to use their mobile device in a normal way and later they were questioned about the usability issues which they answered by filling in the questionnaire. The start and end of the distribution time for newspaper distribution was noticed and reported by the distributors, which helped in calculating the time and efficiency of this newspaper distribution technique and provided a base for understanding the usability issues. The complaints reported by the people determined the accuracy of this newspaper distribution technique.

It was recommended for each participant to be sure that one has the correct address book information in one’s distribution area and they have done a little practice for operating the mobile device, before beginning to perform our system based newspaper distribution techniques.

6.4 Evaluation Design

A within-participants design was used. All 6 participants reported their time taken for manual distribution transcription. Then participants were randomly assigned to two groups of three participants each. The first group performed the experiment with the electronic version of transcription first, followed by multimodal version of distribution transcription, while the second group of participants performed in the reverse order.

For each of the distribution transcription techniques, participants were asked to complete two sessions of 4 buildings of trails each. The buildings were selected on the basis of the number of flats present in that building. A total number of 4 buildings were selected for this purpose. Participants were advised to “Pause” the system and take rest for at least 5 minutes between each building. Both techniques were conducted on different days, such that there is minimum interference between these two distribution techniques. In summary, the design was as follows:

6 participants x
 3 techniques x
 2 sessions per technique x
 4 buildings per session x
 28 addresses per building x
 1 Newspaper per address
 = 4032 newspapers distributed in total.

6.5 Results

This section will give details about the data summary from the evaluation. It will also discuss other evaluation issues such as physical comfort of the distributors, newspaper distribution speed and newspaper distribution complaint rate.

6.5.1 Data Summary

The data was collected from 10 participants for three different distribution transcription techniques and the distributors reported the start and end distribution times. In case of using the manual newspaper distribution technique, the average time calculated for the manual distribution was 4.5 minutes at the rate of delivering an average number of 28 newspapers to 28 addresses per building. While using the electronic version of newspaper distribution transcription, the average time calculated was 4 minutes at the rate of delivering an average number of 28 newspapers to 28 addresses per building. Furthermore, in case of using the multimodal version of newspaper distribution transcription the average time for distribution reduced to 2.8 minutes at the rate of delivering 28 newspapers to 28 addresses per building. Some participants wanted a faster delivery of newspapers therefore they increased the MUNDRE speed, from the options menu, and were able to distribute them in an average time of 2 minutes at the rate of delivering an average number of 28 newspapers to an average of 28 addresses per building.

6.5.2 Physical Comfort

Whilst using the electronic version of newspaper distribution, the participants reported the ease of use as compared to the manual distribution and the time efficiency was obvious. On the other hand, all the participants reported that while using their mobile phone enhanced with multimodal

interaction techniques during their working hours made their work easy, efficient and they completed it with more accuracy.

Moreover, in order to determine the system's usability a survey was conducted and the distributors were asked verbal and non-verbal questions with the help of a questionnaire. The results gained from the answers of distributors show that distributors were excited in using the mobile device during their working hours. Their desire for a change in the current process of the address book handling is obvious from the results. They felt much comfortable in using the multimodal version of distribution transcription. However, they were uncomfortable for using it solely for electronic version of distribution transcription.

The user satisfaction survey conducted after the experiments shows that 4 distributors were unsatisfied with using the electronic version distribution technique. The participants noticed that handling the mobile device and repeatedly making an eye contact with its display did made the system less easily operable and lead to unsatisfactory results. On the other hand, 9 distributors were satisfied by using the multimodal version of the distribution technique, while the remaining one who was unsatisfied, suggested more handful use of customized gestures. However, he was satisfied with the overall performance of the system.

All distributors understood the normal operating procedure for the system correctly and they felt no need for conducting a separate training in order to operate the system normally. This shows that the graphical UI for our system was easy to use and easily memorisable. The distributors were comfortable with the operations of the system on the mobile device. The survey results also reveal that other distribution tasks of the distributors, such as holding the newspapers in the arms and delivering them in the post box, were not affected from the use of system during their working hours. This was truly reflected in case of multimodal version of distribution transcription while the distributors encountered some problems in delivering the newspapers in case of electronic version of distribution transcription.

