ALADIN

The development of self-learning assistive vocal interfaces for people with physical impairments

Dag van de Fonetiek, 15 December 2011

Janneke van de Loo, Guy De Pauw, Jort Gemmeke,
Peter Karsmakers, Bert van den Broeck, Walter Daelemans,
Hugo Van hamme

Universiteit Antwerpen

COMPUTATIONAL LINGUISTICS & PSYCHOLINGUISTICS RESEARCH CENTER / CLiPS

ALADIN - introduction

Demand from health care sector:

Vocal user interfaces for people with **physical impairments**, for controlling devices at home (TV, radio, lights, etc.)

ALADIN - introduction

Problems with existing vocal interfaces:

- high development costs because of large variation in user needs
- lack of robustness of speech recognizer to:
  - user-specific speech characteristics (pathological speech, regional pronunciation variation)
  - environmental noise
- predefined vocabulary & grammar → learning and adaptation required from user

ALADIN - introduction

→ **ALADIN**: Adaptation and Learning for Assistive Domestic Vocal INterfaces

Goal: develop a robust, self-learning domestic vocal interface that adapts to the user instead of the other way around:

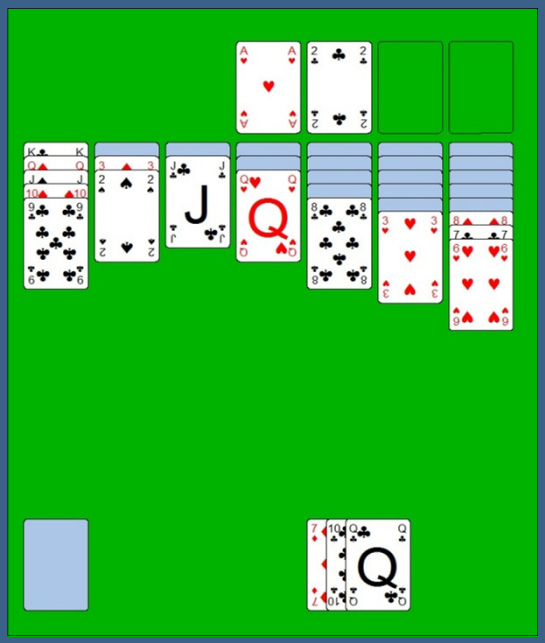
- learn the user's vocabulary & grammar constructs
- learn the user's voice & pronunciation characteristics

How? Unsupervised learning on the basis of training examples: vocal commands + associated controls (actions)

ALADIN - introduction

Target applications:

- control devices such as TV, lights, doors, heating, etc.
- play games, e.g. the card game *patience (solitaire)*

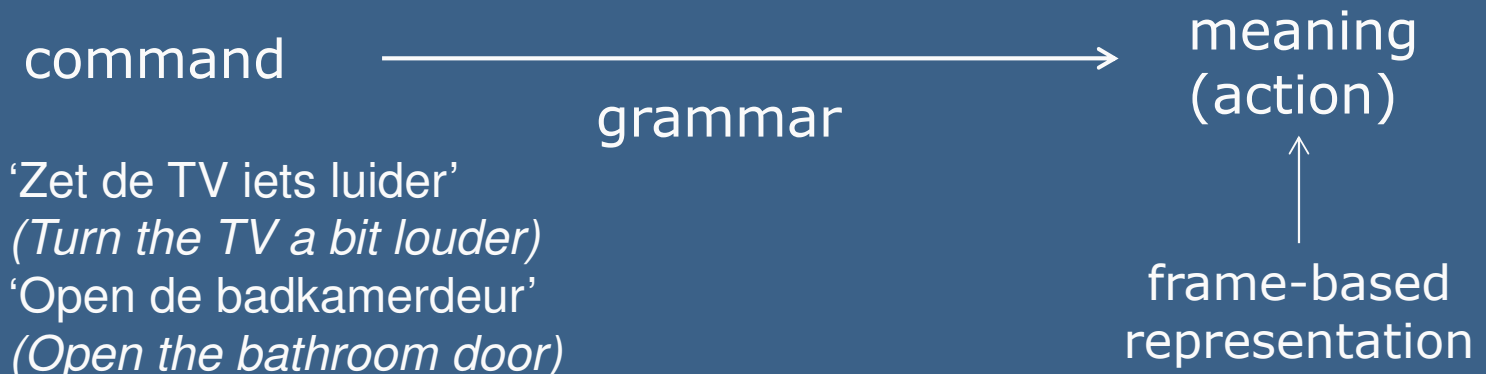


ALADIN - grammar induction

Grammar: maps the structure of the sentence to its meaning – in this case: an action

ALADIN - grammar induction

Grammar: maps the structure of the sentence to its meaning – in this case: an action



'Zet de TV iets luider'
(Turn the TV a bit louder)
'Open de badkamerdeur'
(Open the bathroom door)

ALADIN - grammar induction

Grammar: maps the structure of the sentence to its meaning – in this case: an action

