Brazilian Portuguese Speech Recognition for Navigation on Mobile Device Applications

Proposal

Edwin Miguel Triana Gómez
MSc student of Computer Science

Institute of Mathematics and Statistics
University of São Paulo
Brazil

Outline

- Problem statement
- Objectives
- Background
- Methodology
- Preliminary results
- Implications/Contributions

Problem Statement

Problem Statement

Modern mobile devices

Problem Statement
Mobile devices

- High portability
- Easy manipulation
- Easy transportation
- Higher configurations
- Many connectivity options

Images from:
http://www.umesc.usgs.gov/terrestrial/amphibians/armi/methods.html
http://www.exorcorp.com/page.aspx?pageId=213

Problem Statement

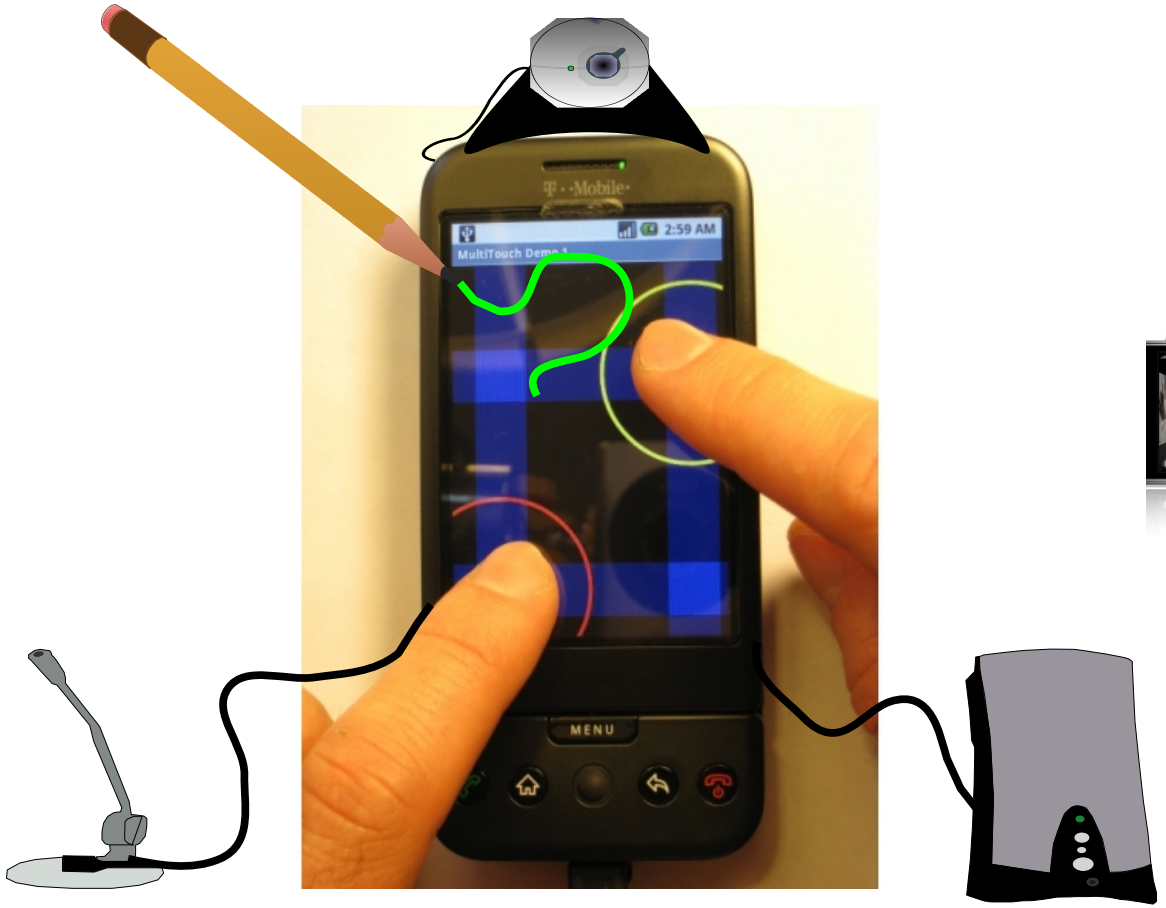
Focus on operation

- Operation requires high attention levels
- Detrimental performance in parallel activities
- Tiny screens -vs- complicated functions

Image by Peter Macdiarmid/Getty Images

Problem Statement

Multimodal interfaces



Images from
http://cdn.androidcommunity.com
http://www.apple.com

Problem Statement
Speech Recognition

Natural skill

No direct attention

Acess to handicapped persons

Image from http://www.talkingfingers.com

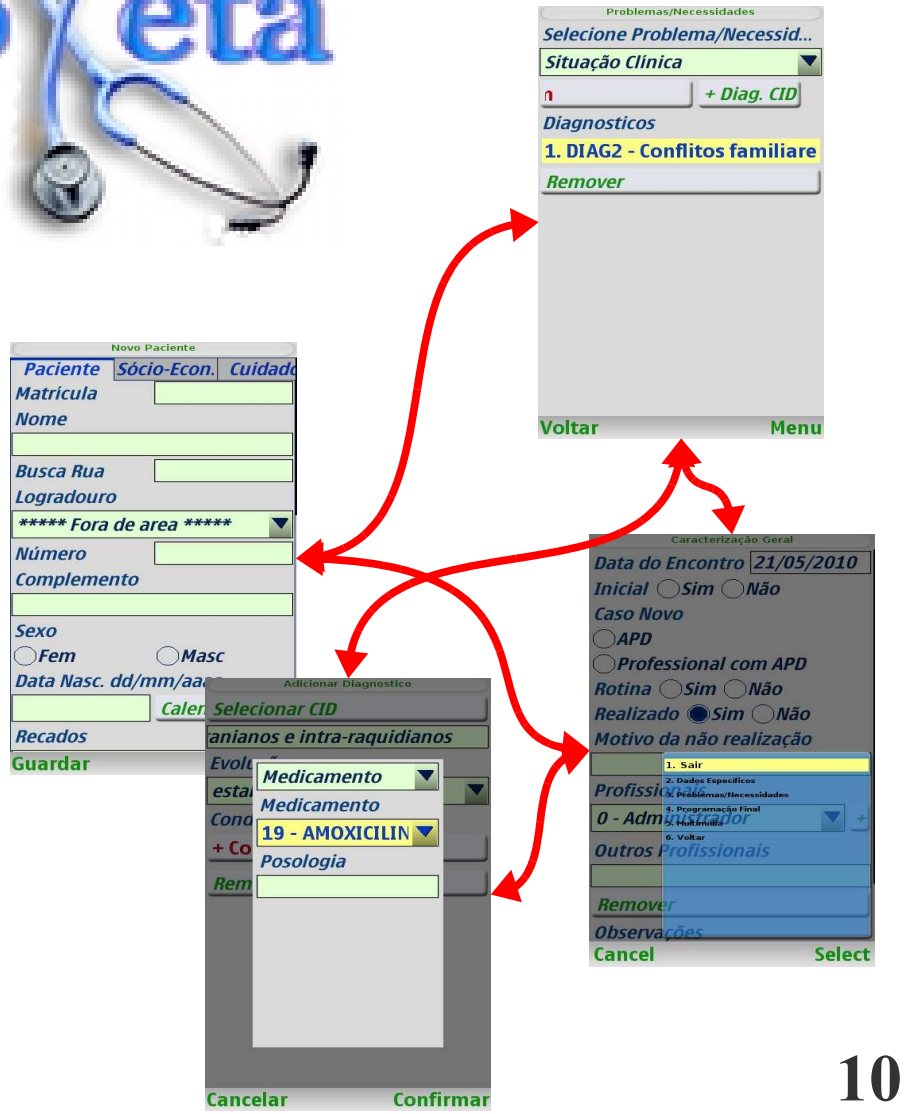
Problem Statement

- Check patient status
  - General
  - Vital signs
  - Related information
- Limited time for a visit
  - More scheduled visits
  - One or more procedures for the same patient

