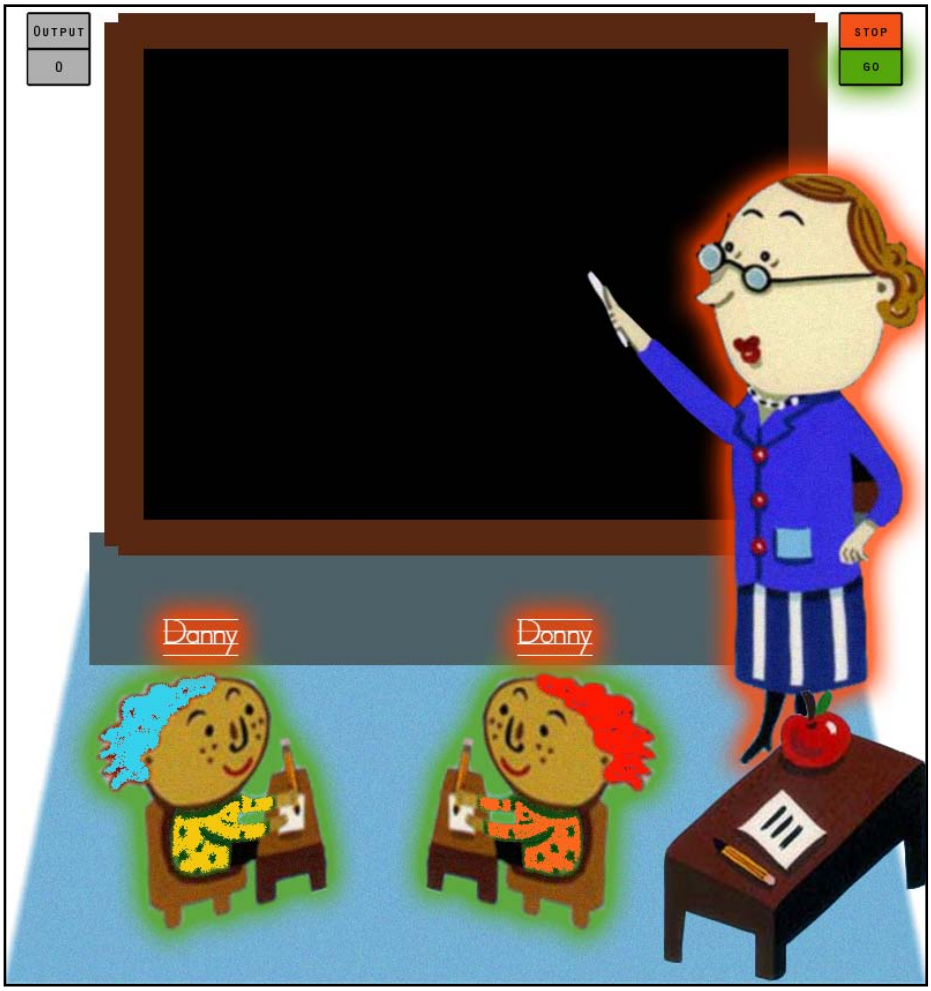
The effect of L2 input quality & quantity on early English L2 acquisition



Jan Willem Chevalking
(student number 0444928)

Master Thesis "Linguistics: The Study of the Language Faculty"
Utrecht University

Supervisor: Dr. Sharon Unsworth, Utrecht University
Independent reader: Dr. René Kager

April, 2010

Contents

1. Abstract

The introduction of English L2 in primary schools is currently expanding rapidly in The Netherlands. The current study investigates factors affecting L2 English acquisition in formal learning contexts such as the Dutch primary school classroom, focusing on acquisition two phonemes, one of which is also found in Dutch phonology ([i]), and another that is new to early English L2 learners ([æ]). Assessment of receptive vocabulary development is also carried out, using a native English Peabody Picture Vocabulary Test 4 (PPVT4), to determine whether or not pupils (age 4-8 yrs.) show improvement across a three-month period with a minimum of 5 and a maximum of 12 hours of English L2 input during that period. PPVT-4 results suggest an effect of age and amount of L2 contact, where higher age groups show better performance and a greater number of hours of L2 input also yields higher PPVT-4 scores. The phoneme discrimination experiments show more accurate discrimination of the /æ/ phoneme after 5 to 12 hours of exposure than before, while no pattern was discerned for the /i/ phoneme. The phoneme discrimination experiments suggest that early English L2 learners have difficulty in discerning small acoustic differences. In this view, the findings recorded here are supported by findings from other studies (Heeren 2006, Gerrits 2001, de Brasileiro 2009, Werker & Logan 1985).

2. Introduction

During my work as a secondary school English teacher at a VMBO school in the Netherlands, I questioned the effectiveness of formal language teaching in which teachers (including myself) guide pupils through a method containing 'real-life, communicative' language learning situations. The pupils I taught were being trained as car-mechanics, welders, carpenters, cooks, salespersons and electricians. It seemed to me that if an English teacher were actually a car-mechanic, he/she would be in the perfect position to present a genuine ‘real-life’ language learning opportunity.

To test this idea, I took two disassembled mopeds into my classroom, instructing students to find out the English term for each part using illustrations which - in turn - would help them interpret the instructions to reassemble a given moped part. Pupils worked enthusiastically on this project, and on testing their vocabulary two weeks after the project it turned out that retention (usually at a class average of 60%) was greater (75%) than would be the case when using a method and complementary wordlist. Unfortunately, the project could not be continued as the curriculum did not allow for side-stepping the method.

Following this, I also tested how second-year pupils would fare on an English VMBO examination paper to establish a measure of my task as a third and fourth-year VMBO teacher. The result showed no significant deviance from the national average for the English VMBO examination (i.e.. class average and national examination values did not differ significantly from the national measure). This strengthened my resolve to investigate factors affecting second language (henceforth L2) acquisition in relation to language teaching, as I had come to believe that the Dutch educational system is regularly swayed by, what seem to be politically - though not always linguistically and/or empirically motivated practise.

One such recent - and particularly interesting - development relates to early English L2 acquisition and English L2 teaching in Primary Schools. This study provides an overview of English in Primary Schools in The Netherlands, followed by a review of studies concerned with English L2 acquisition in schools across Europe and The Netherlands. This leads to a summary of literature regarding factors affecting L2 acquisition in schools, followed by the research design and hypotheses tested in the current study. Results are presented in the prefinal chapter, and these are discussed in chapter 7.

3. Factors affecting L2 acquisition

As an English teacher-trainer working at the Christelijke Hogeschool Ede, I usually start the "Teaching English in Primary School" course with the following set of statements related to English second language acquisition:

- An early start with second language learning leads to higher proficiency levels in learners;
- Young children (age 4-6) are more capable of learning a second language than older children (age 10-12);
- A native English speaker will have a greater learning effect on children than a non-native speaker;
- More English language input leads to higher proficiency levels in learners.

Students are fairly unanimous in claiming that starting early is probably a good idea simply because of the fact that young children are better (second) language learners anyway. More input - it is thought - will no doubt lead to better results, while having a native speaker as an English teacher is not thought to be better by definition. Discussing these statements generally leads to students making more fine-grained distinctions regarding English second language acquisition in primary schools. This process of bold claims being replaced by less clear-cut statements is reflected in the field of foreign language acquisition and it is without doubt a consequence of the number of factors involved in L2 acquisition and the interplay between factors affecting L2 acquisition that complicates the discussion. The current chapter provides an overview of factors affecting L2 acquisition as a background against which the study as described in chapter 6 is set.

Since the early nineties, numerous studies (See Herder & de Bot 2005, Doyé & Hurrell 1997, Blondin et al. 1998 for an overview) regarding foreign language learning in Europe have been conducted i. to draw comparisons between early and late foreign language learners, ii. to compare set standards to actual achievement, iii. to contrast various teaching methodologies, and iv. to gain insight into both teaching and learning processes. In an investigation pooling data from various European studies, Blondin et al. (1998) conclude that "early language learning can have a very positive effect on pupils in terms of language skills, positive attitudes to other languages and cultures and self-confidence. But an early start to foreign language learning is by no means a guarantee for better results than a later start". This conclusion by Blondin et. al. (1998) is based on studies investigating several factors affecting (early) foreign language learning namely, i. linguistic competence, ii. motivation/ attitude and iii. social & cultural awareness. These factors are elaborated by Edelenbos et. al. (2006) who illustrate that language learning may depend on a range of factors as categorised below:

Social factors:	- exposure
	- national & regional infrastructure
	- parents' involvement
School factors:	- continuity
	- time (frequency, intensity)
Teacher factors:	- command of language
	- knowledge of L2 learning methodology/approaches
Learner factors:	- ability
	- socio-economic background
	- starter age

It is clear that these factors are difficult to single out as some of them are interrelated. Johnstone (2003) also reviews a number of studies published in 2002 which illustrate this point: The time factor may (in part) determine the methodology applied to second language education; or a teacher's approach may or may not address a particular learner's learning style which could lead to different learning outcomes. The use of a common first language in teaching a second language (code-switching) could affect language learning outcomes (Turnbull & Arnett, 2002), and the exposure to limited vocabulary could lead to unexpected learner language 'gaps' (Cameron, 2002). These examples - and the list could be extended considerably - illustrate that awareness of the interaction between these factors is key in carrying out applied research in the field of foreign language education. To limit the scope of the current study, the following factors will be taken into account:

- starter age;
- command of language (henceforth) teacher proficiency;
- amount of L2 instruction time
    - frequency;
    - intensity;
    - continuity;
- methodology.

Including the ability and socio-economic background and the social factors mentioned in the previous paragraph would lead to a volume of data beyond the scope of a Master's thesis, so a choice was made to involve schools across which these factors were most likely to be (fairly) constant. Chapter six provides details concerning the subjects involved in this study, suffice it to say that all subjects (n=) are pupils in Protestant Christian schools across the central region of the Netherlands in areas with a similar socio-economic status (RiVM, 2002).

3.1 Starter age

Lenneberg (1967) postulated the concept of a time frame for optimal language acquisition

ranging from two years of age up to puberty. It was hypothesised that a decline in neural plasticity (and completed lateralisation) would lead to incomplete language acquisition if a second language was to be learned after the passing of the critical timeframe. To this day, there is ongoing debate concerning the effect of age on L2 acquisition, and there are numerous studies claiming that there are maturational constraints on L2 acquisition (Johnson & Newport 1989, Hyltenstam 1992, Hyltenstam & Abrahamsson 2003). It is generally accepted that age effects do exist in L2 acquisition, but there is disagreement regarding the age effects being consistent with a critical period. To this end, Singleton (2003) states that the age factor actually comprises a range of age-related factors and that L2 age effects cannot be solely due to neurological predetermination. Other factors play a role in L2 acquisition: Motivation is shown to play a significant role in language learning in studies by loup et. al. (1994) showing post-puberty L2 learners successfully striving for unaccented speech. An exceptional case (age: 22 yrs.) in Moyer (1999) also revealed highly successful L2 (German) phonological acquisition as the result of a high level of motivation and interest. Flege et. al. demonstrated the importance of environmental factors on quality of L2 accent through i. time spent with native-speakers (Flege et. al. 1997) and ii. time spent in the target language country (Riney & Flege, 1998). A re-analysis of the Johnson & Newport (1989) data by Bialystok (1997) confirms that the age factor is indeed intricately linked to the environmental factors. These studies clearly illustrate the complexity of the debate concerning age and L2 acquisition.

Turning now to another distinction with respect to the age factor that is essential to the current study: the age factor in the context of classroom L2 acquisition as opposed to naturalistic L2 learning. The former may be characterised as formal L2 acquisition in the classroom while the latter is generally characterised as L2 acquisition through immersion in the second language environment. In a naturalistic L2 learning context, Snow and Hoefnagel-Höhle (1978) revealed higher initial L2 learning rates in older L2 learners (age: 12-15 yrs. and adult) compared to early L2 learners (age: 3-5 yrs.). The early learners did catch up with the older language learners after a longer period of L2 exposure. These findings reinforced the notion that early L2 learners would ultimately reach higher proficiency levels than older learners in a naturalistic L2 learning context. Larson-Hall (2008) shows a trend in a formal L2 learning context that mirrors naturalistic learning, where 19-year old Japanese students were divided into ‘early’ (8.3 yrs. old) and ‘late’ (12.5 yrs. old) learner groups. The late group outperformed the early group at 800 hours of L2 exposure while the early group consistently outperformed the late learner group after 1500+ hours of exposure in a grammaticality judgement task and a phonemic discrimination task. These findings have not been revealed in earlier studies due to the fact that others (Mayo 2003, Lecumberri & Gallardo 2003) were unable to test pupils with 600 hours of L2 contact or more.

Muñoz (2008) argues that there is asymmetry between naturalistic L2 acquisition and L2 acquisition in a formal context, expressed through differences in amount and quality of L2 exposure, and the Larson-Hall study clearly shows that a large amount of native English L2

input is required to replicate findings from naturalistic L2 acquisition studies.

The current study addresses early L2 acquisition in a formal context, and based on the findings reported in the previous paragraphs it may be concluded that the age factor comprises a range of related factors that will be outlined in the following paragraphs. It is clear, however, that a focus on the amount of L2 contact and quality of L2 input is essential in studies concerning a formal L2 learning context (Singleton 1997, Muñoz 2008).

3.2 Teacher proficiency

The previous paragraph provided an initial insight into factors affecting L2 acquisition. It showed that factors are interrelated and therefore more difficult to isolate. The current paragraph will continue in the same vein to provide further insight into teacher proficiency and its effect on L2 acquisition in a formal L2 learning context (see paragraph 3.1) to shed some light on light on interspeaker differences in the current study (see paragraph 6.1.1.1).

The general term 'proficiency' has been used thus far to differentiate between beginners and more advanced L2 learners. It is generally held that native speakers are proficient in their L1 and that L2 learners of a particular language tend to be less proficient language users than native speakers. The distinction native versus non-native speaker is often used to quantify proficiency in absolute terms, but close inspection reveals varying degrees of proficiency across both native and non-native speakers. In fact, stating that a given speaker is proficient in a language is equivalent to claiming that a speaker shows command of a language across all linguistic domains: phonology, (morpho)-syntax, semantics, and pragmatics. It is likely that there are individual speaker variations in proficiency levels across these domains, and quantifying proficiency is therefore an issue that requires adequate instruments and description in this study.

