

# Universal contrastive analysis as a learning principle in CAPT

Jacques Koreman, Preben Wik, Olaf Husby, Egil Albertsen

Department of Language and Communication Studies, NTNU, Trondheim, Norway

jacques.koreman@ntnu.no, prebenwik@gmail.com, olaf.husby@ntnu.no, egil.albertsen@ntnu.no

## Abstract

Cross-linguistic comparison is a good starting point for computer-assisted pronunciation training (CAPT). A comparison between the segment inventories of a learner's mother tongue (L1) and the target language (L2) can be made on the basis of the IPA categories. Since these are claimed to reflect universal phonemic distinctions, mastering the contrasts in the target language in perception and production ensures communicative effectiveness for learners at the segmental level. The Computer-Assisted Listening and Speaking Tutor (CALST) implements contrastive analysis in two types of exercises for phonetic and abstract listening. In these exercises users can practice with word pairs/sets demonstrating unfamiliar sound contrasts of the target language to improve their perceptual discrimination. Since substitutions for unfamiliar sounds depend on L1, the selection of sound contrasts which are trained in the exercises should also depend on L1. We shall argue for a pragmatic approach to the selection of exercises.

**Index Terms:** computer-assisted pronunciation training, CAPT, CALST, sound contrasts, differential substitution

## 1. Introduction

Pronunciation training is often given limited attention in foreign language classes. One of the reasons is that users with different native languages often have very different challenges when it comes to acquiring a correct, or at least communicatively effective, pronunciation. These varied challenges are difficult to address in class, also because language teachers can at best only have in-depth knowledge of a limited number of foreign languages. We have therefore developed a CAPT system for Norwegian which adapts the learning trajectory to the user's native language (L1) on the basis of a contrastive analysis. The system is called the Computer-Assisted Listening and Speaking Tutor (CALST). It is used in Norwegian classes for foreign students and staff at NTNU and can be downloaded for free [1].

The system uses cross-linguistic comparison as a basis for pronunciation training. Although the Contrastive Analysis Hypothesis, especially its strong (predictive) version, has been criticized [2], it is nevertheless a good starting point in language teaching. Or as Ellis expressed it: "No theory of L2 acquisition is complete without an account of L1 transfer" [3, p. 341]. In classes where learners with the same language background learn a second language (L2), e.g. Norwegian students studying English in Norway, this principle can be implemented successfully. But in classes where the students have mixed native language (L1) backgrounds this is not possible, and L2 teachers often adopt a more constrained implementation of contrastive analysis, focusing on typical pronunciation problems which most foreign learners have when they acquire the L2 – often in

combination with instruction to correct individual pronunciation errors in class.

With the increasing variation in language backgrounds of students in the Norwegian courses at NTNU (and presumably also in language courses in many other countries, given the increasing mobility across countries), variation in the students' L1 backgrounds becomes more and more of a challenge to standard classroom teaching. Computer-assisted pronunciation training (CAPT) systems can use a contrastive approach to guide language learners with varying L1's through relevant instruction and individual exercises, and may thus help to meet this challenge. This article describes a segmental CAPT implementation of the principle of contrastive analysis on a "universal" basis in Section 2.

At present, CALST users work through *all* the exercises for *unfamiliar* L2 sounds. Ideally, a narrower selection of exercises should be made: It is well known that differential substitutions occur across L1's, i.e. an L2 speech sound may be substituted with different sounds depending on the learner's L1. Differential substitutions occur even if the L1's have the same relevant (phonetically close) set of phonemes in their inventories: German and Dutch both have /t/ and /s/, but German speakers often substitute /s/, while Dutch speakers often substitute /t/ for English /θ,ð/ [4]. Since we aim to offer CAPT users an efficient learning trajectory, only exercises for substitutions which actually occur should be offered. The link between contrastive analysis and the selection of appropriate exercises (which would be different for German and Dutch learners of English because of the different substitutions) is discussed in Section 3. We describe possible linguistic approaches to this challenge and their inherent problems, and explain why and how we have decided to implement a pragmatic solution instead.

