# Individualized feedback for lexical stress errors

## Towards a CAPT system for French learners of German

---

Anjana Sofia Vakil

M.Sc. Thesis proposal

Language Science and Technology

*Prepared under the supervision of*

Prof. Dr. Bernd Möbius

Dr. Jürgen Trouvain

UNIVERSITÄT DES SAARLANDES

Saarland University

Department of Computational Linguistics & Phonetics

10 Never, 2014

# Contents


**1 Introduction** 1

1.1 Context: The IFCASL project . . . 1
1.2 Objectives . . . 1
1.3 Thesis overview . . . 2

**2 Background and related work** 3

2.1 Computer-Assisted Pronunciation Tutoring . . . 3
2.1.1 Pronunciation in foreign language education . . . 3
2.1.2 Computer-based and intelligent tutoring systems . . . 3
2.1.3 Survey of existing CAPT systems . . . 3
2.2 Towards CAPT for French learners of German . . . 3
2.2.1 Phonetic and phonological comparison . . . 3
2.2.2 Targeting lexical stress errors . . . 3
2.3 Summary . . . 5

**3 System overview** 6

3.1 Goal and architecture . . . 6
3.2 Tools and technologies . . . 6
3.3 User interface . . . 6

**4 Diagnosis of lexical stress errors** 7

4.1 Related work . . . 7
4.2 Automatic segmentation of nonnative speech . . . 7
4.2.1 Segmentation via forced alignment . . . 7
4.2.2 Evaluation of system accuracy . . . 7
4.2.3 Coping with segmentation errors . . . 7
4.3 Prosodic analysis . . . 7
4.3.1 Pitch . . . 7
4.3.2 Duration . . . 8
4.3.3 Intensity . . . 8
4.4 Comparison of native and nonnative speech . . . 8
4.4.1 Using a single reference speaker . . . 8
4.4.2 Using multiple reference speakers . . . 8
4.4.3 Using no reference speaker? . . . 8

4.5 Summary . . . . . 8

**5 Feedback on lexical stress errors** **9**
5.1 Related work . . . . . 9
5.2 Visual feedback . . . . . 10
5.2.1 Stylized text . . . . . 10
5.2.2 Graphical representations of prosody . . . . . 10
5.2.3 Visualizations of the speech signal . . . . . 10
5.3 Auditory feedback . . . . . 10
5.3.1 Enhanced reference utterance . . . . . 10
5.3.2 Resynthesized learner speech . . . . . 10
5.4 Alternative feedback types . . . . . 10
5.4.1 Metalinguistic feedback . . . . . 10
5.4.2 Interactive feedback . . . . . 10
5.4.3 Implicit feedback . . . . . 10
5.5 Summary . . . . . 10

**6 Conclusion and outlook** **11**
6.1 Thesis summary . . . . . 11
6.2 Future work . . . . . 11

**Bibliography** **12**

# 1 Introduction

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## 1.1 Context: The IFCASL project

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## 1.2 Objectives

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## 1.3 Thesis overview

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

# Background and related work 2

## 2.1 Computer-Assisted Pronunciation Tutoring

In the field of second-language education, pronunciation has traditionally been given less attention than other areas such as grammar or vocabulary (Derwing and Munro, 2005). One reason for this may be that pronunciation is best taught through one-on-one instruction, which is not often possible in the traditional classroom setting. Hence the attraction of Computer-Assisted Pronunciation Training (CAPT) systems, which have the potential to automatically provide highly individualized analysis of learner errors, and feedback on how to correct them and achieve more intelligible and native-like pronunciation in the target language (Witt, 2012).

### 2.1.1 Pronunciation in foreign language education

### 2.1.2 Computer-based and intelligent tutoring systems

### 2.1.3 Survey of existing CAPT systems

## 2.2 Towards CAPT for French learners of German

### 2.2.1 Phonetic and phonological comparison

**Segments**

**Prosody**

**Other factors**

### 2.2.2 Targeting lexical stress errors

Learners of a foreign language typically make a wide variety of pronunciation errors, at both the segmental level (e.g. errors in producing certain individual phones of the target language) and the prosodic level (e.g. errors in the speaker's intonation

contour or the duration of certain syllables or words). As it is not possible to address all of these in an automated system, one of the first aims of this work is to identify a single type of error which is well suited to being addressed via a CAPT system targeting French L1 learners of German as the L2.