Image from: http://www.scientificamerican.com

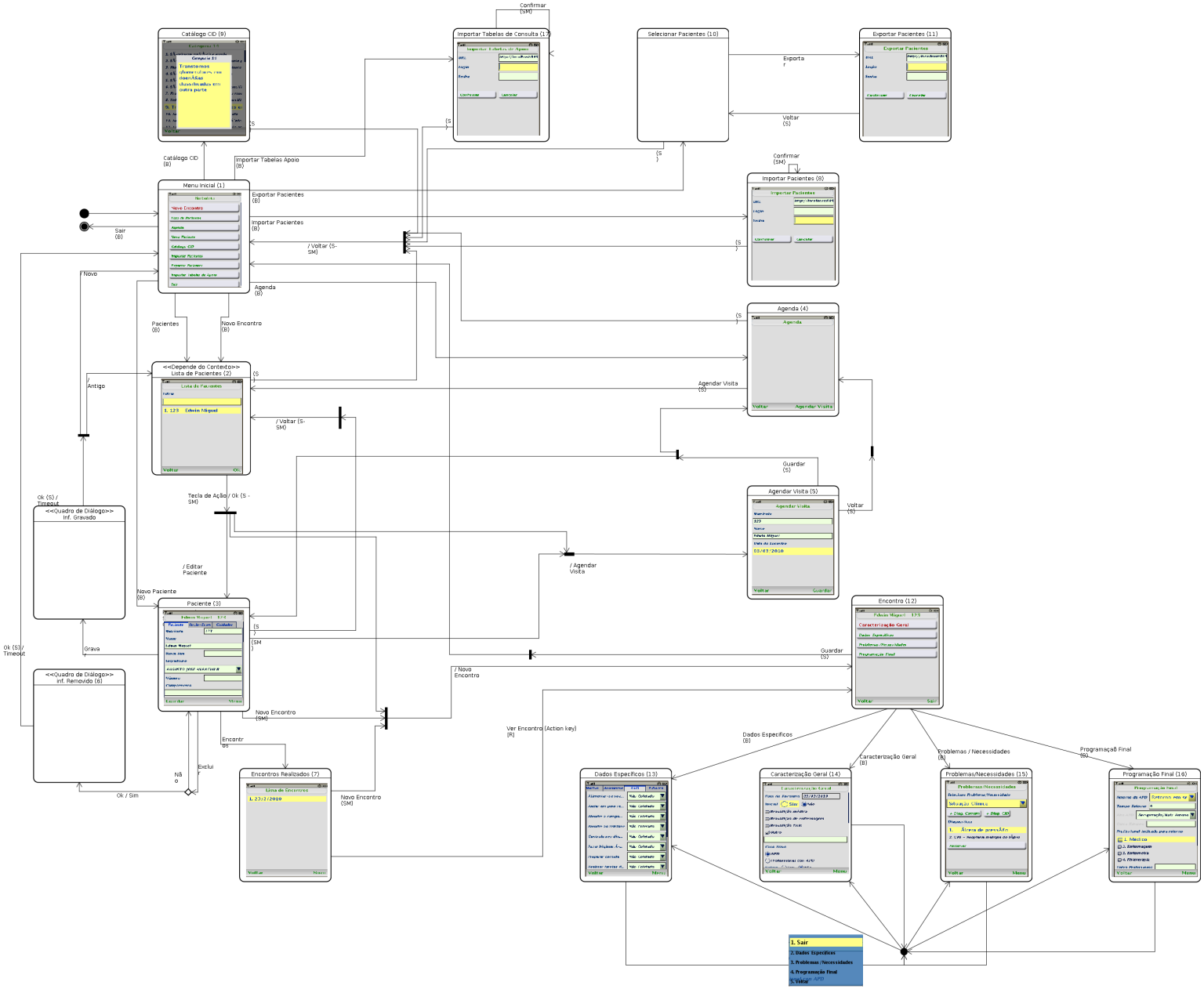
Problem Statement

- Input the data into the device
  - Verify input data
  - Navigate through the system
  - Select the correct interface



Problem Statement

Borboleta



Objectives

Objectives

Verify if the introduction of voice as an alternative input method to navigate through the system will improve the Borboleta system's usability

Objectives

Verify if the introduction of voice as an alternative input method to navigate through the system will improve the Borboleta system's usability

Use a small Portuguese vocabulary to control the Borboleta system, and to open the path for speech recognition of health care notes, requiring a larger specialized vocabulary

Objectives

Verify if the introduction of voice as an alternative input method to navigate through the system will improve the Borboleta system's usability

Use a small Portuguese vocabulary to control the Borboleta system, and to open the path for speech recognition of health care notes, requiring a larger specialized vocabulary

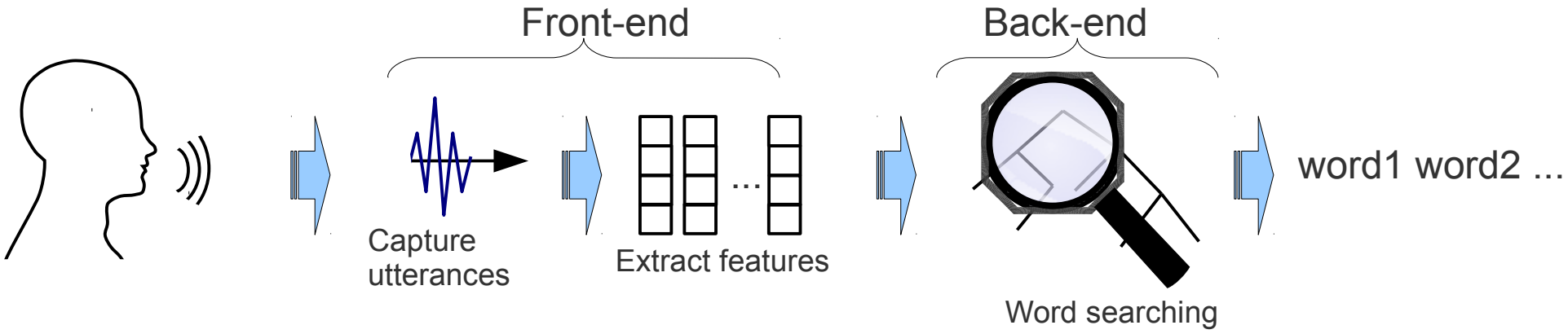
Evaluate and conclude how the speech recognition approaches perform within the Borboleta system

Background
Speech Recognition

Speech Recognition

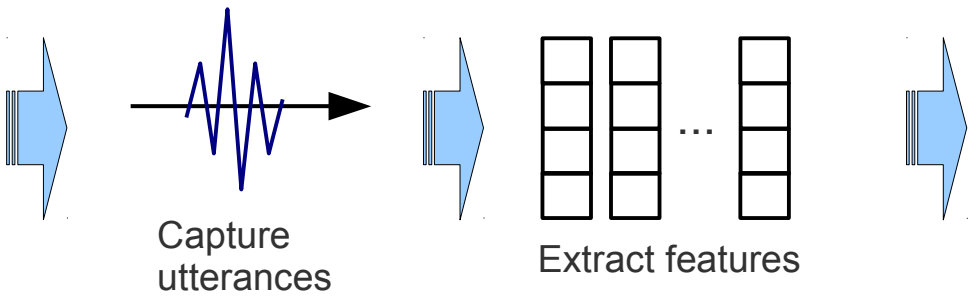
Automatic speech recognition is a process where utterances are translated to character string with an equivalent meaning