Turning to the classroom L2 learning situation, a teacher is the primary L2 source for early learners. There are however related factors as Driscoll et al. (2004) have noted:

i. teacher language skills (proficiency, fluency, competence)
ii. (use of) L2 resources
iii. language teaching skills
iv. knowledge of child language development

Points i. and ii. stand in direct relation to the input that learners receive, while iii. and iv., are directed at motivating and guiding the learning process. The latter points do not relate directly to the input learners receive but certainly play a role in L2 acquisition in a formal context. These factors require attention and need to be controlled for (inasmuch as this is possible) in the current study.

As mentioned earlier, proficiency is often discussed in a native speaker versus non-native speaker context where a native teacher equates to high-quality input and a non-native teacher to lesser quality input: Árva & Medgyes (2000) unsurprisingly point out that native speakers are generally more proficient and confident in the target language. Gilzow &

Rhodes (2000) claim that native or near-native proficiency levels are essential for successful L2 teaching. Hunt et. al. (2005) conclude that native speakers can provide more cultural background related to the target L2 than non-natives, but the latter teacher group is often the class teacher who has more insight into individual pupil's language learning needs, allowing for more flexible incorporation of L2 acquisition into the daily classroom routine. In the same vein, Broekhof (2007) claims that native speaker language instructors are likely to be less capable of relating English language concepts to their Dutch counterparts. These studies exemplify the dichotomy between native and non-native teachers and show that it extends beyond linguistic proficiency. Medgyes (1994) - a Hungarian teacher educator - makes a fitting statement regarding this distinction: an ideal native English L2 teacher is one who is highly proficient in the learners' L1, and an ideal non-native English L2 teacher is one who has achieved near-native proficiency in English. In order to quantify possible differences between native and non-native L2 teachers, Moussu & Llurda (2008) emphasise the importance of including proficiency as an independent variable in research relating to (English) L2 acquisition so that the effect of variations in proficiency become apparent.

Interestingly, Larson-Hall (2008) noted that Japanese early learners taught by a native English teacher performed better on a phonemic discrimination test when contrasted to pupils who had been taught by non-native English teachers only. The effect was not statistically significant due to high variance, but the trend was clear: an early start (age 4-6 yrs.) with a native English L2 teacher showed better phonemic discrimination results compared to early learners who had not had a native English L2 teacher. Larson-Hall points out that (large-scale) replication of the study is required to draw robust conclusions.

3.3 Amount of L2 instruction time

The previous paragraphs alluded to the importance regarding the amount of L2 instruction time available. The current paragraph will provide further details on the influence of this factor on L2 acquisition. Carroll (1975) investigated factors that determine levels of acquired proficiency in a formal L2 acquisition setting, where the L2 is not readily accessible in the pupil's community and was concluded that amount of L2 contact was a key determinant of proficiency. With respect to the formal L2 learning context, Lightbrown (2000) concludes that "the most important reason for incomplete L2 acquisition in foreign language classroom settings is the lack of time available for contact with the language". Extensive instruction such as offered in immersion language learning models, enables learners to attain high proficiency levels. In such cases, learners are exposed to as much as 4000 hours of L2 input (Swain, 1981). These studies illustrate the importance of ample L2 contact.

The concept 'Amount of time' does require closer inspection: Dutch 'Early' and 'Late L2 learners groups are both provided with a similar amount of input, but the way in which the L2 input is presented is very different; 'Early' English L2 learners are being 'taught' through songs, games, and Total Physical Response activities while 'Late' English L2 learners work generally use a method that employs 'cognition-based' learning strategies. This section addresses the following concepts to illustrate the multiple facets regarding the concept of

'amount of L2 instruction time':

- 'frequency' - how often L2 input is provided;
- intensity - how much L2 input is provided over a given amount of time;
- continuity - how L2 input is sustained (throughout Primary School);
- L2 language learning models / methodology - the way in which L2 input is provided (implicit / explicit).

In a study contrasting three intensive English L2 acquisition programmes in Quebec, Collins et. al. (1999) showed that distribution of time (frequency and intensity taken together) affects L2 learning outcomes, provided that the total amount of L2 exposure is kept constant. Measurement of pupils' (age: 11-12 yrs.) progress was made across 300-400 hours of L2 contact. It was found that an intensive five-month language learning period showed more positive L2 learning outcomes compared to the same amount of input spread over a ten month period; A greater intensity and frequency of L2 input across a shorter period of time yielded better learning outcomes. Mayo (2003) found that increased exposure to classroom L3 English (396 hrs. vs. 596 hrs.) led to improved results in a grammaticality judgement task for Spanish/Basque bilinguals. There is, unfortunately, no report of frequency or intensity in Mayo's study, which is an omission that could have provided further insight into L2 contact based on results from Collins et. al. (1999). Canadian immersion studies in the 70s (Barik & Swain, 1976) have shown positive results regarding L2 French acquisition in intensive programmes for early L2 learners (age: 6-9 yrs.). In contrasting extensive, semi-intensive and intensive L2 English programmes at the University of Barcelona, Serrano (2007) shows that students at intermediate L2 levels (CEF[1]: B1 level) show greater improvement when following an intensive 110-hour L2 programme, while there is no such difference between the different programmes for students at advanced L2 levels (CEF: C1 level) for the same length of exposure (110 hours). The extensive programme spanned across a six month period with two-hour L2 sessions twice a week, while the semi-intensive programme spanned across a three-month period with four 2.5 hour L2 sessions a week. The intensive programme spanned a two-month period involving daily 5-hour L2 sessions. The findings from Serrano (2007) mirror the findings by Collins et. al (1999), showing a greater L2 learning effect in intensive programmes. These findings contrast with current L2 learning policies in Europe, as is pointed out in the following chapter. The following paragraph will illustrate the combination of factors involved in formal L2 acquisition and L2 learning outcomes using various L2 learning methods.

3.4 L2 learning methodology

Turning to language learning methodology to gain insight into its influence on other factors (amount/quality L2 input) regarding L2 acquisition. Gilzow & Rhodes (2000) discern

[1]For a complete CEF level description please see http://www.coe.int/T/DG4/Linguistic/Source/Framework_EN.pdf

the following foreign language learning models:

Language learning model / Methodology	% Class time in L2
Total immersion; instruction in a foreign language across the (entire) curriculum / implicit L2 learning;	50-100%
Two-way immersion; instruction in two languages, both spread across the curriculum / implicit L_X/L_Y learning;	at least 50%
Partial immersion; instruction in foreign language spread across some (approx. 50%) of the curriculum / implicit L2 learning;	approx. 50%
Content-based Language Learning (CLIL); instruction pertaining to the regular curriculum and the learned language itself / mostly implicit (some explicit) learning;	15 - 50%
Foreign Language in the Elementary School (FLES); time is spent learning language separate from other curricular activities / explicit learning	10 - 20%

Table 1: Foreign language learning models / methodology used in primary schools

Table 1 shows the differences between these language learning models in terms of amount of language contact: pupils learning a foreign language through an immersion model will spend a considerable portion of their time during certain grades (Barick & Swain 1976, Collins et. al. 1999) communicating in a foreign language, which contrasts with the FLES model where there is generally much less exposure to the foreign language across a longer period of time. The models also contrast in the type of language that pupils are exposed to: the immersion models and the CLIL model (to a lesser extent) provide a very broad range of input ranging from mathematics to - say - geography), while the FLES model focuses on general everyday language. Another important difference between the immersion language learning model and the FLES model is that the former generally has instructors that are native (or near-native) speakers of the target language, while the latter generally makes use of non-native target language speakers (Moussu & Llurda, 2008). L2 learning models therefore differ both in terms of L2 input quantity (frequency/intensity/continuity), and quality (native/non-native). Language learning through immersion has been shown to be effective in Canada, with no significant adverse effects on native language or subject knowledge (Barick & Swain 1976, Clyne et. al. 1995).

The factors affecting L2 acquisition that have been detailed here are central to the current study where the amount of L2 English input for early learners (age: 4-6 years) will be held constant at 30-40 minutes per week across a three-month period. Using teacher-trainees with differing proficiency levels an attempt is made to measure the effect of differing teacher proficiency levels (B1-C1 CEF[2] levels) on pupils across schools. A general measure of L2 development is also made to determine the L2 learning effect across approximately 10 hours of L2 language instruction. The following chapter provides provides details relating to L2

[2]For a complete CEF level description please see http://www.coe.int/T/DG4/Linguistic/Source/Framework_EN.pdf

acquisition in primary schools across Europe and specifically in the Netherlands, to reveal L2 learning effects in a low input and primarily non-native L2 teacher environment.

4. Foreign Language learning in Primary Schools - an overview

This chapter provides an overview of current foreign language learning policy and practice across Europe and specifically in the Netherlands as a practical background relating to the research outlined in the following chapters. Section 4.1 provides an overview of foreign language learning in Primary Schools across Europe and section 4.2 focuses on English L2 language learning in the Netherlands.

4.1 Foreign Language learning in Primary schools across Europe

The 19th century marked the beginning of systematic foreign language teaching in most European countries. Major changes to European educational systems in the 1950s and 1960 led to a widespread acceptance of foreign language teaching in secondary schools and the majority of EU countries changed their foreign language learning policies and introduced foreign language learning to primary schools in the 1970s, 1980s or 1990s as is shown in the figure below:

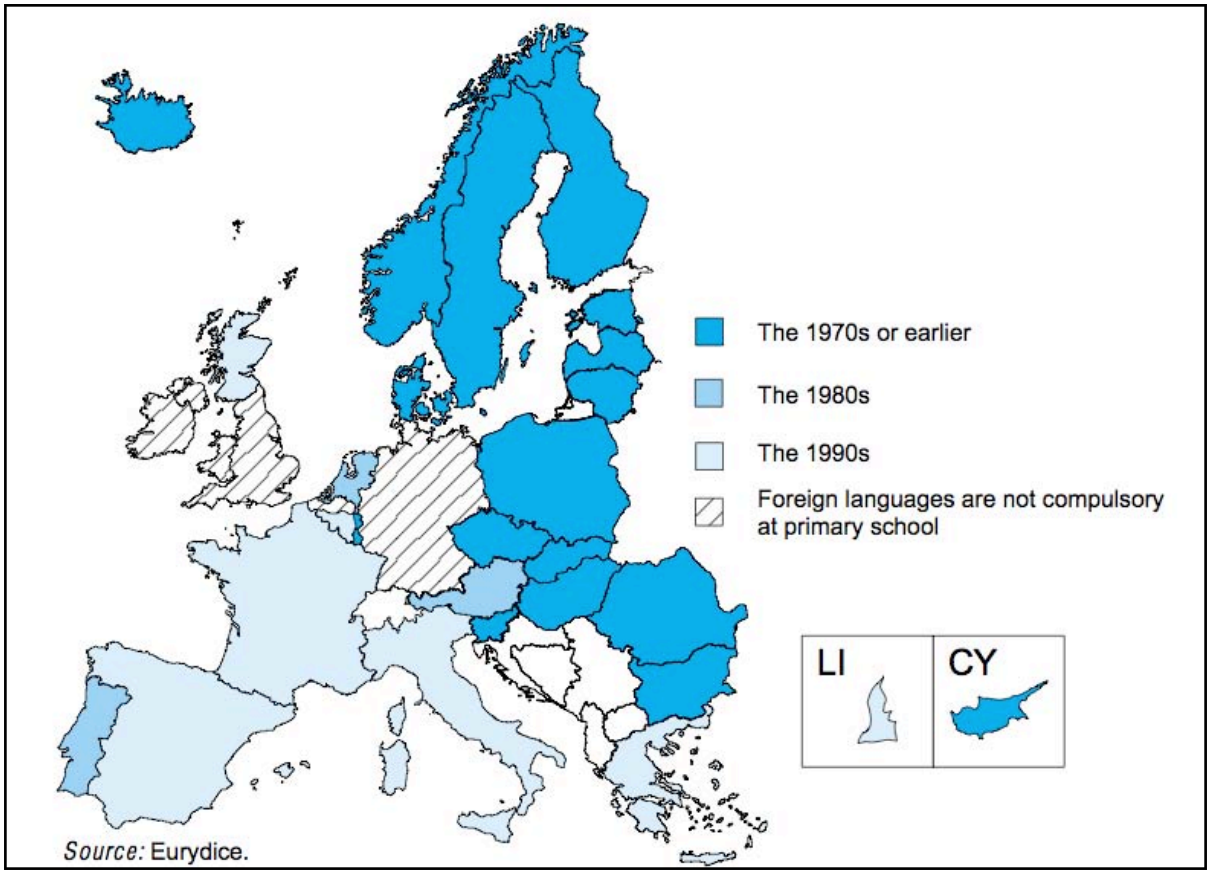


Figure 1: Introduction of the first foreign language in the minimum curriculum at the primary level (eurydice, 2001 pg. 61)

The European Commission's publication 'Teaching and Learning: Towards the learning society' (1995) envisages proficiency in three languages for European citizens. This language learning policy is paralleled by an increased general adoption of foreign language instruction for early learners. Kubanek-German (1998) notes several trends with respect to foreign language teaching in Europe:

- Offering foreign languages to early learners (pre-primary) has become a more common element in curricula since 1990;
- The starting age for foreign language learning is decreasing;
- There is widespread interest in foreign language learning for children across Europe;
- National guidelines across Europe indicate that language learning should be enjoyable and that time should also be allocated to creating language awareness;
- English as a foreign language is dominant;
- Learners' confidence regarding foreign language competence appears to be growing.

Table 1 below provides an overview of time spent annually on foreign language learning per EU country in both Primary and Secondary schools:

Country	**Primary school / hrs. per year**	**Secondary school / hrs. per year**
Belgium	60[1]	1411[3]
Czech republic	35	95*
Denmark	40	200*
Germany	28	190*
Estonia	70	151
Greece	39	140
Spain	57	113
France	32	115
Ireland	NC	NC
Italy	80	88*
Cyprus	21	125
Latvia	35	135
Lithuania	13	132
Luxembourg	351	310
Hungary	28	163
Malta	170	219*

[3]Figure is an average between French, German and Flemish speaking Belgium.

Country	Primary school / hrs. per year	Secondary school / hrs. per year
Netherlands	30[4]	214[2]
Austria	20	125*
Poland	25	65
Portugal	40	101
Slovenia	39	79
Slovakia	***	104
Finland	38	139
Sweden	63	108
UK	***	**
Iceland	23	157
Liechtenstein	18	140*
Norway	39	128
Bulgaria	***	104
Romania	38	105

Table 2: Time spent annually (hours) on foreign language learning in both Primary and Secondary schools. *average number of hours across different school types. **no figures available - schools may allocate flexible number of hours to foreign language learning. ***no figures available as Foreign language is not compulsory.

