

Error selection for ASR-based English pronunciation training in 'My Pronunciation Coach'

Catia Cucchiari¹, Henk van den Heuvel¹, Eric Sanders¹, Helmer Strik²

¹ Centre for Language and Speech Technology, Radboud University Nijmegen, The Netherlands.

² Department of Linguistics, Radboud University Nijmegen, The Netherlands.

[c.cucchiari, h.vandenheuvel, e.sanders, h.strik]@let.ru.nl

Abstract

In this paper we report on a study of pronunciation errors that was conducted within the framework of the project "My Pronunciation Coach", which is aimed at developing an ASR-based system for pronunciation training for learners of English with Dutch as their mother tongue. The aim of this study was to obtain quantitative data on the occurrence of pronunciation errors in Dutch English speech. We present the results of this study and compare them to those of previous investigations. Finally, we discuss the implications of these results for the development of My Pronunciation Coach.

Index Terms: ASR-based pronunciation training, pronunciation errors, second language speech, CAPT.

1. Introduction

Acquiring a correct pronunciation in a second language (L2) is a fundamental skill for successful interaction in the L2 for all learners, irrespective of their educational level or career [1]. Although Dutch learners of English in general achieve relatively good quality levels in pronunciation [2], they appear to have problems with various aspects of the English sound system and many learners would like to sound as native as possible [3] because pronunciation is important for successful interaction and social acceptance.

There are good reasons for giving attention to pronunciation improvement in language teaching programs, but the main problem seems to be that pronunciation correction requires so much time, feedback and practice, that most of the time it is not feasible in a language classroom, in any case not with the necessary amount of individualized attention.

For this reason, interest in Computer Assisted Pronunciation Training (CAPT) applications that make use of Automatic Speech Recognition (ASR) has been growing in the last few years. CAPT can provide a private, stress-free environment in which students can access virtually unlimited input, practice at their own pace and, through the integration of ASR, receive individualized, instantaneous feedback.

An attempt to develop an ASR-based CAPT system that specifically addresses the errors in English pronunciation made by Dutch learners was undertaken at Radboud University Nijmegen by the Centre for Language and Speech Technology (CLST) and the university language centre Radboud in'to Languages through the project 'My Pronunciation Coach' (MPC). In analogy to similar projects aimed at developing ASR-based pronunciation training systems for learners of Dutch as a second language, the Dutch-CAPT project [4], DISCO [5] and Repetitor [6] a first, important stage in MPC was considered to be a thorough analysis of pronunciation errors made by L2 learners. It is this part of our work that will be the topic of the present paper.

In section 2 we give a brief introduction to the MPC project. Section 3 presents a preliminary selection of pronunciation errors made by Dutch learners of English, which was based on existing literature and teacher experience. In section 4 we present a data-driven study on English pronunciation errors made by Dutch learners. In section 4.1 we describe the aim of this study, while method and material are addressed in 4.2. In section 4.3 we present the results of this study and in section 4.4 we relate them to findings from previous studies. In section 5 we discuss the results of our study and consider its implications for the development of the MPC system.

2. The MPC project: background and objectives

Dutch students appear to have problems with different aspects of the English sound system, as has been documented in a number of text books [7], [8] and pronunciation studies [9], [3]. Some studies provide information on which aspects of pronunciation and which sounds appear to be more problematic and also indicate how a correct pronunciation could be achieved [7], [8], [3].

Other authors have addressed the question of whether native-like pronunciation should be the objective of pronunciation instruction or whether intelligibility should be considered to be sufficient [2]. Although many researchers now agree that achieving a native-like pronunciation may not be necessary for all learners, a reasonably intelligible pronunciation is considered to be an essential component of communicative competence [1], [10], [11], [12]. Furthermore, improving pronunciation and trying to sound as native as possible appears to be an important objective for language learners [3].

For these (and other) reasons, we decided to use the knowledge and expertise gained in our Dutch-CAPT and DISCO projects to develop a similar system for Dutch learners of English, 'My Pronunciation Coach' (MPC), which students can use to practice pronunciation anytime and anywhere they want. MPC is carried out in co-operation with the university language centre Radboud in'to Languages. This program is unique in that it aims at providing personalized, detailed and accurate feedback on pronunciation at the level of individual sounds. Different versions will be developed for students with different levels of proficiency and learning targets. For further details, see: [13].