## 2. Learning sound contrasts

The main stress in the CALST system is on listening skills, since it is generally accepted that this is an important prerequisite for correct pronunciation [5], but also because pronunciation skills require automatic error detection, which at present still has clear limitations [6]. CALST has two special features [7]. The first is that for each user, the exercises which the learner is guided through are based on a comparison of the Norwegian segment inventory with the segment inventory of the user's native language. At present, CALST uses a database which contains over 500 languages. Although CALST has Norwegian as the target language, the contrastive analysis tool can be used as a basis for CAPT systems for any language. The contrastive analysis is further described in Section 2.1.

The second special feature of CALST, described in Section 2.2, is determined by a peculiarity of Norwegian: It has no accepted pronunciation standard. For this reason four main dialect variants, each represented by one male and one female speaker (with sub-dialectal variation), are available in CALST.

Instead of forcing the user to exclusively commit to one single dialect as the target dialect, the system offers the possibility of acquiring production skills in one (target) dialect, while the user can develop perception skills in all four dialects. The latter is necessary because learners will hear all dialects in their everyday interactions and pronunciation can vary strongly across dialects, with the sound inventories varying in size from 42 to 53 phonemes [8].

After defining the source and target language/dialect, the user is guided through exercises for sound segments that are likely to present a challenge to learning Norwegian pronunciation. The exercises use minimal pairs which demonstrate sound contrasts. Each exercise starts with a short articulatory explanation of the similarities and differences between the contrasted speech sounds (Section 2.3).

The user can then choose between two exercise types. The first exercise type enables the user to listen phonetically to the unfamiliar sound in contrast with another sound in a minimal pair. After listening to the word pair the user hears a new realization of one of the two words again, and has to decide which of the first two words it resembles. A more detailed description is given in Section 2.4.

The second exercise type, described in Section 2.5, requires more abstract listening skills. In these exercises, the listener only hears one word, and has to select a button on the screen which is labeled with that word.

To facilitate understanding, it is recommended that the reader first download and open CALST [1].

### 2.1. Multi-language contrastive analysis

The tool L1-L2map [9] was developed on the basis of the UCLA phonetic segment inventory database (UPSID, [10]) to enable contrastive analysis of any language pair. The UPSID database, which contains 451 languages, was extended with languages spoken by larger immigrant groups in Norway which are not available in the UPSID database, and we now have access to the segment inventories of over 500 languages. Since foreign speakers have problems with pronouncing known segments in unusual syllable positions [11,12], position in the syllable was also added as a descriptive feature and consonants can be marked for their appearance in syllable onset, nucleus and coda [9]. Since vowels always occur in the syllable nucleus, they are not marked for position.

The visualization of the contrastive analyses in L1-L2map is based on an easily interpretable colour coding in the consonant and vowel charts of the International Phonetic Association [13]. This makes it easy to interpret the analysis results, although typically the information in the charts is not shown to CAPT users. Normally, a CAPT system will use the results from the contrastive analysis to make a selection from all available pronunciation exercises dependent on the user's native language, without actually showing the analysis results. All that CALST requires of the user in order to perform a contrastive analysis is that (s)he specifies his/her native language and the target dialect when starting the program for the first time.

The use of IPA charts in L1-L2map makes it easy for language experts to define the segment inventory of his/her native language in L1-L2map (language expert privileges required). For each phoneme in the charts, a number of phonetic variants are shown when a cell in the charts (or phoneme symbol) is clicked, allowing the language expert to select the phonetic realization of the phoneme which best represents a

“canonical” pronunciation in the language – for consonants usually the realization in the onset of a stressed syllable in an isolated word, and for vowels their pronunciation in isolation. L1-L2map is implemented as a wiki and can be incorporated into any CAPT system as a server-client system. The biggest practical challenge for the wiki is involving language experts for the many languages in the database to contribute with positional information for the consonants or to define the segment inventories for languages which are not yet available.