Table 2 clearly shows that secondary schools spend considerably more time (10-15% of total teaching time, Eurydice, 2005) on foreign language learning than Primary schools, and the number of years of exposure is also greater in secondary schools. Secondary schools also generally employ specialist foreign language teachers. The previous chapter reported the importance of both language contact and teacher proficiency (see also paragraph 4.2.1) and it seems reasonable to suggest that conditions for foreign language learning are currently more favourable in secondary schools. So, while there are trends showing a decrease in learner age there is no notable change in time spent learning L2 at an earlier age across Europe. The following section provides further details regarding English L2 learning in primary schools in The Netherlands.

4.2 English in Primary Schools in The Netherlands

This section is divided into two sub-sections, namely: Late English L2 learning (age: 10-12 yrs.) and Early English L2 learning (age: 4-6 yrs.) in Dutch primary schools.

4.2.1 Late English L2 Language learning in Primary Schools (EIBO)

L2 English was introduced in the Netherlands as a compulsory subject within the Primary School curriculum in 1986 and it is referred to as ‘Engels in het Basisonderwijs’ (EIBO - English in Primary education). It is taught in the final and pre-final years of Primary School to

[4]Figure based on Heesters et. al. (2008)

children between 10 to 12 years of age. This will be referred to as being Late English L2 learning henceforth. Eight percent of all Primary Schools start teaching English a year earlier (Edelenbos & de Jong, 2004). English lessons are (on average) weekly 45-minute sessions spread across a two-year period. Pupils are therefore exposed to a maximum of 60 hours of classroom English at Primary School. Table 2 below summarises the results of studies by Heesters et. al. (2008) and Edelenbosch et. al. (1996), in which (class) teachers were asked to indicate which language they used during English lessons. This provides an estimate of the actual number of hours of English language contact provided in primary schools:

Language of instruction 1996		**Language of instruction 2006**	
% English	**hrs. English**	**% English**	**hrs. English**
26% English	15 hrs. 36 min.	33% English	19 hrs. 48 min.
61% English/Dutch combination	± 18 hrs. 18 min.[5]	33% English/Dutch combination	± 9 hrs. 54 min.[3]
Total:	**± 33 hrs. 54 min.**	**Total:**	**± 29 hrs. 42 min.**

Table 3: estimated amount of English L2 contact throughout Primary School

Table 2 shows an annual total of 30 hrs. of classroom English in primary schools in The Netherlands (European average ± 55 hrs.), which would amount to 60 hours of classroom English during primary school while the estimated amount of English language contact is only 30 hrs. during the entire primary school period. The estimates listed in table 3 are below the European average and Late learners are faced with little actual English language input in primary schools.

By contrast, a preliminary study (Nuhaan, 2009)[6] investigating the amount of English input Dutch children (age: 9 - 12 yrs.) receive outside the classroom on a weekly basis indicates that English L2 contact is much more frequent outside the classroom:

Age	**Medium**		
	TV	**Internet**	**Games**
9-10 yrs.	3 hrs. 14 minutes	1 hr. 35 minutes	3 hrs. 17 minutes
10-11 yrs.	4 hrs. 47 minutes	1 hr. 55 minutes	4 hrs. 25 minutes
11-12 yrs.	5 hrs. 21 minutes	2 hrs. 34 minutes	5 hrs. 3 minutes

Table 4: weekly English L2 input via media among children attending PS ‘de Fontein’ in Barneveld

Aside from differences in quantity of English input there is also a difference in quality of English: English input via media will be native English input, while class teachers English proficiency levels range from B1 to C1 level.[7] This proficiency level estimate is based on

[5]assuming a 50% division between English and Dutch
[6]To access this (unpublished) paper, please contact jwchevalking@che.nl
[7]For a complete CEF level description please see http://www.coe.int/T/DG4/Linguistic/Source/Framework_EN.pdf

Quick Placement Test diagnostic test scores at the Christelijke Hogeschool Ede - Primary teacher training institute and data collected by Edelenbos & de Jong (2004).[8] These level estimates indicate a considerable difference in proficiency between class teachers and native speakers of English. Heesters et. al. (2008) note that roughly one-third of all class teachers have had no specific schooling in L2 English language teaching. Classroom English L2 input is not only provided by (classroom) teachers but also through language learning material. Most methods use websites and/or multimedia material to provide pupils with native (British) English input. The general trend in frequently used methods[9] is to base language learning on themes relating to 'everyday communicative situations'. Methods contain mostly English, while some instructions to exercises are provided in Dutch. To sum up, classroom English L2 input in the EIBO programme ranges from native input, usually from additional materials rather than a live native speaker, to non-native B1 proficiency level input over a maximum of 60 hours spread across grades 7 and 8 at Primary School. From the data and estimates presented thus far, the question rises as to the effect of both quantity and quality of English language input in classroom English on L2 language learners, particularly in view of the factors outlined in the previous chapter.

To monitor L2 learning progress in the EIBO programme, English language-learning goals were constructed as it was introduced into the Dutch curriculum. Pupils were evaluated and language learning goals were adapted as time progressed[10]. They are aimed at:

i. being able to understand basic everyday spoken language;
ii. being able to talk about basic everyday topics;
iii. having a general understanding of simple written text;
iv. being able to use a dictionary

(adapted from: Edelenbos & de Jong, 2004, pg. 19).

There is no clear definition of the terms 'basic', 'everyday', 'general' and 'simple' in terms of lexical frequency, phonological or (morpho-)syntactic structure, or semantic content. Educational publishers take word frequency into account, but they do so to varying degrees; some provide an overview of lexical items in a method together with their corresponding frequencies, others provide a wordlist without reference to a lexical corpus. The content in methods is always directed at the L2 learning goals formulated in the previous paragraph, but methods do not provide a clear overview of (CEF) proficiency levels and how these relate to the material with the exception of material from the Anglia Network.

---

[8]See http://www.nabmvt.nl/publicaties/00013/, pg. 24 and 28 for an overview of English L2 proficiency levels in Dutch Secondary and Higher education.
[9]common methods are: Bubbles (ThiemeMeulenhoff), Hello World (Malmberg), Junior (ThiemeMeulenhoff), Real English – Let's do it! (Bekadidact), The Team (Wolters Noordhoff)
[10]see http://www.minocw.nl/documenten/kerndoelenboekje.pdf pg. 21 for an overview of the current language-learning goals

To date, there have been three nation-wide evaluations (Vinjé 1991, Edelenbosch et. al. 1996, Heesters et. al. 2008) based on the goals as formulated in the previous paragraph. Table 4 provides an overview of the 1996 and 2006 evaluation results (adapted from: van der Schoot, 2008):

Skill	Minimum 1996	Satisfactory 1996	Minimum 2006	Satisfactory 2006
Listening	84	46	95	66
Reading	70	32	92	50
Speaking	64	33	a	a
Rec. Vocabulary	68	34	84	50
Prod. Vocabulary	63	25	a	a

Table 5: % of children realising the set goals for a given skill divided into Minimal and Satisfactory categories. a Denotes unavailability of data due to changed goals nation-wide.

The 'minimum' and 'satisfactory' columns denote two standards that are defined as follows:

- Minimum: indicates a minimum level of attainment of the language learning goals (see 3.1) by Primary School pupils. It is expected that 90-95% of all pupils reach this level on completing primary school;
- Satisfactory: indicates a satisfactory level of attainment of the language learning goals (see 3.1) by Primary School pupils. It is expected that 70-75% of all pupils reach this level on completing primary school (Heesters et. al., 2008, pg. 28).

Again, there is no clear description of what measures up to minimal or satisfactory standards in linguistic terms; instead, to relate language learning goals to the tests, 11 examiners were each asked to rate test items (A1 and A2 level material)[11] and place them into the above-mentioned categories. These evaluations were then pooled and re-evaluated to allow for test-scoring. Definitive categorisation took place after testing children, viewing the results and comparing these to the examiner categorisations[12].

The data in table 5 should be regarded as indicative only; there are no grounds for comparing the 1996 and the 2006 data sets as i. the 2006 pupils were evaluated less strictly, and ii. the evaluated skills differed between the two data sets. There is a trend to be noted:

- Listening consistently shows the highest scores;
- Reading consistently appears as 'second-best';
- Receptive vocabulary consistently shows the lowest score.

[11]For a complete ALTE level description please see http://www.coe.int/T/DG4/Linguistic/Source/Framework_EN.pdf

[12]See Heesters et. al., 2008, pg. 28, 29 and Edelenbos et. al. pg. 25, 26 for a description of test construction

Receptive vocabulary is the lowest scoring skill of the three receptive skills that were tested. This is striking as words are the 'building blocks' of any multi-word utterance - be it spoken or written. Close inspection of sample test-items (Heesters et. al., 2008, pg. 48-51) reveals that vocabulary testing is the least 'life-like', least practised exercise[13], and pupils cannot rely on contextual cues (that they might possibly know) to the extent that they can in a reading or listening reading test.

To sum up the current state of affairs regarding Late L2 English in Dutch Primary schools it appears that i. L2 input quantity is low (max. 30 hrs. / year), ii. L2 input quality ranges from B1 proficiency level up to C2 proficiency level and iii. language learning goals are not being met to the set standards. The following section addresses Early L2 acquisition in Primary Schools, which - unlike Late L2 acquisition - has not been implemented and evaluated to the same extent.

4.2.2 'Early' English L2 Language learning in Primary Schools (VVTO-E)

An increasing number of Primary Schools are offering Early English L2 to learners between 4 to 8 years of age; There are currently 277 Primary Schools[14] that offer Early English L2, compared to fewer than 10 Primary Schools in 1994. Early English L2, or VVTO-E (Vroeg Vreemde Talenonderwijs - Engels), is not a compulsory part of the Dutch Primary School curriculum, but its implementation is encouraged via subsidies and teacher-training programmes (Deelder & Maljers pg. 23, 2006). Early English is generally taught in the first two grades in Primary Schools. Many schools then choose to discontinue L2 English during grades 3 and 4 as reading/writing and mathematics is prioritised. This does raise continuity issues which are currently being debated.

The amount of input provided generally ranges between 30 - 240 minutes per week[15] and depends on school-specific goals / limitations, and priorities within the curriculum. In 2005, 35 out of 54 schools employed a native speaker to teach ‘Early’ English L2[16]. There are no current figures available regarding the number of ‘native’ speakers teaching ‘Early English L2, but the distribution ‘native - non-native’ speakers will no doubt shift towards non-native classroom teachers; there are simply not enough native English teachers.

English is offered in a playful manner through songs, games and TPR (Total Physical Response, i.e. physically responding to language instruction) activities. Early learners (and parents & teachers) are generally content with the way in which ‘Early’ English is being offered Deelder & Maljers pg. 27, 2006). An important consequence of learning an L2 at this young age is that children cannot read or write, thus side-stepping i. inconsistent English grapheme - phoneme mapping which can be problematic for dyslexic pupils (van Berkel,

[13]as is found in the following methods: Bubbles (ThiemeMeulenhoff), Hello World (Malmberg), Junior (ThiemeMeulenhoff), Real English – Let’s do it! (Bekadidact), The Team (Wolters Noordhoff)
[14]see http://www.talenopdebasisschool.nl/downloads/adressenlijst_vvto.pdf for an overview of these schools
[15]see http://www.talenopdebasisschool.nl/downloads/adressenlijst_vvto.pdf for an overview of time spent on English per school
[16]see http://www.minocw.nl/documenten/brief2k-2005-doc-29329.pdf

2007) and ii. a more 'formal' (focus on grammar, wordlists, writing) approach to L2 language acquisition.

There are few published studies focusing on VVTO-E in the Netherlands, but research is currently in progress. An early investigation by Goorhuis-Brouwer & de Bot (2005) showed that there were no adverse effects on first language development in early learners when exposed to 2,5-5 hrs. of native speaker English input a week (Philipsen, 2007). Aarts & Ronde (2006) report that when 6-year-old children are provided with 1¼ hrs. of native speaker English L2 input each week, they show low (productive) proficiency levels by the end of their second year of English. For both studies mentioned here, classroom language instructors were native speakers of English, but the amount time spent teaching English varied between the two studies. It is unfortunate that the former study does not report pupil progress regarding English as this could have provided insight into the effect of varying lengths of English language teaching on pupils. Given that there is considerable variation in length of exposure to English input between VVTO-E schools at present, the possible effects of this variable should be investigated to aid in policy-making regarding VVTO-E in Dutch primary schools. Not all schools employ a native English speaker to teach VVTO-E (unlike the studies outlined above), and the increase in the number of schools offering VVTO-E has led to increasing numbers of class teachers teaching English - as is the case in EIBO. Consequently, the L2 input pupils receive in VVTO-E is varied ranging from B1 CEF level to C1 CEF level. The possible effects of this variation should also be studied to determine the task of teacher-training institutes in preparing their students to teach VVTO-E (and EIBO).

This chapter and the previous chapter reviewed factors affecting L2 acquisition and the status quo regarding L2 acquisition in both Europe and The Netherlands. It was found that time spent in contact with an L2 was a crucial factor in relation to learning outcomes and that teacher proficiency was another key factor involved in L2 acquisition in a formal L2 learning context. Learner age has also been studied extensively in relation to L2 acquisition and various studies implicated that the interplay between factors is as important to understanding and interpreting learning outcomes as the factors in their own right.

Reviewing the European and Dutch formal language learning contexts in the light of results from studies investigating L2 learning factors, it can be stated that L2 contact is minimal (30 hours per year) for Dutch L2 late learners (age 10-12 yrs.). Late L2 learners receive most of their L2 input from sources outside the classroom, and research involving late L2 learners should establish a measure this L2 input. Early learners will also receive some L2 English input outside the classroom, but to a much lesser extent as television series and other multimedia sources for this age group are more often in Dutch. The current study addresses early learners (age: 4-6 yrs.), and these learners are provided with between 30-240 minutes of L2 input per week provided by either a class teacher (proficiency levels ranging from B1 to C1 CEF level - see appendix A for an overview) or a native speaker. The following chapter will elaborate on phoneme development in relation to L2 acquisition in order to establish a measure of child sensitivity to (novel) phonemic distinctions, given L2

input ranging between B1 and C1 proficiency levels across a three-month period.