To guide this selection, we propose a set of three criteria that such an error must meet. First, the error must be produced with a some degree of frequency by French L1 speakers in their production of L2 German, as it would be a misuse of resources to design a system which addresses an error that is seldom made by learners. Secondly, the given error must have a significant impact on the perceived intelligibility of the learner's speech; as the ultimate goal of the system is to help learners communicate more effectively in the L2, an error which is commonly made but nevertheless does not impede understanding of the learner's L2 speech, and thus does not hinder communication in the L2, is not an ideal target. Finally, in order for the CAPT system to provide any meaningful diagnosis of and feedback on the error, it must lend itself to reasonably accurate and reliable detection through automatic processing.

**Figure 2.1:** Criteria for selecting errors to target in a CAPT system.

As illustrated in 2.1, the best error to target with the CAPT system will fulfill all of these criteria, rather than only one or two of the three. For example, vowel quality errors (e.g. an L1 French speaker producing a German /ə/ as [œ]) may occur frequently in the L2 speech and may be relatively easy to detect automatically, but may not have a great impact on the intelligibility of the L2 German speech. On the other hand, equally frequent vowel quantity errors (e.g. the L1 French speaker producing a German long /eː/ as [e]) may have a greater impact on intelligibility in some cases, but may be more difficult to reliably identify automatically.

Analysis of the typical and expected errors described in in terms of these criteria reveals that lexical stress errors are a strong candidate for treatment via CAPT, and

will therefore be the focus of the prototype CAPT system described in this thesis. The remainder of this section justifies the selection of this type of error by describing how it fulfills the aforementioned criteria as well or better than any other error type.