Speech Recognition
Process



Speech Recognition

Front End: Signal processing

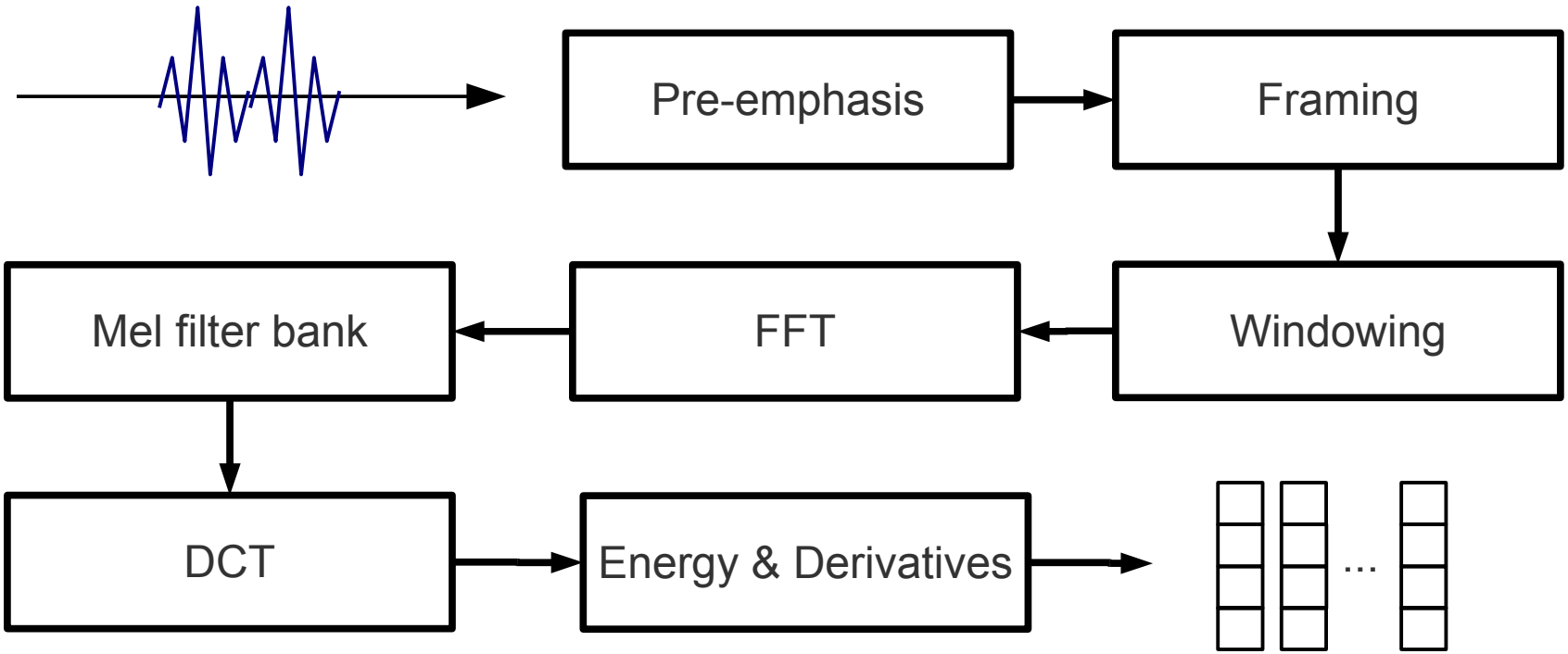


Voice utterances are digitized to waveforms and subsequently transformed to a parametric representation

Methods: Filter-banks, LPC, MFCC

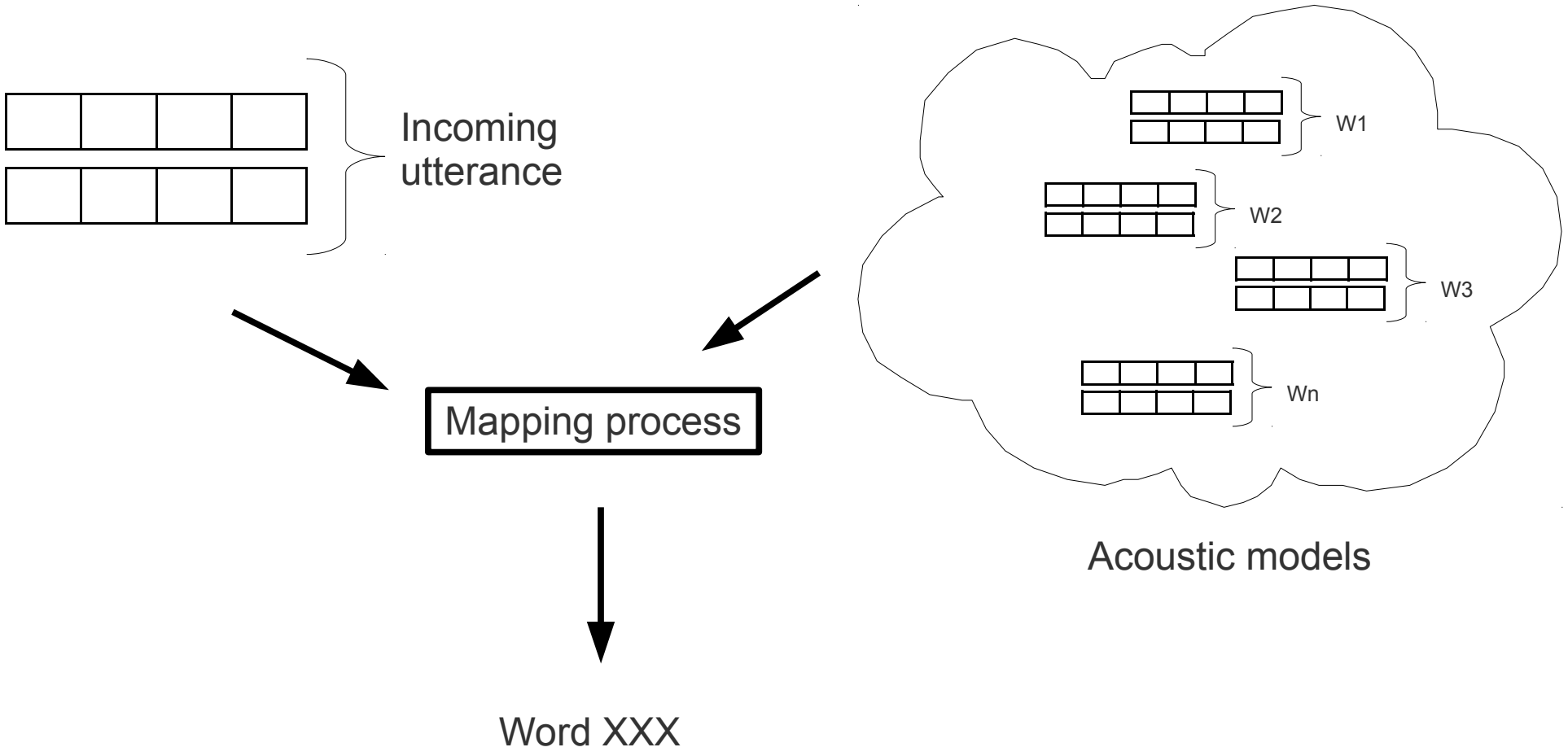
Speech Recognition

Mel Frequency Cepstral Coefficients



Speech Recognition

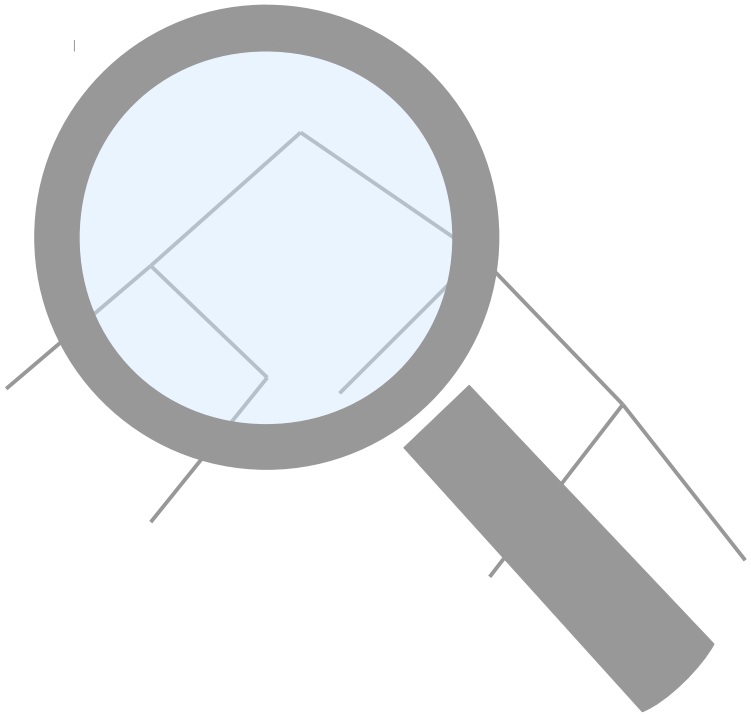
Back End: Speech Recognition Approaches



Speech Recognition

Back End: Speech Recognition Approaches

- Types
  - Isolated word
  - Semi- and Continuous
- Methods
  - Whole word
  - Phoneme
- Techniques
  - Template matching
  - Statistical modeling