5. Speech perception development

5.1 Introduction

Larson-Hall (2008) found that Japanese L2 learners of English starting at age 8 yrs. showed better performance on a phonemic discrimination test than L2 learners starting at age 12.5 yrs., while the two learner groups did not differ significantly in performance on a grammaticality judgement task. Both groups received ≤ 4 hrs. of L2 English input on a weekly basis, so differences found in Larson-Hall (2008) could not be attributed to variation in amount or quality of input. Flege et. al. (1999) also noted that age affects different linguistic domains differently; 240 Korean participants showed stronger foreign accents as their age of arrival (AOA) to the US increased, while scores on a grammaticality judgement task (morhpo-syntax) did not significantly decrease as a function of age. Following from this, Flege et. al. (1999, 1995) hypothesise that the interaction between the L1 and L2 phonological systems changes due to the fact that increasing age entails a more developed L1 phoneme inventory. A number of speech perception studies (Jusczyk 1993, Nittrouer, 1996, Gerrits 2001, Heeren 2006) found that children do not categorise sounds in an adult-like manner until they are between 6 - 9 years old. So both perception of phonemes and production of phonemes are found to be affected by age.

These findings and the underlying processes are elaborated in the current chapter and serve as a foundation to the current study in which the interaction between non-native teacher language production and early L2 learners' perception of English L2 phonemes is studied. This chapter outlines L1 phoneme development, which is followed by L2 phoneme development - discussing differences between L1 and L2 phoneme development as found in the literature. An overview of Dutch and English phoneme inventories is provided to gain insight into the task that Dutch L2 learners of English are faced with. The chapter concludes by stating the research questions that are investigated in the current study.

5.2 L1 phoneme development

In order to learn language an infant has to learn to segment a continuous speech signal into words, syllables and speech sounds. This is no easy task, as the infant has no prior knowledge of word boundaries and no lexicon that serves as a frame of reference from which to infer such boundaries. Segmenting the speech signal into its constituents and discriminating between speech sounds are the (innate) skills that are required to create an 'input lexicon' that acts as a bootstrap for further word recognition (Jusczyk, 1992).

Infants are generally faced with speech sounds from a particular language. At a very young age, infants learn speech sounds particular to the language they are faced with. This process of establishing mental representations of speech sounds (called phonemes) is no easy task and discrimination between phonemes is hampered by variation based on i. position in the speech signal (cue ambiguity) and ii. interference between speech sounds (coarticulation).

This is shown in the examples below:

i. Cue ambiguity: phoneme /p/ is [p] in [læp] but [p^h] in [p^hæt] - English language learners will have to establish that [p] and [p^h] belong to the phonological category /p/, and that [p] and are variants of /p/ (allophones). The English language learner will also have to establish in which contexts each allophone occurs: [p^h] is found in syllable-initial position and [p] is found in other positions.

ii. Coarticulation: observe the differences in pattern between the spectograms for 'mat' - [mæt], 'fat' - [fæt] and 'pat' - [p^hæt]. Each of these words differ only in their initial consonant sound which clearly affects its realisation in terms of intensity (yellow line), pitch (blue line) and formant frequency spread (red dots).

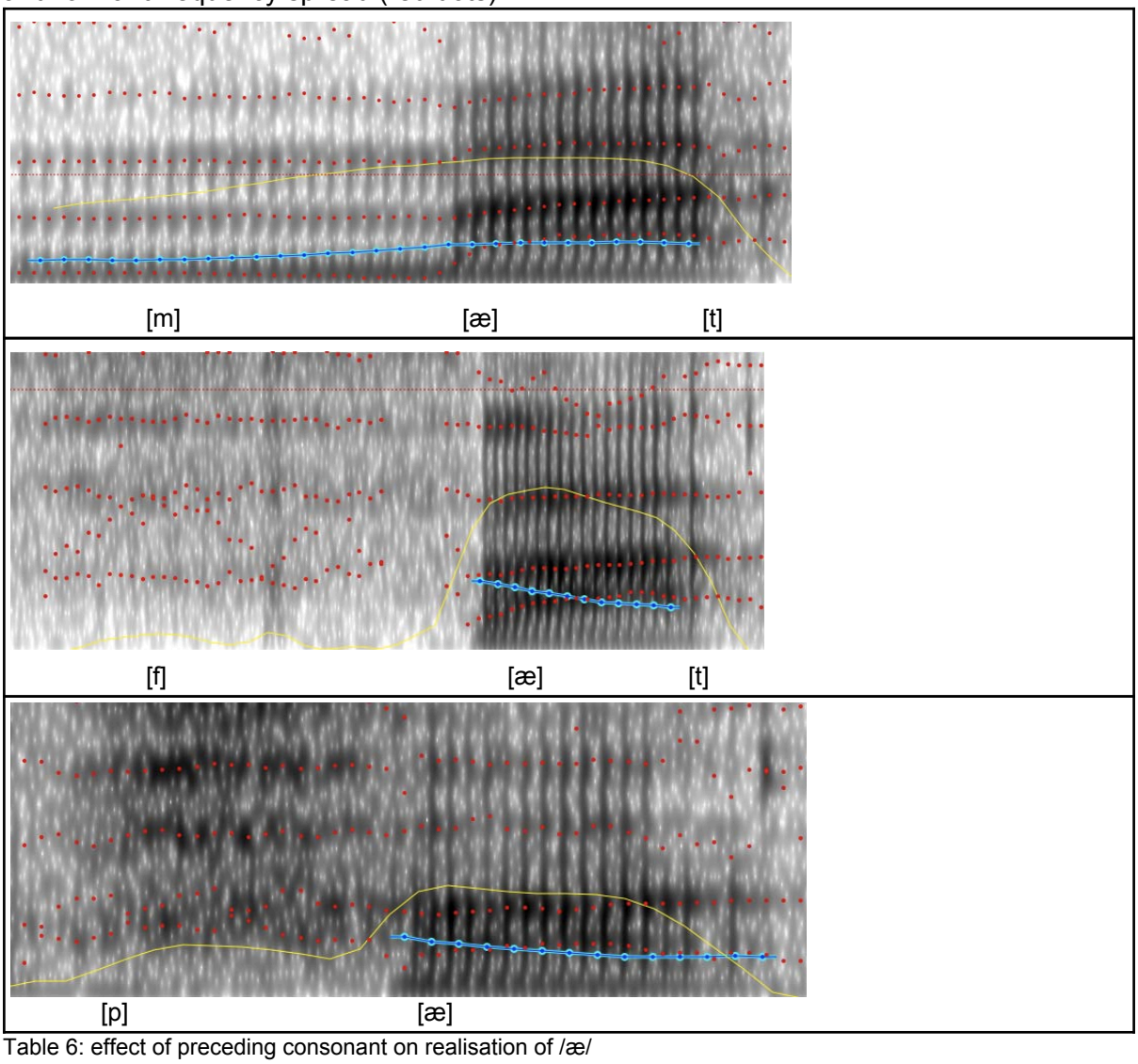


Table 6: effect of preceding consonant on realisation of /æ/

Despite variation[17] in the speech signal, Eimas et. al. (1971) showed that infants (age: 1 month & 4 months) are capable of distinguishing a [+VOICE] - [-VOICE] contrast in [ba] - [pa] in a categorical manner. From an early age infants start to pay particular attention to the structure of their L1, and as their linguistic experience with this L1 increases infants quickly become less sensitive to contrasts that fall outside their L1. Jusczyk (1993) claims that this process first takes place at the prosodic level, and then extends to the phonetic level. Werker et. al. (1981) and Werker & Tees (1984, 2002) showed that young infants (age: 6-8 months) can discriminate between distinctions used in various languages without any experience in any of these languages, and that this ability decreases through experience with a specific language before the end of an infant's first year of life (Werker & Tees, 1984, 2002). By this process, infants start to resemble adult speakers in showing increased difficulty in discriminating between (phonetically) similar sounds that do not contrast phonemically. For example, infants in an English L1 environment start to learn to focus on distinctions that distinguish between meaning as in 'bad' - [bæd] and 'bed' - [bɛd] where /æ/ and /ɛ/ are distinc phonemes in English, while they lose the ability to distinguish between non-contrastive phones as in 'pit' [pɪt] and 'pit' [$p^hɪt$]. Thus, very young infants can detect phonetic contrasts, this ability soon 'diminishes' as the infant starts to perceive change categorically rather than continuously. The illustration below shows categorical perception:

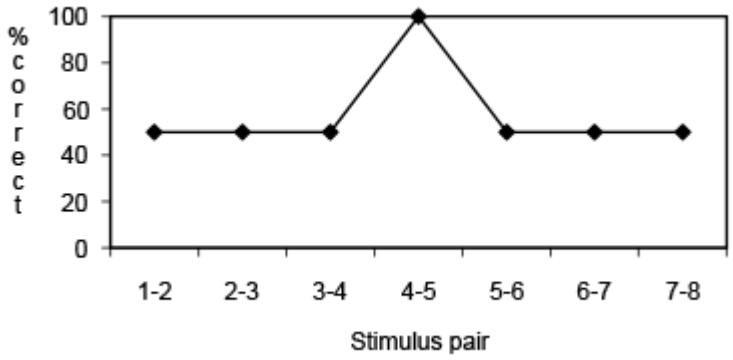


Figure 2: Categorical perception

Suppose that 8 stimuli are created to form a continuum between two phonemes where 1 and 8 are maximally different. The x-axis in the graph represents the 8 stimuli in pairs and the y-axis shows the % of correct discriminations between each pair (50% = chance level). An ideal case of categorical perception is represented here as all pairs are discriminated at chance level except pair 4-5, where there is 100% correct between-category discrimination: a listener's perception near the phoneme category boundary is heightened so that two distinct phoneme categories are perceived.

Thus, for an infant to acquire the phonology of a language it will have to learn which allophones belong to a given phonemic category, and the child will also learn in which sequence these phones occur in the language (phonotactics). For example, English does not permit the sequence 'fben' [fbɛn] or 'busr' [bʌsr]. Both [fb] and [sr] are sequences that are disallowed in English. Infants as young as 9 months show a preference for sequences

[17]See also Strange (1995) for a detailed description of the 'Constancy Problem' as illustrated here.

that are permitted in the native language (Jusczyk, 1992). It is clear, however, that - despite the complexity of the speech signal - infants are capable of discriminating between phonemic contrasts in their native language at an early age. These phonemic contrasts are marked by several acoustic 'cues'. The listed cues are key to the current study:

- Spectral cues - vowels differ in terms of spectral properties which correspond to tongue height and tongue backness in the oral cavity;
- Durational cues - speech sound differ in terms of their length.

Each of these cues can be measured which enables quantification of phones, showing both duration of the phone produced and position of the articulatory apparatus during production of that phone. The table below details the cues and illustrates their values in relation to the position of the articulatory apparatus:

Place / Time	**Cue**	**Position articulatory apparatus**
degree of opening (tongue height)	F1	Hz - closed = low value, open = high value
degree of backness (tongue backness)	F2	Hz - front = high value, back = low value
duration	duration	milliseconds

Table 7: Speech cues

These cues can be represented graphically and plotting F2 values against F1 values reveals the position of the articulatory apparatus in the articulatory space.

The illustration below is a scatterplot of F1 and F2 measurements of 76 speakers of General American English vowel sounds. (33 men, 28 women, 15 children - 1520 vowel tokens in total):

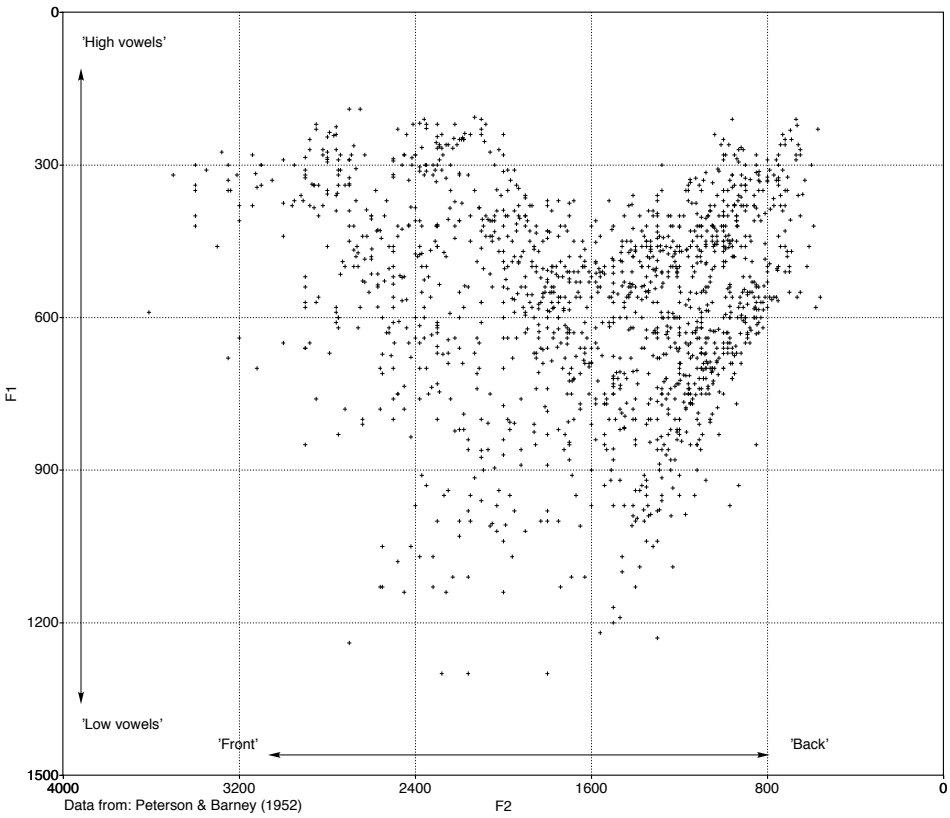


Figure 3: F2 vs. F1 plot GA monophthongal vowels

Acoustic reality as shown in the scatterplot is fuzzy at best, which is the result of inter-speaker differences, speech rate and linguistic context[18]. Variability extends to the durational cue as well, which leads to high variation within phoneme categories. It is truly remarkable that (L2) learners can acquire language in the face of such variability.

Variability aside, Jusczyk (1993) states that children assign different weightings to cues when compared to adults. Nittrouer (1996) showed that children (age: 3 yrs.) are less sensitive to changes in acoustic cue values when compared to adults. Gerrits (2001) elaborated these findings and found that Dutch children (up to 6 yrs. of age) showed adult-like weighting of the durational cue in a vowel contrast, but that spectral cues revealed different weightings between adults and children. The current study focuses on early English L2 learners aged 4 to 7 years, who - based on the findings presented here - have yet to

[18]See Beddor & Gottfried (1995) for an extensive overview of methodological issues that require attention in speech perception research.

establish an adult-like cue-weighting system in their L1. For this reason we now turn to L2 phoneme development.

