Selecting segmental errors in non-native Dutch for optimal pronunciation training[*]

AMBRA NERI, CATIA CUCCHIARINI AND HELMER STRIK

Abstract

The current emphasis in second language teaching lies in the achievement of communicative effectiveness. In line with this approach, pronunciation training is nowadays geared towards helping learners avoid serious pronunciation errors, rather than eradicating the finest traces of foreign accent. However, to devise optimal pronunciation training programmes, systematic information on these pronunciation problems is needed, especially in the case of the development of Computer Assisted Pronunciation Training systems.

The research reported on in this paper is aimed at obtaining systematic information on segmental pronunciation errors made by learners of Dutch with different mother tongues. In particular, we aimed at identifying errors that are frequent, perceptually salient, persistent, and potentially hampering to communication. To achieve this goal we conducted analyses on different corpora of speech produced by L2 learners under different conditions. This resulted in a robust inventory of pronunciation errors that can be used for designing efficient pronunciation training programs.

1. Introduction

Over the last few decades, language teaching has been dominated by learner-centred communicative approaches that emphasize learning through interaction and real-life communication situations (Richards and Rodgers 1986, Savignon 1991). The aim of these approaches is to enable learners to successfully communicate their needs in the second language (L2).[1] Since pronunciation is an obvious component of communication, and since many studies have shown that serious pronunciation problems can hamper communication (see Van Heuven and De Vries 1983, De Graaf 1986, Flege 1999, Van Wijngaarden 2003), or even put the learner at a social and professional disadvantage (see reviews in Eisenstein 1983, Morley 1991, Munro and Derwing 1995), this teaching approach has brought about a renewed interest in pronunciation train-

ing. Not surprisingly, a number of recent studies have shown that tailor-made training is effective in improving both perceptive and productive pronunciation skills (Akahane-Yamada et al. 1998, Derwing et al. 1998, Harless et al. 1999, Lively et al. 1993, Moyer 1999, Bongaerts 1999, 2001, McCandliss et al. 2002, Hardison 2004).[2] Although the specific focus of the training in these studies varied, the general emphasis in current pronunciation training mainly lies in the achievement of fluency and "comfortably-intelligible" (Abercrombie 1991[1949]:93) rather than accent-free pronunciation (see also Kenworthy 1987, Morley 1991, Celce-Murcia et al. 1996). In other words, learners should be able to communicate effectively and efficiently (Kenworthy 1987, Morley 1991) in the L2, in the sense that they should be *easily understood* without provoking frustration or irritation in an interlocutor, but they do not need to sound indistinguishable from native speakers.

Despite this revived interest, the time that is generally available for pronunciation training in traditional classroom instruction has remained relatively limited. In addition, second language (SL) courses, i.e., courses taking place in the country where the L2 is spoken, are characterized by a growing presence of foreign students with different mother tongues, a phenomenon that is becoming common even in *foreign* language (FL) teaching. Consequently, optimal and realistic pronunciation training must not only be geared towards effective and efficient communication, it must also be time-effective, focusing on pronunciation aspects that appear to be the most problematic for a large group of learners of a given L2.

Let us now examine the information that is available for designing pronunciation training that targets those aspects. First of all, we know from research on the mechanisms involved in L2 speech perception and production that the sources of L2 pronunciation errors are often to be found in L1 characteristics that start to become ingrained in the learner's mind during the first months of life, and in universal linguistic tendencies that manifest themselves in implicational orders of acquisition (see the models proposed, for instance, by Eckman 1977, Best 1995, Flege 1995, Kuhl and Iverson 1995, Major 2001, Jusczyk and Luce 2002, Escudero and Boersma 2004). Provided that sufficient information about the languages in question is available, these models allow us to, for instance, predict which phonemes and contrasts of a given L2 will be more difficult to learn for a speaker of a given L1. A similar principle underlies many studies carried out within the framework of Contrastive Analysis research, which, together with Error Analysis, and Interlanguage (IL) research, has yielded inventories of typical errors since the 1950s. However, the problems with most of these studies are that they almost only concern the English language. They do not consider the impact of those errors on communication, and very rarely examine the production of speakers of different L1s at the same time. Similarly, the indicated models have limited value when one common

training programme must be devised for speakers of virtually any possible L1, especially when little is known about the phonetic/phonological system of one or more languages.

More recent studies on native and non-native speech perception and on spoken word recognition in general provide further information indicating that both segmental and supra-segmental factors are important for communication effectiveness and efficiency. Segmental errors, both at phonemic and allophonic/sub-phonemic level, can hinder communication, for instance by slowing down word recognition speed (Derwing and Munro 1997, Rogers and Dalby 1996, Van Alphen 2004, Smith 2005, Shatzman 2006). At the same time, syllabic structure, lexical stress, intonation, and rhythm help a listener to segment the speech stream and recognize words more quickly, or to identify the most important bits of information in an utterance (Cutler 1984, Anderson-Hsieh and Koehler 1988, Anderson-Hsieh et al. 1992, Cutler and Butterfield 1992, Munro and Derwing 1995, Celce-Murcia et al. 1996, Jusczyk and Luce 2002, Hawkins 2003). However, studies on allophonic detail generally look at very specific cases and often use (native) speech that was manipulated artificially, for instance with techniques such as gating and splicing, rather than considering a wide range of phenomena in authentic non-native speech. This approach is necessary because of the large number of factors that can affect auditory intelligibility and communication in general, often in combination with segmentals and supra-segmentals. These factors include: gestures and visual cues, speech rate and style, and voice quality settings (Kenworthy 1987, Morley 1991, Hardison 1996, Munro and Derwing 2001, Van Alphen 2004, Shatzman 2006). Some other important factors to include, such as environmental noise, the listener's degree of familiarity with the talker or the talker's foreign accent, the pragmatic context in which a message is delivered, the redundancy of a message, and the cognitive load it puts on the listener's processing capacity, are, however, independent of the speaker and the speaker's pronunciation skills (Kenworthy 1987, Leather 1998, Munro and Derwing 2001, Van Wijngaarden 2003, Smith 2005).[3]

Moreover, isolating segmental factors from supra-segmental factors and establishing their relative contribution to communication is often difficult since these two levels interact. It is perhaps for this reason that the few studies that attempted to address the issue of pronunciation error gravity hierarchies have failed to demonstrate whether and which level has a stronger impact. These studies also suffer from methodological problems such as the lack of a clear definition and operationalisation of the specific pronunciation dimension investigated, or the use of different scales of measurement for items to be compared (Fayer and Krasinski 1987, Anderson-Hsieh et al. 1992, Derwing and Munro 1997). Consequently, the results of these studies cannot be considered conclusive. More rigorous research has revealed that non-native pronunciation

can be described along a number of dimensions such as intelligibility, comprehensibility, and accentedness, which are correlated but different (Munro and Derwing 1995, Derwing and Munro 1997). However, no study has attempted to establish which specific errors have an impact on these dimensions or what kind of impact they have. The difficulty in this type of investigation is perhaps best illustrated by considering intelligibility. This concept is extremely popular in recent literature, but has hardly ever been isolated and studied. An exception is represented by Munro, Derwing, and colleagues: these authors define intelligibility as "the extent to which the native speaker understands the intended message" (Derwing and Munro 1997: 2). To *globally* measure intelligibility of non-native speech, the authors asked native listeners to orthographically transcribe utterances spoken by non-natives, and they subsequently counted 'errors' in the transcriptions. For the purpose of obtaining a ranked *inventory* of errors hampering intelligibility, rather than a global measure, we cannot just count errors in a transcription, we need to categorize them. Moreover, if we deal with different L1s at the same time, we also need to have comparable material across speakers. But to measure the *actual* intelligibility of an utterance to an interlocutor, it is necessary to use sentences[4] that are semantically unpredictable. In this way, the task of recognizing (and then transcribing) the spoken words is not facilitated by contextual semantic cues. Ideally, words should also have similar amounts of lexical competitors across sentences. Moreover, each sentence presented to a listener should be different, otherwise his/her listening will be facilitated when a given speech stimulus is encountered for the second time. This type of material would not only be difficult to prepare, it would also make it extremely hard to make generalizations across phonetic contexts and speakers, and equally hard to draw any meaningful conclusion as to which segmental errors bear the most responsibility for intelligibility issues. In any case, intelligibility is only one aspect of communication effectiveness and efficiency.

From this overview, it should be clear that very little information is available as to which typical, communication-hampering L2 errors should be tackled in L2 pronunciation teaching, especially for languages other than English. This information is nevertheless much needed in the case of Computer Assisted Pronunciation Training (CAPT) systems: these systems can be useful tools for assisting with self practice and for complementing teacher-fronted pronunciation training. However, they lack the capacity of a human tutor to adapt to student needs and, therefore, must be carefully 'pre-programmed' to target specific problems in speech production. Despite the methodological problems just sketched for this type of research, we believe that it is possible to obtain reliable information on errors that potentially hamper communication and that could be targeted in training to benefit a large group of learners with different L1s. This can be done by examining the non-native production of speakers from a large number of L1s with respect to one L2: the peculiarities of the target language

are likely to result in common problems. These problems can be subsequently pruned on the basis of criteria to obtain a selection of errors that can hamper communication and that appear frequently and persistently in the production of a large number of speakers. This selection can serve as a basis for teachers or CAPT developers to devise efficient pronunciation training that addresses learners with different L1s.

The goal of this study is to propose a method for selecting errors that deserve priority in pronunciation training of Dutch as L2 for speakers of different mother tongues. More precisely, pronunciation errors are listed in an inventory obtained by applying specific criteria in two phases to different types of non-native speech. In the first phase, auditory analyses are carried out by human experts to identify segmental errors. In the second phase, the errors identified are examined in terms of their potential impact on communication, their frequency, and their persistence. Finally, only the errors that meet all or most criteria are included in the final inventory. To limit the scope of this study, the research presented in this paper focuses on the segmental level. This choice also has to do with the application domain of the research within which this work was carried out, i.e. automatic speech recognition for CAPT, which, at its current stage of development, relies primarily on segmental information. Supra-segmental factors that may be responsible for certain erroneous realizations at the segmental level will nevertheless be discussed, since, as we explained, these two levels are often connected.

This paper is organized into three sections each of which describes a separate study on a separate database. The three studies are preceded by a short description of the Dutch language and are followed by a section with general discussion and conclusions. The first study is based on SL-read, i.e., a database consisting of read speech produced by L2 learners of Dutch with different L1s and proficiency levels. This database provided a provisional list of pronunciation errors. The second study is based on SL-xtmp, i.e., fragments of extemporaneous speech with which a second list of errors occurring under more realistic speaking conditions was compiled. Finally, the third study focuses on FL-read, i.e., read speech produced by Italian FL learners who varied with respect to amount of instruction in Dutch. This final analysis made it possible to compile a list of errors made in an FL context and to obtain more detailed information on persistent errors. The results of the three studies are discussed and combined in the final inventory in the concluding section of this paper.

2. The Dutch language and studies available on non-native Dutch pronunciation

Dutch is a West Germanic, Indo-European language that is officially spoken in the Netherlands, Belgian Flanders, Surinam, Aruba, and the Netherlands An-

tilles – though with different varieties – by approximately 21 million people. The Dutch language is taught all over the world. The majority of learners in the Netherlands are immigrants, many of whom are officially required to take a Dutch language course before or upon entering the country and applying for a work permit. Studies on pronunciation problems in non-native Dutch are nevertheless relatively scarce. Some information on general pronunciation problems of non-natives has been collected by speech therapists and language teachers within the Dutch-L2 teaching context, as well as by researchers investigating non-native Dutch for other purposes (Aan de Wiel et al. 1991, Coenen 1991, Tupker 1995, Doeleman 1998). However, these studies have not been systematic. Systematic studies have considered the differences between Dutch and Turkish (Van Boeschoten 1989), and Dutch and Japanese (De Graaf 1986), but comprehensive research on a large group of languages remains insufficient.

The main findings in the available studies point to a common problem with vowels, rather than consonants, both with respect to monophthongs and diphthongs. This may, in part, be due to the relatively high number (Crothers 1978, Maddieson 1984, Lindblöm 1986) of vocalic phonemes of Dutch, which comprises 13 monophthongs, a few marginal vowels found mainly in loan words, and 3 diphthongs. As can be seen in Figure 1, Dutch vowels can be distinguished by their place of articulation, degree of opening (vowel height), and lip rounding, and in some cases by the secondary features length and tenseness. This gives a rough idea of the complexity of this system for an L2 learner. In addition, the difficulties with vowels reported in the literature on L2 learners of Dutch may be due to the fact that learning to articulate new vowels intrinsically requires more effort than learning to articulate consonants (Flege 1988, Strange 1995). The line of reasoning is that the former are produced by movements of the tongue in an "open" and hidden space – the oral cavity – while, for at least some of the consonants, the structures involved in phonation are easier to identify visually and tactilely. The richer proprioceptive feedback for consonants would facilitate long-term retention of these phonemes, with respect to vowels (Fletcher 1983). Moreover, vowels might be more difficult to acquire than consonants because L1-specific patterns seem to emerge earlier for vowels than for consonants (Werker and Tees 1984, Polka and Werker 1994).