The distributors welcomed the use of "New Address Summary" shown on the display of the mobile device. This eliminated the manual correction time significantly and made their task easier, which did not require any repeated calculation of the total number of newspapers. This feature proved more beneficial in case of multimodal version of distribution transcription as the distributors were able to listen to the summary information and it eliminated the required visual attention and reduced the cognitive load for doing the calculations repeatedly during their working hours. However in case of using the electronic version of distribution transcription, the distributors were required to make their visual attention toward the display of the mobile device repeatedly.

However, 4 distributors reported difficulties in reading out the text of the addresses in case of electronic version of distribution transcription. They reported the cause of the problem was small text size. However, they affirmed that contents of the information restricted the text size and it was unavoidable to display the small text with the current amount of address information. On the other hand, text appearance was not a problem in case multimodal version of distribution transcription, as all the text was announced to the distributors. Moreover, 8 distributors needed to navigate between addresses during their working hours. In case of electronic version of distribution transcription they used the keys to open the "Options" menu and then manually get back and forth in between addresses. However, in case of using the multimodal version of distribution transcription they used the gesture facility. All the distributors reported that using gestures to control the navigation between the addresses proved more beneficial and less time consuming. Additionally, it required less visual attention and less cognitive load on the distributor as compared to sole electronic version

of distribution transcription. All the distributors reported that using “Options” menu to navigate between addresses and sometimes was even more time consuming as compared to manual distribution technique.

Some distributors reported that the announcements for the addresses were too quick and sometimes they had to use the gestures to get back to the previous address announcement in case of multimodal version of distribution transcription. Therefore, the distributors wanted to be able to customize the distribution speed more precisely in terms of time in seconds. One of the distributors reported that he was unable to comprehend some of the announcements completely. He complained about the pronunciation of the synthesized voice and reported that it was not natural. Later he was notified about the technological constraints for pronouncing the information right. As we already know that most mobile manufacturers till now, do not provide support for all languages in their high quality text to speech synthesizers.

All the distributors reported that their efficiency for distributing the newspapers increased remarkably in case of using the multimodal version of distribution transcription. They appreciated the use of gestures to navigate between addresses and for pausing and restarting the system as well. They commented that this kind of functionality will ease their life during their working hours and they will be able to newspapers distribution will become more interesting. The usability results discussed here prove that using multimodal version of distribution technique was more beneficial as compared to other distribution techniques. The distributors were more satisfied and welcomed the use of such interfaces on mobile devices.

6.5.3 Newspaper distribution speed

Newspaper distribution speed was calculated as the number of newspapers delivered per minute. The timing in our experiments just started when the distributor just reaches outside the first building and looks at the address book and ends when one has delivered all the newspapers. As a result, the time of distribution naturally includes the time for looking at the address book, in case of manual distribution, and looking at the mobile phone, in case of electronic version of distribution; however, this time is eliminated automatically in case of using the multimodal version of MUNDRE, where the distributor only hears the announcements and keeping on distributing the newspapers smoothly.

The average distribution speed was calculated for each distribution technique separately. The average distribution speed was 6 newspapers delivered per minute (dpm) in case of using the manual distribution technique, and was 7 newspapers dpm in case of using the electronic version of newspaper distribution; however, the distribution speed increased remarkably when the multimodal version of MUNDRE was taken into use, and the average distribution speed increased to 10 newspapers delivered per minute.

6.5.4 Newspaper distribution complaint rate

The complaint rate for newspaper distribution determines the number of undelivered newspapers or wrong newspapers delivered in that particular distribution area. The complaint rate for all the three newspaper distribution techniques was calculated and it was only 8% in case of manual distribution, 6% in case of electronic version of newspaper distribution transcription and 2% in case of multimodal version of newspaper distribution transcription.

The complaint rate results reveal that the distribution technique where multimodal version of distribution transcription is used, the performance of distributors increased significantly and they were able to distribute the newspapers more accurately. This also proves that merely using the electronic version of newspaper distribution transcription is not much helpful in lowering down the complaint rate.

