Segmental errors in Dutch as a second language:
how to establish priorities for CAPT

Ambra Neri
University of Nijmegen
A.Neri@let.kun.nl

Catia Cucchiarini
University of Nijmegen
C.Cucchiarini@let.kun.nl

Helmer Strik
University of Nijmegen
H.Strik@let.kun.nl

Abstract

In this paper we report on a study that was carried out to obtain an inventory of segmental errors in the Dutch of adult learners with different mother tongues (L1s). The errors observed were subsequently examined in detail to select a number of errors that should receive priority in Computer Assisted Pronunciation Training (CAPT) for Dutch as L2.

1 Introduction

Current advanced CAPT systems are able to provide feedback on pronunciation quality. As many pronunciation errors are due to interference from the L1, some of these systems target one specific language pair (e.g. [1], [2], [3]). These systems are trained on a selection of typical L1-L2 errors. In this way, precise predictions can be made on possible errors, thus making the error identification procedure more efficient and boosting the effectiveness of the training. However, this approach is not always feasible, e.g. when no information is available on typical L1-L2 errors. Moreover, an L1-specific approach is not always desirable because the number of potential users will be limited to the speakers of the chosen language pair. Ideally, a CAPT system should benefit as many learners as possible and, at the same time, identify frequent problems with precision.

Within the framework of our research to develop a CAPT system for Dutch (L2), we wanted to train classifiers to automatically detect the most important errors [4]. To this end, we needed an inventory of segmental errors in nonnative Dutch speech. For this reason, we carried out an experiment to obtain objective annotations of frequent errors. We then carefully studied these annotations and selected a number of errors that, according to predetermined criteria, should first be addressed in a CAPT system.

This paper is organized as follows. First, a short overview of the literature on segmental errors in Dutch as L2 is given. Second, preliminary analyses and the main experiment are described which were carried out with the purpose of producing objective phonetic annotations of segmental errors. Finally, the results are presented and discussed, and a final list of errors to target in CAPT for Dutch as L2 is proposed.

2 Literature on segmental errors in nonnative Dutch

Systematic studies on pronunciation errors in nonnative Dutch are scarce. Some information has been collected by speech therapists and language teachers within the Dutch as L2 teaching context [5], [6], [7]. Other studies considered the differences between Dutch and one or more other languages (e.g. [8], [9], [10]), but comprehensive research is still lacking.

In general, these studies indicate a common problem with vowels, both the monophthongs and the diphthongs, which seem to be more problematic than consonants. Among consonants a common problem seems to be the velar/uvular voiceless fricative /x/, a famous shibboleth sound of the Dutch language.

3 Methodology

In order to look into more detail at frequent pronunciation problems, we analysed Dutch nonnative speech auditorily and made annotations of the most salient deviations from canonical, native Dutch. However, this task is time-consuming, costly and to a certain degree subjective. Therefore, we first carried out auditory analyses with a small set of speech material to establish whether the first author (henceforth annotator1) would be a suitable annotator for a larger set of speech material, i.e. able to annotate segmental deviations in a way that would be similar to what other human experts would do. The main experiment consisted in the annotation and analysis of a larger set of speech material that was produced by 31 nonnative speakers.

3.1 Speech material

The speech material used in the current study is a subset of DL2N1, a previously collected database (see [11]). DL2N1 consists of two sets of five Dutch phonetically-rich sentences that were read aloud and recorded over the telephone by 20 Dutch native speakers and 60 nonnative speakers with different L1s and different levels of proficiency in Dutch. The speech material was orthographically transcribed and evaluated both by machine (ASR) and by different groups of human experts (phoneticians and speech therapists) on several pronunciation aspects (e.g. overall pronunciation, segmental quality, fluency,

and speech rate). For the experiment reported here, we selected a subgroup of 460 sentences by 31 nonnative speakers who had received relatively low scores on overall pronunciation quality (see Table 1 for a distribution according to the L1s). For the preliminary analyses, two different subsets of 45 sentences produced by the same nine nonnative speakers (one for each L1 group) were used.