Speech Recognition
Template Matching

- When a person says a word several times, each signal representation will follow a pattern.
- Acoustic pattern: Template
- Find the closest template
- Dynamic Time Warping (DTW)
- Euclidean distance

Speech Recognition
Statistical Approach

- Speech variations like pitch, length, internal noise, environmental noise and co-articulation effects at word boundaries make sentences harder to recognize
- Uses Acoustic and Language modeling
- Hidden Markov models (HMMs) is the most used technique

Speech Recognition

Statistical Approach - HMM

$$\underset{i}{\operatorname{argmax}}\{P(w_i|O)\} \longrightarrow P(w_i|O) = \frac{P(O|w_i)P(w_i)}{P(O)}$$

~~$P(O|w_i) = P(o_1, o_2 \ldots |w_i)$~~

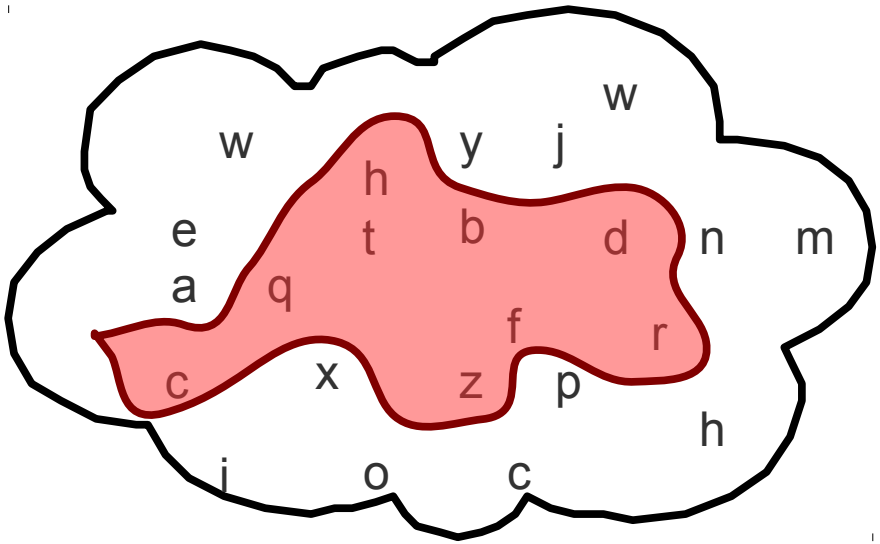
HMM

$$P(O|w_i) = P(O|M_i)$$

Speech Recognition

Statistical Approach – Language Model

- *P(Wi)* = Language Model
- Reduce the search space
- Probabilistic model of the all possible sequences in the system
- CFG and N-gram



Speech Recognition

Software - Sphinx

- From Carnegie Mellon University (1988)
- HMM-based - Speaker independent
- Sphinx-(X): I, II, III, 4 and Pocketsphinx
- Sphinx-based projects:
  - Lumenvox
  - Microsoft Whisper
  - Jvoice-xml
  - Gnome-voice-control
  - Zanzibar

Speech Recognition
Software - HTK

- From Cambridge University (1989)
- HMM Tool Kit
- Became part of the ECRL in 1995
- Microsoft purchased the ECRL in 1999
- Source code released in September 2000
- Current version 3.4

Speech Recognition
Software - Julius

- From Kyoto University (1991)
- Copyrighted by Nagoya Institute of Technology in 2005
- Widely compatible with HTK
- Target platform: Linux
- There is a speech recognition system for Brazilian Portuguese

Word Error Rates

Task	Vocabulary Size	WER (%)
Connected digits	11	0.6
Resource Management	1000	2.7
Wall Street Journal	5000	7.2

WER for Sphinx 4[1]

1. Dmitry Zaykovskiy. Survey of the speech recognition techniques for mobile devices. SPECOM'200, pages 88–93, 2006

Background
Brazilian Portuguese Corpora

Brazilian Portuguese Corpora

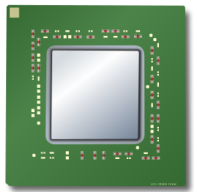
- Corpus: Data set of utterances and their transcriptions
- Available Brazilian Portuguese corpora
  - OGI-22 – WER = 19.9%[1]
  - Spoltech - WER = 18.6%[1]
  - GlobalPhone
  - VoxForge
- **Common vocabulary domain**

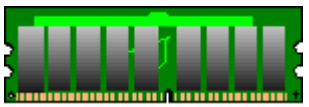
1. Nelson Neto, Patrick Silva, Aldebaro Klautau, and Andre Adami. Spoltech and OGI-22 Baseline System for Speech Recognition in Brazilian Portuguese. International Conference on Computational Processing of Portuguese Language, 5190/2008:256–259, 2008.