With respect to consonants (see Table 1), studies indicate that a common problem is the velar/uvular voiceless fricative /x/, a well-known shibboleth sound of the Dutch language, which has been shown to be a rather uncommon sound among many languages (Maddieson 1984).[5]

Another characteristic of the Dutch language that might be of interest is the fact that there is not always a direct correspondence between pronunciation and orthography. This might cause spelling pronunciation problems especially for speakers of transparent languages, such as Italian, Spanish, Turkish, and Croatian (Erdener and Burnham 2005). With respect to syllabic structure, Dutch
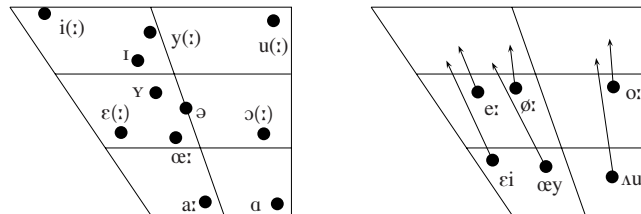
Figure 1. *The Dutch vowel system (Gussenhoven 1999: 76)*

Table 1. *The Dutch consonants (based on Gussenhoven 1999: 74)*

	Bilabial	Labio-dental	Alveolar	Post-alveolar	Palatal	Velar	Uvular	Glottal
Plosive	p b		t d	(c)		k		(ʔ)
Nasal	m		n	(ɲ)				
Fricative		f v	s z	(s) (ʒ)			χ	ɦ
Tap			ɾ					
Approximant		ʋ			j			
Lateral approximant			l					

presents a vast repertoire that includes complex syllables. These can contain consonant clusters of up to three consonants in initial position (as in the word *schreeuw* [sxreːw] – 'scream') and four in final position (as in *herfst* [hɛrfst] – 'autumn') the pronunciation of which requires a certain degree of articulatory effort. Dutch is considered a stress-timed language, and its syllables can have different duration. Syllables containing schwa are unstressed and much shorter than syllables with a full vowel nucleus carrying word stress. Moreover, vowels in unstressed syllables can be reduced or even deleted (Booij 1999, Binnenpoorte 2006). These findings clearly indicate that there are complex characteristics of the Dutch phonetic-phonological system that might represent a problem for learners of different L1s.

3. Study 1: SL-read

3.1. *Goal*

The goal of this study was to obtain an initial inventory of perceptually salient, frequent, and persistent errors that might hamper communication, made by

adult L2 learners of Dutch with different L1s. This study was based on a database of read speech produced by adult non-native speakers living in the Netherlands.

3.2.   *Method*

3.2.1.   *Speakers.*   The speakers involved in this study constitute a subgroup of the original 80 native and non-native adult Dutch speakers who took part in a previous experiment (see Cucchiarini et al. 2000a). All the non-native speakers were living in the Netherlands at the time of the recordings and all reported having learnt Dutch for at least a few months. For this study, 54 non-native speakers were selected to obtain a sufficiently varied group with respect to mother tongue, self-reported proficiency level (based on a standard Dutch test), and gender. On average, these speakers had spent three years in the Netherlands, with a maximum of 14 years.

As Table 2 shows, proficiency levels are not represented equally in this study. Only nine low-proficiency speakers are included in this sample, representing 16 % of the total speakers. The fact that the majority of the speakers are intermediate and proficient speakers implies that the errors identified in this material will be more representative of errors likely to persist over a longer span of time and over a greater amount of instruction, in line with one of the criteria that we adopted for selecting errors (see Section 3.2.3).

From Table 2, it is also evident that some languages were grouped together (e.g., Japanese and Chinese) for our analyses in a way that may not accurately reflect the differences in the phonetic/phonological systems they represent. This choice has to do with the impossibility of finding a sample of non-native speakers that would be perfectly balanced for language proficiency and mother tongue, and that would still be of sufficient size.

3.2.2.   *Material.*   For this initial study, a database of read speech in the form of phonetically rich sentences was chosen. Such material appeared particularly appropriate for the purpose of this investigation for a number of reasons. First of all, the homogeneity in content ensures that when human annotations are made, as in this case (see Section 3.2.3), the annotator is not influenced by oral production factors lying outside the domain of segmental quality. Different grammatical or lexical skills might induce the listener to annotate fewer errors for a more proficient speaker or vice versa. More importantly, such material is characterized by a homogeneous and complete phonemic makeup, which enables comparisons for all Dutch phonemes across speakers.

The speech material in the original database consisted of four sets of five sentences each. The four sets have very similar phonemic makeup and sentence length, and each contains all the Dutch phonemes. Each phoneme appears at

Table 2. *Distribution of the speakers across L1 and proficiency level (low, intermediate, high)*

L1	Low	Inter.	High	Total
Arabic	1	3	2	6
Chinese/Japanese	1	1	1	3
Turkish	1	2	0	3
Italian	0	2	4	6
Spanish	1	2	2	5
English	2	2	2	6
German	1	3	2	6
French	1	3	1	5
Swedish/Norwegian	1	3	0	4
Polish/Russian	0	2	2	4
Serbo-Croatian	0	2	4	6
Total	9	25	20	54

least once in a set, with the more common phonemes appearing more often. The duration of each set is approximately 30 seconds. The sentences were read and recorded over the telephone line. The acoustic conditions of the recordings varied considerably since the subjects were calling from home, but the fragments are generally of good quality (for further details on the sentences, see Cucchiarini et al. 2000a). For the present investigation, only three sets of sentences were considered, for a total of 15 sentences per speaker, in order to limit the task for the annotator.

3.2.3. *Procedure.* The procedure to obtain the final inventory of phonemic errors consisted of two different stages in which errors meeting the following five criteria were selected:
1. Perceptually salient
2. Frequent
3. Common across speakers of various L1s
4. Persistent
5. Potentially hampering to communication.
These criteria stem from practical considerations. First of all, in line with the focus of current pronunciation training approaches, only errors that are perceived by a human listener as clearly deviant from standard Dutch – in other words, as obvious errors – should be considered (criterion 1). With respect to frequency, addressing infrequent errors will, obviously, have little impact on overall pronunciation performance, and will, therefore, not contribute signifi-

cantly to improving communication (criterion 2). In order to be useful for as many learners of Dutch as possible, our inventory should primarily contain errors shared by speakers with different mother tongues (criterion 3). Similarly, it is wiser to spend time and effort in correcting errors that appear to persist over time and possibly over a certain amount of instruction, rather than errors that simply disappear through exposure to the L2 (criterion 4). In this study, which is cross-sectional, errors found in the production of advanced learners are considered to be representative of persistent errors. Finally, perceptually salient deviations that are unlikely to hamper communication, such as deviations that Dutch listeners are familiar with because they are found in certain well-known non-standard Dutch varieties, ought not to be prioritised (criterion 5).

The first phase of the procedure consisted in auditory analyses of the speech material. Here, an expert was asked to annotate perceptually salient errors, in line with criterion 1. The second phase consisted in an evaluation of the errors identified: The annotations were examined from a quantitative and a qualitative point of view, on the basis of criteria 2, 3, and 5 (criterion 4 had already been met indirectly through the selection of the speakers). More specifically, the quantitative analysis was designed a) to identify overall error frequency patterns, i.e. the most frequently mispronounced target phonemes and their corresponding realizations in the whole database, and b) to determine whether those errors were also produced by each individual L1 group, both in terms of frequency and types of realization. In the qualitative analysis, factors such as the nature, possible source, and potential impact of the errors were considered to further refine the list and arrive at a robust final inventory of errors (criterion 5). In other words, only errors meeting all or most of the five criteria qualified to be included in the final inventory.

3.2.3.1. *Annotations.* Auditory analyses were conducted on the 810 fragments (15 sentences × 54 speakers) to produce annotations of perceptually salient deviations from the canonical Dutch pronunciation. For this task, the annotator was provided with canonical SAMPA (Speech Assessment Method Phonetic Alphabet, Wells 2004) phonetic transcriptions of the recordings, serving as the correct reference. These transcriptions were obtained automatically through a lexicon-lookup procedure based on the verbatim orthographic transcriptions of the fragments. The annotator, one of the authors, was well aware of the ultimate purpose of this study. She listened to each sentence as often as she wished and edited the phonetic transcription whenever she noticed salient discrepancies – which she felt should be corrected (criterion 1) – between the learner's spoken utterance and the transcription. The editing procedure consisted of annotating deletions of phonemes, adding inserted segments, and replacing mispronounced target phonemes with their actual realizations. More specifically, in case of substitutions or insertions, the annotator was allowed

to replace the target phoneme with a different Dutch phoneme or allophone or with a foreign phoneme from a short list, or else to add a diacritic symbol. As a wide range of mispronunciations could occur in the non-native speech samples, the list of foreign phonemes was intended as a reference to help the annotator identify possible non-Dutch sounds and to limit the number of symbols for the possible realizations to a manageable size.

Because of the costs of such labour-intensive annotations, it was decided to have one expert carry out the task. However, it is well known that phonetic transcription or even annotation is a relatively subjective task. Since excessive subjectivity in the annotations could compromise the robustness of any inventory based on them, we took measures to minimize this problem. We therefore chose one annotator (henceforth ANN) whom we considered qualified as expert because she had completed the same type of task using a consensus procedure with other experts in the past. We also conducted preliminary investigations to assess the reliability of ANN's annotations. These investigations are described in greater detail in the following sections.

Preliminary investigations on annotator reliability: first check

To determine whether the degree of subjectivity in our annotations would be acceptable, we asked ANN and five other Dutch experts to annotate errors for a subset of the speech fragments and then compared the annotations. This subset consists of two different groups of 45 sentences produced by nine non-native speakers distributed across various L1s. The annotations followed the procedure described in 3.2.3.1. No calibration took place to provide the annotators with examples of what they should consider typical "perceptually salient errors". Instead, all annotators received instructions explaining the procedure and simply requesting them to annotate what they considered "the most serious errors to be corrected in the subjects' speech". This phrase was considered preferable to more detailed instructions since in this way the annotations would reflect more truly what native listeners (subjectively) perceive as a pronunciation error. At the same time, this phrase was intended to make it clear for annotators that they should only consider obvious, perceptually salient errors that really deserve to be targeted in pronunciation training.

The annotations produced by the five experts were subsequently compared pairwise with those of ANN for the same subset. Each pair of transcriptions was aligned automatically, and the disagreements were located and tallied by an alignment program. The following (fictitious) example with IPA (International Phonetic Alphabet) phonetic symbols shows the algorithm output with the aligned annotations of two annotators Exp_x and Exp_y for the speech fragment *Ik wou al [...]* ('I wanted already [...]'), whose available reference transcription was /ɪk/ /ʋʌu/ /ɑl/:

Annotation by Exp_x		i:	k		ʊ	ʌu		ɑ	l	l	
Annotation by Exp_y		ɪ	k		ʊ	u		a:	l	0	
Distances	0.0	0.5	0.0	0.0	0.0	1.5	0.0	0.5	0.0	2.0	0.0
Disagreements	0	1	0	0	0	1	0	1	0	1	0

This program uses an adaptation of the standard dynamic programming algorithm which aligns two sequences of elements, minimizing the cumulative distance between them. This occurs on the basis of distance measures between the various symbols, taking into account the effect of diacritics on basic symbols. The distances are specified in terms of phonetic features in two feature matrices – one for consonants and one for vowels – where each phone is defined by a unique combination of feature values (see Vieregge et al. 1984, Cucchiarini 1996). The number of (dis)agreements for each pair yielded by the program (see bottom row in our example) was used to calculate percentage agreement by the following formula:

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

In other words, this statistic tells us to what degree ANN and another annotator agreed in considering the same target phonemes correct and in indicating exactly the same realization for any mispronounced phoneme in the transcriptions.

Preliminary investigations on annotator reliability: second check

The reliability of ANN as annotator was also examined from a different point of view. As we briefly mentioned, the original SL-read database was collected for other experiments prior to this study. More precisely, it was built in the framework of research aimed at developing measures to automatically assess L2 learner pronunciation. For that study, nine raters (three phoneticians and six speech therapists) used a 10-point scale to evaluate a portion of the speech fragments for overall pronunciation quality, segmental quality, fluency, and speech rate. Based on the high inter-rater reliability measured (Cucchiarini et al. 2000b), the raters' scores can be considered good indicators of the quality of the subjects' pronunciation. For the present analysis, we calculated the mean of the standardized scores assigned to each speaker by all the raters along the dimension "segmental quality" and compared it to the number of errors – relative to the total number of target phonemes – annotated by ANN for each

speaker. The rationale behind this analysis was that any strong negative correlations found between the scores for segmental quality and the number of errors annotated would constitute additional evidence for the relative objectivity of ANN's annotations.

3.3. *Results*

3.3.1. *Preliminary investigations on annotator reliability.* The degree of agreement between ANN and the five experts who annotated a small portion of the database was generally very high, with a mean for the five pairs of $M = 89.3\%(SD = 2.57)$, indicating that the errors identified by ANN were also identified by the other annotators to a considerable extent.

Analogously, the analysis of the scores assigned to each speaker for segmental quality and the percentage of errors annotated by ANN for each speaker (total phonemes divided by total erroneous phonemes) yielded a strong negative correlation ($r = -0.79$, $p < 0.01$). In other words, those speakers who had received a low score on segmental quality from the raters were also found to produce many segmental errors in the annotations by ANN.

These results indicate that ANN's annotations can be considered as sufficiently reliable for the rest of the analyses.

3.3.2. *Annotations: Quantitative analysis.* The first overall phenomenon that we noticed with regard to the more frequently mispronounced phonemes is that erroneous vowels outnumber erroneous consonants, both in relative and absolute terms. While the vowels in SL-read only represent 40 % (n = 14, 218) of the target phonemes, 8 % of the vowels were deleted or substituted (n = 1, 151), as opposed to 2 % for the consonants (n = 452), which represent 60 % (n = 21, 159) of the target phonemes.