L1 groups	# speakers
Arabic	4
Chinese/Japanese	3
Turkish	3
Spanish/Italian/Portuguese	5
Russian/Polish/Serbo Cr/Bulgarian	5
Am/BrEnglish	2
German	4
French	3
Swedish/Norwegian	2

Table 1: Distribution of the 31 speakers according to the variable 'L1 group'.

3.2 Annotation procedure

The annotators who participated in the current study were annotator1 and five Dutch expert listeners whose annotations of a subset of the material were used to check whether they were in agreement with those by annotator1.

All annotators were given SAMPA phonetic transcriptions of the selected speech material. These transcriptions were obtained through a lexicon-lookup procedure based on the verbatim orthographic transcriptions. The annotators were asked to listen to each sentence as often as they wished, and to edit the phonetic transcription by annotating what they considered the most serious discrepancies in terms of a limited inventory of phonetic symbols. A list of foreign sounds was given to the annotators to help them identify possible non-Dutch sounds and to keep the number of symbols for the possible realizations to a manageable size, given that all kinds of mispronunciations could be present in the nonnative speech samples.

3.3 Preliminary auditory analyses: Agreement among annotators

Preliminary auditory analyses were carried out in order to check the objectivity of the annotations by annotator1: annotator1 and two annotators annotated a set of 45 sentences by 9 speakers, and annotator1 and three other annotators annotated a different set of 45 sentences by the same speakers. The annotators' transcriptions were then compared pairwise with each other and with the annotations by annotator1.

Each pair of transcriptions was aligned automatically by using the Align program [12], which uses an adapted version of the standard dynamic programming algorithm, and aligns two sequences of elements minimizing the cumulative distance between them. Distance measures between the various symbols are calculated on the basis of articulatory features defining vowels and consonants. The number of errors that Align yielded for each pair was then used to calculate percentage agreement, which is computed by the following formula:

$$\text{percentage agreement} = \frac{\#\,\text{agreements}}{\#\,\text{disagreements} + \#\,\text{agreements}} \times 100$$

3.4 Experiment: Annotations and analysis of errors

3.4.1 Overall trends

First of all, we examined the errors annotated by annotator1, globally. We started by looking at what type of sound was more problematic in the sense that it led to insertions, deletions or substitutions more frequently. Further zooming in onto our data, we then tried to identify the most frequent consonantal and vocalic errors for all the speakers.

In our analysis, we tried to identify the nature of the errors we found, and we examined the realizations in which they generally resulted.

3.4.2 Common pronunciation problems

We also wanted to establish whether and to what degree the overall picture was reflected in each single L1 group. Therefore, we looked for important L1-specific problems and their causes.

3.4.3 Priorities in CAPT for Dutch as L2

Finally, we brought together all our findings to draw up a list of errors that should receive priority in CAPT for Dutch as L2. In previous work [13] we defined four criteria for selecting the errors to be addressed in CAPT: frequency, persistence, perceptual relevance, and automatic detectability. In this study, we mainly considered the first three, while the fourth one will be addressed in a companion paper [4].

4 Results and discussion

4.1 Preliminary auditory analyses: Agreement among annotators

The agreement between annotator1 and the various annotators was generally high ($\bar{x}$ =89.3%, SD=2.6), and did not differ significantly (T-test, p=0.22) from the agreement between the other annotator pairs ($\bar{x}$ =91.1%, SD=3.3). These results indicated that the annotations by annotator1 could be assumed to be objective and could thus be used for further research.

4.2 Experiment: Annotations and analysis of errors

4.2.1 Overall trends

With regard to the sounds that are mispronounced more frequently, we found that vowels are

mispronounced more often than consonants (see Table 2). This trend is in compliance with the data available in the literature and presented in section 2.

		consonant	vowel	sounds
Incorrect	Count	329	682	1,011
	% of incorrect	32.5%	67.5%	100.0%
correct	Count	12,342	7,758	20,100
	% of correct	61.4%	38.6%	100.0%
Total	Count	12,671	8,440	21,111
	% of total	60.0%	40.0%	100.0%

Table 2: Distribution of mispronounced consonants and vowels for the entire set.

The more specific data on sounds that were frequently mispronounced are summed up in Table 3.