Background
Speech Recognition on Mobile Devices

Speech Recognition on Mobile Devices

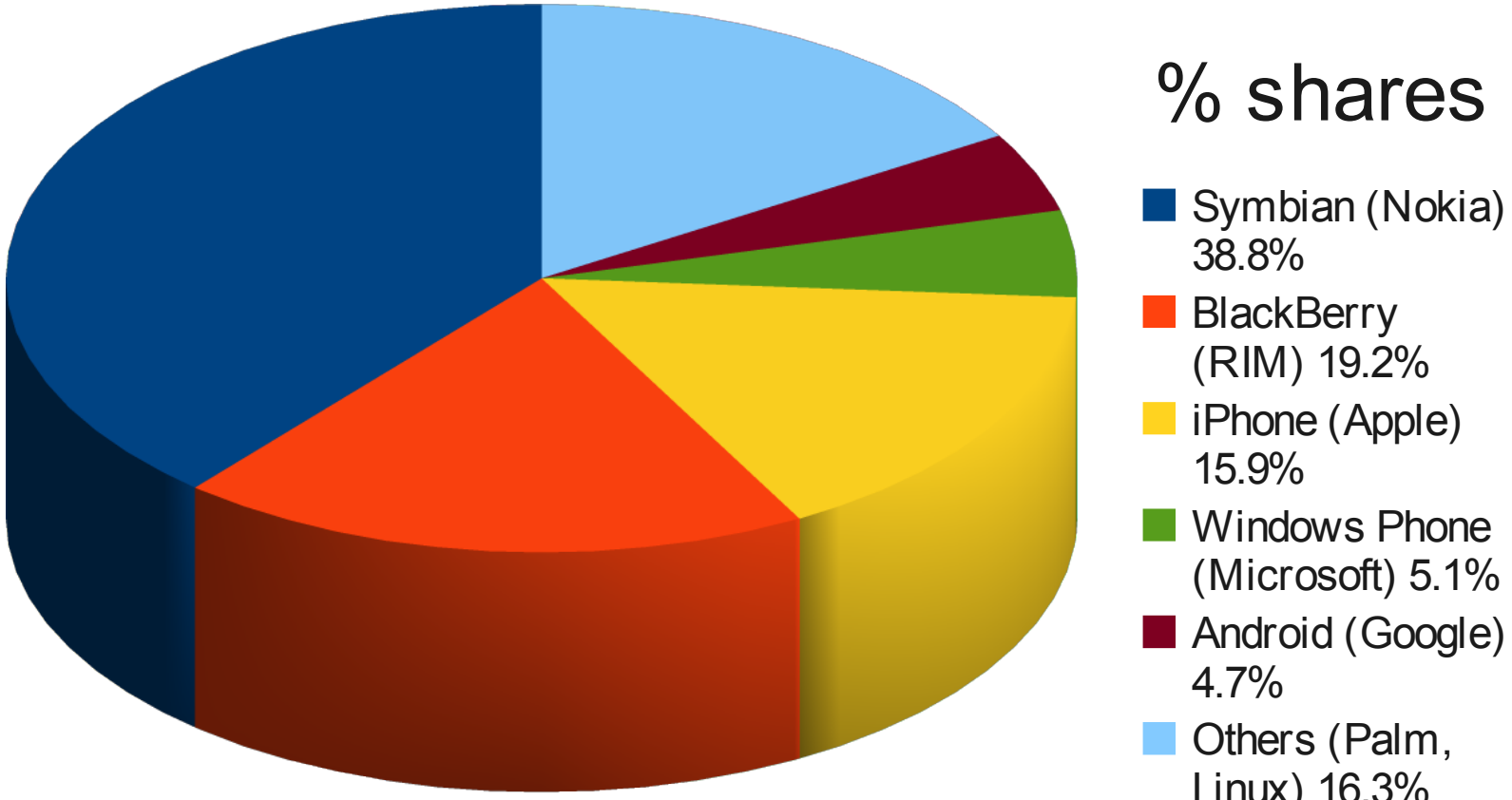
Mobile devices and platforms





Speech Recognition on Mobile Devices

Mobile devices and platforms

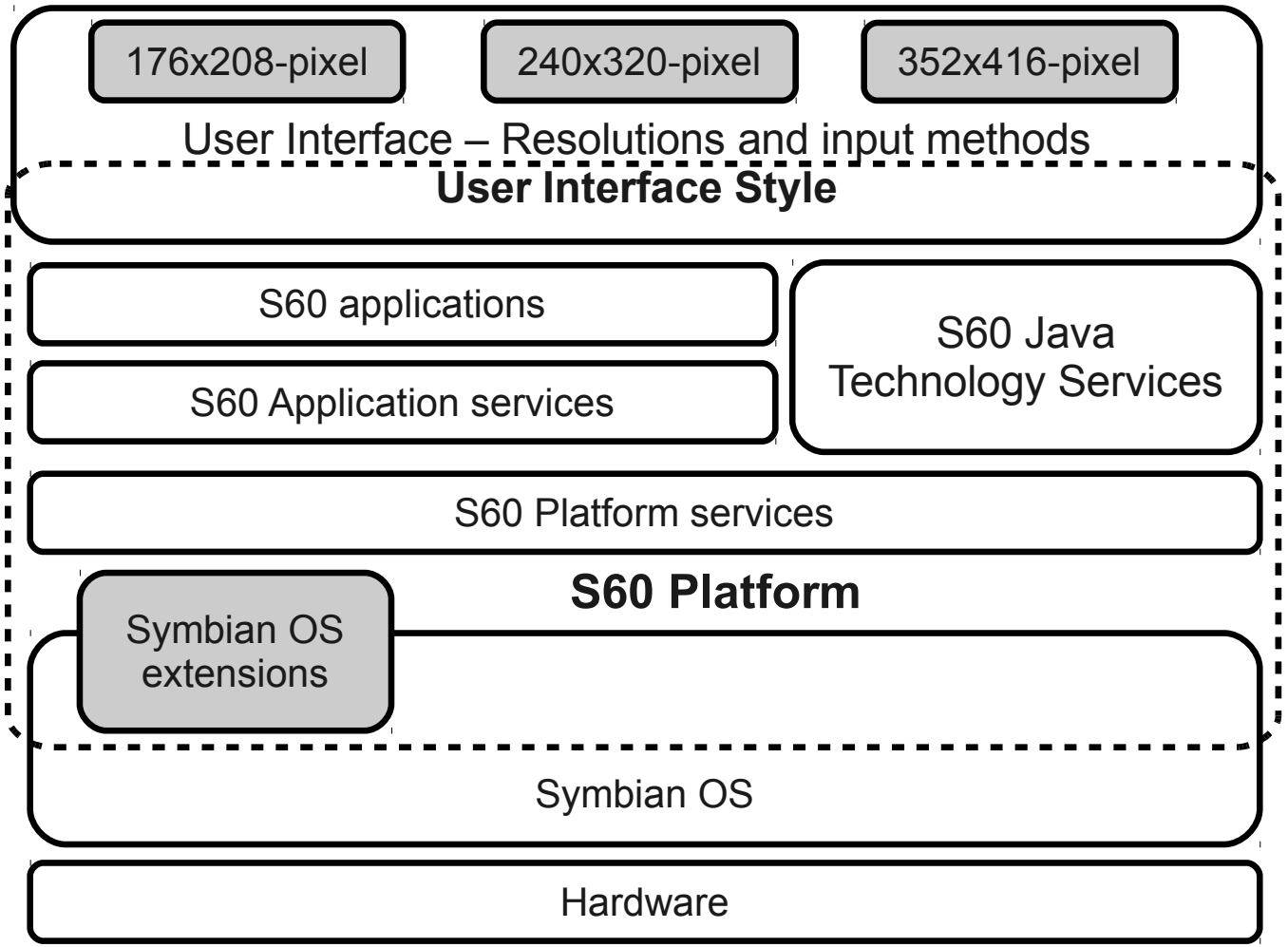


Mobile OS market Q1 2010

ttp://www.canalys.com/pr/2009/r2009112.html

Speech Recognition on Mobile Devices

Symbian

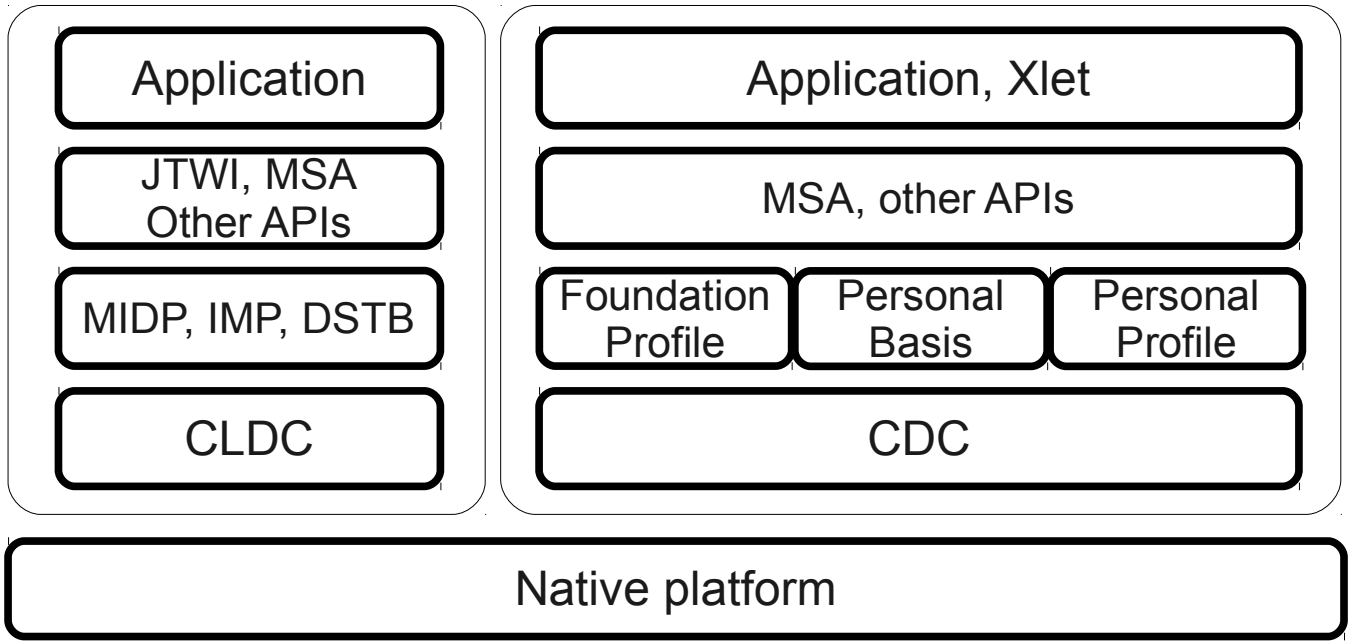


Speech Recognition on Mobile Devices
POSIX libraries for Symbian

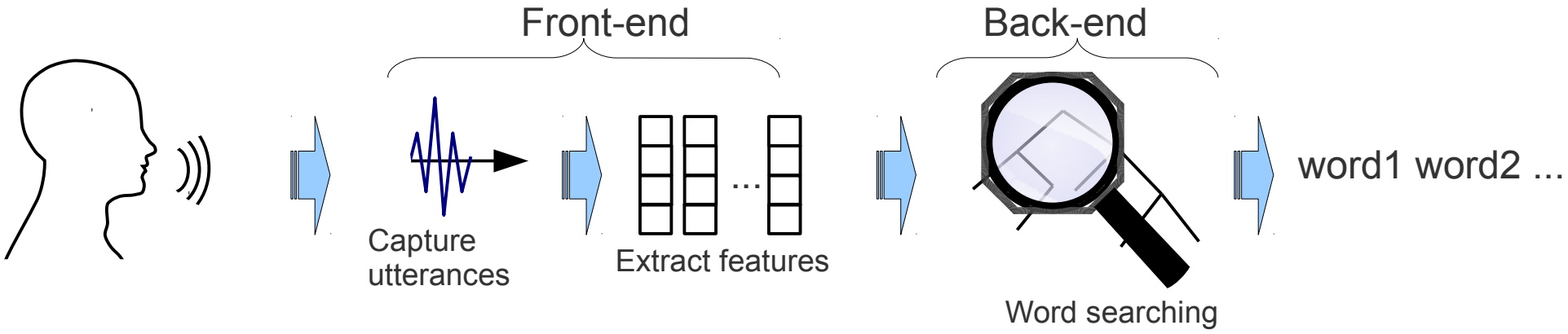
Implementation	Supported libraries
P.I.P.S	libc, libm, libpthread, libdl
Open C	P.I.P.S, libz, libcrypt, libcrypto, libssl, libglib
Open C/C++	Open C, IOStreams, STL, Boost