5.3 L2 phoneme development

The introductory paragraph to this chapter showed findings from various studies that distinguished between (morpho-)syntactic and phonological performance in early and late L2 learners. The study by Flege et. al. (1999) reported age-related effects on phonological production and Larson-Hall (2008) reported age-related effects on phonological perception. These studies do not, however, attempt to establish how L2 phonemes are acquired. Flege (1999) does hypothesise that the interaction between the L1 and L2 phonological systems changes due to the fact that increasing age entails a more developed L1 phoneme inventory. L2 learners - be they early or late L2 learners - are faced with the task of segmenting an L2 acoustic stream into its constituents. Werker & Tees (1984) found that 10-month old infants had attuned their perception of the acoustic stream to their L1, and given the fact that realisation of phonemes is language-specific the question rises as to how an L1 phonemic inventory affects acquisition of an L2 inventory.

Assuming that the L1 phoneme inventory is the initial state - while recognising that cross-linguistic influence is not limited to matching and substituting L1 and L2 inventories (Leather, 2003) - of the learners' L2 phoneme inventory, an L2 learner will transfer his/her i. L1 vowel and consonant category set, ii. L1 segmental boundaries and iii. L1 rules for phonetic realisation of phonemic categories. Given that natural languages differ phonemically and allophonically, the learner will have to establish which sounds in the L2 are phonemic and which sounds are allophonic in the L2. In relation to the L1 vowel and consonant category set, Escudero (2005) illustrates between-language perceptual differences in the graph below:

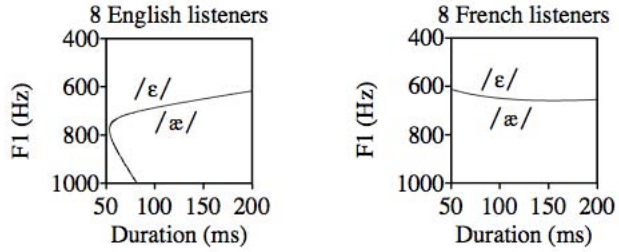


Fig. 1.6. Categorization of CF /æ/ and /ɛ/ by CF (left) and CE (right) listeners (adapted from Escudero & Boersma 2004a).

Figure 4: Escudero (2005) - Categorisation of /æ/ and /ɛ/ in a CVC context by Canadian French and Canadian English listeners

The curved lines in both graphs represent the category boundary between the phonemes involved, the y-axis represents the F1 spectral cue (tongue height) and the x- axis repesents the durational cue. It is clear that Canadian English L1 listeners use both the F1 and the

durational cue to differentiate between /æ/ and /ɛ/, while the Canadian French listeners used mainly F1. The category boundary, and the perceived distribution of the two phonemes were shown to be different for the two listener groups. These differences do not only hold between languages, but also between different varieties of the same language (Escudero & Boersma, 2003, 2004). Given these findings, the following situation could arise in developing an L2 phoneme inventory:

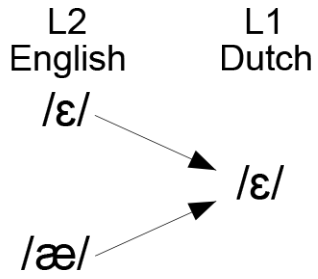
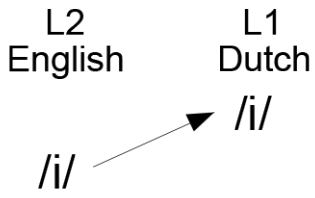
	
The L1 has a single phoneme category and the L2 has two, which may lead to misperception of the /ɛ/ - /æ/ contrast; the two L2 categories are mapped onto a single L1 category.	The L1 and L2 both have 'similar' categories with (slightly) differing boundaries.

Table 8: phoneme categorisation - possibilities

The example illustrates Dutch L1 speakers failing to perceive (or produce) /æ/, which is part of the English but not the Dutch phoneme inventory. On hearing /æ/, Dutch L1 learners of English map this phoneme (which is clearly non-contrastive in the Dutch phoneme inventory) onto /ɛ/, which is also part of the Dutch phoneme inventory. The cues used to differentiate between /æ/ and /ɛ/ for English L1 speakers are clearly dissimilar to those in Dutch. The example illustrated in the right column is similar to the Canadian English / Canadian French example in that both languages show the same phoneme /i/, but an underlying difference in cue weighting may (or may not) lead to differences in category boundaries between the languages.

Previous research reported that native speakers of Dutch learning English as a second language, experience difficulties in perceiving an /æ/ from an /ɛ/ (the scenario in the left column in the table above). The reason being that English /æ/ does not have an immediate Dutch counterpart, and it is suggested that L2 language learners consequently map the unknown L2 phoneme /æ/ onto the phonetically nearest L1 phoneme /ɛ/, in a process called 'inter-lingual identification' (Flege 1988). Weber and Cutler (2004) confirmed this perceptual difficulty of the /æ/~/ɛ/ contrast for Dutch learners of English by means of an eye-tracking experiment; The discrimination difficulties turned out to be asymmetrical in phoneme identification tasks, where /æ/ was identified as /ɛ/, while /ɛ/ was less likely to produce an

/æ/ response (Weber & Cutler, 2004). Flege (1987) claims that ‘similar’ phonemes (i.e. phonemes that occur in both L1 and L2 phoneme inventories) cannot be produced by L2 learners in a native-like fashion. Given these findings and in view of L2 phoneme development as outline here, the English and Dutch vowel inventories are investigated in the following section to provide details regarding the phonemes used in the current study.

5.4 Dutch and English phoneme inventories

Given the studies by Flege et. al. (1987, 1988) and Weber & Cutler (2004) and their focus on vowels, this section addresses the Dutch vowel system and contrasts it to the English (Standard Southern British English variant) vowel system.

5.4.1 The Dutch vowel system

Booij (1995) provides an overview of Dutch phonology, stating that there are several varieties of modern Dutch: northern standard Dutch spoken in the Netherlands and southern standard Dutch which is spoken in Belgium. Aside from these varieties, the Dutch language landscape is strewn with regiolects (intermediate language varieties between standard Dutch and specific dialect). This study involves speakers of northern standard Dutch; the vowel description below will therefore address northern standard Dutch only.

NSD vowel	**CVC example**
/y/	'vuur' (fire)
/ɪ/	'tik' (knock/tap)
/e/	'beet' (bite)
/ø/	'reus' (giant)
/ɑ/	'bak' (bake)
/a/	'taak' (task)
/u/	'boek' (book)
/ʏ/	'hut' (hut)
/ɔ/	'bot' (bone)
/o/	‘boot’ (boat)
/ə/	'bottən' (bones)
/ɛi/	'zijn' (to be)
/œy/	'tuin' (garden)
/ɔu/	'koud' (cold)

Table 9: Dutch vowels adapted from: The phonology of Dutch (Booij, 1995)

Adank et al. (2004) provide a description of the vowels of northern (and southern) standard Dutch including formant and durational values. Data for the F2 vs. F1 plot below was taken from Adank et. al. (2004), and represents average F1 / F2 values (N=10; female) for Dutch monophthongal vowel sounds. Data was obtained by recording subjects reading out sentences containing the target phonemes in three contexts (CVC, CVCV, V). As can be seen in the scatterplot, the boundaries of the acoustic space are marked by the /i/, /u/ and /a/ in an 'upside-down' triangle shape where /a/ corresponds to the lowest tongue position (highest F1 value), /i/ to the frontmost vowel (highest F2 value) and /u/ to the backmost vowel (lowest F2 value).

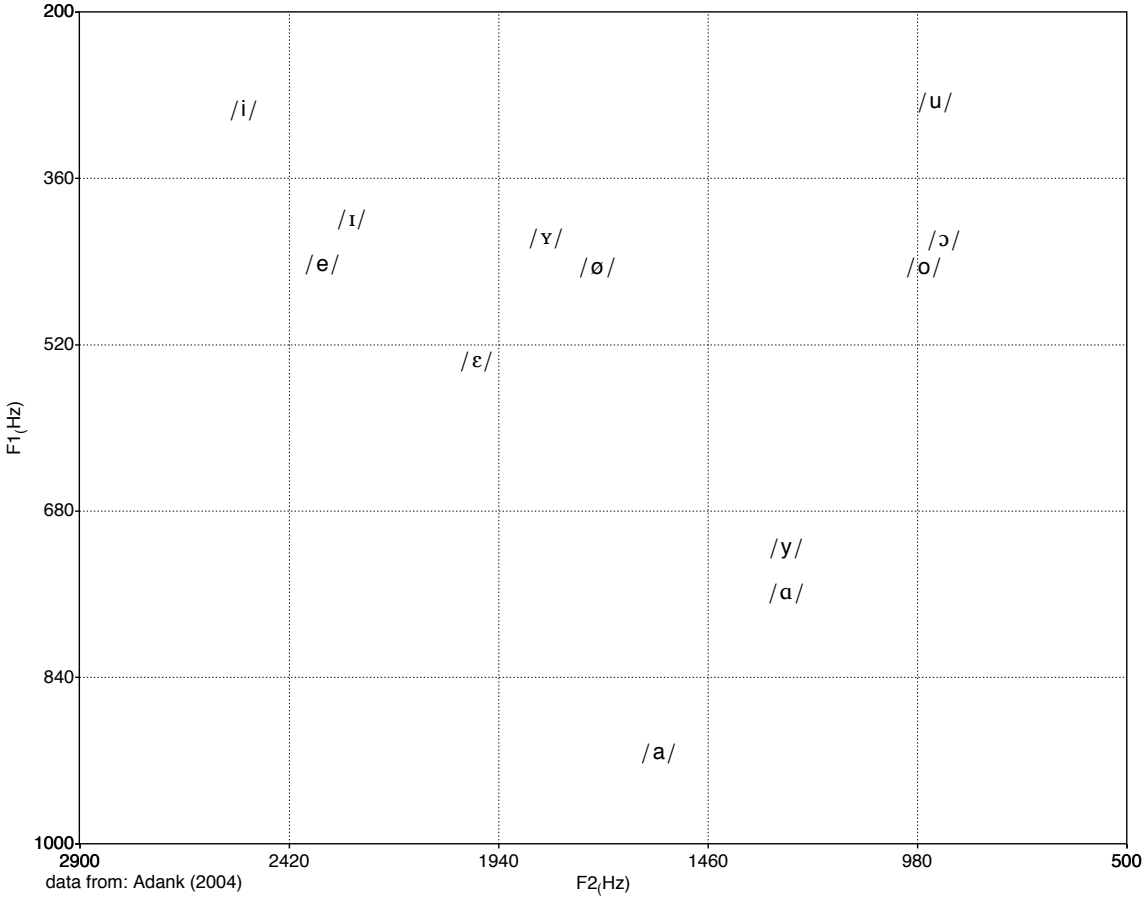


Figure 5: Dutch (NSD) F2 vs. F1 values representing acoustic space.

5.4.2 The English vowel system

English vowels are generally divided into three separate inventories: Standard Southern British English (SSBE), Scottish Standard English (SSE) and General American (GA). The current study will focus on the SSBE variant as the native speakers involved in this study may be classified into this category. The following phonemes are part of the SBBE vowel inventory (Giegerich, 1992):

SBBE vowel	CVC(C) example
/i/	'beat'
/ɛ/	'bet'
/æ/	'bat'
/ɪ/	'bit'
/e/	'bait'
/ɑ/	'bath'
/u/	'boot'
/ʊ/	'put'
/ʌ/	'but'
/ɔ/	'bought'
/ɒ/	'cot'
/ɜ/	'bird'
/aɪ/	'bite'
/aʊ/	'bout'
/ɔɪ/	'noise'
/ɪə/	'beard'
/ɛə/	'scarce'
/ʊə/	'gourd'

Table 10: SBBE vowels adapted from: English Phonology: An Introduction (Giegerich, 1992)

Deterding (2004) provides an analysis of monophthongal vowels in the MARSEC[19] database. The measurements are taken from connected speech (e.g. newsreading), with no description of vowel context (CVC, CCVC or otherwise).

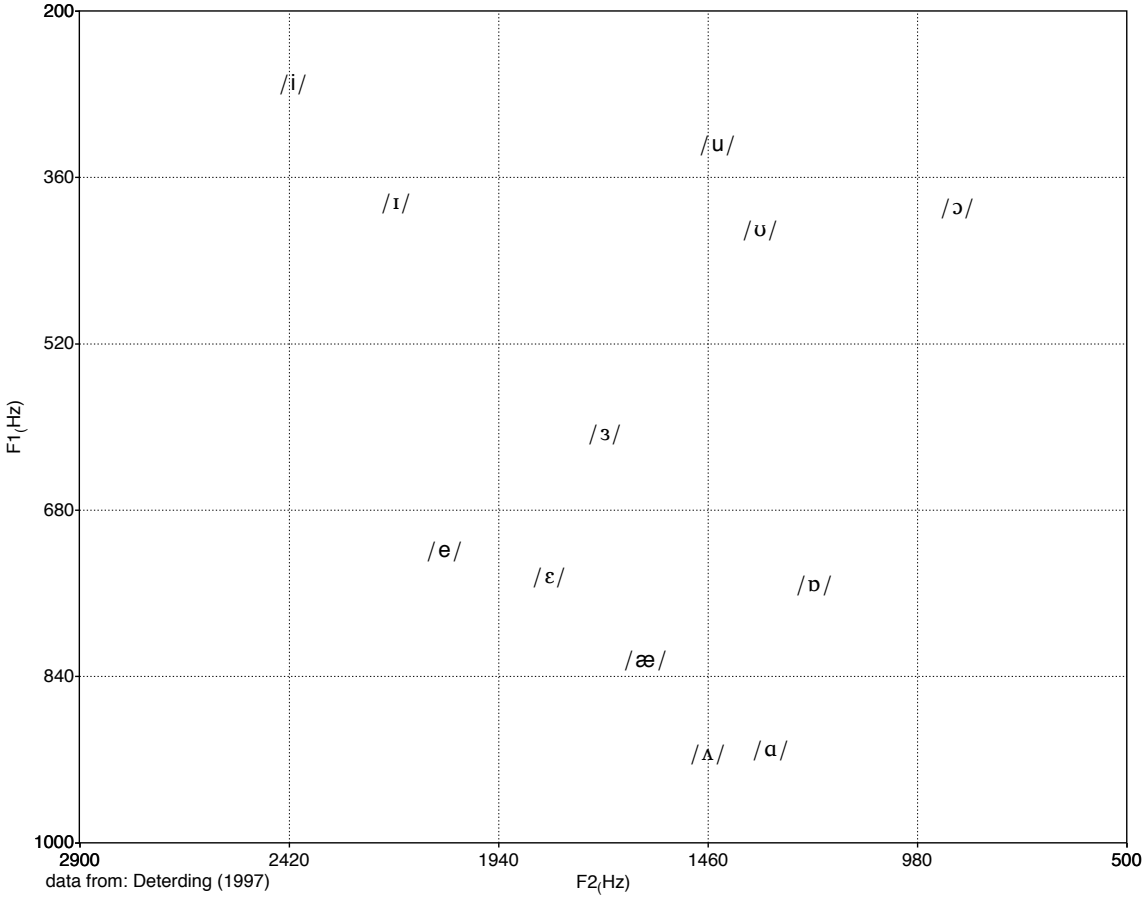


Figure 6: F2 vs. F1 values representing acoustic space in female SBBE speakers (data from Deterding, 1997)

5.4.3 Comparison between Dutch and English vowels

The aim of this section is to ensure selection of vowel sounds that differ in terms of F1, F2 and durational cue values, where Northern Standard Dutch (black) vowels are plotted alongside Standard Southern British English (red) vowels. The scatterplot below shows that /i/, /ɔ/ and /ɪ/ are similar in terms of F1 and F2 value as the NSD and SSBE phonemes are in close proximity to each other, while /ɛ/ is clearly different. The /æ/ phoneme is much closer to Dutch /a/ than Dutch /ɛ/. Faced with these data, it is remarkable that /æ/ is generally mapped to /ɛ/ rather than /a/ (Flege, 1988). It is possible however, that variations in (recording) data could account for some differences or similarities between NSD and SSBE.