#### Frequency of production

#### Impact on intelligibility

Warren et al., 2009, Magen, 1998

#### Feasibility of automatic detection

Engwall, 2012; Delmonte, 2011

## 2.3 Summary

# 3 System overview

3.1 Goal and architecture

3.2 Tools and technologies

3.3 User interface

# 4 Diagnosis of lexical stress errors

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## 4.1 Related work

Lots of Nancy references

Duong et al., 2011

Probst et al., 2002

## 4.2 Automatic segmentation of nonnative speech

### 4.2.1 Segmentation via forced alignment

### 4.2.2 Evaluation of system accuracy

### 4.2.3 Coping with segmentation errors

Mesbahi et al., 2011, Orosanu et al., 2012

## 4.3 Prosodic analysis

### 4.3.1 Pitch

### 4.3.2 Duration

### 4.3.3 Intensity

## 4.4 Comparison of native and nonnative speech

Probst et al., 2002

### 4.4.1 Using a single reference speaker

**Manually selecting a reference**

**Automatically selecting a reference**

### 4.4.2 Using multiple reference speakers

### 4.4.3 Using no reference speaker?

Duong et al., 2011

## 4.5 Summary

# Feedback on lexical stress errors

5

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

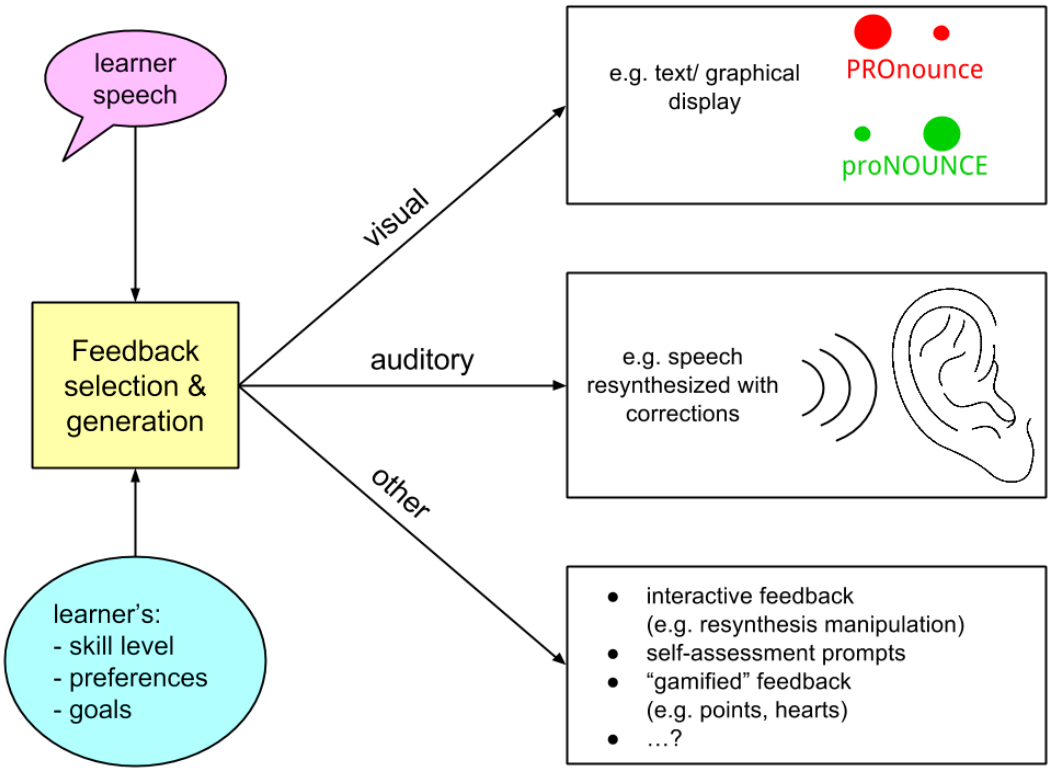


**Figure 5.1:** Delivery of prosody feedback in different modalities.

## 5.1 Related work

Sitaram et al., 2011 Bonneau and Colotte, 2011

## 5.2 Visual feedback

### 5.2.1 Stylized text

### 5.2.2 Graphical representations of prosody

### 5.2.3 Visualizations of the speech signal

## 5.3 Auditory feedback

### 5.3.1 Enhanced reference utterance

### 5.3.2 Resynthesized learner speech

## 5.4 Alternative feedback types

### 5.4.1 Metalinguistic feedback

### 5.4.2 Interactive feedback

### 5.4.3 Implicit feedback

## 5.5 Summary

# 6 Conclusion and outlook

## 6.1 Thesis summary

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## 6.2 Future work

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

# Bibliography


Bonneau, Anne and Vincent Colotte (2011). “Automatic Feedback for L2 Prosody Learning”. In: *Speech and Language Technologies*. Ed. by Ivo Ipsic. 1977. InTech.

Delmonte, Rodolfo (2011). “Exploring Speech Technologies for Language Learning”. In: *Speech and Language Technologies*. Ed. by Ivo Ipsic. InTech.

Derwing, Tracey M and Murray J Munro (2005). “Second Language Accent and Pronunciation Teaching: A Research-Based Approach”. In: *TESOL Quarterly* 39.3, pp. 379–397.

Duong, Minh, Jack Mostow, and Sunayana Sitaram (2011). “Two methods for assessing oral reading prosody”. In: *ACM Transactions on Speech and Language Processing* 7.212, pp. 1–22.

“Proceedings of the International Symposium on Automatic Detection of Errors in Pronunciation Training (IS ADEPT)” (2012). In: ed. by Olov Engwall. Stockholm, Sweden: KTH, Computer Science and Communication.

Magen, Harriet S (1998). “The perception of foreign-accented speech”. In: *Journal of Phonetics* 26.4, pp. 381–400.

Mesbahi, L, D Jouvet, A Bonneau, and D Fohr (2011). “Reliability of non-native speech automatic segmentation for prosodic feedback.” In: *SLaTE*.

Orosanu, Luiza, Denis Jouvet, Dominique Fohr, Irina Illina, and Anne Bonneau (2012). “Combining criteria for the detection of incorrect entries of non-native speech in the context of foreign language learning”. In: *SLT 2012 - 4th IEEE Workshop on Spoken Language Technology*.

Probst, Katharina, Yan Ke, and Maxine Eskenazi (2002). “Enhancing foreign language tutors – In search of the golden speaker”. In: *Speech Communication* 37.3-4, pp. 161–173.

Sitaram, S, J Mostow, Y Li, A Weinstein, D Yen, and J Valeri (2011). “What visual feedback should a reading tutor give children on their oral reading prosody?” In: *SLaTE*.

Warren, Paul, Irina Elgort, and David Crabbe (2009). “Comprehensibility and prosody ratings for pronunciation software development”. In: *Language Learning & Technology* 13.3, pp. 87–102.

Witt, Silke M (2012). “Automatic error detection in pronunciation training: Where we are and where we need to go”. In: *Proceedings of the International Symposium on Automatic Detection of Errors in Pronunciation Training (IS ADEPT)*, pp. 1–8.