7 Discussion

Previously many new interaction techniques had been introduced for gesture based systems. These systems were equipped with the sensor subsystems. Those sensor subsystems were large in size and difficult to handle, such as the systems developed by Tanaka [2004] for mobile music making on a PDA device, the MESH hardware introduced by Hughes et al. [2004] for information navigation and gaming using tilt of the mobile device, the Scan and Tilt system for mobile museum guides [Mäntyjärvi et al., 2006] used the iPAQ PDA equipped with EcerTech's 2D acceleration sensor hardware and the TiltText [Widgor & Balakrishnan, 2003] hardware used a Motorola i95cl was equipped with 2-axis accelerometer's hardware connected to the phone via a serial cable. However, by the advancement of mobile hardware technology, these sensors subsystems started to become embedded in the mobile devices, which made it easy for the users to handle and port the device. Recently, Vajk et al. [2007] used the Nokia 5500's embedded accelerometer to demonstrate using a mobile phone as a "Wii-like" controller for playing games on large public displays. We used a similar approach for selecting the hardware (Nokia N95) for our system, which could be easily handled and ported by the newspaper distributor.

Gestures are used as a fast and natural modality for interacting with the mobile device; Widgor and Balakrishnan [2003] used gestures for text input, Mäntyjärvi et al. [2006] used tilt gesture to select and navigate information appearing on the iPAQ PDA for guiding the user in the museum, Hinkley et al. [2000] and Bartlett [2000] used gestures to scroll the display of the mobile device. The results of their experiments reveal that gestural mobile interaction methods are fast, natural and provide an intuitive way for building eyes free interfaces and can be easily adapted by the users. We also used the same approach and used gestures in our system to navigate the information for newspaper distribution. We introduced a simple and easy to remember intuitive set of gestures for the distributor. This approach introduced a new interaction technique in this kind of working environment and supported the notion of reducing the cognitive load on the distributor. The results of our experiments reveal that distributors found it easy to navigate the information during their working hours and it proved to have an increased efficiency of their work.

However, the results of the experiments conducted by Widgor and Balakrishnan [2003] also reveal that there is a need to precisely determine the tilting parameters in order to determine the cost of tilt movement. We also need a thorough analysis for the tilt movements used in our system in order to determine their cost and to make them more efficient for the distributor.

Linjama et al. [2005] augmented the gestural control with vibrotactile feedback to offer a new dimension to multimodal interaction and concluded that haptic interfaces must be designed to be multimodal such that visual haptic content are in sync and support each other; we also incorporated vibrotactile feedback to enhance the user interaction and to keep it in sync with the information being displayed and announced for the distributor.

Turunen and Hakulinen suggested that adding speech user interface to the mobile devices overcomes the problem of small screen displays in these devices [Turunen & Hakulinen, 2007]. We used a similar approach and we added the text to speech synthesis as an output modality. Incorporating text to speech as an audio feedback to the distributors and vibration as a haptic feedback proved beneficial for them. Our results show that incorporating the multimodal interaction in the system proves to have better usability and the overall efficiency for completing the distribution task has been increased as compared to the electronic and manual version of newspaper distribution techniques.

This discussion is based on the feedback from the distributors. It provides a base for arguments favouring the use of multimodal user interfaces on mobile devices in such working environments. These interfaces are a key for building eyes free interfaces. This thesis work practically demonstrates the use of such interfaces and supports the idea of incorporating multimodalities in the mobile devices and reduces the cognitive load on the users of such systems and resultantly more efficient and better results are achieved in this kind of working environment.

8 Conclusions and future work

Our system offers simple and easily memorisable gestures requiring less visual and mental attention from its users. Although the functionality offered by these gestures is limited in number, however they offer the users to interact in a natural and non-irritating manner during the newspaper distribution. Although getting input is the primary purpose of gesture interaction. However, defining customized gestures for our system can be done in the future. A thorough statistical analysis of the skill improvement phase for gestures and incorporation of customized gestures in our system is needed. Studying those results will reveal important conclusions for using multimodal user interfaces in such working environments.