### 2.2. The pronunciation of Norwegian

Many languages have a pronunciation standard, although they may vary in the rigidity with which speakers use that standard. The choice of using a dialectal versus a more standard pronunciation will often depend on the formality of the situation. In contrast, Norwegian speakers use their dialect largely independently of social status and context, so that even the Prime Minister, for example, speaks to the nation in his own dialect. As a result of migration, different dialects are often spoken within the same area. This means that foreigners need to be able to deal with a variety of dialectal pronunciations.

To select one dialect as the target variant for production, i.e. as a role model, the user specifies the target dialect when starting CALST the first time. Since any learner of Norwegian will have to become familiar with the sometimes quite different pronunciation variants of *all* the different dialects in order to become communicatively effective, perception exercises in CALST offer the possibility of switching between dialects. The dialects are implemented as different talking heads on the right-hand side of the CALST window. For legibility of the text in the figures, these are not shown in the figures in this article.

### 2.3. Articulatory explanations

As a first step in the selection of the sound contrast exercises, the CALST user chooses a phoneme which is unfamiliar from his/her L1, e.g. the voiced palatal nasal consonant /ɲ/. After that, a sound contrast is selected, e.g. /ɲ-ŋ/, i.e. the contrast between the voiced palatal nasal and its velar counterpart. The list of sound contrasts is set up on the basis of experience collected from classroom practice.

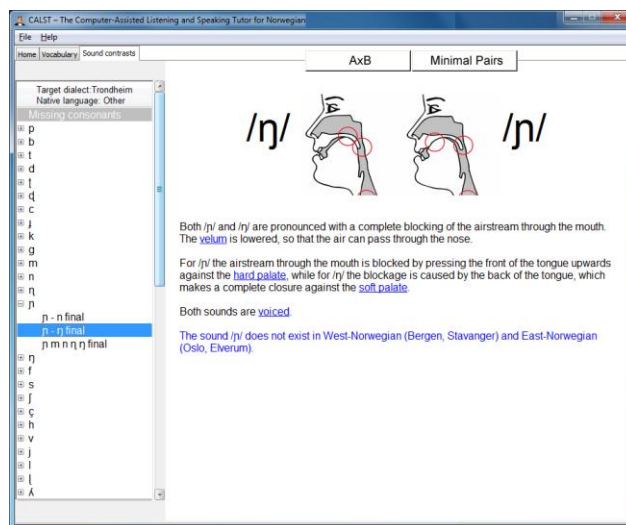


Figure 1: Explanatory text for the articulatory /ɲ-ŋ/ contrast

Drawn sagittal cross-sections for the contrasted sounds are shown at the top of the CALST window with circles to indicate place of articulation (here: palatal versus velar), manner of articulation (oral versus nasal, i.e. raised or lowered velum) and voicing (circle around the glottis), cf. Figure 1. A short articulatory description of the sounds is given for these three dimensions, always following the same pattern: first, the commonalities between the two sounds are described, and then the difference(s). If the sounds do not occur in all dialects of Norwegian, this is commented on at the end of the description. For vowels, the point of maximum constriction is indicated by a red circle.

Like L1-L2map, the articulatory descriptions follow IPA standards: For consonants, the dimensions manner and place of articulation and voicing are used, while vowels are described on the basis of degree of opening, the front-to-back dimension and lip rounding as well as length. Although Norwegian has long consonants in stressed syllables, the length is not phonemic and depends on the length of the preceding vowel for syllable-final consonants: Long vowels are followed by short consonants, and short vowels are followed by longer consonants. Long and short consonants thus occur in a complementary distribution. The length feature is not distinctive for consonants and therefore not important for communicative effectiveness.

## 2.4. Phonetic listening

CALST users will normally start with phonetic listening exercises to become familiar with a sound contrast (see Figure 2 for a single trial of a phonetic listening exercise). These exercises are implemented as ABX exercises, i.e. the user first listens to one word (e.g. [k<sup>h</sup>at:], while the button for <katt> on screen is highlighted) and then to the other word in the minimal pair ([k<sup>h</sup>at:], button for <kart> highlighted). Then a third word is spoken by the tutor (implemented as a talking face, not shown in the figure) while the middle button on the screen is highlighted, and the user is required to click on a button with the text label for the word spoken, i.e. <katt> or <kart>. Note that the exercise is somewhat misleadingly called AXB on account of the visualization on the screen, where the middle button represents the *third* word (in real AXB the second word is the test word).