Mispronounced vowels					
Cano-nical	Count	% of mispro-nounced vowels	% of same phoneme	realized as	Count
/@/	150	22%	5.1%	deleted	50
				/e/-/e:/	34
				/E/	11
/A/	110	16.10%	13.8%	/a/-/a:/	89
/Y/	55	8.10%	27.1%	/u/	40
/9y/	51	7.50%	22.5%	/Au/	22
				/Oi/	16
/a:/	50	7.30%	7.8%	/A/-/A:/	39
/y/	42	6.20%	22.3%	/u/	38
/2:/	42	6.20%	21.7%	/y/	19
Mispronounced consonants					
Cano-nical	count	% of mispro-nounced consonants	% of same phoneme	realized as	Count
/r/	79	24%	5.4%	deleted	31
				/6/	21
				/l/	12
/x/	38	11.60%	5.9%	/h/	16
				/g/	13

Table 3: Most frequently mispronounced vowels and consonants (SAMPA).

A few observations should be made on Table 3. First, it only shows the most frequent errors, given the impossibility to display all data within this paper. Second, it contains more vowels than consonants because, as we have just seen, vowels resulted in more mispronunciations - a trend that should also be reflected in our final list. Third, the problematic sounds are ranked on the basis of absolute frequency of mispronunciation ('count' and '% of mispronounced vowels/consonants') rather than relative frequency ('% of same phoneme'). The reason for this choice is that the relative frequency may be very high even if the sound is marginal in the language (i.e. a relative frequency of 50% mispronunciation may be obtained when one of two occurrences of a sound is mispronounced), whereas

in our case we want to identify the sounds that give rise to considerable numbers of mispronunciations, hence our interest in absolute frequencies.

4.2.2 Common pronunciation problems

An examination of the data on vowels and consonants for the individual L1 groups confirms the overall trend according to which vowels are more problematic than consonants. Looking at the specific mispronunciations within vowels and consonants, we noticed that the overall trend observed for the vowels was well reflected within the single L1 groups, while the results for consonants tended to indicate more L1 specific patterns, i.e. more variation in the frequently mispronounced consonants across the various L1s.

On the whole, Scandinavian and English subjects produced fewer errors with vowels. This may be explained by the fact that these languages, like Dutch, have complex vocalic systems, with length as a distinctive feature. The English speakers also appeared to have fewer and different errors from all other groups. Closer inspection of the data revealed that the two English speakers had both relatively high proficiency levels in Dutch which, of course, may explain their different 'error behaviour'.

When we looked at the correlations between the number of errors (% within the same L1 group) and the two variables overall pronunciation and segmental quality, we found strong negative correlations (Pearson's r of -0.89 and -0.87 respectively, $p<0.01$). Finally, we discovered a trend indicating that the L1s that are typologically closer to the Dutch language resulted in lower percentages of erroneous realizations, with the Germanic languages (English, Swedish/Norwegian, and German) at the lower end.

4.2.3 Priorities in CAPT for Dutch as L2

At this stage it is important to look at individual errors in order to draw up a list that can be used for our CAPT system. With respect to vowels, we see that the most frequent problems for all L1 groups (except for the English speakers) are /@/, /A/, and /Y/. /A/ and /Y/ are clearly mispronounced because of structural differences between the Dutch vocalic system – which comprises 13 monophthongal and 3 diphthongal sounds, with length and lip-rounding as distinctive features - and the majority of the L1s considered here. Moreover, when mispronounced, /A/ and /Y/ are often replaced by other Dutch sounds, which can lead to serious problems in the communication (a different meaning). Therefore, these errors should definitely be included in our final list. /@/ is important too, but to a lesser degree: of all changes, 50 were deletions, mainly occurring in word-final position. In 23 of these cases, the deletion occurred after a /Z/, in French loan-words (e.g. 'etalage'). This suggests that the cause of the error is due to the fact that the speaker ignores the Dutch pronunciation of the word, rather than to a structural difficulty articulating the 'schwa'. Besides, the

deletion of /@/ in these words is unlikely to lead to serious communication problems. Finally, the /Z/ sound is very infrequent in normal Dutch; therefore this type of error will be infrequent, too, in normal, spontaneous (non-read) Dutch. Moreover, 45 of the incorrect realizations are /E/, /e:/ or /e/ substitutions. Given that all those realizations occur when the 'schwa' sound is represented by the grapheme [e], these results point to interference from the orthographical level, which may not occur or may not occur as frequently in spontaneous speech. Finally, a portion of these substitutions is legitimate in native Dutch too (e.g. 'een' pronounced as /@n/ is the English article 'a/an', pronounced as /e:n/ it means 'one'), and was only annotated here because the original transcriptions were based on a canonical pronunciation lexicon with no pronunciation variants.