Our system used the accelerometer sensor for determining the tilt of the mobile device; incorporating any other sensor technologies to determine the related distributor's activities and causing the system to automatically change its state accordingly will certainly enhance the usability of the system. An example of this additional sensor technology is making use of Bluetooth sensor technology, which can activate the key sensor to transmit the identity of the distributor whenever he is near the door building and desires to unlock it.

Timely activation and deactivation of different input/output modalities within the mobile device proved to be a challenging task. Synchronizing all such modalities plays a major role in the implementation such multimodal user interfaces. The rich user interface offered a greater degree of freedom to the distributors and they were able to perform other work related tasks more easily and effectively. This proved to be the main factor for overall performance improvement and low complaint rate for the distributors using our system.

We compared the three newspaper distribution techniques to prove that the development of our system supported the incorporation of multimodalities. Although technological constraints in mobile devices still hinder the efficient use of all human interaction channels, the application development using the existing technologies remains a challenge.

References:

- [Bartlett, 2000] Joel F. Bartlett, “Rock ‘n’ Scroll is Here to Stay”, IEEE Computer Graphics and Applications, 2000, pp. 40-45
- [Chang & O’Sullivan, 2005] A. Chang and C. O’Sullivan, “Audio-Haptic Feedback in Mobile Phones”, ACM CHI 2005, pp. 1264-1267
- [Coulton et al., 2007] P. Coulton, W. Bamford, F. Chehimi, P. Gilbertson, O. Rashid, “Using In-build RFID/NFC, Cameras, and 3D accelerometers as Mobile phone sensors”, Mobile Phone Programming and its Application to Wireless Networking, 2007, pp. 381-396
- [Crossan & Smith, 2004] A. Crossan and R. M. Smith, “Variability in Wrist-Tilt Accelerometer Based Gesture Interfaces”, Mobile HCI 2004, pp. 144-155
- [Eysenck & Kaene, 2005] M. W. Eysenck and M. T. Keane, “Cognitive Psychology, a student’s handbook”, 5th edition, 2005
- [Häkkinen & Mäntyjärvi, 2006] J. Häkkinen and J. Mäntyjärvi, “Developing Design Guideline for Context-Aware Mobile Applications”, Proceedings of the 3rd International Conference on Mobile technology, applications and systems, 2006
- [Harrison et al., 1998] B.L. Harrison, K.P. Fishkin, A. Gujar, C. Mochon, R. Want, “Squeeze Me, Hold Me, Tilt Me! An Exploration of Manipulative User Interfaces”, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1998, pp. 17-24
- [Hinkley & Horvitz, 2001] K. Hinkley and E. Horvitz, “Toward More Sensitive Mobile Phones”, Proceedings of 14th Annual ACM Symposium on User Interface Software and Technology, 2001, pp. 191-192
- [Hinckley et al., 2000] K. Hinckley, J. Pierce, M. Sinclair, E. Horvitz, “Sensing Techniques for Mobile Interaction”, ACM UIST 2000, pp. 91-100
- [Hughes et al., 2004] S. Hughes, I. Oakley, S. O’Modhrain, “MESH: Supporting Mobile Multi-modal Interfaces”, ACM UIST’04, 2004
- [Kela et al., 2005] J. Kela, P. Korpipää, J. Mäntyjärvi, S. Kallio, G. Savino, L. Jozzo, S. D. Marca, “Accelerometer-based gesture control for a design environment”, Proceedings of Ubiquitous Computing 2006, pp. 285-299
- [Linjama & Kaaresoja, 2004] J. Linjama and T. Kaaresoja, “Novel, minimalist haptic gesture interaction for mobile devices”, Proceedings of the 3rd Nordic Conference on Human-Computer Interaction, 2004, pp. 457-458
- [Linjama et al., 2005] J. Linjama, J. Häkkinen, S. Ronkainen, “Gesture Interfaces for Mobile Devices – Minimalist Approach for Haptic Interaction”, Proceeding of CHI 2005 workshop, 2005
- [Mäntyjärvi et al., 2004] J. Mäntyjärvi, J. Kela, P. Korpipää, S. Kallio, “Enabling fast and effortless customisation in accelerometer based gesture interaction”, Proceedings of the 3rd International Conference on Mobile and Ubiquitous Multimedia, 2004, pp. 25-31