'Zet de TV iets luider'
(Turn the TV a bit louder)
'Open de badkamerdeur'
(Open the bathroom door)

'Leg de harten drie op de klaveren vier'
(Put the three of hearts on the four of clubs)
'Leg de harten vier in kolom drie'
(Put the four of hearts in column three)

Patience data collection

- 8 participants, playing patience using voice commands
- commands are executed by the experimenter
- half of the participants: Wizard-of-Oz
- each participant 2 x 30 mins (with at least 3 weeks in between)

Participants:

- 4 men, 4 women
- 4 higher educated, 4 lower educated
- ages between 23 and 73

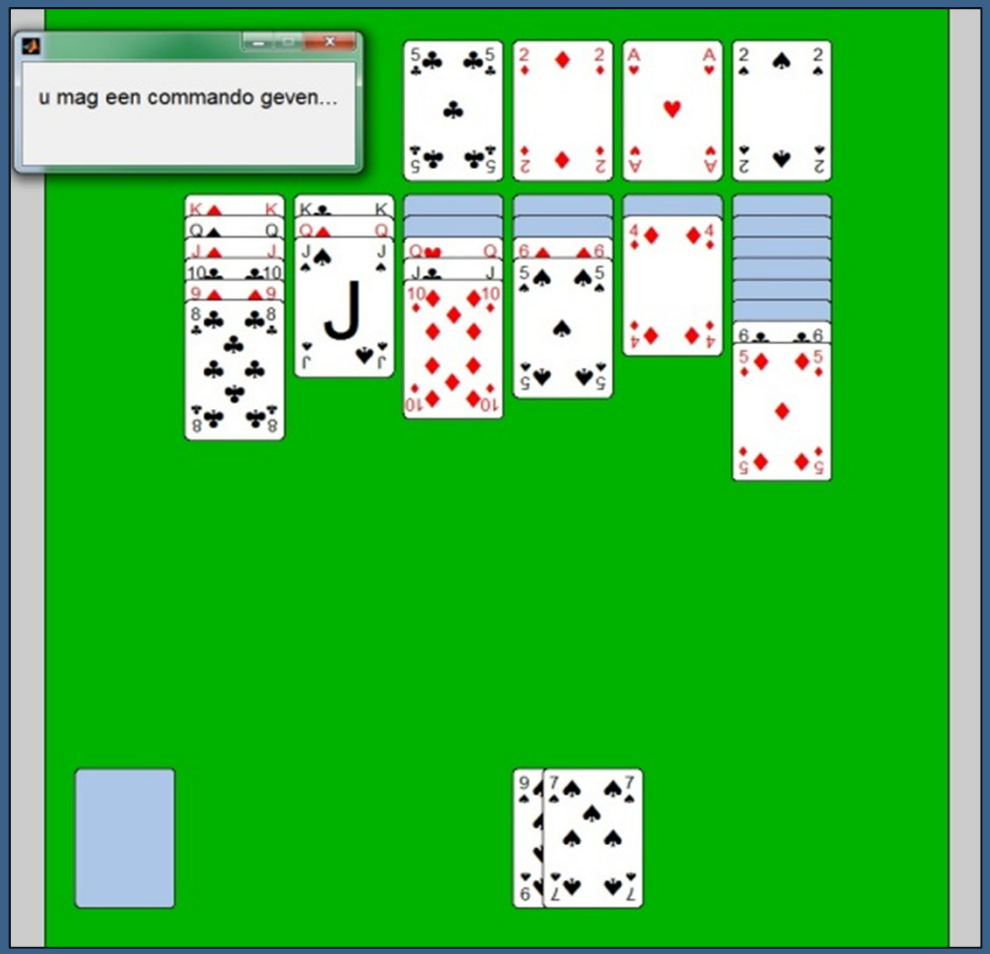
Patience data collection

Resulting files per move:

- Recorded: commands (audio)
- Generated by Patience program (Matlab):
  - Frame description of action
  - Complete workspace (state of the game)

The commands are orthographically transcribed and annotated.