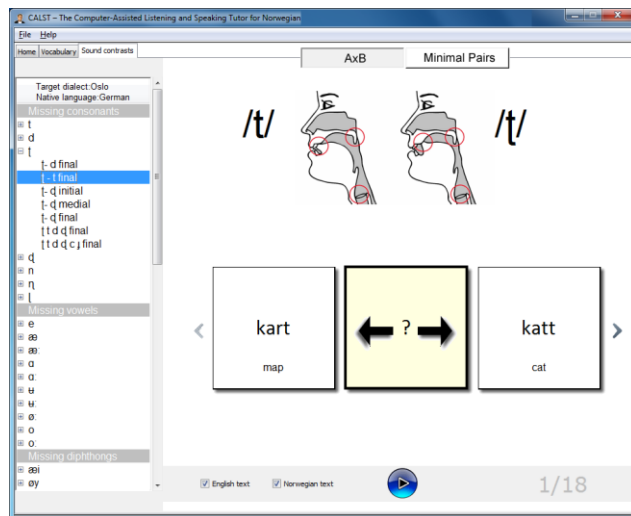


Figure 2: Single trial for an ABX exercise for the /t-t/ contrast

To prevent users from focusing on irrelevant acoustic properties to distinguish between the words, two recordings are available for each word, and the same recording is never used twice in a single trial. Irrelevant acoustic properties can be non-distinctive differences in fundamental frequency or even coincidental background noises, although great care was taken to prevent such differences in the recording of the words.

## 2.5. Abstract listening

After each ABX exercise, the learner is expected to take a minimal pair exercise. This exercise differs from the ABX exercises in that only one word is spoken by the tutor, and the user is required to click on one of two buttons on the screen, labeled with the words in the pair. Since the user only hears a single word and cannot rely on an acoustic comparison, (s)he has to use an internalized representation of the phonemes to decide which of the buttons to click. This exercise is therefore more advanced than ABX exercises.

After taking several related minimal pair exercises, for example contrasting sets of consonants which may be confused across L1's, such as the voiced and voiceless dental and retroflex plosives, the user can take *sound set* exercises. As in the minimal pair exercises, the user only hears a single word, and clicks on the corresponding button on the screen. In sound set exercises, the buttons are not labeled with words, but with sound symbols with the text "as in <WORD>" underneath (see Figure 3). The reason for labeling the buttons with sound symbols is that it is usually not possible to find minimal sets, i.e. word sets that are only distinguished by the phonemes which are the focus of the exercise. This also requires that the language learner listen to the sounds at an even higher level of abstraction, since the sounds occur in different contexts.

Segmental differences can thus be implemented in CAPT systems in a relatively straightforward manner on the basis of a "universal" contrastive analysis. The direct comparison of any L1 with the segment inventory of L2 is the basis for selecting exercises which are directly relevant for the user. It may be possible to extend this approach to prosodic-phonetic and other linguistic properties.