3. Preliminary selection of pronunciation errors to be addressed in MPC

In developing Dutch-CAPT we formulated a number of criteria for error selection in CAPT systems [4], among which error frequency, salience and persistence. These criteria were also adopted to carry out a first selection of pronunciation errors to be addressed in MPC. This selection was based on

Radboud in'to Languages's teaching experience throughout the Netherlands and research data presented in various studies [8], [14], [15], [9]. The relevance of the selected errors is not only dependent on the effect mispronunciation can have on intelligibility, but also the possible negative attitude a Dutch English pronunciation can evoke [11], [16], [15]. In this section the errors selected on the basis of Van den Doel's error hierarchy ([15] p. 292) and Gussenhoven and Broeders's substitution tables ([14] p. 113, 171) are presented.

The Dutch English vowel pronunciation errors selected are displayed in Table 1, where column 2 shows the RP vowels selected, followed by the condition or the context in which the error occurs. Column 4 describes the Dutch vowel used to replace the original RP vowel, together with an example Dutch word. In the right-hand column examples of words with the selected RP vowels are shown. If there is no condition specified, the error can be applied to all conditions.

Table 1: Vowel Errors in Dutch English Pronunciation

	RP	Condition	Dutch	Example
1	/ɪə/	before /r/	/i/ tien	beer, idea
2	/æ/	+fortis consonant + lenis consonant	/ɛ/ pet /ɛ:/ serre	bat bad
3	/ʌ/	spelling with o spelling with u	/ɔ/ sok /ʏ/ bus /ə/ bedacht	other bus unwise
4	/u:/		/u/ soep	soup
5	/ʊ/		/u/ goed	good

For consonants, pronunciation errors are shown in Table 2, which is organized like Table 1.

Table 2: Consonant Errors in Dutch English Pronunciation

	RP	Condition	Dutch	Example
1	/b/	word-final	/p/ Rob	hub
2	/d/	word-final	/t/ bad	bad
3	/g/	word-final	/k/ lik	big
4	[p ^h] [t ^h] [k ^h]	initial voiceless plosives	/p/ pak /t/ tak /k/ kat	pack tap cap
5	/tʃ/		/ʃ/ sjaal	chips
6	/dʒ/		/tʃ/	bridge
7	/dʒ/		/ʃ/ sjaal	jam
8	/w/		/v/ wie	wine
9	/ð/		/d/ dak	the, this
10	/θ/		/s/ sap	booth
11	/θ/		/s/ sap /t/ tap	thirty, three
12	/s/		/ʃ/ sjaal	socks
13	/z/	word-final	/s/ sap	jazz

4. Pronunciation errors by Dutch learners of English: a data-driven study

4.1. Aim

Tables 1 and 2 show a preliminary selection of errors to be addressed in MPC that was based on existing literature and teacher experience. This can be considered a top-down, knowledge-based approach to error selection, which is extremely useful to obtain an initial idea of common errors

and provide information on the relevance of the errors in terms of salience, persistence, and stigmatization. However, a knowledge-based approach can provide only an impression of the frequency of each error. For more objective and precise information on this point a bottom-up, data-driven approach is required. To develop systems like the one envisaged in MPC, both approaches are necessary. In particular, a data-driven study is required to obtain objective and quantitative data on the frequency of occurrence and the contexts of the various errors, which, in turn, is necessary to develop error detection algorithms and to design useful pronunciation exercises. For this purpose, we decided to use speech recordings of Dutch students of English that had been made at the Radboud University in Nijmegen.

4.2. Method and material

The material we used for analyzing errors in English pronunciation by learners with Dutch as mother tongue (L1) consists of short stories read aloud by Dutch students of English. These stories are used to test the pronunciation level of the students in their first and second year at the Department of English Language and Culture of the Radboud University Nijmegen.

There were 7 different stories varying in length from 200 to 350 words. They were especially selected for the purpose of pronunciation testing. This material can be described as "found data", since it was not collected specifically for the purpose of the MPC project, but for another purpose, namely testing pronunciation for a university English course.

For each story we had the orthographic transcript which was made available to two judges (students) who checked and corrected the orthographic representations. After this process had been completed, the pronunciation was checked. This was done by two Dutch students of English in the last phase of their studies, who were trained in phonetics and had specifically been selected for this task by English teachers.