We then identified the most frequent vowel and consonant errors and the corresponding most frequent occurrences across all speakers. For the vowels, we selected eight phonemes representing 79 % of the mispronounced vowels (see Table 3a) and 57 % of all the mispronounced vowels and consonants combined. For the consonants, we selected five phonemes, representing 58 % of the mispronounced consonants (see Table 3b) and 16 % of all the mispronounced phonemes. More vowels than consonants were selected since vowels were mispronounced more frequently. The results are summarized in Table 3a (vowels), and Table 3b (consonants).

Having identified the most frequent errors across all L1 groups, we examined each individual L1 group to determine whether and to what extent the error behaviour each of them exhibited reflected the overall results. The general tendency toward more problems with vowels than with consonants was found in each individual L1 group as well. With respect to the error frequency pattern,

Table 3a. *Most frequently mispronounced vowels (MIS V) and corresponding most frequent realizations (Real)*

TV	#MIS	% of MIS V	Real	# Real
/ə/	243	21.1	Deleted	91
			[e]-/eː/	65
			/ɛ/	25
/ɑ/	191	16.6	[a]-/aː/	157
/œy/	96	8.3	/ʌu/	46
			[ɔi]	24
/ʏ/	94	8.2	/u/	69
			/y/	12
/aː/	84	7.3	/ɑː/-[ɑː]	73
/øː/	68	5.9	/y/	31
			/oː/	11
			/u/	9
/ɛi/	68	5.9	[ei]	36
			[ɑi]-[ai]	18
/y/	66	5.7	/u/	56
Total	910	79.1		

Note. TV stands for target vowel, MIS stands for 'mispronounced'. # indicates absolute counts, while % indicates the percentage relative to all mispronounced vowels. Square brackets indicate that a realization does not constitute a Dutch phoneme.

we compared the list of the eight most frequently mispronounced vowels identified for *all* speakers with the eight most frequently mispronounced vowels of *each* L1 group, and we did the same for the consonants. The results show that the frequency pattern for mispronounced vowels is very similar across the various L1 groups, while the pattern for mispronounced consonants is more heterogeneous.

We then examined the error patterns in each L1 group to see whether these reflected the overall error pattern. In other words, we tried to establish whether each common error was also made in each individual L1 group with the same frequency pattern. The individual results generally confirm the overall tendency, especially for the vowels (see Table 4a and Table 4b). While six out of the eleven L1 groups produce all of the frequent, common vowel errors, the ratio is 3/11 for the consonants. The other groups, with the exception of

Table 3b. *Most frequently mispronounced consonants (MIS C) and corresponding most frequent realizations*

TC	#MIS	% of MIS C	Real	# Real
/r/	79	17.5	Deleted	38
			[ɐ]	18
			/l/	16
/t/	59	13.1	Deleted	34
			/d/	17
/x/	38	11.9	[ɡ]	20
			/h/	17
			/k/	5
/w/	36	8	/u/	17
			/f/	11
/h/	35	7.7	Deleted	21
			/x/	13
Total	247	58.2		

the Swedish/Norwegian speakers, produce at least six of the eight errors selected for the vowels and three of the five errors selected for the consonants. In general, about 85 % of the frequent consonant errors are also found in the individual groups; for vowels, this percentage is even higher, 90 % (see Table 4a and 4b).

Thus far, we have only considered errors in terms of substitutions and deletions. However, in our analysis, we also examined insertions as this phenomenon may reveal a problem with specific phonemes or with a specific phonetic context. For reasons of limited space, we will only show a summary table of the most common illegal insertions and the contexts in which they occur most frequently (see Table 5).

As Table 5 shows, the most frequent illegal insertion is that of the schwa phoneme, which appears in word- and syllable-final position after a consonant and within consonant clusters such as in *afspraak* ('meeting') and is realized as, for instance, [ɑfəsəpraːk]. Schwa insertion is a common phenomenon in native Dutch, but not in these contexts, as will be explained later. This insertion was made by all L1 groups with the exception of the German speakers. Another frequent insertion found in all groups, except for the Arabic and the Polish/Russian speakers, is that of the glottal fricative /h/, which frequently occurs at the beginning of words. An even more important problem in this list seems to be the pronunciation of the velar nasal /ŋ/, which in Dutch is represented by the graphemes *ng* as in the word *lang* [lɑŋ] – 'long'.[6] The insertions of [ɡ], /x/,

Table 4a. *Occurrence of most frequent vowel errors (overall trend) within individual L1 groups*

L1	Target vowel							
	/ə/	/ɑ/	/œy/	/ɤ/	/aː/	/øː/	/ɛi/	/y/
	realized as							
	Deleted [e]-/eː/ /ɛ/	[a]-/aː/	/ʌu/ [ɔi]	/u/ /y/	[ɑ]-/ɑː/	/y/ /oː/ /u/	[ei] [ɑi]-[ai]	/u/
Arabic	✓	✓	✓	✓	✓	✓	✓	✓
Chinese/Japanese	✓	✓	✓	✓	✓	✓	✓	No
Turkish	✓	✓	✓	✓	✓	✓	✓	✓
Italian	✓	No	✓	✓	✓	✓	No	✓
Spanish	✓	✓	✓	✓	✓	✓	✓	✓
English	✓	✓	✓	✓	✓	✓	✓	✓
German	✓	✓	✓	✓	✓	✓	✓	✓
French	✓	✓	✓	✓	✓	No	✓	No
Swedish/Norwegian	✓	✓	No	✓	✓	No	✓	No
Serbo-Croatian	✓	✓	✓	✓	✓	✓	✓	✓
Russian/Polish	✓	✓	✓	✓	✓	✓	No	✓

and /k/ after this sound, which account for 27 % of all the insertions, clearly indicate a problem with the pronunciation of /ŋ/. However, this problem is not found among the Chinese/Japanese, the English, and the Swedish/Norwegian speakers.[7]

3.3.3.   *Annotations: Qualitative analysis.*   Our analysis of the data reveals a number of common problems in the pronunciation of Dutch phonemes by non-native speakers with different L1s. Firstly, vowels are clearly more problematic than consonants, regardless of the L1 of the speaker. This tendency is in accord with the data available in the literature (De Graaf 1986, Van Boeschoten 1989, Aan de Wiel et al. 1991, Coenen 1991) and is most likely attributable to a structural characteristic of the Dutch language, which presents a relatively high number of vowel phonemes, as we already saw. Moreover, when examining the individual error frequency patterns displayed by the various L1 groups, we observed stronger similarities for vowels than for consonants. Taken together, these initial indications have important implications for compiling a list of errors to target in pronunciation training: vowels clearly deserve priority over consonants.

With respect to the specific errors identified, a number of considerations are in place. Let us start with the vowels. The most frequent error concerns

Table 4b. *Occurrence of most frequent consonant errors (overall trend) within individual L1 groups*

Target	/r/	/t/	/x/	/w/	/h/
Realized as	Deleted [ɐ] /l/	Deleted /d/	[g] /h/ /k/	/u/ /f/	Deleted /x/
Arabic	✓	✓	✓	✓	No
Chinese/Japanese	✓	✓	✓	✓	✓
Turkish	No	✓	✓	✓	No
Italian	✓	✓	✓	No	✓
Spanish	No	✓	✓	✓	✓
English	No	✓	✓	✓	No
German	✓	✓	✓	✓	No
French	✓	✓	✓	✓	✓
Swedish/Norwegian	✓	✓	No	No	No
Serbo-Croatian	✓	✓	✓	✓	✓
Russian/Polish	✓	✓	✓	No	✓

Table 5. *Most frequent insertions (INS) and corresponding context (overall)*

INS	#INS	% of INS	Context
/ə/	75	28.8	59 post-consonantal, word-/syllable final; 16 within consonant clusters
[g]	36	13.8	36 after /ŋ/
/h/	22	8.5	9 in *het*
/x/	18	6.9	15 after /ŋ/
/k/	17	6.5	15 after /ŋ/
Total	168	64.6	

the schwa, the mid-central vowel found mainly in unstressed syllables. Of the erroneous realizations of this phoneme, 37 % consist of deletions, of which 33 % occur after the (marginal) voiced palato-alveolar fricative [ʒ] in French loan words (e.g., *garage* – garage – incorrectly realized as [xaraːʒ] instead of [xaraːʒə]). The schwa occurs in many languages and is produced with the articulators in resting position and these French loan words are found in other languages with a different pronunciation. Therefore, these deletions may simply indicate that many non-native speakers ignore the Dutch pronunciation of these particular words (and possibly transfer it from another language), rather than revealing a problem articulating this phoneme. If this is correct, this error is likely to disappear naturally once the speaker becomes more familiar

with these loan words. In addition, this fricative is very infrequent in normal Dutch, in which it represents 0.05 % of consonants (based on the *Blinde Biliotheek* and *Netherlands National News Agency/Algemeen Nederlands Persbureau [ANP]* databases from the *Spoken Dutch Corpus/Corpus Gesproken Nederlands [CGN]*, see Oostdijk 2002). In the phonetically rich sentences used for this study, it represents 1 % of the consonants because it was a requirement of our analysis to study all the Dutch phonemes within a relatively small sample of speech. It follows that this particular deletion may be much less frequent in more spontaneous Dutch. The schwa is also frequently (37 % of the incorrect realizations) substituted by /ɛ/ and /e(ː)/ when the phoneme is represented by the grapheme *e*. It seems plausible that part of these errors may be due to the inability of some non-native speakers to apply vowel reduction rules, since the schwa is the sound to which full vowels can be reduced in unstressed position in Dutch, especially /e/ (Booij 1999: 20, 133), and since this problem was observed in other studies on Germanic languages (e.g., Flege and Bohn 1989). However, close examination of the cases in which illegal substitutions take place reveals a different phenomenon, namely interference from the orthographic level. This problem, however, may occur only sporadically in spontaneous speech[8] (see Sections 4.3.1 and 4.3.2), thus reducing the weight of this error for our list.

In contrast, /ɑ/, /aː/, /ɤ/ and /y/ are mainly substituted by legitimate phonemes of the Dutch language, potentially leading to serious communication problems whenever the context does not contribute to disambiguating the meaning of the utterance. The case of /ɑ/-/aː/ seems to be bidirectional, in the sense that the two phonemes are often confused. The diphthongs and the /øː/ lead to substitutions with several different (neighbouring) sounds.

For the consonants, we identified five phonemes that seem to be problematic for most of the speakers: /r/, /t/, /x/, /ʋ/, and /h/. Problems with /r/ were identified for 8 of the 11 L1 groups, indicating that most groups could benefit from additional training with this phoneme. However, most of the errors are produced by the Chinese/Japanese speakers – responsible for most of the deletions and /l/ substitutions – and by the German group – with various vocalizations of /r/ in postvocalic position. Moreover, this sound has different realizations in native Dutch, too. It is realized as uvular approximant [ʁ] (the most common variant in today's standard Dutch, according to Vieregge and Broeders 1993), as alveolar flap [ɾ], uvular fricative [χ], or uvular roll [ʀ], depending on its position as well as on individual and regional variation (Booij 1999: 8, Van de Velde and van Hout 2001). The relative importance of this error within the list obtained could thus be scaled down.

The plosive /t/ is mainly deleted in word-final position, especially in consonant-clusters such as those in the words *belooft* [bəloːft] '(he) promises' – and *gelokt* [χələkt] 'lured'. These deletions seem to be due to assimilation

processes across word boundaries, which do not generally represent legal processes of Dutch connected speech. They may also be motivated by a problem in the realization of this consonant in coda consonant clusters (Tarone 1980, Lin 2001), rather than with the articulation of this plosive. It is also true that these errors occur in read speech, which may display peculiar phenomena such as the anticipation of certain sounds leading to the deletion or the assimilation of preceding sounds. Furthermore, /t/ deletion is quite common in some Dutch dialects (Goeman 1999) and is a typical process of fast speech in standard Dutch (Booij 1999). A few deletions found in our database (e.g., in the word *niet* [nit] – 'not') represent processes of Standard Dutch too. However, the deletion of word-final /t/ might seriously compromise the grammaticality of an utterance, and possibly its comprehensibility, because word-final /t/ has a morphological function in Dutch verbs. More precisely, it marks (singular second and third) person number in Dutch present tense verbs, as in the *belooft* [bəloːft] example, while its absence can indicate (present tense) first person number as in *beloof* [bəloːf] '(I) promise'. It can also signal mood and tense, as in the contrast [bəloːft]/[bəloːf] – 'promised' (past participle) or 'promise' (present imperative). In contrast, the reason why this plosive is sometimes replaced by its voiced counterpart /d/ seems to be an orthographic confusion in 10 out of 17 cases, all occurring in the word *hard* – 'hard' [hɑrt] – pronounced as [hɑrd]. It cannot be excluded, though, that some of these substitutions stem from factors such as resyllabification induced by the reading task, for instance resulting in *hard als* – 'as strong as' – being pronounced as [hɑr | dɑls] rather than [hɑrt | ɑls], or an L1 preference for open syllables. For the remaining cases, however, no systematic pattern emerges and no phonological process in Dutch justifies the various realizations found. In consideration of the fact that the phoneme seems to be problematic across all L1 groups and that a number of /t/ deletions found are not licit and may hamper communication, we believe that the realization of /t/ should be considered in our provisional list. However, we should only focus on the realization of this phoneme in word-final position or within coda consonant clusters.