- [Mäntyjärvi et al., 2006] J. Mäntyjärvi, F. Paterno, Z. Salvador, C. Santoro, “Scan and Tilt – Towards Natural Interaction for Mobile Museum Guides”, *Mobile HCI 2006*. pp. 191-194
- [Mäntyjärvi et al., 2007] J. Mäntyjärvi, F. Paterno, C. Santoro, “Incorporating Tilt-Based Interaction in Multimodal User Interfaces for Mobile Devices”, “Task Models and Diagrams for User Interface Design”, Volume 4385/2007, 2007, pp. 230-244
- [Neerinx et al., 2008] M.A. Neerinx, A.H.M. Cremers, J.M. Kessens, D.A.V. Leeuwen, K.P. Truong, “Attuning speech-enabled interfaces to user and context for inclusive design: technology, methodology and practice”, *Universal Access in the Information Society*, Long paper, 2008
- [Raisamo, 1999] Roope Raisamo, “Multimodal Human-Computer Interaction: a constructive and empirical study”, *Academic Dissertation*, University of Tampere, 1999
- [Rekimoto, 1996] Jun Rekimoto, “Tilting Operations for Small Screen Interfaces”, *ACM UIST 1996*, pp. 167-168.
- [Rekimoto, 2001] Jun Rekimoto, “GestureWrist and GesturePad: Unobstrusive Wearable Interaction Devices”, *Fifth International Symposium on Wearable Computers 2001 (ISWC’01)*, 2001, pp. 21
- [Segal & Fusella, 1970] Segal, S.J. and Fusella, V., “Influence of imaged pictures and sounds on detection of visual and auditory signals”, *Journal of Experimental Psychology*, 1970, pp 458-464
- [Tanaka et al., 2007] A. Tanaka, G.Valadon, C. Berger, “Social Mobile Music Navigation Using The Compass”, 4th International Mobile Music Workshop, 2007
- [Tanaka, 2004] Atau Tanaka, “Mobile Music Making”, *Proceedings of the 2004 conference on New Interfaces for Musical Expression*, 2004, pp. 154-156
- [Treisman & Davies, 1973] Treisman, A.M. and Davies A., *Divided Attention to ear and eye*. In S.Kornblum (Ed.), *Attention and performance*, Vol. IV. London: Academic Press, 1973
- [Tsukada & Yasummura, 2002] K. Tsukada and M. Yasummura, “Ubi-Finger: Gesture Input Device for Mobile Use”, *Proceedings of APCHI 2002*, Vol. 1, 2002, pp.388-400
- [Turunen & Hakulinen, 2007] M. Turunen and J. Hakulinen, ”Spoken and Multimodal Communication Systems in Mobile Settings”, *Verbal and Nonverbal Communication Behaviours*, 2007, pp. 227-241
- [Vajk et al., 2007] T. Vajk, P. Coulton, W. Bamford, R. Edwards, “Using a Mobile Phone as a “Wii-like” Controller for Playing Games on a Large Public Display”, *International Journal of Computer Games Technology*, 2008
- [Widgor & Balakrishnan, 2003] D. Wigdor and R. Balakrishnan, “TiltText: Using Tilt for Text Input to Mobile Phones”, *ACM UIST 2003*, pp. 81-90
- [Wikipedia-Speech Synthesis, 2008] en.wikipedia.org/wiki/Speech_synthesis
- [Wikipedia-Multimodal Interaction, 2008] en.wikipedia.org/wiki/Multimodal_interaction

[Williamson et al., 2007] J. Williamson, R. Murray-Smith, S. Hughes, “Devices as Interactive Physical Containers: The Shoogle System”, CHI’07 Extended Abstracts on Human Factors in Computing Systems, 2007, pp. 2013-2018

[Williamson et al., 2007] J. Williamson, R. Murray-Smith, S. Hughes, “Shoogle: Excitatory Multimodal Interaction on Mobile Devices”, ACM CHI 2007, pp. 121-124