To make it easier for the transcribers to evaluate the students' pronunciation, a phonemic representation was first created by lexicon lookup in CELEX[17]. The phonemic transcription was corrected if necessary. The transcribers judged every sound and indicated clear pronunciation errors. If a phoneme had been substituted, inserted or deleted, the correct symbol was annotated, if a phoneme had been pronounced incorrectly, but could not be represented by another phoneme, the phoneme was marked by an asterisk (*). If an accepted pronunciation variant had been pronounced, this was annotated as a pronunciation variant and not counted as an error. A subset of the error annotations made by the students was checked by two experienced English pronunciation teachers and judged as very good.

In total the database contains speech material of 226 students reading 617 stories. 58 students read stories both in their first and their second year. The average number of words is 257 and the average length is 98 seconds. The total number of phonemes is about 520.000 (200.000 vowels and 320.000 consonants).

4.3. Results

In Table 3 the number of errors found by the transcribers is shown. The second row indicates the number of errors of all students in the first year, the second year and both years together. The third row shows these results only for those students who took the pronunciation test in both first and second year.

Table 3: Overall frequency of pronunciation errors in the two groups of students.

	1st year	2nd year	average
all students	2.26% 5910 of 261,829	1.22% 3245 of 265,421	1.74%
students in both years	1.86% 1967 of 105,837	1.30% 1644 of 126,170	1.56%

About 1.58% of the vowels and 1.82% of the consonants were mispronounced in the full data set. These percentages may appear to be extremely low, and they probably are, but this does not necessarily imply that pronunciation training is completely superfluous in such cases. Especially this group of students, who want to sound as native as possible, strive to eliminate any traces of non-nativeness from their speech, so even errors that are not very frequent, but that might give them away as non-natives. Whereas the vowel errors were distributed over the whole set, the consonant errors were clearly confined to a restricted set. Table 4 and 5 show a ranked list of vowels and consonants, respectively, displaying the most frequent pronunciation errors (those with a minimum absolute frequency of 100). We also show the relative frequency of occurrence and the relative position in this list (as reflected by percentages of erroneous pronunciations relative to the total number of occurrences of the phoneme).

Table 4: Frequency of vowel mispronunciations

	V	#Err	Rel. freq	Realized as
1	/ʌ/	738	1 (6.0%)	/ə/ (312), /ʊ/ (270)
2	/ʊ/	259	2 (3.4%)	/ʌ/ (186)
3	/ɔ:/	250	4 (3.1%)	/ə/ (77), /ɑ:/ (57), * /ɔ:/ (53), /ʊ/ (31)
4	/ə/	238	9 (0.6%)	/æ/ (65), /ʊ/ (36) /ɑ:/ (27), /ɔ:/ (27)
5	/æ/	220	7 (1.1%)	/ɛ/ (102), /ə/ (43)
6	/ɑ:/	201	3 (3.2%)	/ɔ:/ (120), /æ/ (38)
7	/ɛ/	137	6 (1.2%)	/æ/ (67), /ɪ/ (38)
8	/əʊ/	116	5 (2.2%)	* /əʊ/ (46), /ɔ:/ (33)
9	/ɪ/	108	10 (0.3%)	/ə/ (50), /i/ (22)
10	/u:/	106	8 (0.9%)	/əʊ/ (59)

Table 5: Frequency of consonant mispronunciations

	C	#Err	Rel. pos.	Confused with
1	/ð/	1555	2 (8.0%)	/d/ (1436), /t/ (75), /θ/ (24)
2	∅	733	1 (-)	r-insertion after vowel
3	/z/	929	3 (5.7%)	/s/ (887)
4	/t/	471	5 (1.2%)	/d/ (414)
5	/d/	375	4 (1.3%)	/t/ (313)

The asterisks in the tables (*) indicate pronunciations that do not lead to a phoneme substitution, but that are still erroneous. The most common vowel mispronunciation seems to concern the vowel /ʌ/. This phoneme does not exist in Dutch. According to [7] it is often realized as /ɔ/ when it is represented by the grapheme o as in “other” or as /ʏ/ when it is represented by the grapheme u as in “bus”. /ə/ is another possible mispronunciation in the latter case. Our data indicate that this is indeed the case: /ʌ/ is realized as /ə/ in the majority of mispronunciations. We also found many confusions of /ʊ/ into /ʌ/. Note that both vowels are not part of the Dutch vowel inventory. A more detailed analysis of these cases might shed light on the underlying problems, for example whether these

errors have to do with difficulties in articulating the correct speech sound or with associating the right phonemes to the corresponding graphemes. .