Speech Recognition on Mobile Devices

Java Micro Edition

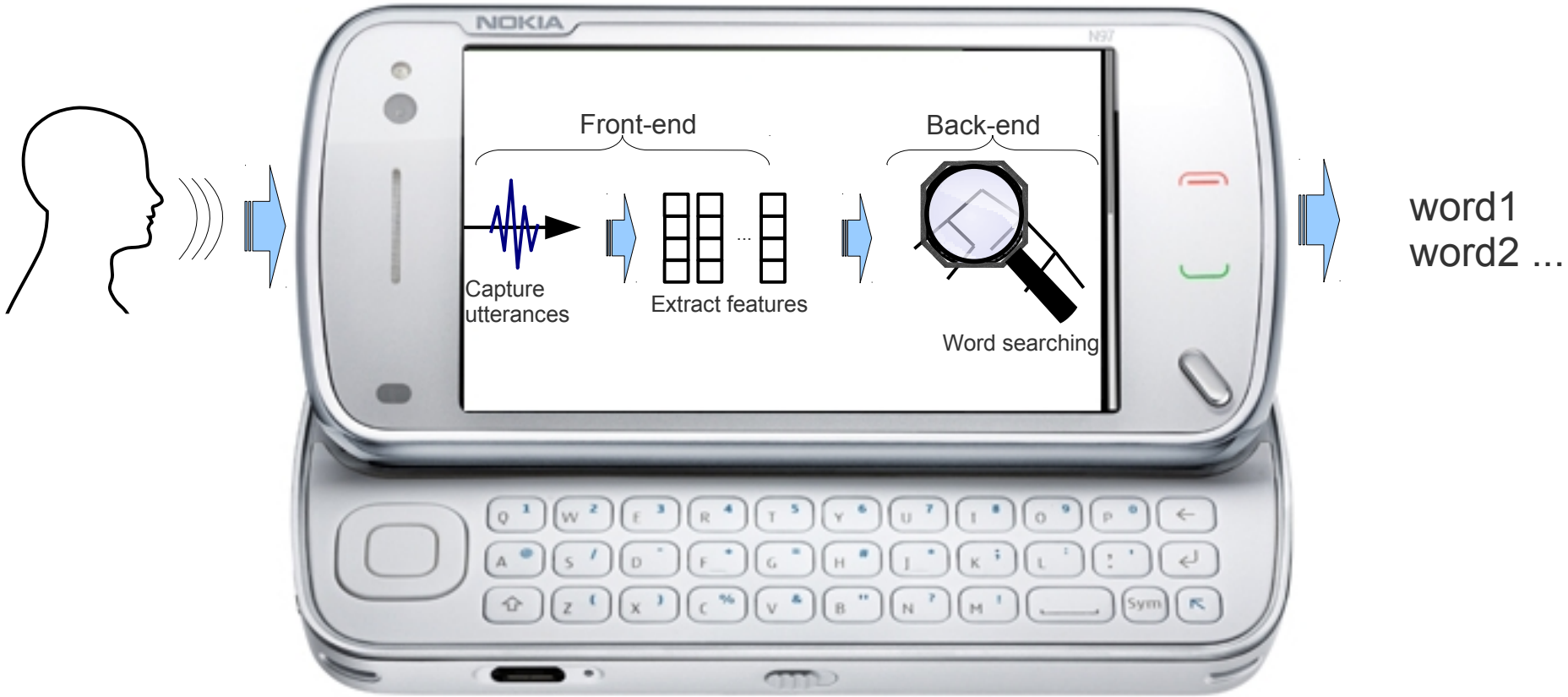


Speech Recognition on Mobile Devices
Approaches



Speech Recognition on Mobile Devices

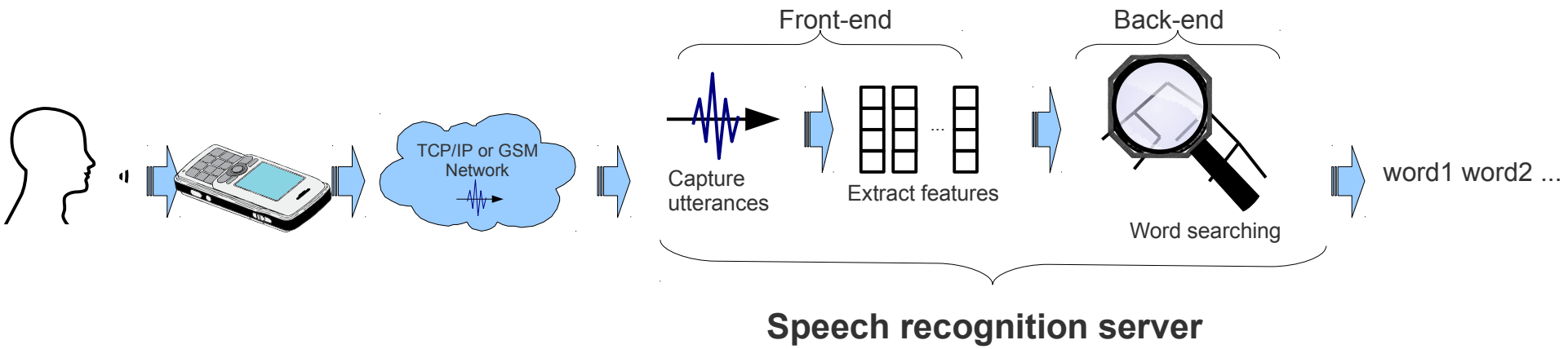
Embedded Speech Recognition



http://conversations.nokia.com/wp-content/uploads/ovi-store-landscape.jpg

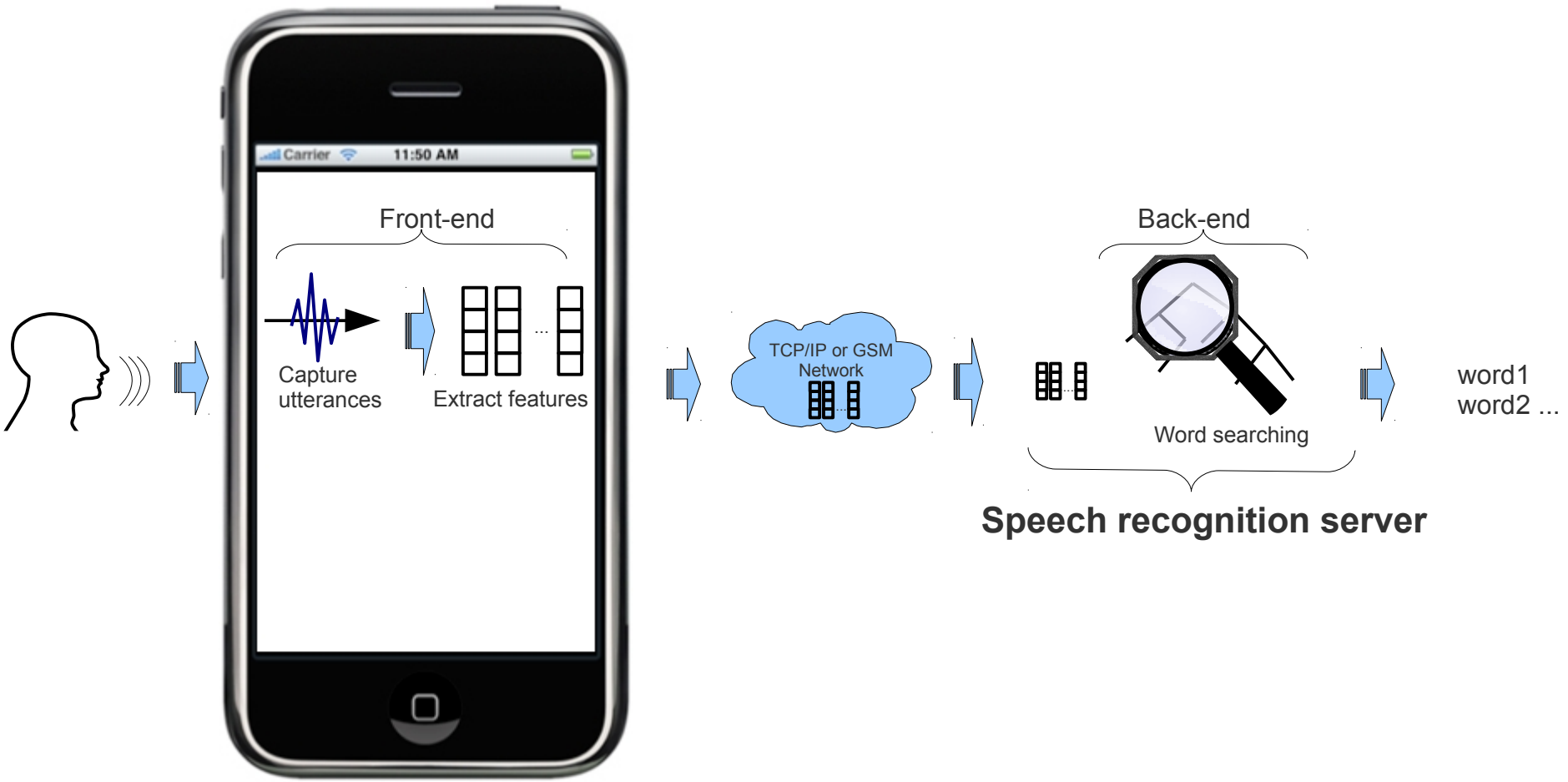