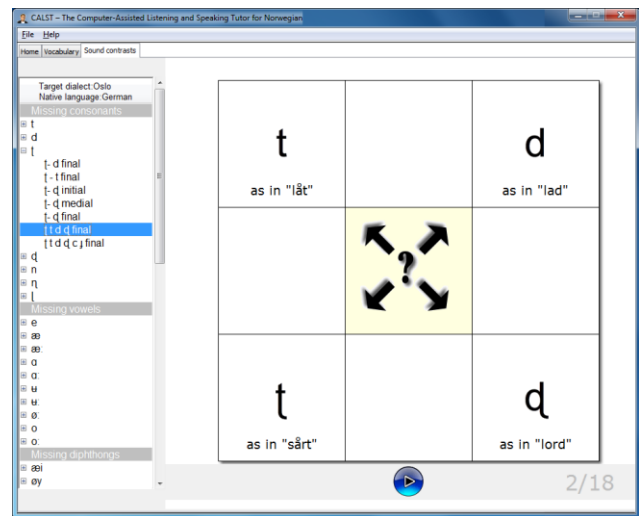


Figure 3: Single trial from a sound set exercise for /t-d-t-d/

### 3. Exercise selection

Presently, each unfamiliar phoneme in L2 gives access to a set of exercises in CALST in which that particular phoneme is contrasted with *several* phonetically similar phonemes in different syllable positions (in onset and coda, for consonants). When L2 learners learn a new sound, however, the errors they make depend on their L1, and learners with different L1's may substitute different sounds. This is called differential substitution. Since our aim is to offer an efficient learning trajectory, the coupling between the contrastive analysis and the selection of exercises should take this into consideration. Exercises for unfamiliar phonemes should only contrast these with the phoneme(s) learners are likely to use as a substitute.

Two different approaches can be used for selecting exercises on the basis of actually occurring (differential) substitutions. The first is based on a linguistic mapping of the phonemes in L2 onto the L1 system on the basis of the phonological features of the L1. In section 3.1, we shall discuss these phonological approaches, analyzing their usefulness for our purpose and presenting some partly speculative ideas which will have to be verified or falsified in future research. The second approach is based on observations of actual errors and is described in section 3.2. This latter approach is now being implemented in CALST as part of a user logging system.

#### 3.1. Linguistic selection

Differential substitution is a well-known phenomenon in L2 learning. The differential substitutions for English /θ,ð/ have been investigated particularly extensively, and will be used as a basis for discussion. It is well known that these consonants, which are unfamiliar for many foreign learners of English, may be replaced by /t,d/, /s,z/ or /f,v/ depending on the L2 learner's native language (e.g. in Russian, German and Finnish, respectively). Several explanations have been offered in the literature as to why L2 learners make different substitutions even though they have similar phoneme inventories (at least with respect to the substituted phonemes). Particularly underspecification theory [14] and optimality theory [15] offer good explanations of the phenomenon. Both however must rely on external evidence for the underspecification patterns or the constraint rankings which explain the differential substitutions. This evidence can come from phonological processes in the L1 (including the phonological adaptations of loanwords to L1 phonology) or from substitution errors in foreign language learning. The latter makes the problem cyclic, since we want to use the underspecification pattern for *predicting* phoneme substitutions in the L2. Evidence from L1 phonology may solve the problem, but [14] suggests that such evidence is often absent. This would create a learning problem, since it is not clear how a child can learn the underspecification patterns for its native language, and by implication the substitutions in an L2 cannot be predicted on the basis of the L1 underspecification patterns.

In the explanations for differential substitutions which were mentioned above, we must have access to detailed knowledge of the L1 phonology in order to predict the possible misperceptions by L2 learners and their repair strategies when dealing with an unfamiliar sound in L2. As Weinberger writes, the segmental phonologies of languages do not always "contain the requisite rules to direct us in the construction of an optimal underspecified matrix" [14]. Therefore, "...while [the theory of underspecification] is fundamentally correct insofar as it simplifies the task of the first language learner, it has the

problem of generating ambiguous matrices. That is, without native language evidence, multiple underspecified formulations are logically possible," and the ambiguous matrices can therefore not always predict which differential substitutions actually occur. With that reservation, incomplete or ambiguous underspecification matrices can still be used to predict the set of *possible* or *likely* L1-dependent substitutions. This can narrow down the possible substitutions, and thus constrains the number of exercises which an L2-learner with a given L1 needs to take. Clearly, generating underspecification matrices or OT constraint rankings for many languages is a laborious task and exceeds the scope of our present project.

Another method for the selection of sound contrast exercises may be based on an approach to generating underspecification matrices which makes claims to universality. In similarity with other underspecification- and OT-based explanations, the fully underspecified lexicon (FUL) approach to speech perception assumes different feature specifications for different languages [16]. However, beyond a basic set of contrasts (features) which is present in all languages, FUL claims that "[a]ll other features depend on the phonological systems of individual languages. The assignment does not depend on whether any feature is active in a phonological rule, but only if it is necessary to establish a phonemic contrast." This means that the set of underspecified features for any language can be defined solely on the basis of the sound contrasts occurring in the language. Since this would enable the generation of an underspecification matrix for each language solely on the basis of the phoneme inventories which CALST already uses (see Section 2.1), this makes the FUL model very compatible with the multi-language approach of CALST. This approach will be investigated further in the future.