[19]see P. Roach, G. Knowles, T. Varadi, and S. Arnfield. (1993) for additional information

Given these observations, the current study will focus on the following phonemes and the formant and durational values as in tables 11 and 12:

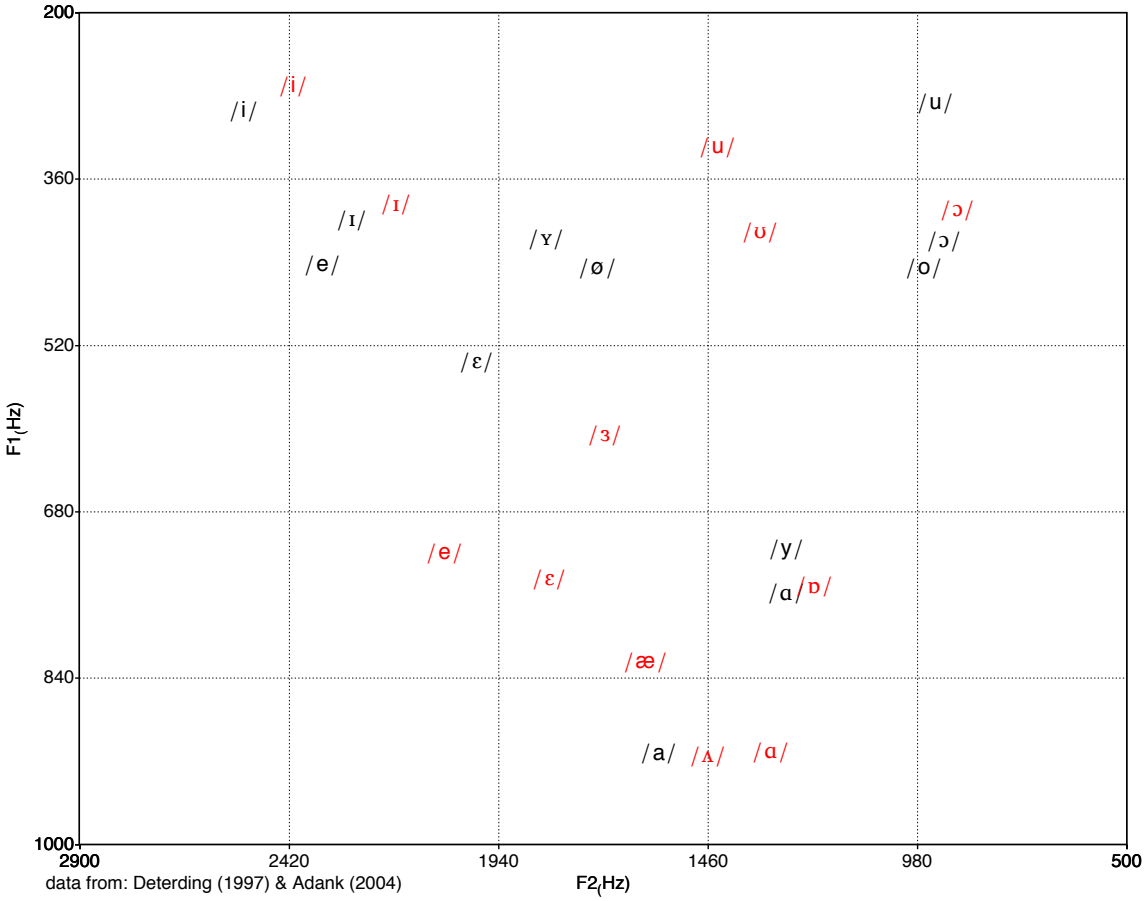


Figure 7: SBBE (red) and NSD (black) vowel contrast

The Dutch /i/ vowel will be contrasted to the similar SSB English /i/ vowel so as to gain insight into which cue is responsible for its identification as belonging to either the NSD /i/ phoneme category or the SSBE /i/ phoneme category.

NSD vowel	CVC example	F1 (Hz)	F2 (Hz)	Duration (ms)
/i/	'Piet' (Peter)	294	2524	92

Table 11: Dutch target vowels spectral/durational cues values (N=20;female) data from Adank et. al., 2004. Cue values taken from measurements using a CVC context.

The table below shows that /ɛ/ and /æ/ differ along tongue-height and tongue backness as well as duration. Early learners in the current study are tested on their perception of these phonemes to establish whether or not the findings from Flege (1988) and Weber & Cutler can be replicated with Dutch early learners of L2 English, and whether a perceptual shift may be noted across a three-month period of English L2 non-native input.

SBBE vowel	CVC(C) example	F1 (Hz)	F2 (Hz)	Duration (ms)
/i/	'beat'	270*	2410*	125**
/ɛ/	'bet'	744*	1823*	132*
/æ/	'bat'	824*	1602*	167*

Table 12: SBBE target vowels average spectral/durational cues values (N=20) *data from Broersma (2005) & **Flege (1990)

5.5 Research questions, hypotheses, predictions

Limitations with respect to the amount of L2 input available to early English learners combined with the logistic dilemma of recruiting qualified native L2 instructors in the Netherlands is likely to affect implementation of EU language learning policy in Dutch primary schools. In this view, it is essential to investigate teacher(trainee) English L2 proficiency and its effect on early L2 learners in a (minimal input) classroom situation. Specifically, the study outlined here will attempt to establish a relationship between quality of phonological input to a perceptual shift in early L2 learners for the phoneme pairs /æ/ - /ɛ/ and /i/ - i/. This requires i. measurement of the input that early learners receive to establish between-teacher differences, and ii. measurement of perceptual discrimination in early English learners on the target phonemes. A measure of the early English L2 learners' receptive vocabulary is also made using a PPVT-4 test to gain an indication of lexical development over the same period of time.

The aim of this study is to:

i. Assess the suitability of the perceptual discrimination task for early learners aged 4 - 6 years,
ii. Test whether there is a relationship between quality of phonological input and a perceptual shift in early L2 learners, given a constant amount of L2 input,
iii. Monitor receptive vocabulary development, given a limited amount of L2 input.

Given previous findings by Gerrits (2001), Heeren (2006) and de Brasileiro (2009) it is unsure as to which cues are likely to be relied upon in the perceptual discrimination task: Gerrits (2001) concludes that 4 and 6-year-olds weighted spectral cues less heavily than older children and adult, while there was no difference in child and adult duration ue reliance. Gerrits concluded that durational cue reliance was acquired before spectrum cue reliance. Heeren (2006) questioned these results and found that 7-year-olds but not 5-year-

olds showed adult-like durational cue reliance. De Brasileiro (2009) concludes that children differ in their cue reliance when compared to adults, and that monolingual Dutch children at the age of 6;5 yrs. showed spectral cue reliance but not durational cue reliance in an adult-like manner. Possible explanations for the distinction between child- and adultlike cue reliance are offered by Sussman (1993, 2001) and who claims that differences are the result of developmental differences in the auditory system. Nittrouer (1996) however, argues that these differences are attributed to child linguistic development; children develop cue weighting on the basis of linguistic experience which leads to a gradual fine-tuning into the child's L1 where a child attends to increasing levels of phonological detail as the cue weighting system develops.

Heeren (2006) notes that children have difficulties in processing small acoustic differences, and she points out the importance of short and positive test experiences when testing children. Walley & Flege (1999) also point out that children (age: 5 yrs.) seems to have more difficulty processing non-native sounds than native sounds and it is suggested that this difficulty arises from processing non-native sounds as native sounds. In this view, children in the current study are likely to have difficulty discriminating between /æ/ and /ɛ/, and following Weber & Cutler (2004) they are likely to interpret /æ/ as being /ɛ/. This interpretation is asymmetrical as /ɛ/ is much less likely to be interpreted as being /æ/. The current study will attempt to replicate this finding and we will also attempt to determine whether or not a shift toward a greater number of 'correct' identifications is possible - given a minimal amount of non-native L2 input. It is hypothesised that the shift is limited to the extent of the input provided (i.e. L2 learners will not shift perceptual boundaries beyond the L2 input provided). The early L2 learners are also tested on their receptive vocabulary skills using the PPVT-4 instrument to monitor lexical development across a three month period. Based on previous findings, it is expected that early learners will significantly increase their receptive vocabulary scores.

The following chapter details the experiments carried as out in the current study and also shows the results obtained.

6. Experiments & Results

To test the hypotheses stated in the previous chapter, the following target language properties will be assessed using a pretest - posttest design:

1. Lexical development using a digital version of the Peabody Picture Vocabulary Test-4;
2. Perceptual discrimination in early English learners for the English phoneme contrast /æ/ - /ɛ/ and Dutch-English phoneme contrast /i/ - /i/;

The time span between pre- and posttest is 3 months during which pupils receive between 5-12 hours of instruction.[20] To establish whether or not there are differences across pupil groups based on phoneme realisations by their student-teachers, the student-teacher's phoneme realisations are measured and contrasted to native speaker phoneme realisations. Details concerning this measurement are provided in section 6.1.1. The perceptual discrimination task is described in section 6.1.2, and a overview of the PPVT-4 digitalisation is provided in section 6.1.3.

Section 6.2 provides on overview of the participants involved in both the perceptual discrimination task and the PPVT-4, and section 6.3 addresses experimental procedures. The final section in this chapter details the results from both the perceptual discrimination task and the PPVT-4.

6.1 Methodology

6.1.1 Native and non-Native English vowel realisations

This section describes the measurement of the target vowels that early learners receive as their L2 input. This is done to establish an index of the /æ/ and /i/ phoneme realisations, contrasting native speaker spectral and durational values to those of non-native student-teacher values. Given the findings by Schutter & Chevalking (2009, ms.) it is expected that there will be a significant between- student-teacher-group differences in the realisation of the phoneme /æ/; Schutter & Chevalking (2009) recorded 3 native British English speakers (age: 21-35 years) and 10 non-native English speakers at B1 proficiency level[21] (age: 18-63) using a CVC imitation task, which yielded a significant native - non-native group difference ($p < .05$ for F1, F2 and duration cue values). No measurement was made for /i/ in the previous experiment, but given the fact that the average spectral values for Dutch /i/ are similar to English /i/ (see Chapter 5, table 11 and 12), it is expected that differences between the native and non-native group will be most salient in the duration cue.

[20] The effect of variation in L2 instruction is taken into account during data analysis.
[21] See Appendix A for a CEF level overview or http://www.coe.int/t/dg4/Linguistic/Source/Framework_EN.pdf for a detailed level description.

6.1.1.1 Participants

The non-native English participants taking part in this experiment are third-year Primary School teacher-trainees participating in an Early English L2 minor course at the Christelijke Hogeschool Ede. Their proficiency levels range from B2 to C1[22], with a discrepancy between receptive (reading/listening) and productive skills (speaking/writing) as shown in the table below:

	Receptive skill level	**Productive skill level**
Teacher 1 - female	B2/C1	B2
Teacher 2 - female	B2/C1	B2
Teacher 3 - female	B2/C1	B2
Teacher 4 - female	B2	B2
Teacher 5 - female	C1	C1
Teacher 6 - female	C1	C1
Teacher 7 - female	B2/C1	B2
Teacher 8 - female	C1	C1
Teacher 9 - female	B2	B2
Teacher 10 - female	B2/C1	B2
Teacher 11 - female	C1	C1

Table 13: Teacher-trainee proficiency levels (CEF)

One female native SSBE speaker was also recorded to establish a native non-native speaker contrast measure.

6.1.1.2 Materials

The participants were recorded in a quiet room using a Samson C03U USB condenser microphone and Audacity 1.3.6 (recording at 441000 kHz, 32 bit) on a Macbook (2.4 Ghz, Mac OS 10.5) laptop. The recordings were analysed using PRAAT 5.1.18. To obtain several samples of each phoneme across different contexts, participants were asked to read aloud a standard text named 'Comma Gets a Cure'.[23] The text is part of the International Dialects of English Archive and was specially composed using J.C. Wells' standard lexical set which allows researchers to examine a reader's English pronunciation across a wide variety of phonemic contexts. Recording a standard text such as 'Comma Gets a Cure' has the advantage of being able to make comparisons with recordings from the IDEA database to cross-examine possible phonetic context effects, should they be noted during analysis. The following target tokens were selected from the text: [ænɪməl], [hæpɪ], [tɹæp], fli:s], [stɹi:t],

[22]These levels were established using a combination of self-evaluation and tutor evaluation provided by The Common European Framework and the Portfolio in the Swiss National Science Foundation project (Schneider, Günther & North, Brian, 2000): Fremdsprachen können – was heisst das? Chur/Zürich, Rüegger.
[23]See http://web.ku.edu/~idea/readings/comma.htm for further details

[si:]. Three tokens for phoneme /ɛ/ were also obtained from the SBBE participant to contrast non-native /æ/ realisations to native /ɛ/.