Speech Recognition on Mobile Devices

Network Speech Recognition



Speech Recognition on Mobile Devices

Distributed Speech Recognition



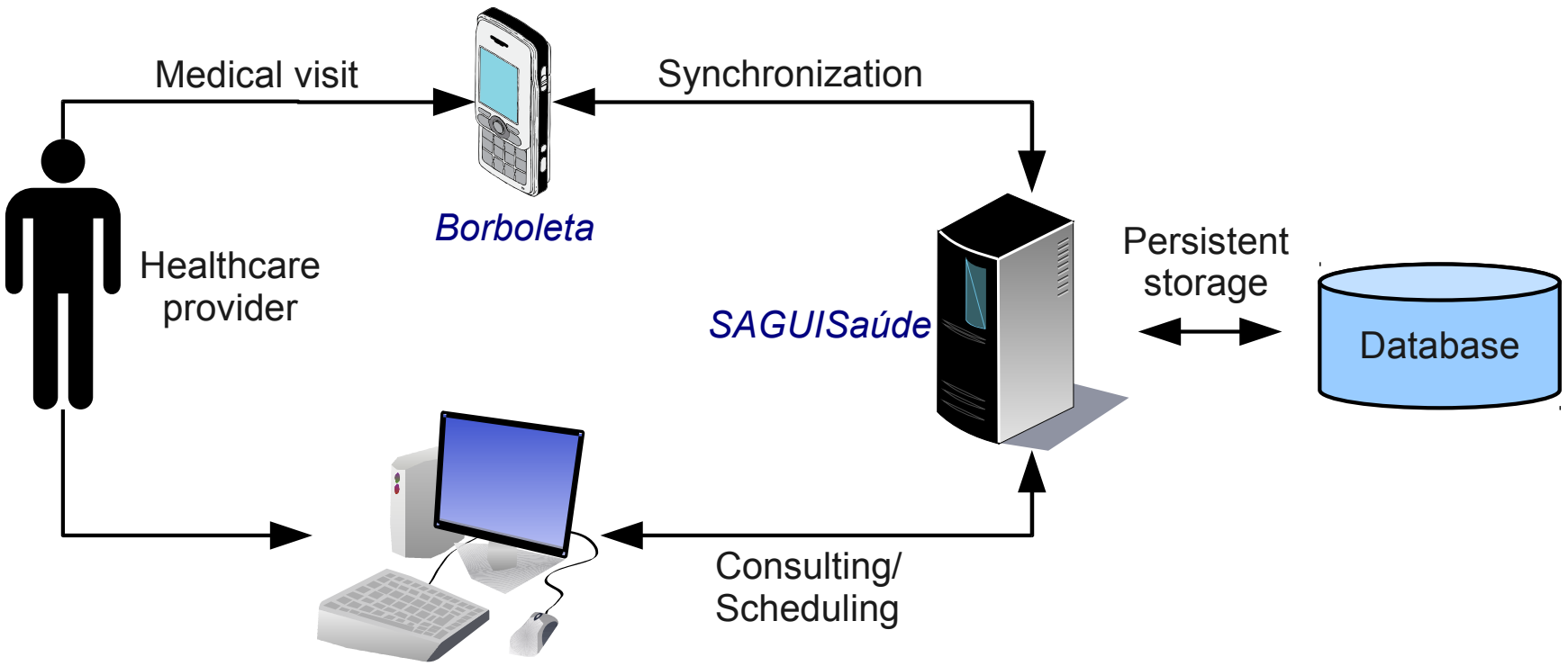
http://developer.apple.com

Speech Recognition on Mobile Devices
Current Applications

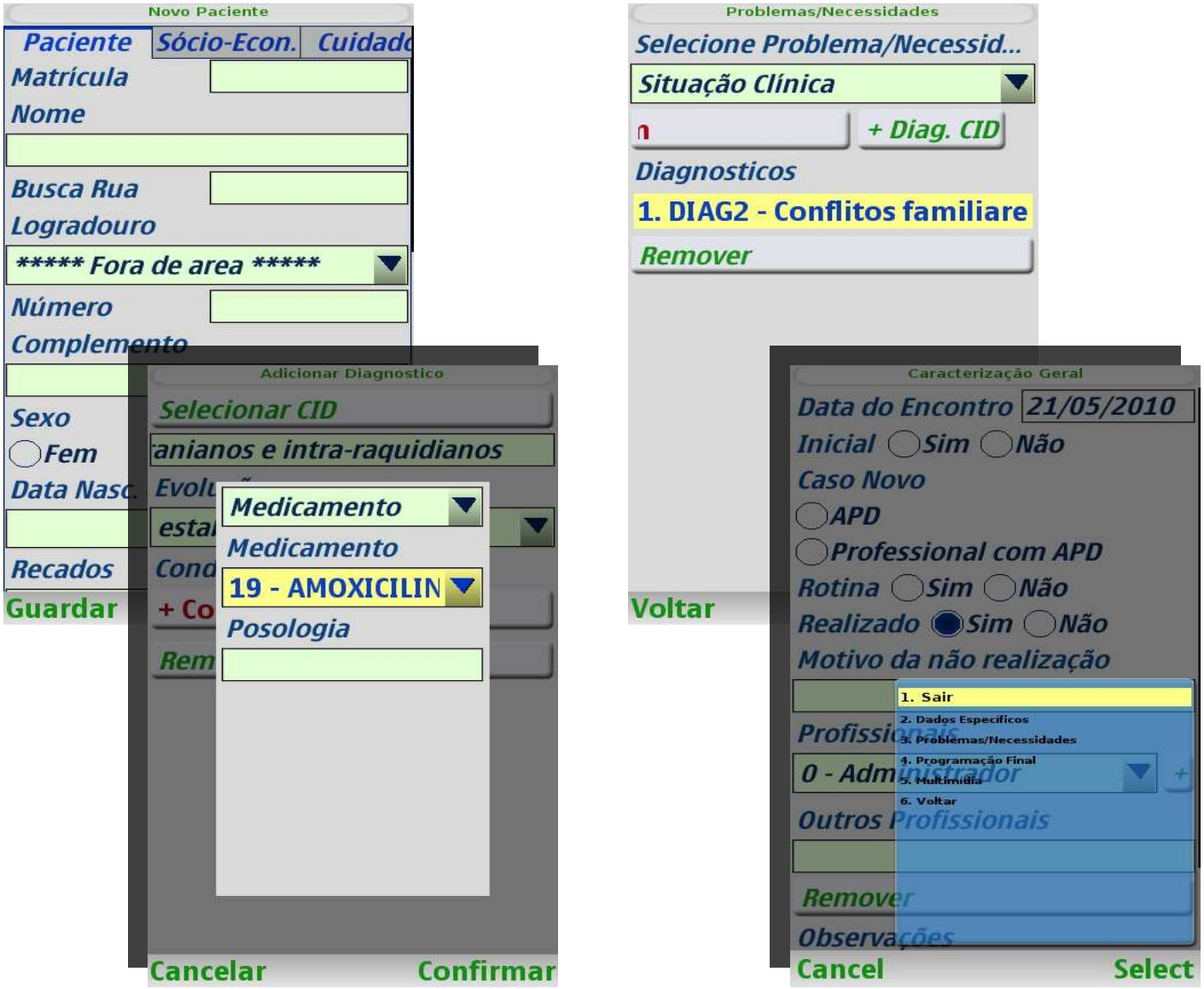
Software	Developer	Words	WER
PocketSphinx	CMU	994	14%
PocketSUMMIT	MIT	2000	19%