A much more straightforward case is represented by the velar/uvular fricative /x/, long a famous Dutch shibboleth found in relatively few languages in the world (Maddieson 1984). Addressing it within pronunciation training will, therefore, not only help learners to communicate more effectively, it will probably help them to 'sound more Dutch' too. The problem concerning the semiconsonant /ʋ/, by contrast, may be less serious because the [u] (or more precisely [β] or [w]) realization is found in some national (Brabants) and international (Suriname) varieties of Dutch (Rietveld and Van Heuven 1997, Booij 1999). The number of speakers of these varieties is relatively high in the Netherlands, and this accent is, therefore, well-known to Dutch people. So, this realization may be perceived as striking and reveal some non-standard

Dutch accent, but it does not seem likely to lead to communication problems. Therefore, it does not seem necessary to include this error among priority errors. Doubts arise also with respect to the priority that should be given to the glottal fricative /h/, because the errors relative to this phoneme only appeared in 6 of the 11 L1 groups considered here. It may nevertheless be useful to target specifically the /h/-/x/ substitution since it may help to develop a better representation and production of the latter fricative.

To conclude our analysis, let us consider the illegal insertions. The fact that many schwas are inserted in post-consonantal, word- and syllable-final position may be explained by syllabic factors, i.e., by a universal tendency towards a simple CV syllable structure (documented in other studies, e.g., Tarone 1980,[9] Eckman 1981, Carlisle 1991), contrasting with the variation in the Dutch syllabic repertoire (Booij 1999). At the same time, this factor may coincide with a difficulty in pronouncing consonant clusters: as already mentioned, Dutch syllables can contain uncommon consonant clusters of up to three consonants in initial position and four in final position, the pronunciation of which requires a certain degree of articulatory effort. Problems in the production of L2 consonant clusters resulting in medial vowel epenthesis have often been observed (Broselow 1983, Hancin-Bhatt and Bhatt 1998, Davidson et al. 2004), especially within formal tasks without linguistic context, such as list reading (Major 1987, Weinberger 1987, Lin 2001). The picture is complicated by the fact that schwa epenthesis does occur in Dutch, but only in non-homorganic, coda consonant clusters, on the basis of (optional) rules governed by speech style (Kuijpers et al. 1996, Booij 1999, Swerts et al. 2001), adding up to the possible confusion in the learner's representation of the Dutch syllable structure. While some of the articulatory difficulties involved in pronouncing Dutch consonant clusters may disappear with the acquisition of a certain degree of skill, thus naturally alleviating the problem of schwa epenthesis, schwa epenthesis can seriously disrupt the stress pattern of a word or sentence by creating new syllables and modifying existing ones. This, in turn, can compromise the communication flow since stress patterns play a crucial role in the listener's recognition of words (Cutler 1984). Consequently, schwa epenthesis should be included in our inventory of priority errors.

With respect to the glottal fricative /h/, we observed that 40 % of the insertions occur at the beginning of the word *het* (the Dutch definite article and pronoun for neuter nouns). This word, normally pronounced [ət], has a marked variant which is pronounced as [hɛt], i.e., with initial /h/. However, this variant was not included in the stimuli for this database. Therefore, this error seems to pertain more to the lexical level, rather than to pronunciation training, and should be omitted from our list. Finally, a number of illegal insertions of [g], /x/ and /k/ were observed after the velar nasal /ŋ/[10] as in [bəʋeːxɪŋgən]/ [bəʋeːxɪŋxən]/[bəʋeːxɪŋkən] instead of [bəʋeːxɪŋgən] for *bewegingen*, 'move-

ments'. Since this phoneme is represented in Dutch by the grapheme sequence ng, it is possible that most of these insertions originate from orthographic interference. Support for this hypothesis comes from the fact that this sound is found in many languages, but it often represents an allophone of /n/ before the velar stops [g] and /k/ (Greenberg 1978), rather than an independent phoneme. While the problem is considerable in terms of the overall frequency of errors it generates, we must point out that it does not occur in the Chinese/Japanese, the English, or the Swedish/Norwegian groups, and it is present only to a limited extent in the production of the Polish/Russian and German speakers. Consequently, targeting this problem is likely to benefit only a limited group of learners. As a compromise, we may include it in our inventory, while assigning it very low priority.

3.4.   *Conclusions of Study 1*

In this study, we analysed annotations of perceptually salient pronunciation errors in the Dutch of adult non-native speakers with different L1s. On the basis of this analysis, we have produced a list of target phonemes that should receive priority in pronunciation training. For the vowels, which are more problematic, the following phonemes, listed according to their priority, should be addressed first because they are frequently mispronounced by learners of many L1s and might hamper communication: for the monophthongs /ɑ/, /aː/, /ʏ/, /y/, /øː/, and /ə/, and for the diphthongs /œy/ and /ɛi/. For the consonants, /x/ and, to a lesser degree, /t/ should be addressed by focussing on the context in which they occur, while targeting /r/ and /ŋ/ may only be relevant for a limited number of learners. Since the speakers involved in this study had been living in the Netherlands and had received a certain degree of instruction in Dutch, we can hypothesize that the errors identified also represent problems that persist over time and that are resistant to (a certain amount of) instruction. This hypothesis needs nevertheless to be confirmed by data in which these two factors are studied more systematically.

   Another important issue should be considered at this point. The data just described stem from an analysis of read speech, which has sometimes been criticized for several reasons as an artificial basis for assessing pronunciation quality. Firstly, pronunciation elicited with a formal reading task has been shown to result in different types of speech errors and in greater segmental accuracy with respect to more spontaneous oral tasks, since the cognitive load is low when the learner is simply focusing on the form rather than on the content (Dickerson 1975, Lin 2001, Cucchiarini et al. 2002). The oral production thus obtained would not truly reflect the actual production of a learner engaged in a real-life speaking task. In addition, recent findings (Lin 2001) seem to indicate that specific reading tasks with different units of elicitation and analysis

– such as word lists as opposed to passages, and singleton consonants versus consonant clusters, respectively – may result in different pronunciation error patterns.[11] Reading a list may also result in a distorted prosodic pattern, potentially affecting the quality of some phonemes as well. Moreover, read speech may be influenced by orthography to a larger degree than spontaneous speech (Young-Scholten 1997), and this influence is likely to differ among readers of different writing systems, i.e., with a different phonemic awareness (Bassetti 2006). Finally, the way in which phonetically rich sentences are constructed, in particular the requirement to include infrequent phonemes within a relatively small portion of speech, makes it to some degree artificial. Consequently, to achieve greater robustness in our selection and to confirm some of the hypotheses that we advanced on the actual occurrence of certain errors, an analysis of spontaneous speech is desirable.

4.   Study 2: SL-xtmp

4.1.   *Goal*

The goal of this study was to obtain an inventory of errors that are made by learners of Dutch in a spontaneous speaking context and that meet the criteria indicated in Section 3.2.3. To obtain the inventory, auditory analyses were carried out and the errors annotated were examined following a procedure closely resembling the one described in Study 1.

4.2.   *Method*

4.2.1.   *Speakers.*   The speakers in this study were 43 learners of Dutch representing a subgroup of 60 immigrants who took part in a test of Dutch L2 proficiency. This test, known as *Profieltoets*, is normally administered to immigrants who have followed 500 to 600 hours of a Dutch language course, in compliance with the immigration regulations of the Netherlands. The speakers involved in this study originally served as subjects for an experiment aimed at investigating automatic pronunciation assessment for L2 learners of Dutch (Cucchiarini et al. 2002). Because of the context within which the speech was collected, where protection of the candidates' privacy was essential, the only explicit information provided to the authors about these subjects is their level of proficiency (low-proficiency and intermediate). However, this database was chosen for its ecological validity since it contains an authentic sample of speech produced by real learners of Dutch that had to be officially evaluated for its quality. The subjects were selected by employees of the institute (*CITOgroep*) where the test was developed in order to provide variety with respect to gender, mother tongue, and proficiency level. For the analysis described in this study, 23 speakers at the low-proficiency level and 20 speakers at the intermediate level were selected semi-randomly.

4.2.2. *Material.* The database used in this study consists of extemporane-ous speech. This type of material was chosen to compensate the shortcomings of Study 1. The test used to elicit the speech fragments contained in the SL-xtmp database was a speaking test from the 1998 *Profieltoets.* For this test, the candidates had to answer a number of questions that differed for the two proficiency levels. For the low-proficiency speakers, eight so-called short tasks were selected in which the candidates were required to answer simple novel questions such as "You want to buy a Dutch dictionary. You go to a bookshop, but you see seven different dictionaries. What can you ask to the shop assis-tant?" Their answers had an average duration of 15 seconds. The eight ques-tions selected for the intermediate group were cognitively more demanding and resulted in answers of approximately 30 seconds in which the candidate had to make a choice and motivate it. As a result, more speech data were available for the intermediate speakers. In the absence of any information on the amount of exposure the speakers had to Dutch, we assumed that the errors made by intermediate speakers would be more representative of persistent errors. Con-sequently, having more speech of intermediate speakers was considered to be a positive fact since it made the database more representative of "persistent" errors, in accordance with criterion 4.

The advantages of extemporaneous speech should be clear from the discus-sion in Section 3.4. The task in the *Profieltoets* speaking test, in particular, specifically requires the candidates to pretend to speak as if in an everyday-life situation: this ensures even greater realism in the speech, making the latter more likely to reflect errors made in real-life situations. However, this type of material is characterized by heterogeneity. In this specific case, the frag-ments differ both in length and content for each item, also making the type and amount of phonemes vary by speaker across individual items.

Two more characteristics of this material deserve mention here. First, the test was administered in a language laboratory to several subjects at the same time; therefore, the recording conditions were rather adverse, with a good deal of background speech. Second, probably because some speakers experienced some degree of anxiety, which is known to occur frequently during tests and especially while carrying out oral tasks in the L2 (Horwitz et al. 1986, MacIn-tyre and Gardner 1991, 1994), many false starts and hesitations were produced, especially among the low-proficiency speakers. Some candidates also spoke rather softly and did not articulate the sounds clearly, making some recordings difficult to understand and a few close to incomprehensible. For this reason, 17 of the utterances produced by the low-proficiency speakers, and 8 of the utter-ances produced by the intermediate speakers, plus one speaker altogether were eliminated from the original set. Measures were also taken to ensure that the poor quality of some of the remaining recordings would not affect the results of our analyses (see Section 4.2.3).

4.2.3. *Procedure.*   The procedure used in this study closely resembles the
one adopted in Study 1: it was guided by the criteria sketched in Section 3.2.3
and it consisted in initial auditory analyses, followed by a two-step analysis
of the annotations. The 319 fragments selected were, thus, first analysed audi-
torily by ANN and annotations of the most salient deviations from the Dutch
canonical pronunciation were made on the basis of criterion 1. This time, how-
ever, an additional criterion, "the error should be clearly audible", was adopted,
because of the poor acoustic conditions of a considerable number of record-
ings. In other words, only clearly audible errors were edited, while doubtful
cases were ignored, thus counting as correct. While this approach means that
more errors could actually be present in the recordings than in the annotations,
such a conservative criterion seemed more appropriate for the purpose of the
study: the risk of overlooking a few errors appeared wiser than counting errors
not actually produced.

Once again, the analyses of the annotations were both quantitative and qual-
itative, and resulted in a final list of frequent and persistent errors that might
hamper communication. However, these analyses differed slightly from those
of Study 1 on account of the differences in the material and the speakers of
SL-xtmp. First of all, the fact that we were dealing with semi-spontaneous
speech and with adult learners meant that we might have to reckon with in-
stances of avoidance, i.e. the phenomenon whereby learners avoid producing
difficult structures for fear of making a mistake (Schachter 1974, Onwuegbuzie
et al. 1999), as hypothesized by Piske et al. (2001). If some target phonemes
are regularly avoided, possibly as a result of *lexical* avoidance, certain errors
may be overlooked just because they are not in the data, while the reason for
their absence may lie exactly in the difficulty of producing the target phoneme.
For this purpose, we first ran a check on the completeness of our data, by
comparing the frequencies of the target phonemes in SL-xtmp with the fre-
quencies in databases of spoken Dutch, which contain a total of 3.7 million
sounds (*Blinde bibliotheek* and *ANP* from the *CGN*-Oostdijk 2002-and read
speech from the *Institute of Phonetic Sciences/Instituut voor Fonetiek Amster-
dam [IFA]* database-Van Son et al. 2001). Second, since we had no information
on the specific L1s in the databases, we could not draw inferences about L1-
specific errors, and thus only examined overall error trends across all speakers.
However, we did have information on the *proficiency level* of the speakers,
which we used to focus our analysis primarily on persistent errors. As indi-
cated in Section 3.2.3, we assumed that errors produced by the more proficient
group would be representative of persistent errors.

4.3. *Results*

4.3.1. *Annotations: Quantitative analysis.* First of all, when we checked the completeness of the speech data in terms of target phonemes, we could not discern any clear cases of avoidance in either proficiency group: the target phonemes produced by low-proficiency beginner and intermediate speakers show similar frequencies to those found in a large database of spoken Dutch.

We then started exploring our data by looking at the number of errors annotated for the two proficiency groups. As explained, more material was available for the intermediate speakers as a result of the different task and proficiency of this group. It is, therefore, not surprising that, in absolute terms, these speakers produced twice as many errors (n = 370) as the low-proficiency speakers (n = 185). These numbers are probably also an indirect consequence of the additional criterion that was adopted in the annotation of this database: since the more proficient speakers generally articulated more clearly than the low-proficiency speakers, it was easier for the annotator to *hear* pronunciation errors in their speech and to annotate the errors as such. However, the percentages of errors (relative to all phonemes) identified for each group are equal (each 2.3 % of all the target phonemes). Moreover, a bias towards the more proficient speakers would indirectly enable us to meet criterion 4. Consequently, in the remainder of these analyses, we consider the error behaviour displayed by all the speakers.