6.1.1.3 Procedure

Permission was obtained from all participants to use the recording for research purposes. Participants were then provided with the text 'Comma Gets a Cure' to read through to become familiar with the text while the researcher ensures that the recording equipment is in place. The participant then read the text aloud at a steady pace to prevent speech errors. On completion, the participant is asked a few personal questions in English. This is also recorded to control for within-speaker pronunciation differences.

Target items were then extracted from the recordings and the vowel segments were analysed using Praat: F1 (in Hz), F2 (in Hz) spectral values and durational values (in Ms) were thus obtained. The following section provides an overview of the results.

6.1.1.4 Results

The data collected from each participant consisted of the tokens described in section 6.1.1.2 and two additional tokens (one CVC token for /æ/ and one CVC token for /i:/) taken from the 'normal' speech recorded after the 'read aloud' task. A total of 8 tokens per participant was extracted, and F1, F2 and duration values were measured using Praat; each vowel section of the token was selected and measured for duration, and the formant values were measured across the selected duration. These values were then averaged and compared to single point measurement taken by the experimenter, to ensure accurate measurement of cue values. Measurements were carried out twice: once by the experimenter and once by an assistant. This was done to prevent experimenter-bias. It was found that the measurements did not differ significantly between the experimenter and the assistant ($p < .05$), and the values were entered into SPSS to compare participant cue values and establish differences. To meet the assumptions for an ANOVA, tests for normality (Shapiro-Wilk) and homogeneity of variance (Levene) were carried out. The Shapiro-Wilk test revealed that the measurements for the F1, F2 and duration cues for both [æ] and [i:] were normally distributed ($p > .05$)[24] for all but one subject (Subject 3; $p < .05$ for the [i:] duration cue).

[24] H_0: samples are normally distributed population

The Levene's test revealed the following:

Test of Homogeneity of Variances				
	Levene Statistic	df1	df2	Sig.
F1 Value (Hz) - /æ/	1.290	11	36	.270
F2 Value (Hz) - /æ/	2.217	11	36	.036
Duration - /æ/	1.441	11	36	.198
F1 Value (Hz) - /i:/	1.187	11	36	.330
F2 Value (Hz) - /i:/	1.650	11	36	.126
Duration - /i:/	9.000	11	36	.000

Table 14: Homogeneity of variances test

Table 14 shows that the Null-hypothesis of equal variances assumed is rejected for two variables namely: F2 value [æ], and Duration [i:]. Based on these two tests a non-parametric test (Kruskal-Wallis) is used to establish whether or not tokens are produced similarly across participants. The results are presented in table 15:

Kruskal-Wallis Test Statistics						
	F1 - /ae/	F2 - /ae/	Duration- /ae/	F1 - /i:/	F2 - /i:/	Duration - /i:/
Chi-Square	36.293	24.202	21.603	30.665	23.904	13.331
df	11	11	11	11	11	11
Asymp. Sig.	.000	.012	.028	.001	.013	.272

Table 15: Testing equality of cue value medians between participants

The null-hypothesis that the tokens are produced similarly across participants is rejected for F1 - /æ/, F2 -/æ/, Duration - /æ/, F1 -/i:/, F2 - /i:/, but not for duration -/i:/. Thus, except for Duration - /i:/ all cue values are significantly different across participants ($p < .05$). Note that F1 cue values provide the most reliable indication of between-subject differences ($p \leq .001$), while duration is the least reliable indication of between subject differences ($p < .028$). A Bonferroni post-hoc test was carried out for each of the cues for each phoneme, to establish differences between participants. The tables on the following pages provide an overview of the post-hoc analyses:

	NS	NNS 1	NNS 2	NNS 3	NNS 4	NNS 5	NNS 6	NNS 7	NNS 8	NNS 9	NNS 10	NNS 11
NS	X				*					*		
NNS 1		X										
NNS 2			X									
NNS 3			*	X	*					*		
NNS 4					X							
NNS 5		*	*		*	X				*		
NNS 6					*		X					
NNS 7			*		*			X		*		
NNS 8					*				X			
NNS 9										X		
NNS 10					*					*	X	
NNS 11												X
*. The mean difference is significant at the 0.05 level.												

Table 16: Bonferroni post-hoc analysis F1 cue /æ/ phoneme

The table above illustrates that teacher-trainee 2, 4 and 9 produce significantly different F1 values for the /æ/ phoneme, compared to other teacher-trainees and the native speaker. Teacher-trainees 4 and 9 show this difference most clearly. These speakers were found to be at B2 proficiency levels across productive and receptive skills. The following table illustrates between-speaker differences for the F2 cue value of the /æ/ phoneme:

	NS	NNS 1	NNS 2	NNS 3	NNS 4	NNS 5	NNS 6	NNS 7	NNS 8	NNS 9	NNS 10	NNS 11
NS	X											
NNS 1		X										
NNS 2			X									
NNS 3				X						*		
NNS 4					X							
NNS 5						X						
NNS 6							X			*		
NNS 7								X				
NNS 8									X	*		
NNS 9										X		
NNS 10											X	
NNS 11												X
*. The mean difference is significant at the 0.05 level.												

Table 17: Bonferroni post-hoc analysis F2 cue /æ/ phoneme

The table above shows that differences in /æ/ production are sustained for teacher-trainee 9, while other differences concerning the F2 cue value do not reach significance.

	NS	NNS 1	NNS 2	NNS 3	NNS 4	NNS 5	NNS 6	NNS 7	NNS 8	NNS 9	NNS 10	NNS 11
NS	X											
NNS 1		X										
NNS 2			X									
NNS 3				X								
NNS 4					X							
NNS 5						X						
NNS 6							X					
NNS 7								X				*
NNS 8									X			*
NNS 9										X		
NNS 10											X	
NNS 11												X
. The mean difference is significant at the 0.05 level.												

Table 18: Bonferroni post-hoc analysis Duration cue /æ/ phoneme

Table 18 shows few differences, where teacher-trainee 11 only differs significantly from teacher-trainees 7 and 8. No other differences were significant. The following tables detail the teacher-trainee and native speaker differences for the /i:/ phoneme:

	NS	NNS 1	NNS 2	NNS 3	NNS 4	NNS 5	NNS 6	NNS 7	NNS 8	NNS 9	NNS 10	NNS 11
NS	X											
NNS 1		X										
NNS 2			X	*			*		*	*	*	*
NNS 3			*	X								
NNS 4					X		*		*			*
NNS 5						X	*		*			*
NNS 6			*		*	*	X					
NNS 7								X				
NNS 8			*		*	*			X			
NNS 9			*							X		
NNS 10			*								X	
NNS 11												X
. The mean difference is significant at the 0.05 level.												

Table 19: Bonferroni post-hoc analysis F1 cue /i:/ phoneme

The table above shows that F1 cue values for the /i:/ phoneme, teacher trainee 2 differs significantly from most other teacher-trainees. The native speaker does not differ significantly from any of the teacher-trainees. Differences do not - however - correlate to differences in proficiency levels.

For the F2 cue value, no between-speaker differences reached significance. A Tamhane's T post-hoc analysis revealed the following significant differences for the duration cue:

	NS	NNS 1	NNS 2	NNS 3	NNS 4	NNS 5	NNS 6	NNS 7	NNS 8	NNS 9	NNS 10	NNS 11
NS	X			*								
NNS 1		X										
NNS 2			X									
NNS 3	*			X						*		
NNS 4					X							
NNS 5						X						
NNS 6							X					
NNS 7								X				
NNS 8									X			
NNS 9				*						X		
NNS 10											X	
NNS 11												X
. The mean difference is significant at the 0.05 level.												

Table 20: Bonferroni post-hoc analysis F1 cue /i:/ phoneme

To test whether or not there is a significant difference between the SSBE native speaker participant and the non-native teacher trainees, a Mann-Whitney test was carried out:

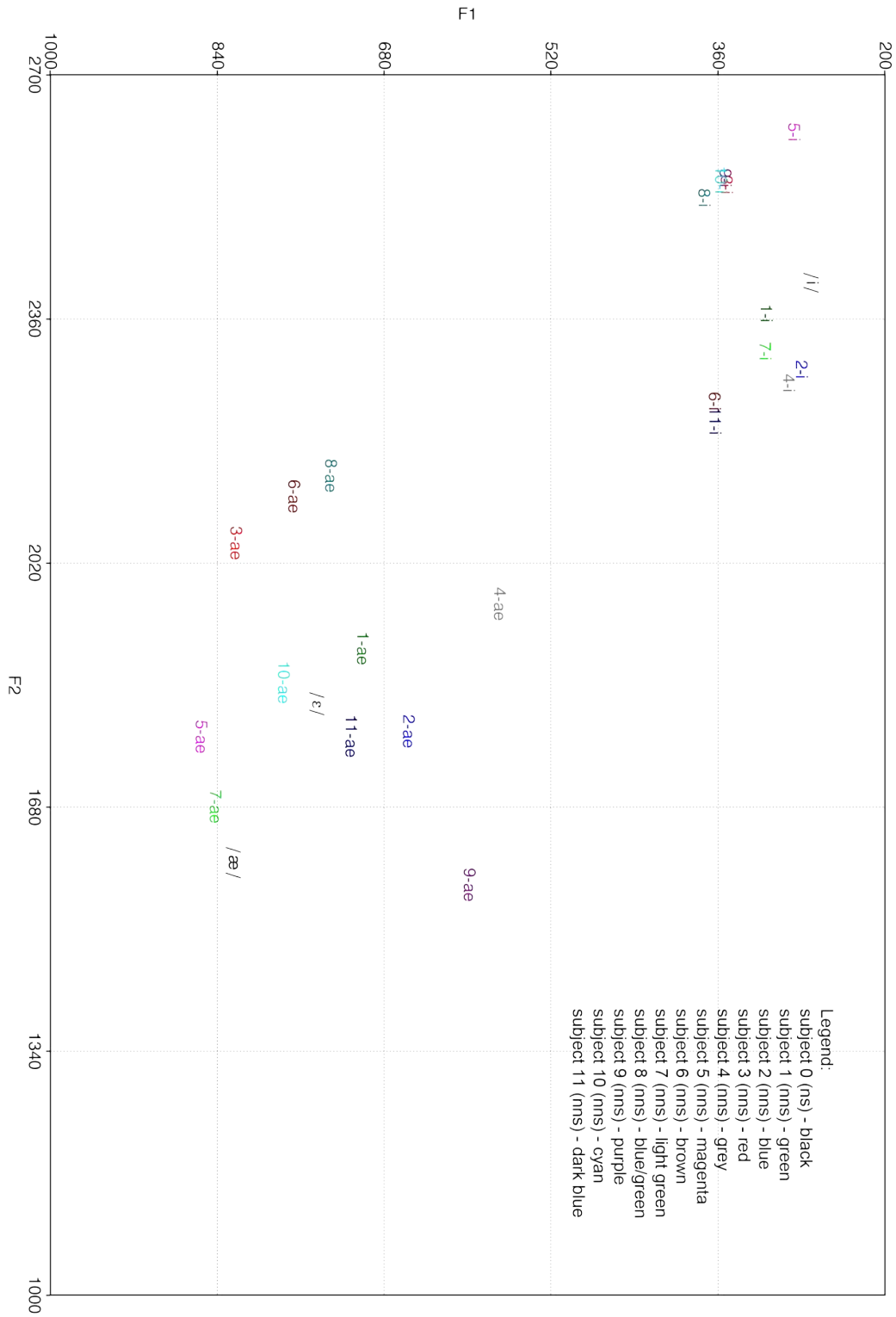
Test Statistics						
	F1 - /ae/	F2 - /ae/	Duration - /ae/	F1 - /i:/	F2 - /i:/	Duration - /i:/
Mann-Whitney U	74.000	78.000	69.500	66.000	80.000	38.000
Wilcoxon W	1064.000	88.000	1059.500	76.000	1070.000	1028.000
Z	-.522	-.373	-.691	-.821	-.298	-1.866
Asymp. Sig. (2-tailed)	.602	.709	.490	.412	.765	.062
Exact Sig. [2*(1-tailed Sig.)]	.627[a]	.733[a]	.505[a]	.437[a]	.787[a]	.063[a]

Table 21: Mann-Whitney test SBBE participant vs. non-native Teacher-trainees. a: Not corrected for ties.

Table 16 shows that the SBBE participant's cue values are not significantly different from non-native teacher-trainees. That is not to say that there are no significant between participant differences (see above), but that the measurements provide no reliable indication of a native - non-native distinction. An increased number of participants and different target tokens (e.g. CVC structures without liquidae [l], [ɹ], or nasals [n] could have provided more reliable between subject differences. As it stands, it may be concluded that there are significant differences between participants' cue values but these differences are less clear-

cut in comparison to the CVC imitation task carried out by Schutter & Chevalking (2009). The measurements taken here, however, are a more accurate representation of natural speech. A post-hoc analysis (Tamhane - equal variances not assumed) revealed a significant difference ($p < .05$) between the native speaker, the teacher-trainees at C1 CEF level and the teacher-trainees at B2 level for the durational cue for the /i/ phoneme only. All other differences were non-significant.

Representing the spectral cue values in a F1 vs. F2 plot illustrates participant differences/ similarities (see the following page). Spectral values for /i/ show relatively little spread (in both F1 and F2 values) when compared to those for /æ/. Note that the non-native spectral cue values for /æ/ approximate /ɛ/ more closely than they do native /æ/, which (visually) confirms the findings by Flege (1988) and Weber & Cutler (2004).