With respect to the consonants, we see that /r/ and /x/ are the most frequent errors across all L1 groups. The problems with /r/, however, mainly concern the Asian group – responsible for most of the deletions and the /l/ substitutions - and the German group – with several vocalizations of the /r/ in postvocalic position - rather than all the speakers. Moreover, /r/ can have many different realizations in native Dutch too; therefore it should not be the first priority of a CAPT system to focus on such a problematic sound. The fricative /x/, on the other hand, is a famous shibboleth sound in Dutch, i.e. when mispronounced it is perceptually relevant. As a matter of fact, problems with this sound have also been observed in the literature. Since it is not found in some of the L1 phonetic systems considered here (e.g. Italian, French, etc.), we can indeed assume that learners will have problems articulating this new sound. Therefore, it should be taken up in our list.

5   Conclusions

In this study we have seen that some L1-specific errors can be identified, and that languages that are typologically closer to the L2 tend to result in fewer errors. However, we have also identified clear, common error patterns across speakers of various L1s. Our idea is that while developing a CAPT system such common errors should be addressed first, because this makes it possible to build a system that caters for learners of different L1s. On the basis of our analyses and of considerations on the nature of the errors observed, we suggest that a CAPT system for Dutch as L2 should at least address the following sounds: /A/, /Y/, and /x/. Moreover, the data on typical realizations of those sounds could be used to train specific classifiers for pronunciation error detection (see [4]). In addition, frequent L1-specific deviations of the sort identified here could also be studied in more detail and addressed in L1-specific versions of the same CAPT, thus adding extra value to the system.

References

[1] Menzel, W. et al. 2000. Interactive Pronunciation training, *ReCALL* 13: 67-78.

[2] Carey, M.J. 2003 *An L1-specific CALL pedagogy for the instruction of pronunciation with Korean learners of English.* PhD Disseratation. Maquairie University: Sydney, Australia. http://www.shlrc.mq.edu.au/~michael/carey_phd.pdf

[3] Mak, B. et al. 2003. PLASER: Pronunciation Learning via Automatic Speech Recognition, *Proc. HLT-NAACL 2003 Workshop on Building Educational Applications using Natural Language Processing.* Edmonton, Canada. 23-29.

[4] Truong, K, Neri, A, Cucchiarini, C., Strik, H. 2004. Automatic pronunciation error detection: an acoustic-phonetic approach. This volume.

[5] Aan de Wiel, M.C., Van den Brink, G.A., Struijk van Bergen, S.G.E. 1991. Diagnostiek van uitspraakproblemen van tweede-taalverwervers van het Nederlands. *Levende Talen* 466: 507-511.

[6] Coenen, J. 1991. *Uitgesproken Nederlands.* Groningen: Wolters-Noordhoff.

[7] Tupker, Y. 1995. *Uitspraak Onderwijzen.* Manuscript. Tilburg: Katholieke Universiteit Brabant.

[8] De Graaf, T. 1986. De uitspraak van het Nederlands door buitenlanders. *Logopedie en Foniatrie* 58: 343-347.

[9] Van Boeschoten, J.A. 1989. *Verstaanbaarheid van klanken in het Nederlands gesproken door Turken.* Dordrecht: ICG Printing.

[10] Doeleman, R.1998. *Native reactions to nonnative speech.* Tilburg: Tilburg University Press.

[11] Cucchiarini, C., Strik, H., Boves, L. 2000. Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America* 1072: 989-999.

[12] Cucchiarini, C. 1996. Assessing transcription agreement: methodological aspects. *Clinical Linguistics and Phonetics* 102: 131-155.

[13] Neri, A., Cucchiarini, C., Strik, H., Boves, L. 2002. The pedagogy-technology interface in Computer Assisted Pronunciation Training. *Computer Assisted Language Learning*, 15: 441-467.