A first, global examination of the errors indicated that vowels (n = 9,458), representing 39 % of the target phonemes, are much more frequently mispronounced (n = 354, 4 % of all vowels) than consonants (n = 139, 1 % of all consonants). In this respect, it is worth mentioning the extreme difference in the error patterns displayed by the two proficiency groups: for the low-proficiency speakers, the number of mispronounced consonants (n = 75, 46.6 % of all mispronounced phonemes) almost equals that of the mispronounced vowels (n = 86, 53.4 %), while the ratio for the intermediate speakers is 64/268 (19.3 %/ 80.7 %). Table 6a and Table 6b present an overview of the specific vowel and consonant errors and their corresponding realizations for all the speakers in this database. For the vowels, we selected the nine most frequently mispronounced vowel phonemes, representing 79 % of the mispronounced vowels and 57 % of all the mispronounced phonemes. For the consonants, we selected the three most frequent errors, representing 67 % of the mispronounced consonants and 18 % of all the mispronounced phonemes.

As indicated in Study 1, another source of information on problems with Dutch sounds is illegal insertions. The results for the most frequent insertions, and the contexts in which they occur, are indicated in Table 7.

As the table shows, the most frequent illegal insertion is that of schwa, which appears in word- or syllable-final position after a consonant and within conso-

Table 6a. *Most frequently mispronounced vowels and corresponding most frequent realizations (overall)*

TV	#MIS	% of MIS V	Real	# Real
/aː/	47	13.3	/ɑ/	31
			[a]	7
			[ɑː]	6
/y/	43	12.1	/u/	38
/ɑ/	37	10.5	[a]	25
			/aː/	8
/œy/	29	8.2	/ʌu/	27
/ɛi/	29	8.2	[ei]	17
			[e]-/eː/	8
/ɔ/	28	7.9	[o]	10
			/oː/	4
			/ɔː/	4
/ɛ/	24	6.8	[e]	17
/i/	24	6.8	/ɪ/	19
/ɣ/	20	5.6	/u/	18
Total	281	79.4		

Table 6b. *Most frequently mispronounced consonants and corresponding most frequent realizations (overall)*

TC	#MIS	% of MIS C	Real	# Real
/x/	52	37.4	/k/	25
			[ɡ]	9
			/h/	7
/s/	25	18	[ʃ]	17
/h/	16	11.7	/x/	15
Total	93	66.9		

nant clusters. The insertion of the voiced plosive [ɡ] only occurs after the velar nasal /ŋ/, either in syllable- or word-final position, indicating a systematic problem with the velar nasal phoneme.

Table 7. *Most frequent illegal insertions and relative context*

INS	#INS	% of INS	Context
/ə/	24	38.7	16 post-consonantal, word-/syllable final, 7 within consonant clusters
[g]	13	21	13 after /ŋ/
Total	37	59.7	

4.3.2. *Annotations: Qualitative analysis.* The overall results show that vowels tend to be mispronounced much more frequently than consonants, in agreement with the data available in the literature. A further check of this trend in the two proficiency groups indicates that the percentage of mispronounced consonants slightly decreases as a function of proficiency level, whereas the percentage of erroneous vowels doubles.

The interpretation of the specific errors identified for vowels is quite straightforward, and since it broadly coincides with the results from Study 1, it need not be repeated here. On the whole, many vowels are confused with sounds that are similar from a phonetic and phonological point of view, but differ both in quantity and quality. In line with our hypothesis (Sections 3.3.3 and 3.4), the errors relative to /ə/ are so infrequent in this database that they do not even appear in Table 6a.

With respect to consonants, we observe that the uvular/velar fricative /x/ is the most frequently mispronounced phoneme, and it displays exactly the same realizations found in Study 1. Moreover, a closer examination of the context in which /s/ is mispronounced reveals that 50 % of the [ʃ] realizations are in fact a substitution of the entire consonant cluster [sx], represented by the graphemes *sch*, and that they mainly occur in word-initial position. This may indicate, once more, a difficulty with the fricative /x/, rather than with the /s/, or at any rate with these sounds in combination because they require a shift from an alveolar (coronal) to a velar/uvular (high). Another fricative, the glottal /h/, appears in the list of most frequent errors and is erroneously realized as /x/ in initial position, either confirming a tendency to confuse /h/ and /x/, possibly as a consequence of interference from the L1, or indicating a form of hypercorrection with respect to the velar/uvular fricative. Finally, a review of the illegal insertions detected in this database reveals no new trends with respect to Study 1.

4.4. *Conclusions of Study 2*

In this study, we obtained a list of perceptually salient, frequent, and persistent errors made by learners with different L1s that are likely to be produced in spontaneous speech. This list can, in fact, be considered to be underrepresenta-

tive because of the restrictive criterion applied in the annotation, according to which only "clearly audible" errors were annotated. However, this list shows some clear trends that broadly validate the data available in the literature and in Study 1. For the vowels, which pose more problems than consonants in the long term, the following phonemes, ranked according to their priority, should be addressed first of all: for the monophthongs /aː/-/ɑ/, /y/-/ʏ/, /ɔ/, /ɛ/, and /i/, and for the diphthongs /œy/ and /ɛi/. On the basis of the types of substitutions observed, we can hypothesize that a pronunciation training programme that contrastively addresses pairs of confusable vowels, such as /aː/-/ɑ/ and /y/-/ʏ/ may turn out to be particularly effective and efficient in that it may solve two problems simultaneously. For the consonants, /x/ should be addressed and contrasted to the other (clusters containing) fricatives such as /h/ and [sx], given the confusion observed in the realization of these sounds. The characteristics of the subjects and the material used in this study give us reason to believe that these errors are those that typically persist over a certain amount of time and instruction in Dutch. However, this hypothesis needs to be confirmed by data in which time and instruction are studied more systematically. Within the framework of a cross-sectional study as this one, an FL teaching context would seem the appropriate environment for this purpose because, in this case, both factors can be controlled more easily: they tend to coincide and to be restricted to the learning environment.

5.   Study 3: FL-read

5.1.   *Goal*

The purpose of this study is twofold: to obtain an inventory of pronunciation errors made by a group of SL learners of Dutch that meet the criteria indicated in Section 3.2.3 and to gain insight into error behaviour across several years of instruction and exposure to Dutch. The latter goal has to do with the fact that the FL learning context makes it possible to more easily quantify and control the amount of instruction and exposure to the FL. It should be pointed out, however, that the study described here is not a longitudinal study: it examines the oral production of different speakers at various levels at one point in time. Annotations of Dutch (FL) errors from a database of read speech were made and analysed.

5.2.   *Method*

5.2.1.   *Speakers.*   The speakers involved in this study were 20 Italian students aged 19 to 25 attending a Dutch course at the School for Interpreters and Translators (SSLMIT) of the University of Trieste (Italy). The choice of Italian students was mainly dictated by pragmatic reasons: given the difficulty of collecting a database of FL-Dutch in several countries, we chose one language and

one specific group of subjects for which we knew we could collect a reasonably balanced database that would be likely to yield a conservative list of errors (for reasons that are explained later). Moreover, Italian is an Indo-European language, like Dutch, but it belongs to a different branch. The students were selected across five years of instruction to obtain a sample that would be as homogeneous as possible (see Table 8). All the subjects were Italian native speakers who started studying Dutch at university after the age of 18, and who were already highly proficient in at least another second language, either English or German. Foreign language proficiency either in English, German, or French was a requirement for all Italian native speakers who enrolled at the university, and was measured with an admission test on listening and writing skills. All of the students had followed at least 150 hours of lessons in Dutch during their first year and had spent from 0 to 8 months in a Dutch-speaking country. In the first two years, the students shared a number of courses meant to introduce them to both interpreting and translating, while the students in the following years were specializing in one of those lines of study. This distinction was used to pool subjects into two main proficiency groups for part of the analyses reported on here: Group A included the students from the first two years, while Group B included the students from the three last years (see Table 8).

To interpret the results of this study correctly, a few more characteristics of these subjects should be mentioned. They all had some degree of talent for FL learning and a strong motivation for learning Dutch.[12] They were facilitated in learning Dutch, at least initially, by the fact that they already spoke English, German, or both fluently. Moreover, they were being trained to become professional translators or interpreters, i.e., to achieve a high level of FL proficiency at the end of their studies (although the emphasis of the Dutch course lay primarily in receptive rather than productive skills). Finally, they all expressed a desire to improve their pronunciation in Dutch, though they considered grammatical accuracy more important. On account of these characteristics, we can reasonably assume that the errors identified for this sample represent an underestimate of a more representative sample of FL learners with the same or a similar language background, both in terms of the quantity of errors produced and their persistence over time. In other words, the error pattern displayed by these subjects in spoken Dutch can be considered rather conservative.

5.2.2. *Speech material.* For this study, the same elicitation material was used as in the SL-read study, i.e., phonetically rich sentences to be read aloud after optional rehearsal sessions. Originally, recordings were also made of semi-spontaneous speech consisting of a description of five pictures representing everyday scenes and objects, but this task turned out to be too problematic for these subjects. Therefore, the material obtained in this way was discarded from

Table 8. *Distribution of the Italian subjects across the variables years of instruction (yrs), broad proficiency group (Broad PG), gender (M=male, F=female), and time spent in a Dutch-speaking country*

# yrs	Broad PG	# M	# F	Total	months.weeks in Dutch-speaking country
1	A (low)	2	2	4	0.0–0.3
2		2	2	4	1.2–2.0
3	B (high)	1	3	4	1.0–3.1
4		0	4	4	3.0–6.0
5*		0	4	4	2.0–8.0

* The fifth year is indicated with an asterisk, because the courses officially lasted four years, but the students who did not complete their exams within four years generally continued following courses or practising at the university.

the study.[13] The recordings we used were made in an office of the SSLMIT building, and were conducted under the supervision of an investigator who explained the meaning of certain words beforehand, if required, to avoid production problems resulting from possible comprehension problems. For the annotations, two sets of sentences were selected, for a total of 200 sentences (see Section 5.2.3).

5.2.3. *Procedure.*   The procedure we adopted here followed the criteria outlined in Section 3.2.3 with the obvious exception of criterion 3 (common across speakers of various L1s), and consisted in obtaining annotations by means of auditory analyses, followed by a further examination of the errors found. However, some differences with respect to Study 1 and Study 2 deserve mention. In this case, five phoneticians annotated the material during several sessions by means of a consensus procedure. In other words, during each session at least two phoneticians were present, who would annotate only those errors that they both considered perceptually salient and potentially hindering to communication. This consensus procedure generally results in more objective transcriptions (Shriberg et al. 1984), but has the disadvantage of higher costs since it involves more experts. To counterbalance this effect, it was decided to limit the amount of speech to be annotated.

   In accordance with the double goal of this study, quantitative and qualitative analyses were conducted to study whether and how error behaviour changes as a function of augmented instruction and exposure to the FL, as well as to retain "persistent" errors for our final inventory. In the previous studies, the latter was ensured by the fact that more material was available for the more proficient groups, either through a higher number of speakers in that group (Study 1) or as an indirect consequence of the elicitation task (Study 2). In this case,

the speakers were perfectly balanced with respect to the number of years of instruction and amount of speech produced. Moreover, the students within the same year of instruction had spent a similar amount of time in a Dutch speaking country. Consequently, we first of all exploited this characteristic by examining global trends in error behaviour at each year of instruction. We subsequently merged the speakers in these groups into the two broad proficiency Groups A and B (see Table 8) and we compiled lists of frequent vowel and consonant errors for each group, following the procedure used in Study 1 and Study 2. We did this to highlight the contrasts owing to the relatively small size of our sample and to the similarities among the students in these two groups. For the final inventory, following a qualitative analysis of the raw data and in line with criterion 4, we considered only the results obtained for the more proficient students (Group B in Table 8).

5.3. *Results*

5.3.1. *Annotations: Quantitative analysis.* As can be seen in Table 9, the number of errors decreases as a function of instruction time. Further analyses of error behaviour in the individual groups revealed a strong negative correlation between years of instruction and number of errors – represented as the sum of illegal insertions and erroneous consonants and vowels per year of instruction – $(r = -0.91 p < 0.05)$. It appears that the number of pronunciation errors decreases as a result of instruction, or more generally as a result of the amount of exposure to the Dutch language. Vowels are much more frequently mispronounced than consonants, and this disparity increases as the years of instruction increase. As Table 9 and Figure 2 show, the number of erroneous consonants decreases considerably (by 85 %) from the first to the fifth year, whereas the rate of erroneous vowels is merely halved. To confirm this result is a significant, negative correlation between years of instruction and number of

Table 9. *Raw counts and relative frequency (with respect to the same type of target sound) of occurrence of errors (consonants, vowels, and insertions) in each year-of-instruction group*

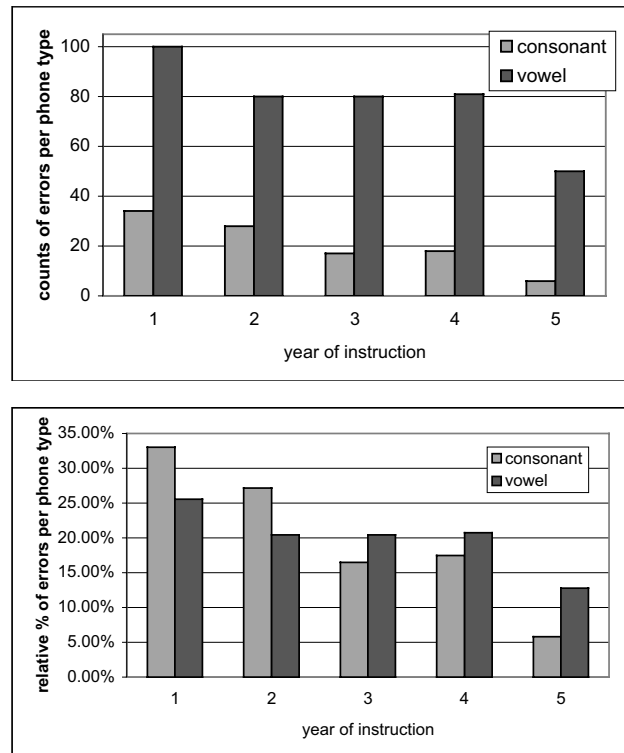| Errors	# years									
	1		2		3		4		5*	
	#	%	#	%	#	%	#	%	#	%
C	34	3	28	2.6	17	1.6	18	1.7	6	0.6
V	100	13.2	80	10.9	80	10.9	81	11.1	50	6.8
INS	39		22		23		16		5	
Total	145		130		120		115		61	

Figure 2. *Counts (top) and percentages (bottom) of vowel and consonant errors for single instruction years. The percentages represent the portion of errors relative to the same type of phoneme.*

consonant errors ($r = -0.75$, $p < 0.01$), based on the data available per speaker. The correlation between years of instruction and number of vowel errors is also significantly negative, but less strong ($r = -0.61$, $p < 0.01$), thus confirming that this type of errors does not decrease as linearly as consonants as a function of increased instruction/exposure to Dutch.[14]

To more clearly discern differences in error behaviour patterns across several years of exposure to the target language, we compiled lists of the most frequent errors in Group A and Group B and subsequently compared them. An initial examination indicates that vowels remain much more problematic than consonants: 211 vowels were mispronounced by Group B, versus only 41 consonants, while the ratio of vowels to consonants in the target set of phonemes was 1,987/3,214 (38 %/62 %).