Among the consonants the mispronunciation of /ð/ as /d/ is by far the most frequent error in absolute terms, typically occurring in function words like *the, that, they, this*. The insertion of /r/ after a vowel is very frequent, too. This phenomenon does not occur in RP, but is common in American English. Whether it should be considered as an error is of course dependent on which variety a student intends to learn. The voiceless realization of /d/ as /t/ and /z/ as /s/ is observed word-finally, as envisaged in Table 1. The phenomenon of final devoicing of obstruents in Dutch phonology ([18] p. 32) might account for these mispronunciations. Similar errors were found for /b,g,v/, but appear to be far less frequent. The opposite phenomenon, the voiced realization of /t/ as /d/ was also observed. Since this process was found between sonorants, in words like *little*, it might indicate an attempt by the transcribers to represent the American pronunciation of /t/ in this position as alveolar flap, given that the corresponding symbol /ɾ/ had not been included in the symbol set.

5. Comparison between knowledge-based and data-driven selection

As explained above, Tables 1 and 2 show a preliminary, knowledge-based selection of pronunciation errors based on existing literature and teacher experience. The data-driven account in the previous section provides quantitative and objective data that can be used to make important choices in developing MPC.

If we compare the data in Tables 1 and 2 with those in Tables 4 and 5, respectively, we can determine whether the information found in the literature and the teachers’ impressions are corroborated by objective data. Three of the five vowel errors in Table 1 were also found in our database as shown in Table 4, while the first and the fifth errors in Table 1 did not appear in the database. This might be due to the relatively high pronunciation quality of the students in our sample, which also appears from the generally low frequency of pronunciation errors and the impressions of the teachers.

On the other hand, we did find mispronunciations that had not been included in Table 1. For instance, it seemed that the sound /ʊ/ was often mispronounced as /ʌ/ and that /ɔ:/ had very different realizations in the database (see Table 4). Another relatively frequent error was the realization of /ə/ as a full vowel. This might indicate difficulties in realizing weak forms when these are required ([7] p. 44-47).

For the consonants we observe the voiceless realization of English voiced obstruents in our database as well. Likewise, the errors in the realization of /ð/ as /d/ are also present. The errors in affricates as listed in Table 2 were not found in our dataset, presumably because these are usually made by less proficient speakers than those in our database. The loss of aspiration in initial voiceless plosives was not marked in our database and cannot be diagnosed for that reason. The mispronunciation of /t/ as /d/ in our database is not in the frequent list of errors derived from the literature; it can be assumed that it is considered a less serious error for less proficient speakers of English as it is relevant for the pronunciation of British English only.

6. Implications for MPC

The study reported on in this paper has some important implications for the development of the MPC system, which will be discussed below.

First of all, the database that was created on the basis of the “found speech” available at the Department of English Language and Culture and the annotations made by the transcribers constitutes a good starting point for MPC. This database can be used not only for analyzing pronunciation errors, as we have done in this paper, but can also be employed for training the speech recognizer and the algorithms for pronunciation error detection, which are required for the MPC system.

In connection with this latter use of the database, an important consideration has to be made. As we have seen in the previous section, the students in this database make relatively few mistakes, which concern a relatively limited set of English phonemes. It is therefore important to determine to what extent this specific database can be used when developing versions of MPC for target groups with lower pronunciation quality than the students in the present investigation. Since we are aware of this possible drawback, we are now trying to collect speech material from other, less proficient target groups as well. However, to some extent this database can be employed, for instance by applying the technique of pronunciation error simulation that we applied previously in our Dutch-CAPT system [4], [19].

Second, the present study has confirmed the occurrence of pronunciation errors that could be expected based on the existing literature and expertise, but has also provided useful information on errors that were not included in the preliminary selection, but that are apparently made even by such advanced students. It seems that these errors should also be addressed in MPC, although further analyses are obviously required to gain more insight into the nature of these errors and the specific contexts in which they occur.

