

# DISCO: Development and Integration of Speech technology in Courseware for language learning

Catia Cucchiarini, Joost van Doremalen and Helmer Strik

Department of Language and Speech, Radboud University Nijmegen, The Netherlands

c.cucchiarini@let.ru.nl, j.vandoremalen@let.ru.nl, h.strik@let.ru.nl

## 1. Abstract

- Increased mobility
- Emphasis on plurilingualism → increasing demand for language training

Speaking proficiency: important for all language learners.  
Optimal training: 1 learner - 1 teacher, esp. for feedback on oral proficiency.  
But: time-consuming and costly.

Computer Assisted Language Learning (CALL) systems with Automatic Speech Recognition (ASR) offer new perspectives for language tutoring.

We investigate whether ASR can be used to improve speaking proficiency in the DISCO project.

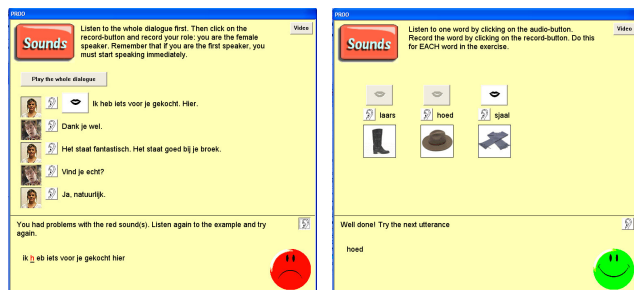
## 2. ASR for pronunciation training

### ASR for pronunciation training: Dutch-CAPT

- Dutch-CAPT: Computer Assisted Pronunciation Training for learning Dutch pronunciation (see figures below)
- Design: based on a thorough study of existing CALL systems and analyses of 3 databases of 116 learners with different mother tongues.
  - Focus: 11 problematic Dutch phonemes
  - Results on pedagogical effectiveness: language learners who used this system only four times for about 30 to 60 minutes improved more than a control group that did not use the system

Results indicate that ASR can be used to improve pronunciation of problematic sounds.

Mean scoring accuracy of error detection was 86% (SD 4%)



### Reason to extend Dutch-CAPT

Speaking a second language requires more than pronunciation alone.

## 3. Other aspects of oral proficiency

### Morphology

- Problems with morphology are persistent in L2 learning.
- Phonetic-phonological properties play a prominent role in this learning process.
- Difficulties in learning Dutch verbal morphology are related to perception and production of L2 phonemes such as schwa and /t/.

### Perception

Distinguishing between:

1) /mak/, 2) /makt/ 3) /makə(n)/  
Necessary for understanding the agreement paradigm

a) /makt/ and b) /maktə/  
Necessary for understanding the tense system

### Production

Trying to pronounce /lopt/

Learners may produce:

1) /lopət/ 2) /lopətə/ 3) /loptə/

### Syntax

In syntax problems have been observed with word order, finite verb position, and pronominal subject omission

Example: de jongen loopt naar huis  
the boy walks home

- \* loop(t) naar huis (typically Moroccan)
- \* walks home
- \* naar huis lopen (typically Turkish)
- \* home walk
- \* loopt naar huis de jongen
- \* walks home the boy

### How to improve morphology and syntax in oral proficiency

- Knowledge of grammar is not always conducive to correct oral production.
- L2 learners need to practice speaking and receive feedback online.
- Need for more comprehensive training programs for speaking proficiency.

### → The DISCO project

### Aim

To develop and test a prototype of an ASR-based CALL application for training oral proficiency for Dutch as a second language, which provides intelligent feedback on various aspects of speaking, such as pronunciation, morphology and syntax.

### Main Challenge in DISCO:

To develop exercises that are suitable for training pronunciation, morphology and syntax and that elicit spoken responses which can be assessed automatically.

### Procedure

This will be achieved by using:

- Dialogues and scenarios illustrating so-called “crucial practice situations” which correspond to realistic situations in which learners might find themselves in Dutch society and in which they have to interact with other citizens.
- Students can play a dialogue and can answer questions based on simple prompts (i.e. vocabulary to be used):
  - they formulate sentences themselves
  - speaker output is constrained
- For detecting morphological and syntactic errors, response expansion software can be used, which takes appropriate responses as input and expands them to form lists of correct and incorrect responses.
- Two phases can be distinguished:
  - speech recognition: sequence of words (+ errors)
  - Segmental analysis: sequence of segments (+ errors)

## 4. Activities so far

- Analysis of morphological and syntactic errors based on recently collected speech material of Dutch L2 learners
- Design of system and choice of contexts from existing course material
- Development of exercises for practicing morphology and syntax
- Testing speech recognizer on recently collected speech material of Dutch L2 learners

### Acknowledgement

The DISCO project is carried out within the STEVIN programme which is funded by the Dutch and Flemish Governments (<http://taaluniversum.org/taal/technologie/stevin/>).