The most frequent vowel errors, making up for approximately 78 % of the vowels that were mispronounced, are listed in Table 10a (Group A) and Table 10b (Group B); the most frequently mispronounced consonants are listed in Table 11a and Table 11b, while insertions are presented in Tables 12a and 12b.

A comparison of Table 10a and Table 10b clearly shows that the diphthong /ɛi/ is the only phoneme to disappear from the list of most frequently mispronounced vowels of the more proficient group. In fact, /ɛi/ can still be found in the erroneous production of that group (not listed in this paper), but its frequency of occurrence is so low that it can be discarded (in line with criterion 2). The other vowels from the list of Group A are problematic for Group B, too, though with different rankings. In addition, two new vowels appear among the frequent errors of the more proficient group, the /aː/ and the /eː/, which are confused with their short counterparts or with /ɑ/ and /ɛː/-/ɪ/ respectively, as if the more advanced interlingual system were characterized by even more confusion.

Consonants display quite a different trend. Initially many consonants are mispronounced with similar frequencies (see Table 11a). Later on, the number of mispronounced consonants decreases, until only a few instances worth considering remain (see Table 11b). The specific realizations of those phonemes will be examined in the qualitative analysis.

The number of illegal insertions in the more proficient group decreases, though the errors identified are similar in the two groups.

5.3.2. *Annotations: qualitative analysis.*   The results obtained on the basis of the annotations of FL-read show that, despite the limited attention devoted to pronunciation training within the subjects' Dutch curriculum, the Italian speakers in this study tend to produce fewer errors as a function of increased instruction time/exposure to the target language. They also produce more errors in vowels than consonants, a trend that becomes more accentuated as exposure to Dutch increases. These data, which are in concordance with the findings from SL contexts, can be taken as an indication that, unless specific instruction is provided, vowels will remain problematic even after prolonged exposure to Dutch. For these subjects, in particular, this result is by no means surprising since Italian comprises only five to seven vowel phonemes (/a/, /e/, /ɛ/, /i/, /o/, /ɔ/, /u/) depending on the regional variety, and it does not have vowel duration as a distinctive feature, which explains the problems at quantitative and qualitative level found here (Canepari 1985, Nespor 1993, Bertinetto and Loporcaro 2005).[15] With respect to the frequent mispronunciations of /ə/ as full vowel /eː/-/ɛ/, we found that 25 % of the cases may either be due to a failure to apply vowel reduction in the indefinite article *een* in the stimuli, or to orthographic interference, and that 50 % seemed only attributable to orthographic interfer-

Table 10a. *Most frequently mispronounced vowels and corresponding most frequent realizations for the lower proficiency group (Group A)*

TV	# MIS	% of MIS V	Real	# Real
/ə/	36	20	[e]-/eː/	25
			Deleted	8
/ɛi/	27	15	[ei]	27
/œy/	20	11.1	/ʌu/	12
			[ɔɪ]	4
/ɑ/	20	11.1	[a]-/aː/	19
/ʏ/	18	10	/y/	9
			/u/	6
/ɛ/	12	6.7	[e]-/eː/	8
/ɪ/	10	5.6	/i/	10
Total	143	79.4		

Table 10b. *Most frequently mispronounced vowels and corresponding most frequent realizations for the higher proficiency group (Group B)*

TV	# MIS	% of MIS V	Real	# Real
/ə/	32	15.2	[e]-/eː/	12
			Deleted	11
			/ɛ/	5
/aː/	27	12.8	/ɑ/	21
			[a]	5
/ɑ/	23	10.9	[a]-/aː/	21
/ɪ/	20	9.5	/i/	19
/ʏ/	19	9	/y/	9
			/u/	6
/eː/	18	8.5	/ɛ/	8
			[e]	4
			/ɪ/	4
/ɛ/	14	6.6	[ə]	7
/œy/	11	5.2	/ʌu/	4
Total	164	77.7		

Table 11a. *Most frequently mispronounced consonants and corresponding most fre-*
quent realizations for the lower proficiency group (Group A)

TC	# MIS	% of MIS C	Real	# Real
/h/	19	30.6	Deleted	15
			/x/	4
/r/	8	12.9	Deleted	5
/d/	5	8.1	/t/	4
/n/	5	8.1	Deleted	5
/t/	5	8.1	Deleted	5
/ʋ/	5	8.1	/f/	5
/x/	5	8.1	[g]	6
Total	52	83.9		

Table 11b. *Most frequently mispronounced consonants and corresponding most fre-*
quent realizations for the higher proficiency group (Group B)

TC	# MIS	% of MIS C	Real	# Real
/h/	10	24.4	Deleted	5
/d/	9	22	/t/	9
/x/	8	19.5	[g]	6
Total	27	65.9		

ence. It thus seems that the importance of these errors in a final inventory may
need to be adjusted, as posited in Study 1.

By contrast, the problems with consonants decrease dramatically, and the
most proficient learners eventually produce very few errors. This suggests that
consonants deserve lower priority than vowels in pronunciation training, both
in terms of ranking and number of target phonemes. With respect to the type of
errors made, we can again resort to the characteristics of Italian to explain them.
For instance, the frequent deletion of the fricative /h/, which systematically oc-
curs in word-initial, prevocalic position, is most likely attributable to interfer-
ence from the L1 through the orthographic level. In Italian, the grapheme *h* in
that position merely serves a lexical function, without actually corresponding
to any real sound. The substitution of [g] for the fricative /x/ may also result
from orthographical interference, since it occurs only when the target sound is
represented by the grapheme *g*. Nevertheless, this substitution may also arise
from a genuine difficulty producing a novel phoneme not found in Italian. How-
ever, since 5 of these 8 substitutions occur initially in the word *garage*, it also
seems plausible that the students were simply not aware of the correct pro-

Table 12a. *Most frequent illegal insertions and the context in which they occur for the lower proficiency group (Group A)*

INS	# INS	% of INS	Context
/ə/	34	55.7	34 post-consonantal, word final
/t/	8	13.1	5 after ambysyllabic /t/ (no degemination applied)
/b/	4	6.6	4 after ambysyllabic /b/ (no degemination applied)
/x/	4	6.6	4 after /ŋ/
Total	50	82	

Table 12b. *Most frequent illegal insertions and the context in which they occur for the higher proficiency group (Group B)*

INS	# INS	% of INS	Context
/ə/	20	45.5	19 post-consonantal, word final
[g]	11	25	11 after /ŋ/
/t/	10	22.7	8 after ambisyllabic /t/ (no degemination applied)
Total	105	93.2	

nunciation of this word in Dutch and may have transferred the pronunciation from Italian (or from another L2 they spoke such as English or French). In that case, this fricative could even be excluded from our final list of frequently mispronounced consonants.

With respect to illegal insertions, we first notice a sharp drop from 39 to only 5 inserted sounds from the first to the fifth year of instruction, indicating that this problem may disappear almost completely as a result of generic instruction/exposure to the target language. At least 2 of the insertions produced by the more proficient group seem to be attributable to Italian phonology. The schwa in post-consonantal word-final position is most probably due to a tendency in Italian (and in most languages) to prefer open CV syllables. Similarly, the presence of an extra /t/ after an ambisyllabic /t/ seems to be due to a failure to apply the Dutch phonological degemination rule, which is understandable since Italian allows lengthening of consonant duration, i.e., gemination, for such orthographical representations (Bertinetto and Loporcaro 2005, Payne 2005). Finally, the insertion of a velar plosive after /ŋ/ can be explained, once again, by a difficulty in pronouncing this phoneme in word-final or pre-vocalic position since [ŋ] occurs in Italian only as an allophone of /n/ before the velar plosives /k/ or /g/. Another possible source of confusion may lie in the orthography since the Dutch velar nasal is represented by the graphemes *ng*.

5.4. *Conclusions of Study 3*

The purpose of this study was, on the one hand, to obtain an inventory of perceptually salient, frequent, persistent, communication-hampering errors made by FL learners of Dutch, and on the other hand to determine whether and how error behaviour changes as a function of increased exposure to the target language. With regard to the first point, vowels were found to be much more problematic than consonants. The most frequent problems concerning the /ə/ substitutions are likely due to an effect of orthography as seen in the previous studies and should, therefore, receive less priority in a pronunciation-training programme. In general, it often happens that a certain target vowel is replaced by another phoneme or by a similar phone. As a consequence, it may be a good strategy to present these phonemes contrastively in pronunciation training and to point learner attention to both the spectral and durational features that differentiate them. In particular, the following sounds should be considered: /ɑ/-/aː/, /ʏ/-/y,u/, /eː/-/ɛ/, [e], /ɪ/-/i/, and /œy/-/ʌu/ for the diphthongs. With regard to the second point, we identified a general trend indicating that problems with vowels are also resistant to prolonged exposure to the target language. The Dutch diphthongs, perhaps because they are not as numerous as the monophthongs, are either problematic only at the beginning (e.g., /ɛi/) or become less problematic as exposure to the FL increases (see /œy/).

Consonant errors disappear almost entirely from the production of the more proficient group even though the phonemic systems of Italian and Dutch display several differences with respect to consonants. As a matter of fact, only 3 consonants – /h/, /d/, and /x/ – are mispronounced in the production of the students in their fifth year of instruction, yielding a total of 6 errors. The implication of this result for the design of an optimal pronunciation-training programme is that consonants should receive little explicit attention. Of course, this conclusion is based on analyses that were biased toward identifying persistent errors, and on data from relatively experienced and talented language learners.

Similarly, only 5 sounds are "illegally" inserted, thus reducing the importance of this pronunciation problem, too. In this case, a session on degemination and on the realization of the velar nasal may be sufficient to ensure that a large number of illegal insertions do not occur. On the other hand, avoiding schwa epenthesis in post-consonantal, word-final position may be more difficult to automatize for a student even when attention is paid to the prosodic features of the target language, because, in adult learners, the latter are known to be more resistant to readjustments (Hirst and Di Cristo 1998).

6. Discussion and conclusions

Now that separate inventories are available of perceptually salient, frequent, persistent, potentially communication-hampering errors made by different learners of Dutch for different types of speech and learning contexts (see Table 13), it is possible to compare these inventories to compile a final list of errors to address in pronunciation training. Specifically, we wish to merge the results for three groups: read speech (Study 1, SL-read) and extemporaneous speech (Study 2, SL-xtmp) by L2 learners of Dutch, and read speech (Study 3, FL-read, Group B) by FL learners of Dutch. In consideration of the specificity of the speakers of FL-read, who were all Italian and fairly proficient speakers, we will first combine the former two groups, and subsequently compare the new inventory with the conclusions we drew for FL-read.