A third important implication concerns the content to be developed for MPC for the target group of university students like those in the present study. Since these students make pronunciation errors on a limited set of phonemes, it follows that the exercises in MPC should focus on these specific difficulties and target a limited number of problematic English phonemes. This is particularly important to be able to develop an automatic system for pronunciation training that can accurately identify errors without producing too many false alarms [4]. For this purpose, it is also important to have training material that contains a sufficient number of errors to train the error detection algorithm. However, the technique of error simulation could also be applied in this case, especially because we already have good indications of which errors occur in which contexts.

In addition, experiments will have to be conducted in the near future to calculate important measures of pronunciation error detection accuracy such as precision and recall for the errors made by such advanced language learners. As discussed in [4] detectability is another important criterion in error selection.

To summarize, the present study has provided valuable language resources and useful information that can be used not only to develop and improve the MPC system, but also for other research purposes and applications.

7. Acknowledgements

The MPC project is funded through an STW Valorisation grant by the Netherlands Organisation for Scientific Research (NWO), number 11795. We are indebted to Liesbet Korebrits,

Michiel Sanders, Edith Schouten, and Simone Barnhoorn of Radboud University Nijmegen, the language and communication centre of the Radboud University Nijmegen. Special thanks go to Warda Nejari who carried out the top-down selection of pronunciation errors. Finally, we thank the Department of English Language and Culture of the Radboud University Nijmegen, and in particular Monique van der Haagen and Hedi Kamara, for providing the speech material employed in this study.

8. References

- [1] Abercrombie, D. “Teaching pronunciation”, in A. Brown [Ed], *Teaching English Pronunciation: A Book of readings*, 87-95, Routledge: London, 1991[1956].
- [2] Wang, H., and Van Heuven, V., “Mutual intelligibility of Chinese, Dutch and American speakers of English”. *Linguistics in the Netherlands*, 213-224, 2003.
- [3] Koet, T., Polder “English in Dutch ears”. PhD Thesis, Utrecht/Amsterdam, 2007.
- [4] Cucchiari, C., Neri, A., and Strik, H., “Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback”, *Speech Communication*, 51, 853-863, 2009.
- [5] Strik, H., Cornillie, F., Colpaert, J., van Doremalen, J., Cucchiari, C., “Developing a CALL System for Practicing Oral Proficiency: How to Design for Speech Technology, Pedagogy and Learners”, *Proc. SLATE-2009 Wroxall Abbey*, 2009.
- [6] Sanders, E. and Van den Heuvel, H., “Automatic Pronunciation Error Detection in Repetitor”, *Proc. L2WS, Tokyo*, 2010.
- [7] Gussenhoven, C., and Broeders, A. “The pronunciation of English. A course for Dutch learners”. Groningen: Wolters-Noordhoff-Longman, 1976.
- [8] Collins, B., and Mees, I., “The Phonetics of English and Dutch”. (5th ed.). Leiden: Brill, 2003.
- [9] Van der Haagen, M., “Caught between norms. The English pronunciation of Dutch Learners”. The Hague: Holland academic graphics, 1998.
- [10] Morley, J. “The pronunciation component in teaching English to speakers of other languages”, *TESOL Quarterly* 25, 481-519, 1991.
- [11] Munro, M.J., Derwing, T.M., “Foreign accent, comprehensibility and intelligibility in the speech of second language learners”, *Language Learning* 45, 73-97, 1995.
- [12] Celce-Murcia, M., Brinton, D.M., Goodwin, J.M., “Teaching Pronunciation”, CUP, Cambridge, 1996.
- [13] <http://lands.let.ru.nl/~strik/research/MPC.html>.
- [14] Gussenhoven, C., and Broeders, A., “English pronunciation for student teachers”. (2nd ed.). Groningen: Wolters-Noordhoff, 1997.
- [15] Van den Doel, R., “How friendly are the natives? An evaluation of native-speaker judgements of foreign-accented British and American English”. Lot: Utrecht, 2006.
- [16] Nejari, W., Gerritsen, M., Korzilius, H., Van der Haagen, M., “Reactions to Dutch accented English”. Submitted.
- [17] R. Baayen, R. Piepenbrock, and H. van Rijn. *The CELEX Lexical Database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, USA, 1993.
- [18] Booij, G. *The Phonology of Dutch*. Clarendon Press, Oxford, 1995.
- [19] Kanters, S, Cucchiari, C., Strik, H. *The Goodness of Pronunciation Algorithm: a Detailed Performance Study*, *Proc. SLATE, Wroxall Abbey*, 2009.