Phonological explanations on the basis of the phoneme inventories of L2 and L1 are faced with yet another problem: In some languages the substitutions are known to be position-dependent, as are the substitutions for English /θ,ð/ by Dutch learners [17,3]. All substituted phonemes /t,d,s,z,f/ (and possibly /v/) can occur in all positions in Dutch, with the exception that only voiceless phonemes occur in final position due to final devoicing in Dutch. One must therefore either assume that there is a position-dependent phonological specification or assume a phonetic explanation.

A phonetic explanation of different substitutions could possibly be that some speakers of Dutch are so familiar with English that they are aware of phonetic variation of the phonemes in English. English dental fricatives can start with an occlusion in word-initial position, whereas they are generally continuant in other positions. This may affect the substitutions applied by L2 learners, although we are aware that this statement is very speculative. It is not unreasonable to assume familiarity of Dutch learners with the acoustic quality of English, because for example television series in English are not usually dubbed, but subtitled. This is not the case for instance in German, where we should therefore expect less variation in the observed substitutions.

To summarize, we point out that the problems intrinsic to ambiguous underspecification matrices apply equally to the incompleteness of an OT description of languages with respect to a ranking of all operating constraints. Although a linguistic approach to predicting differential substitutions is to be preferred because it allows us to select contrastive exercises solely on the basis of phoneme inventories, this is presumably not feasible within our CAPT approach, since it requires a complete

underspecified feature matrix or a complete OT constraint set for each L1 in the L1-L2map database which forms the basis for the CALST system. To be used for exercise selection within the multi-lingual approach of CALST, a linguistic solution would have to be universal. Since no such solution as yet exists, we opt for a more pragmatic approach.

### 3.2. Pragmatic selection

Projects like the Speech Accent Archive [17], in which foreign speakers are recorded and their substitutions are transcribed and categorized, represent a descriptive basis for differential substitutions in a given L2. Such databases also give interesting insights into the completeness of the substitutions. Hanulíková and Weber point out, for example, that the segments substituted for English /θ/ by Dutch and German speakers differ acoustically from the perceptually closest phoneme, although they are target-like and accepted as good exemplars of the specific categories of the substituted phoneme [4]. For the purpose of CALST, where we aim for (at least) communicatively effective pronunciation and perception, a phonemic representation of the substitutions suffices. In a small pilot project, we have therefore started collecting and transcribing foreigners' pronunciation of a short Norwegian text in which all Norwegian phonemes are represented in several contexts.

On the basis of observational data on actually occurring substitutions in L1-L2 pairs, it will be possible to select exercises which are useful for learners with a specific L1 to train unfamiliar sounds from an L2, Norwegian in our case. Such a pragmatic approach to the selection of sound contrast exercises for L2 depending on L1 can make use of the CALST system itself. At the moment, CALST is being converted from a downloadable program into a web-based system where users' progress will be logged in a database. This is useful for the learner, since the system keeps track of the satisfactorily completed exercises and those (s)he still needs to take, and also (in the case where the system is used in combination with classroom teaching) for the teacher, who can monitor students' progress and errors, and use that as a basis for pronunciation teaching in class.

Information which is stored in the database will also reflect the differential substitutions which actually occur. With sufficient data for a given L1, only those exercises can be selected which train the user to hear distinctions which are difficult for an L2-learner to perceive. ABX or minimal pair exercises with no or few mistakes for a given L1 can be discarded. This will help to make the system more efficient for future L2-learners.

In CAPT systems for other languages than Norwegian the same strategy can be used, as long as the system has a similar structure to CALST.

## 4. Discussion

One can ask oneself whether it is pedagogically optimal to only offer training for challenges which foreign language learners are confronted with, since including exercises which are easy (e.g. sound contrasts which occur in both L2 and L1) may help to motivate the learner. Nevertheless, we have focused on a maximally efficient learning trajectory with sound contrast exercises for unfamiliar sounds from L2.