Table 13. *Summative table of results from three studies. T Ph stands for Target Phoneme. Phonemes between brackets were either only problematic for certain L1s or were likely an artefact of the elicitation task. Del indicates a deletion.*

Study 1 – SL-read		Study 2 – SL-xtmp		Study 3 – FL-read	
T Ph	realized as	T Ph	realized as	T Ph	realized as
/ɑ/	[a], /aː/	/aː/	/ɑ/, [a], [ɑː]	/ɑ/	[a], /aː/
/aː/	/ɑ/, [ɑː]	/ɑ/	[a], /aː/	/aː/	/ɑ/, [a]
/ɤ/	/u/, /y/	/y/	/u/	/ɤ/	/y/, /u/
/y/	/u/	/ɤ/	/u/	/eː/	/ɛ/, /ɪ/, [e]
/øː/	/u/, /oː/, /y/	/ɔ/	[o], /oː/, /ɔː/	/ɛ/	[ə]
/ə/	Del, [e], /eː/, /ɛ/	/ɛ/	[e]	/ɪ/	/i/
		/i/	/ɪ/	(/ə/)	[e], /eː/. /ɛ/, Del
/œy/	/ɑu/, [ɔi]	/œy/	/ɑu/	/œy/	/ɑu/
/ɛi/	[ei], [ɑi], [ai]	/ɛi/	[ei], [e], /eː/		
/x/	[g], /h/, /k/	/x/	/k/, /h/, [g]	/h/	Del
/t/	Del, /d/			/d/	/t/
(/r/)	Del, [ɐ], /l/			/x/	[g]

INS	context	INS	context	INS	context
/ə/	Post-consonantal	/ə/	post-consonantal	/ə/	post-consonantal
*[g]	after /ŋ/	[g]	after /ŋ/	[g]	after /ŋ/
/h/	in het			/t/	
*/x/	after /ŋ/				
*/k/	after /ŋ/				

6.1. *SL inventories*

A comparison of SL-read (Tables 3a/b) and SL-xtmp (Tables 6a/b) immediately reveals the existence of numerous common trends and very few differences. Vowels remain the most problematic group of phonemes. They tend to be confused and to display extremely similar realizations in both databases. They also display similar rankings in terms of the frequency of occurrence. If we consider the differences, we first observe that the schwa, the most problematic vowel in the read-speech annotations, does not appear among the 9 most frequently mispronounced vowels in SL-xtmp. This lends support to the hypothesis of the "orthography-induced artefact", whereby the problems found in SL-read for /ə/ may be largely due to orthographic interference, thus suggesting that schwa can be ranked among minor errors or even be eliminated from our list. Another difference is represented by the absence of /øː/ from the vowel errors of SL-xtmp and the slightly different frequency rank of the vowel errors in the two databases. These minor discrepancies may be due to the different speech type and content, and thus different phonemic make-up, of the elicitation material in the two databases. The vowel /øː/ illustrates this point well: in the database used to verify the completeness of SL-xtmp (Oostdijk 2002), this phoneme represents 0.4 % of the 1.4 million vowels; in SL-xtmp it represents 0.1 %; and in SL-read, 2.3 %. As explained in Section 3.3.1, this is a direct consequence of the phonetically rich, read sentences used here.

Having made these considerations, if we now merge the two lists, retaining only those sounds which appear in both lists and taking into account the rankings and the realizations of the individual sounds in each list, we obtain the following ranking of problematic vowels: /ɑ/, /œy/. /aː/, /y/, /ɛi/, /ʏ/. In terms of realizations, /ɑ/ and /aː/ are generally confused and should therefore be tackled together, contrastively. A similar method could be used for the phonemes /y/ and /ʏ/, which often result in /u/, and for the diphthong /œy/ which in most cases is confused with /ʌu/.

Comparison of the consonants reveals that the uvular/velar fricative /x/ and the glottal fricative /h/ occur in both lists of frequent errors, displaying exactly the same erroneous realizations. Moreover, the two phonemes are often confused with each other, with the glottal fricative being sometimes realized as the velar/uvular fricative, possibly as a result of hypercorrection towards the latter. Examining the contexts in which /s/ is mispronounced (Study 2), we observe that many [ʃ] realizations are in fact substitutions for the word-initial consonant cluster [sx], graphically represented as *sch*. This may indicate, once more, a difficulty with the fricative /x/, rather than with the /s/, or at any rate with the articulation of this sequence of sounds. By contrast, the other errors found in Study 1, whose importance in a pronunciation-training programme already appeared questionable, do not appear in Study 2. This suggests that these errors

may be less important for more spontaneous speech contexts than first thought. Therefore, if we collapse the two inventories of consonant errors, we retain only the fricatives /x/ and /h/, realized as /k/-[ɡ]-/h/ and /x/ respectively.

Finally, a review of the illegal insertions in SL-xtmp again reveals strong commonalities with SL-read with regard to schwa and the velar stop [ɡ] after the velar nasal /ŋ/. Other insertions found in Study 1, namely that of /k/ and /x/ after the velar nasal, do not appear in SL-xtmp, indicating that they may have been due to orthographical factors, as hypothesized. In this case then, we are left with the schwa and the [ɡ], which occur in the same contexts in both databases and point to a problem with the Dutch syllabic structure and the velar nasal /ŋ/ respectively.

6.2. *SL+FL inventories*

Comparing this inventory with the data obtained from FL-read, and in particular with the data on persistent errors made by the more proficient speakers (in Group B), we notice a considerable degree of agreement. Study 3 confirms that Dutch vowels are much more problematic than consonants, and that they remain so despite prolonged exposure to Dutch. The inventory of Study 3 also displays vowel and consonant errors (with realizations) that are very similar to those just selected. The insertion of the velar stop [ɡ] after the velar nasal /ŋ/ is also found in this database. On the basis of this final comparison, we can propose that the inventory of errors in Table 14 be given priority in a pronunciation training programme.

While more data from L1s other than those considered in these studies are needed to fully validate the robustness of this inventory, we can reasonably conclude that the list presented in Table 14 can provide both teachers and CAPT developers with a sound basis for designing efficient pronunciation training for L2 learners of Dutch with different L1s. Given the conservative nature of

Table 14. *The final inventory*

T Ph	realized	INS	CONTEXT
/ɑ/	/aː/	/ə/	post-consonantal + word-/syllable final; within consonant clusters
/aː/	/ɑ/		
/y/	/u/		
/ɣ/	/u/		
/œy/	/ʌu/	[ɡ]	after /ŋ/
/ɛi/	[ei]		
/x/	/k/, [ɡ], /h/		
/h/	/x/		

this inventory, along with the high degree of similarity evidenced in the three studies, we also expect it to be relevant for a large number of FL learners, although it may not function as efficiently as an L1-specific version. We believe that the procedure adopted to build this inventory can be successfully applied to other languages, thus benefiting the larger SL teaching and learning community. However, our inventory should be seen as an indication. It should not preclude the language teachers from adopting a different ranking of errors on the basis of specific needs. (For instance, given a large number of Japanese and/or Chinese and/or Korean learners within one class, a teacher could decide to include /r/ as target consonant even though we omitted it from our final list). Moreover, this inventory ought to be integrated with information on supra-segmental errors because, as we indicated, supra-segmentals have been shown to facilitate communication and because they interact with segmentals. Finally, the pedagogical value of this inventory is obviously dependent on the method within which it is applied. For the purpose of our research, we have developed a CAPT system that provides meaningful and engaging multimedia input in Dutch, opportunities for oral practice, and feedback on segmental quality based on automatic speech recognition technology. This CAPT system only provides feedback on the 11 problematic phonemes. An experiment conducted with immigrants learning Dutch has shown that the group receiving automatic feedback by means of this CAPT system made significantly larger improvements in segmental quality for those 11 phonemes than the group of learners who did not receive feeback (see Neri et al. 2006).

Radboud University Nijmegen
<A.Neri@let.ru.nl>
<C.Cucchiarini@let.ru.nl>
<H.Strik@let.ru.nl>

Note

* The present research was supported by the Dutch Organization for Scientific Research (NWO). We would like to thank our colleagues at the Department of Linguistics who helped with the annotations. We would also like to thank A. M. Janssen-van Dieten, T. Bongaerts, I. van de Craats, W. H. Fletcher, and two anonymous reviewers for their comments on earlier versions of this paper.
1. Throughout this paper the abbreviation L2 is used to refer to second language in general, irrespective of whether the L2 is taught in the country where the L2 is spoken or outside the country. Whenever it is necessary to make a distinction between these two contexts, the abbreviations SL (second language) and FL (foreign language) are used.
2. Moyer (1999) and Bongaerts (1999, 2001) only provided indirect evidence of the effect of specific training on pronunciation.

3. These factors often interact with other factors involved in communication in complex ways. For instance, allophonic errors seem to be more disruptive for communication in the case of listener familiarity with the talker (Smith 2005). A very slow non-native speaking rate is often found to be highly accented, possibly because it enhances error saliency, and is considered tedious because it slows the speech processing flow excessively. However, accelerating non-native speech can reduce the saliency of individual errors, but it also eliminates perceptually salient features, hampering the intelligibility of the whole utterance (Munro and Derwing 2001).

4. Using single words would not yield information on errors occurring at word boundaries. Here, coarticulation phenomena occur, which carry important information and which, at the same time, are obviously difficult for L2 learners.

5. For more exhaustive descriptions of the phonetic-phonological system of the Dutch language, see Gussenhoven (1992), Booij (1999), and Rietveld and Van Heuven (1997).

6. Only rarely does *ng* correspond to an [ŋx] sequence (Booij 1999: 80).

7. Although for most of these languages the lack of errors with the velar nasal might have to do with its presence in those languages in the same context as in Dutch or even in more marked position (e.g., Cantonese has syllable-initial /ŋ/), problems with the realization of /ŋ/ were found among speakers of languages such as German which have /ŋ/ in similar contexts and with the same orthography with which it is represented in Dutch.

8. A study by Young-Scholten (1997) found a strong influence of orthography on the pronunciation of segments in a second language test. In the same study, the author also evidenced an effect of orthography when learning: the orthographic representation of a word is more or less stored in the learner's mind as a result of learning through the written medium, especially in the case of literate adult learners. Consequently, even in spontaneous speech, spelling pronunciation errors may not disappear entirely.

9. In this study, which addressed English as spoken by Chinese and Portuguese speakers, Tarone did discern a tendency towards a CV syllable structure, but she found that this structure was achieved more often by deleting a consonant than by inserting a vowel.

10. Dutch phonological rules do allow /k/ insertions between the velar nasal /ŋ/ and the stop /t/ in words like *zingt* /zɪŋt [zɪŋkt] (Booij 1999: 137), but no such cases were listed among the errors in our database.

11. Lin (2001) suggests that epenthesis may occur more often in formal situations (e.g., reading word lists), which require more attention to form than to content in order to make target consonants more audible. On the other hand, deletion (and replacement) should more frequently occur in less formal tasks where more attention is paid to content and where the loss originating from the deletions would be less serious because of the presence of linguistic context (e.g., spontaneous speech).

12. This claim is not based on results stemming from specific tests. It is a logical assumption based on the fact that a) these students had chosen to study at this particular university, b) they had passed the entry exam, c) they were studying Dutch, although it was not a compulsory subject.

13. This elicitation technique is frequently used to study characteristics and differences in spoken production because it allows elicitation of speech that is semi-spontaneous, but comparable across speakers. For this study, however, this elicitation technique turned out to be inappropriate. First of all, students in an FL context are not always familiar with names of everyday objects in the FL, nor are they used to describing object positions in the FL. Moreover, in the case of the SSLMIT students, teaching was geared towards abstract terminology which the students might need as future interpreters and translators, which is difficult to elicit by means of pictures.

14. Although instruction and mere exposure to the Dutch language are two rather different factors, as pointed out by a reviewer, for the purpose of this study we treat them as one factor because exposure to Dutch was practically limited to the lessons at the university (during which very little attention was devoted to pronunciation), while the time spent in a Dutch speaking country by each student was very limited and it increased parallel to the years of instruction (see Table 8).

15. In addition, these results seem to fit in Best's Perceptual Assimilation Model (Best 1995) and in Flege's Speech Learning Model (Flege 1995) models of L2 speech *perception*. According to these models, new L2 sounds that are similar to L1 phonemes will be initially assimilated by the L1 category of the latter. As the learner becomes more proficient, this category will be gradually split into two different ones, possibly by relying on the correct L2 contrast. Here the less proficient group only has problems realizing /ɑ/ correctly, i.e., they probably map the new (and similar) /ɑ/ to the Italian (and Dutch) /aː/, (thereby, collapsing two Dutch phonemes into a single category); as more input is available, they start to create a new, distinct phonetic category for /aː/, but the boundaries between these two categories are still unstable.

References

Aan de Wiel, Margreet C., Gertrude A. van den Brink, and Saskia G.E. Struijk van Bergen (1991). Diagnostiek van uitspraakproblemen van tweede-taalverwervers van het Nederlands. *Levende Talen* 466: 507–511.

Abercrombie, David (1991 [1949]). Teaching pronunciation. In *Teaching English Pronunciation: A Book of Readings,* Adam Brown (ed.), 87–95. Routledge: London.

Akahane-Yamada, Reiko, Erik McDermott, Takahiro Adachi, Hideki Kawahara, and John S. Pruitt (1998). Computer-based second language production training by using spectrographic representation and HMM-based speech recognition scores. *Proceedings of the 5th International Conference on Spoken Language Processing, Sydney, Australia, 1998* [CD-Rom] Paper 0429.

Anderson-Hsieh, Janet, Ruth Johnson, and Kenneth Koehler (1992). The relationship between native speaker judgments of non-native pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning* 42: 529–555.

Anderson-Hsieh, Janet and Kenneth Koehler (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning* 38: 561–613.

Bassetti, Benedetta (2006). Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language and Literacy* 9: 95–114.

Bertinetto, Pier Marco and Michele Loporcaro (2005). The sound pattern of Standard Italian as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association* 35: 131-151.

Best, Catherine T. (1995). A direct realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research,* Winifred Strange (ed.), 171–206. Timonium, MD: York Press Inc.

Binnenpoorte, Diana M. (2006). Phonetic transcriptions of large speech corpora. Doctoral dissertation, Radboud University Nijmegen, The Netherlands.

Bongaerts, Theo (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In *The Critical period Hypothesis and Second language Acquisition*, David Birdsong (ed.), 133–159. Mahwah, NJ: Lawrence Erlbaum.

— (2001). Age-related differences in the acquisition of L2 pronunciation: The Critical Period Hypothesis revisited. Paper presented at the EUROSLA preconference workshop on the Age Factor in L2 acquisition, Paderborn, Germany.

Booij, Geert (1999). *The Phonology of Dutch.* Oxford: Clarendon Press.

Broselow, Ellen (1983). Nonobvious transfer: On predicting epenthesis error. In *Language Transfer in Language Learning*, Larry Selinker (ed.), 269–280. Rowley, MA: Newbury House.

Canepari, Luciano (1985). *Introduzione alla fonetica*. Torino, Italy: Einaudi.

Carlisle, Robert S. (1991). The influence of environment on vowel epenthesis in Spanish/English interphonology. *Applied Linguistics* 12: 76–95.