Background

Borboleta
General Architecture



Borboleta
Interface



Methodology

Methodology
Activities

1. Literature Review
2. Software Review
3. Data Collection
4. Initial Training and Testing
5. Software Design and Development
6. Evaluation

Methodology
Activities

1. Literature Review
   - State of the Art
2. Software Review
   - Speech recognition robustness and accuracy
   - Software compatibility with mobile devices
   - Software architecture
   - Source code availability, license and community activity

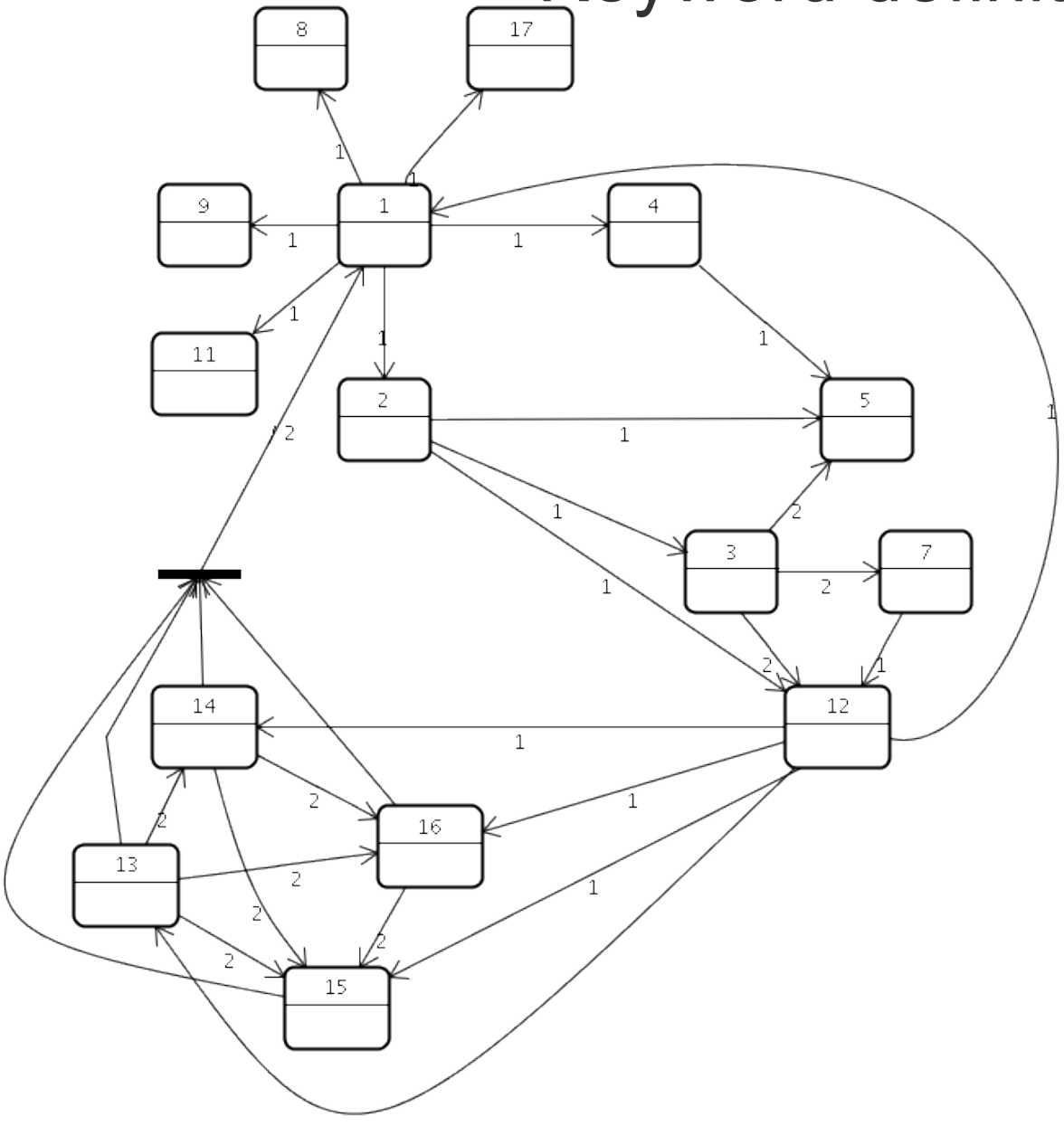
Methodology
Activities – Data collection

Data Collection

- Keyword definition
- Utterance Collection
  - 525 samples (35 samples/word)
- Data preparation

Methodology

Keyword definition



1. Main menu
2. Patient list
3. Patient
4. Appointment book
5. Schedule visit
6. Delete patient confirmation
7. Visit history
8. Import patients
9. CID categories
10. Select patients
11. Export patients
12. Residential visit
13. Specific data
14. General description
15. Problems/Needs
16. Final program

Methodology

Keyword definition

Portuguese	English
Menú Inicial	Main menu
Dados específicos	Specific data
Caracterização geral	General description
Problemas e necessidades	Problems and needs
Programação final	Final program
Encontros realizados	Visit history
Encontro	Residential visit
Novo encontro	New residential visit
Paciente	Patient
Agenda	Appointment book
Agendar visita	Schedule visit
Catálogo CID	International Disease Catalog(IDC)
Sair	Exit
Voltar	Back
Fechar	Close

Methodology

Activities - Initial Training and Testing

- Initial testing on desktop
- Get the system accuracy
  - Whole-word
  - Phoneme
- 70% Training (367 samples)
- 30% Testing (158 samples)

Methodology
Software design and development

- Borboleta was developed for Java ME
- PocketSphinx is developed in C++
- Challenges
  - Port PocketSphinx to Symbian (OpenC++)
  - Modify the capture protocols for Symbian compatibility
  - Define PocketSphinx – Borboleta communication
  - Reduce resource usage

Methodology
Activities

- Evaluation
  - System performance
    - Response time
    - Resource usage
  - System accuracy (WER)
  - Preliminary usability evaluation
  - Feedback in development
  - Final usability evaluation

Methodology

Workplan

Activity	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Fev	Mar
Literature Review													
Software evaluation													
Keyword definition													
Utterance collection													
Data preparation													
Training and Testing													
Software Design													
Software Development													
Accuracy tests													
Usability tests													
Thesis writing													
Thesis preparation & defense													
Corrections													
Paper writing													

Completed task

Current task

Incomplete task

Preliminary Results

- State of the art defined for SR mobile devices and Brazilian Portuguese
- Borboleta navigation and cost diagrams created
- Keyword collection done (525 samples)
- Java ME software developed to collect samples
- Infrastructure for large vocabulary SR is being prepared

Significance/Contributions

- Multimodal interfaces for Borboleta
- Foundations for subsequent research in speech recognition for Brazilian Portuguese (medical domain)
- SR performance and accuracy results
- SR advances closer to real life cases

Thanks!