CALST users are directed to ABX (phonetic listening) and minimal pair/sound set exercises (abstract listening) for those

speech sounds that are predicted to be problematical on the basis of a contrastive analysis (although each user does have access to the complete exercise set for all phonemes in the target language). These problematical speech sounds are practiced in exercises for all contrasts that can be relevant for learners with any L1 background. In order to limit the number of exercises a learner takes, a pragmatic solution will be adopted to the selection of exercises which only considers actually occurring differential substitutions. These are obtained from logged exercise results for learners with the same L1. Exercises where learners with the same L1 made no or few mistakes can be taken off the list of exercises for all learners with that L1.

Over time, the collection of data for each L1 will also allow us to set up an underspecification matrix for that specific L1, at least in as far as it is relevant for substitutions in Norwegian. We hope such matrices will also be relevant for the selection of exercises in other languages, where they can be used for a linguistic prediction of (some) substitutions. In this way, the approach may also contribute to a "universal", or at least multi-lingual, approach to exercise selection.

## 5. Acknowledgements

This work was financed in part by Norgesuniversitetet, project number P54/2009 and in part by VOX project 2010/59.

## 6. References

- [1] The Computer-Assisted Listening and Speaking Tutor (CALST), <http://calst.hf.ntnu.no>, 2012.
- [2] Odlin, T., *Language Transfer*. Cambridge University Press, 1989.
- [3] Ellis, R., *The Study of Second Language Acquisition*. Oxford University Press, 1994.
- [4] Hanulíková, A. and Weber, A., "Production of English interdental fricatives by Dutch, German, and English speakers", in K. Dziubalska-Kolaczyk, M. Wrembel, and M. Kul (eds.), *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010, Poznań (Poland): 173-178*, 2010.
- [5] McAllister, R., "Perceptual foreign accent: L2 users' comprehension ability", in A. James and J. Leather (eds.), *Second-language speech: Structure and process*, 119-132, 1997.
- [6] Wik, P., *The Virtual Language Teacher*. Ph.D. thesis, KTH School of Computer Science and Communication, 2011.
- [7] Husby, O., & Øvregaard, Å., Wik, P., Bech, Ø., Albertsen, E., Nefzaoui, S., Skarpnes, E. & Koreman, J., "Dealing with L1 background and L2 dialects in Norwegian CAPT", *Proc. of the workshop on Speech and Language Technology in Education (SLaTE2011)*, Venice (Italy), 2011.
- [8] Husby, O., Høyte, T. Nefzaoui, S., Nordli, I., Robbins, S. and Øvregaard, Å., *An introduction to Norwegian dialects*. Tapir Akademisk Forlag, 2008.
- [9] Koreman, J., Bech, Ø., Husby, O. & Wik, P., "L1-L2map: a tool for multi-lingual contrastive analysis", *Proc. 17th Int. Congress of Phonetic Sciences (ICPhS2011)*, Hong Kong, 2011.
- [10] Maddieson, I., UPSID: UCLA phonological segment inventory database. Phonetics Laboratory, Department of Linguistics, 1980.
- [11] J. Flege, "Second Language Speech Learning Theory, Findings, and Problems", in W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, 1995.
- [12] Setter, J. and Jenkins, J., "State-of-the-Art Review Article", *Language Teaching*, 38(1): 1-17, 2005
- [13] *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press, 1999.

- [14] Weinberger, S., "Minimal segments in second language phonology", in J. Leather and A. James (eds.), *New Sounds 90, Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second-Language Speech*. Amsterdam: University of Amsterdam.
- [15] Lombardi, L., "Second language data and constraints on Manner: explaining substitutions for the English interdental", *Second Language Research* 19(3): 225–250, 2003.
- [16] Lahiri, A. and Reetz, H., "Distinctive Features: Phonological Underspecification in Processing", *J. of Phonetics*, 38, S. 44-59.
- [17] Weinberger, S. "Speech accent archive", <http://accent.gmu.edu/>. George Mason University, 2003.