Celce-Murcia, Marianne, Donna M. Brinton, and Janet M. Goodwin (1996). *Teaching Pronunciation.* Cambridge: Cambridge University Press.

Coenen, Josée (1991). *Uitgesproken Nederlands*. Groningen: Wolters-Noordhoff.

Crothers, John (1978). Typology and universals of vowel systems. In *Universals of Human Language. Vol. 2 Phonology*, Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.), 93–152. Stanford: Stanford University Press.

Cucchiarini, Catia (1996). Assessing transcription agreement: Methodological aspects. *Clinical Linguistics and Phonetics* 102: 131–155.

Cucchiarini, Catia, Helmer Strik and Louis Boves (2000a). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America* 107: 989–999.

— (2000b). Different aspects of pronunciation quality ratings and their relation to scores produced by speech recognition algorithms. *Speech Communication* 30: 109–119.

— (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America* 111: 2862–2873.

Cutler, Anne (1984). Stress and accent in language production and understanding. In *Intonation, Accent and Rhythm. Studies in Discourse Phonology,* Dafydd Gibbon and Helmut Richter (eds.), 77–90. Berlin/New York: de Gruyter.

Cutler, Ann and Sally Butterfield (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language* 31: 218–236.

Davidson, Lisa, Paul Smolensky and Peter W. Jusczyk (2004). The initial and final states: Theoretical implications and experimental explorations of richness of the base. In *Constraints in Phonological Acquisition,* René Kager, Joe Pater and Wim Zonneveld (eds.), 321–368. Cambridge: Cambridge University Press.

De Graaf, Tjeerd (1986). De uitspraak van het Nederlands door buitenlanders. *Logopedie en Foniatrie* 58: 343–347.

Derwing, Tracey M. and Murray J. Munro (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition* 19: 1–16.

Derwing, Tracey M., Murray J. Munro and Grace Wiebe (1998). Evidence in favour of a broad framework for pronunciation instruction. *Language Learning* 48: 393–410.

Dickerson, Lonna J. (1975). The learner's interlanguage as a system of variable rules. *TESOL Quarterly* 9: 401–407.

Doeleman, Rianne (1998). *Native Reactions to Nonnative Speech*. Tilburg: Tilburg University Press.

Eckman, Fred (1977). Markedness and the Contrastive Analysis Hypothesis. *Language Learning* 27: 315–330.

— (1981). On the naturalness of interlanguage phonological rules. *Language Learning* 31: 195–216.

Eisenstein, Miriam (1983). Native reactions to nonnative speech: A review of empirical research. *Studies in Second Language Acquisition* 13: 23–41.

Erdener Doğu V. and Denis K. Burnham (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning* 55: 191–228.

Escudero, Paola and Paul Boersma (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26: 551–585.

Fayer, Joan M. and Emily Krasinski (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning* 37: 313–326.

Flege, James E. (1988). Using visual information to train foreign-language vowel production. *Language Learning* 38: 365–407.

— (1995). Second-language speech learning: Findings and problems. In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research,* Winifred Strange (ed.), 233–273. Timonium, MD: York Press Inc.

— (1999). Age of learning and second language speech. In *Second Language Acquisition and the Critical Period Hypothesis,* David Birdsong (ed.), 101–131. Mahwah, NJ: Lawrence Erlbaum Associates.

Flege, James E. and Ocke-Schwen Bohn (1989). An instrumental study of vowel reduction and stress placement in Spanish accented English. *Studies in Second Language Acquisition* 11: 35–62.

Fletcher, Kathryn L. (1983). New prospects for speech by the hearing impaired. In *Speech and Language: Advances in Basic research and Practice 9,* Norman Lass (ed.), 1–41. New York: Academic Press.

Goeman, Ton (1999). T-deletie in Nederlandse Dialecten. Kwantitatieve analyse van structurele, ruimtelijke en temporele variatie. Doctoral dissertation, The Hague: HAG.

Greenberg, Joseph H. (1978). Some generalizations concerning initial and final consonant clusters. In *Universals of Human Language. Vol. 2 Phonology*, Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.), 243–279. Stanford: Stanford University Press.

Gussenhoven, Carlos (1999). Dutch. In *Handbook of the International Phonetic Association, Part II, Illustrations of the IPA*, 74–77. Cambridge: Cambridge University Press.

Hancin-Bhatt, Barbara and Rakesh M. Bhatt (1998). Optimal L2 syllables: interactions of transfer and developmental effects. *Studies in Second Language Acquisition* 19: 331–378.

Hardison, Debra M. (1996). Bimodal speech perception by native and nonnative speakers of English: Factors influencing the McGurk Effect. *Language Learning* 46: 3-73.

— (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology* 8: 34–52, Retrieved July 25, 2005, from http://llt.msu.edu/vol8num1/hardison.

Harless, William G., Marcia A. Zier and Robert C. Duncan (1999). Virtual Dialogues with Native Speakers: The Evaluation of an Interactive Multimedia Method. *CALICO Journal* 16: 313–337.

Hawkins, Sara (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics* 31: 373–405.

Hirst, Daniel and Albert Di Cristo (1998). A survey of intonation systems. In *Intonation Systems: A Survey of Twenty Languages,* Daniel Hirst and Albert Di Cristo (eds.), 1–44. Cambridge: Cambridge University Press.

Horwitz, Elaine K., Michael B. Horwitz and Jo Ann Cope (1986). Foreign language classroom anxiety. *The Modern Language Journal* 70: 125–132.

Jusczyk, Peter W. and Paul A. Luce (2002). Speech perception and spoken word recognition: Past and present. *Ear and Hearing* 23: 2–40.

Kenworthy, Joanne (1987). *Teaching English Pronunciation*, Longman Handbooks for Language Teachers. Harlow, UK: Longman.

Kuhl, Patricia K. and Paul Iverson (1995). Linguistic experience and the "perceptual magnet effect". In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research,* Winifred Strange (ed.), 121–154. Timonium, MD: York Press Inc.

Kuijpers, Cecile, Wilma van Donselaar and Anne Cutler (1996). Phonological variation: Epenthesis and deletion of schwa in Dutch. In *Proceedings of ICSLP 1996,* Vol. 1, 149–152. Philadelpia, USA.

Leather, James (1998). Second-Language speech research: An introduction. *Language Learning* 49, Supplement 1: 1–56.

Lin, Yuh-Huey (2001). Syllable simplification strategies – a stylistic perspective. *Language Learning* 51: 681–718.

Lindblöm, Björn (1986). Phonetic universals in vowel systems. In *Experimental Phonology,* John J. Ohala, and Jeri J. Jaeger (eds.), 13–44, Orlando, FL: Academic Press.

Lively, Scott E., John S. Logan and David B. Pisoni (1993). Training Japanese listeners identify /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94: 1242–1255.

MacIntyre, Peter D. and Robert C. Gardner (1991). Language anxiety: Its relationship to other anxieties and to processing in native and second languages. *Language Learning* 41: 513–534.

— (1994). The effects of induced anxiety on three stages of cognitive processing in computerized vocabulary learning. *Studies in Second Language Acquisition* 16: 1–18.

Maddieson, Ian (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Major, Roy C. (1987). A model for interlanguage phonology. In *Interlanguage Phonology: The Acquisition of a Second Language Sound System*, Georgette Ioup and Steven H. Weinberger (eds.), 101–124. New York: Newbury House/Harper and Row.

— (2001). *Foreign Accent. The Ontogeny and Phylogeny of Second Language Phonology*. London: Lawrence Erlbaum Associates.

McCandliss, Bruce D., Julie A. Fiez, Athanassios Protopapas, Mary Conway, and James L. Mc-Clelland (2002). Success and failure in teaching the [r]–[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, and Behavioral Neuroscience* 2: 89–108.

Morley, Joan (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly* 25: 481–519.

Moyer, Alene (1999). Ultimate attainment in L2 phonology. The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition* 21: 81–108.

Munro, Murray J. and Tracey M. Derwing (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45: 73–97.

— (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech. The role of speaking rate. *Studies in Second Language Acquisition* 23: 451–468.

Neri, Ambra, Catia Cucchiarini and Helmer Strik (2006). ASR-based corrective feedback on pronunciation: Does it really work? In *Proceedings of Interspeech 2006 – ICSLP,* 1982–1985. Pittsburgh, PA, USA.

Nespor, Marina (1993). *Fonologia*. Bologna: Il Mulino.

Onwuegbuzie, Anthony J., Phillip Bailey and Christine E. Daley (1999). Factors associated with foreign language anxiety. *Applied Psycholinguistics* 20: 220–239.

Oostdijk, Nelleke H. J. (2002). The design of the Spoken Dutch Corpus. In *New Frontiers of Corpus Research*, Pam Peters, Peter Collins and Adam Smith (eds.), 105–112. Amsterdam: Rodopi.

Payne, Elinor M. (2005). Phonetic variation in Italian consonant geminates, *Journal of the International Phonetic Association* 35: 153–182.

Piske, Thorsten, Ian R. A. MacKay and James E. Flege (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215.

Polka, Linda and Janet F. Werker (1994). Developmental changes in the perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20: 421–435.

Richards, Jack C. and Theodore S. Rodgers (1986). *Approaches and Methods in Language Teaching: A Description and Analysis*. Cambridge: Cambridge University Press.

Rietveld, Toni and Vincent J. Van Heuven (1997). *Algemene Fonetiek*. Bussum: Coutinho.

Rogers Catherine L. and Jonathan M. Dalby (1996). Prediction of foreign-accented speech intelligibility from segmental contrast measures. *Journal of the Acoustical Society of America* 100: 2725 (A).

Savignon, Sandra J. (1991). Communicative Language Teaching: State of the art. *TESOL Quarterly* 25: 261-277.

Schachter, Jacqueline (1974). An error in error analysis. *Language Learning* 24: 205–214.

Shatzman, Keren B. (2006). Sensitivity to detailed acoustic information in word recognition. Doctoral Dissertation, University of Nijmegen (MPI Series in Psycholinguistics, 37), The Netherlands.

Shriberg, Lawrence D., Joan Kwiatkowski and Kit A. Hoffmann (1984). A procedure for phonetic transcription by consensus. *Journal of Speech and Hearing Research* 27: 456–465.

Smith, Rachel (2005). The role of fine phonetic detail in word segmentation. Unpublished Doctoral Dissertation, University of Cambridge, UK.

Strange, Winifred (1995). Cross-language studies of speech perception and production in second-language learning. In *Speech Perception and Linguistic Experience,* Winifred Strange (ed.), 3–45. Timonium, MD: York Press Inc.

Swerts Marc, Hanne Kloots, Steven Gillis and Georges de Schutters (2001). Factors affecting schwa-insertion in final consonant clusters in standard Dutch. In *Proceedings Eurospeech 2001*, 75–78. Aalborg, Denmark, September 2001.

Tarone, Elaine (1980). Some influences on the syllable structure of interlanguage phonology. *International Review of Applied Linguistics* 4: 143–163.

Tupker, Yvonne (1995). Uitspraak Onderwijzen. Manuscript. Tilburg: Katholieke Universiteit Brabant.

Van Alphen, Petra M. (2004). Perceptual relevance of prevoicing in Dutch. Doctoral Dissertation, University of Nijmegen (MPI Series in Psycholinguistics, 25), The Netherlands.

Van Boeschoten, Joost A. (1989). *Verstaanbaarheid van klanken in het Nederlands gesproken door Turken*. Dordrecht: ICG Printing.

Van de Velde, Hans and Roeland van Hout (eds.) (2001). *'r-atics. Sociolinguistic, phonetic and phonological characteristics of /r/*. Brussels: Etudes et Travaux, ILVP-ULB.

Van Heuven, Vincent J. and Jan W. de Vries (1983). Verstaan, begrijpen en waarderen van buitenlandse uitspraak. *Tijdschrift voor Tekst- en Taalkunde* 3: 180–191.

Van Son, Rob, Diana Binnenpoorte, Henk van den Heuvel, and Louis Pols (2001). The IFA Corpus: A phonemically segmented Dutch "Open Source" Speech Database, *Proceedings EUROSPEECH 2001*, Vol. 3: 2051–2054. Aalborg, Denmark.

Van Wijngaarden, Sander J. (2003). The intelligibility of non-native speech. Doctoral dissertation, Free University, Amsterdam.

Vieregge, Wilhelm H. and Ton Broeders (1993). Intra- en interspeaker variation of /r/ in Dutch. In *Proceedings of Eurospeech '93*, 267–270. Berlin.

Vieregge, Wilhelm H., Toni Rietveld, and Carel I. E. Jansen (1984). A distinctive feature based system for the evaluation of segmental transcription in Dutch. In *Proceedings of the Tenth International Congress of Phonetic Sciences,* Marcel P. R. van den Broecke and Andrew D. Cohen (eds.), 654-659. Dordrecht: Foris Publications.

Weinberger, Steven H. (1987). The influence of linguistic context on syllable simplification. In *Interlanguage Phonology: The Acquisition of a Second Language Sound System*, Georgette Ioup and Steven H. Weinberger (eds.), 401–417. New York: Newbury House/Harper and Row.

Wells, John (2004). SAMPA. Computer Readable Phonetic Alphabet. Retrieved August 15, 2004, from http://www.phon.ucl.ac.uk/home/sampa

Werker, Janet F. and Richard C. Tees (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7: 49–63.

Young-Scholten, Martha (1997). Second-language syllable simplification: Deviant development or deviant input? In *New Sounds 97: Proceedings of the Third Symposium on the acquisition of Second-Language Speech*, James Leather and Allan James (eds.), 351–360. Klagenfurt: University of Klagenfurt.