Patience data: example



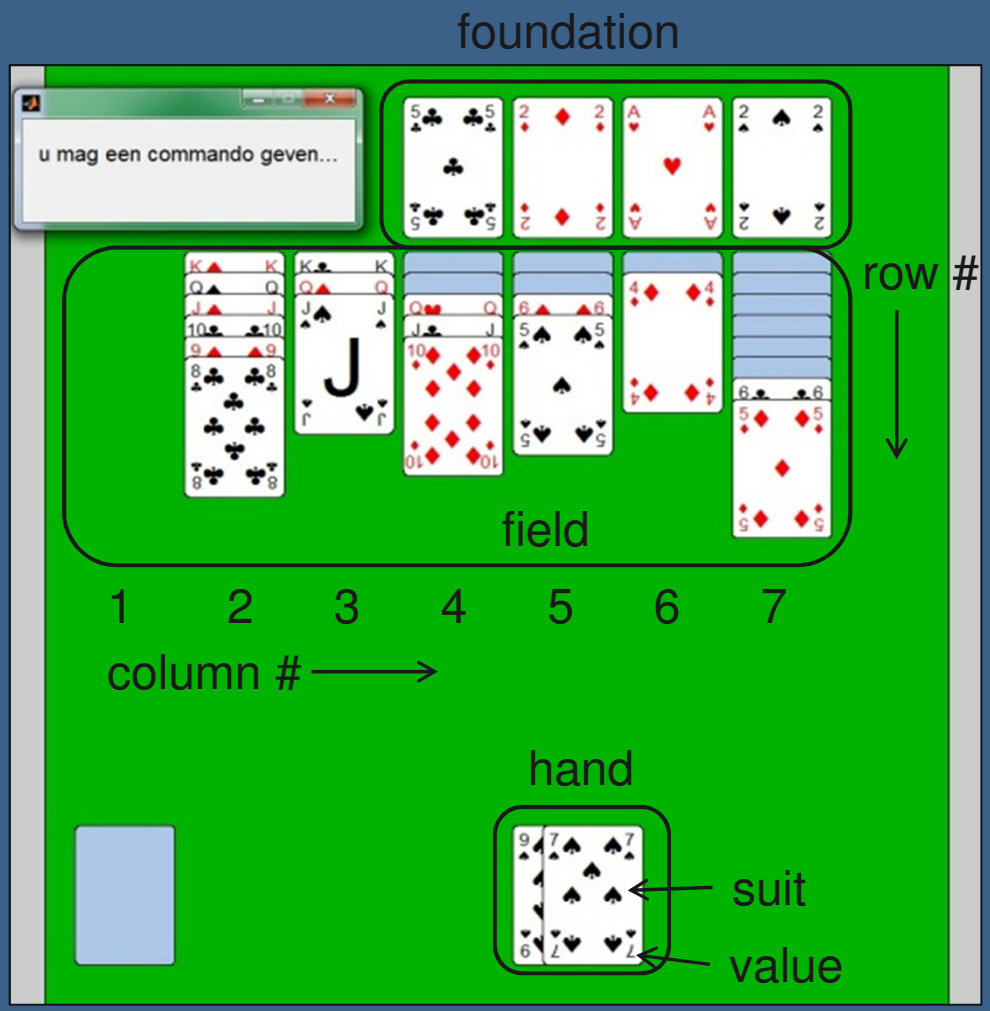
⟵ situation on screen

command:

‘de ruiten vier op de schoppen vijf’

(the four of diamonds on the five of spades)

Patience data: example



← situation on screen

command:

'de ruiten vier op de schoppen vijf'

(the four of diamonds on the five of spades)

Generated by the program: frame description of action

Action frame: 'MoveCard'

Frameslot	Value
From_Suit From_Value From_Foundation From_FieldCol From_FieldRowFaceUp From_FieldRowAll From_Hand	Diamonds 4 - 6 1 2 -
To_Suit To_Value To_Foundation To_FoundationEmpty To_FieldCol To_FieldColEmpty To_FieldRowFaceUp To_FieldRowAll	Spades 5 - - 5 - 2 4
HorizontalMovement_Distance HorizontalMovement_Direction	1 Left

'de ruiten vier op de schoppen vijf'

Grammar Induction

Sequence tagging / chunking approach:

Identify words, or chunks of words, referring to specific frame slots

de	*ruiten*	*vier*	*op*	*de*	*schoppen*	*vijf*
O	FromSuit	FromValue	O	O	ToSuit	ToValue

de	*schoppen*	*koning*	*op*	*de*	*lege*	*plaats*
O	FromSuit	FromValue	O	O	ToColEmpty	ToColEmpty

de	*harten*	*drie*	*afleggen*	*bovenaan*
O	FromSuit	FromValue	ToFoundation	ToFoundation

Grammar Induction

Initially: experiments with text data (transcriptions)

Supervised grammar induction (tagging):
→ training data: commands annotated with "frame slot tags"

de	*ruiten*	*vier*	*op*	*de*	*schoppen*	*vijf*
O	FromSuit	FromValue	O	O	ToSuit	ToValue

→ task: assign "frame slot tags" to commands in test set

Later: *Unsupervised* grammar induction

Supervised grammar induction

Preliminary experiments

- Data:
  - participants 1–4
  - around 270 utterances per participant
    except participant #2: 171 utterances
- Division train/test (per participant):
  - training data: first *n* utterances
  - test data: the rest
- Tool: MBT (Memory-Based Tagger Generator and Tagger[1])

[1] http://ilk.uvt.nl/mbt/

Supervised grammar induction

Average results participants 1-4:

Train size (#utterances)	Tag Accuracy(%) Known words	Unknown words
25	93.9	53.7
50	96.1	53.4
100	97.7	29.3

Future work

- Supervised experiments:
  - learning curves
  - add 2nd layer of tags referring to frame slot values, experiment with shared learning of values
- Unsupervised experiments
  - initially with text as input
  - later with hypotheses from word finding module as input

Questions?