F1
200
360
520
680
840
1000
F2
2700
2360
2020
1680
1340
1000
/ i /
/ ɛ /
/ æ /
5-i
9-i
10-i
8-i
1-i
7-i
2-i
4-i
6-i
11-i
8-ae
6-ae
3-ae
4-ae
1-ae
10-ae
11-ae
2-ae
5-ae
7-ae
9-ae
Legend:
subject 0 (ns) - black
subject 1 (nns) - green
subject 2 (nns) - blue
subject 3 (nns) - red
subject 4 (nns) - grey
subject 5 (nns) - magenta
subject 6 (nns) - brown
subject 7 (nns) - light green
subject 8 (nns) - blue/green
subject 9 (nns) - purple
subject 10 (nns) - cyan
subject 11 (nns) - dark blue

6.1.2 Perceptual discrimination task

The effect of student-teacher phoneme realisations (as shown in the previous section) on early English learner groups is determined through a phoneme discrimination experiment. This section addresses the following methodological issues that need to be considered prior to carrying out such a speech perception experiment:[25]

- phonetic context
- stimuli type
- experimental task

Phonetic context is a key issue that requires consideration in speech perception research. Beddor & Gottfried (1995) point out the importance of presenting stimuli in as natural a context as possible, which requires embedding a target phoneme in a larger context. Placing a target phoneme in a larger (linguistic) context does provide participants with more information than is necessary to complete the discrimination task, and participants may in fact rely on cues other than those under study is such a context is provided. Strange et. al. (2001, pg. 1691) point out that "discrimination of American English vowels by adult Japanese speakers would vary significantly as a function of the contexts in which they were produced and presented". Embedding a target phoneme thus adds to the naturalness of the stimulus, and results in complete representation of a target phoneme (i.e. not its peak/mean formant values), but contextual cues also distract from the target phoneme and its duration and/or spectral values. In other words, presenting a target phoneme in isolation forces participants to rely on an abstract representation of a given phoneme. Conversely, this does not represent the whole L2 input that L2 learners are faced with.

Another key issue to take into consideration is the type of stimulus (natural or synthetic) used in an experiment. Synthetic stimuli may not represent all the relevant properties of a given phoneme, while natural stimuli do not enable control over phoneme variations (Beddor & Gottfried (1995): cues cannot be controlled independently. Werker & Lalonde (1988) show that while participant accuracy may differ between using natural or synthetic stimuli, similar cross-language differences are obtained using both stimulus types.

Experimental task is yet another key issue in speech perception research. Discrimination tasks are used to measure a listeners' ability to distinguish between stimuli. An ABX discrimination task, where listeners indicate whether stimulus X is identical to either response A or B, is commonly used in categorical perception research (e.g. Pisoni, 1975; Schouten & Hessen, 1992). A potential disadvantage of the ABX discrimination task lies in its design: There is an unequal time difference between responses A and B in relation to stimulus X, and careful consideration is required in determining the interval between stimuli/ responses in this design. Werker & Logan (1985) point out that long interstimulus intervals

[25]For a review of methodological issues in cross-language speech perception, see Beddor & Gottfried (1995)

(1500ms) test (adult) participants' phonemic perception, whereas shorter inter stimulus (500ms) intervals test phonetic perception. The ABX design is also reported to have high memory demands in contrast to a 4IAX task that is often used in speech perception research. Heeren (2006) showed that the 4IAX task was too difficult for 6-year-old Dutch children - especially when cue differences between stimuli was small - while Brasileiro (2009) successfully employed an XAB experimental design (which is an adaptation of the ABX design) for subjects ranging from 3;5 to 7;1 years old. Subjects in Brasileiro's study were presented with three tokens: stimulus X followed by responses A and B, where subjects had to decide whether X matches A or B. In the experiment, X is a randomly selected token from a synthetic vowel continuum and A and B are beginning and end points of the same continuum.

The current study aims to establish cue reliance for Early English L2 learners, and in order to test participans' phoneme representations, vowel sounds will be presented in isolation. In order to test cue reliance, a synthetic vowel continuum is generated ranging from /æ/ to /ɛ/ and from English /i/ to Dutch /i/. These stimuli are presented to participants using an XAB experimental similar to Brasileiro (2009). To motivate children (Heeren, 2006) to complete the task, the experiment is presented in a 'game' format. The following sections will outline the experiment in further detail.

6.1.2.1 Materials

Flege et. al. (1997) investigated the effect of English language experience on non-native speaker production and perception of English vowels. It was found that experienced non-native subjects produced and perceived English vowels more accurately than inexperienced non-native subjects. Non-native subjects' degrees of accuracy in producing and perceiving English vowels were also related. Results for the speech perception experiment (Flege, 1997) were obtained by manipulating F1 and F2 formant values in 11 steps ranging from /i/ to /ɪ/ and from /æ/ to /ɛ/. Vowel duration was varied in three steps for each contrast, while the F1 and F2 continuum was divided into 11 steps. Escudero & Boersma (2004) use a design where all cues are manipulated in an equal number of steps as shown in the ilustration:

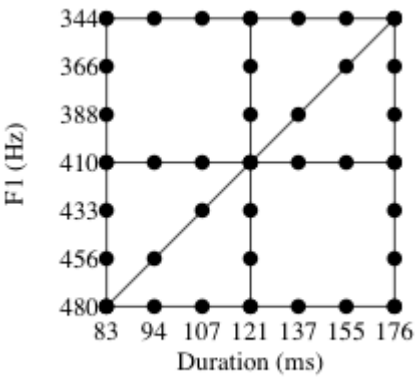


Fig. 1. The 37 stimuli.

The stimuli in Boersma and Escudero's (2004) experiment were synthetic vowels representing a continuum between /i/ and /ɪ/. The vowels were synthesised using data from natural speech measurements where the (average) F1 value for /ɪ/ was 480 Hz and the (average) F1 value for /i/ was 344 Hz. The F2 value for /ɪ/ was 1890 Hz and 2328 Hz for /i/. Duration values were 83 ms and 176 ms for /ɪ/ and /i/ respectively. These averages were used as beginning and end points for the /i/ - /ɪ/ continuum, and a total of 37 stimuli were created to represent the continuum. This study will use a similar method to create stimuli (i.e.

cue values are increased in equal steps), but the number of stimuli will be greatly reduced to accomodate the shorter concentration span in children. The cue value differences between stimuli in the continuum will therefore be much larger, which is in keeping with Heeren's (2006) finding that children are less sensitive to small cue value differences compared to adults.

The previous chapter showed cue values for SSBE as measured by Broersma (2005). Given the fact that the current study involved only one native speaker, data collected by Broersma (2005) will be used as beginning and end-points for the /æ/ - /ɛ/ continuum, and data from Adank (2004) and Broersma (2005) will be used as beginning and end points for the /i/ - /i:/ continuum. The SSBE values used here have been cross-checked against measurements by Schutter & Chevalking (2009) and the SSBE measurements in section 6.1.1, to ensure that the data sets are comparable in terms F1, F2 and duration cue values. These data are repeated in the table below:

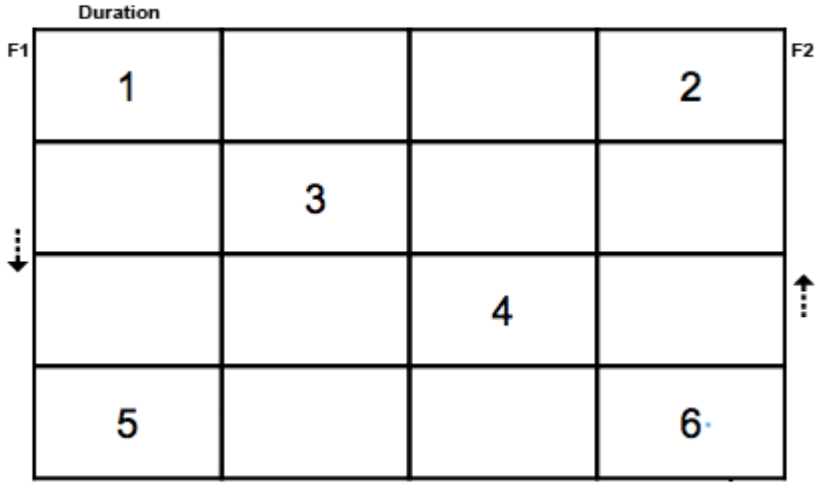
/i/ - /i:/ continuum	**CVC example**	**F1 (Hz)**	**F2 (Hz)**	**Duration (ms)**
/i/	'Piet' (Peter)	294	2524	92
/i:/	'beat'	270	2410	125
/æ/ - /ɛ/ continuum	**CVC(C) example**	**F1 (Hz)**	**F2 (Hz)**	**Duration (ms)**
/æ/	'bat'	824	1602	167
/ɛ/	'bet'	744	1823	132

Table 22: Start and end-point continuum values

A Praat script (see Appendix C) is used to create vowel continua, where F1, F2 and duration values vary in four equal steps. To model female speech as closely as possible a fundamental frequency was set to taper from 200 Hz at vowel onset to 150 Hz at vowel offset while F3, F4 and F5 values are generated as covariates of F2 values using the following algorithm:

f3 = max (2500) f2 + 1000
f4 = max (3500) f3 + 400
f5 = max (4000) f4 + 600

The generated continuum is represented as follows:



As can be seen from the figure, the number of stimuli used in the current experiment is much lower and the differences between cues is therefore much greater. Brasileiro (2009) notes that children make less use of speech cues when compared to adults and Heeren (2006) shows that children are less sensitive to small acoustic differences. Cue value differences in the stimuli for the current study are therefore divided into fewer steps, yielding larger differences between stimuli. A consequence of this will be a less detailed image of child phoneme perception, but the likelihood of noting changes in sensitivity across a set period is maximised.

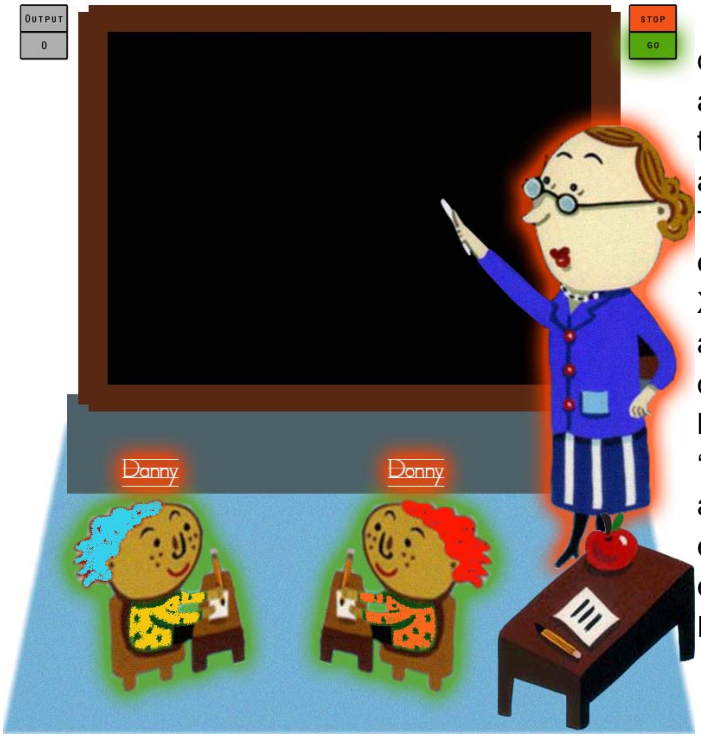
The tables below show actual cue values as generated using the Praat script:

	stimulus	**F1**	**F2**	**duration**
/ɛ/	1	744	1823	0.132
	2	744	1823	0.167
	3	770.14	1746.92	0.143
	4	796.80	1673.28	0.154
	5	824	1602	0.132
/æ/	6	824	1602	0.167

Table 23: /ɛ/ - /æ/ continuum cue values

	stimulus	F1	F2	duration
/i:/	1	270	2410	0.125
	2	270	2410	0.092
	3	277.92	2447.52	0.113
	4	285.92	2485.52	0.102
	5	294	2524	0.125
/i/	6	294	2524	0.092

Table 24: /i:/ - /i/ continuum cue values



These stimuli are placed into an XAB design, where X is a random choice of any of the six stimuli as shown in the tables on this and the previous page, and A and B are either stimulus 1 or 6. To add a playful element to the experiment, the task is set up such that X represents the teacher and A and B are the pupils. Participants have to decide who imitates the teacher best, by clicking on either 'Danny' or 'Donny'. Mouth movements are animated to add to the overall game effect. Section 6.1.2.3. will describe the exact procedure used in testing early English L2 learners.

6.1.2.2 Participants

The table below provides an overview of the participants involved in the current study:

Teacher-trainee*	Subjects	Gender	Age	Group	L1	L2	Hrs. instruction
Participant 1	N = 14	M - 9 F - 5	avg: 4;9 min: 4;0 max: 6;1	1, 2	14 Dutch	2 German	5
Participant 2	N = 13	M - 6 F - 7	avg: 4;6 min: 4;0 max: 5;0	1, 2	12 Dutch 1 Turkish	1 Dutch 1 Turkish	6

Participant 3	N = 24	M - 11 F - 13	avg: 6;7 min: 5;1 max: 7;7	2, 3	24 Dutch		6
Participant 4	N = 15	M - 8 F - 7	avg: 4;7 min: 4;0 max: 6;0	1, 2	15 Dutch	3 English	10
Participant 5	N = 21	M - 10 F - 11	avg: 6;3 min: 5;0 max: 7;3	1, 2, 3	20 Dutch 1 Berber	1 Dutch	6
Participant 6	N = 5	M - 2 F - 3	avg: 6;8 min: 6;0 max: 8;1	3, 4	5 Dutch		6
Participant 7	N = 9	M - 3 F - 6	avg: 4;6 min: 4;2 max: 5;1	1, 2	6 Dutch 1 Moroccan 1 Turkish 1 Bosnian	1 French 2 Dutch 1 English 1 Moroccan	12
Participant 8	N = 10	M - 4 F - 6	avg: 5;5 min: 5;0 max: 6;0	2	10 Dutch		7

Table 25: Participant data. *denotes teacher trainee subjects as shown in table 13

The subject numbers in the leftmost column correspond to the subject numbers as detailed in section 6.1.1.1. A total of 111 ‘Early’ English L2 learners (53 male; 58 female) participated in the perceptual discrimination task. The mean age was 5;6 years, ranging between 4;0 years to 8;0 years. The L1 is Dutch for a large portion (106 children), and 13 children speak an L2 at home. Most children are monolingual Dutch speakers, who get little English input outside the classroom. All children were asked as to whether they watched English TV programmes or played English spoken computer games, and 16 children claimed to receive English input on a regular basis (more than 1 hour each week) via media, primarily through a TV series called ‘Dora’ and through cartoons. The overall average of English input via media amounted to 15 minutes per week. This should, however, be regarded as a rough estimate as parents were not consulted on this matter.

6.1.2.3 Procedure

The ‘Danny & Donny’ game setup is an adaptation of the perceptual discrimination task carried out by de Brasileiro (2009). The game features a teacher (female character), and two pupils named ‘Danny’ and ‘Donny’. These characters are introduced at the start of the experiment using animation and an English voiceover. The aim of the task is then explained to participants. Participants were asked to listen to what the teacher said and then say which of the pupil characters (Danny or Donny) repeated the teacher most accurately. A trial run preceded the actual experiment to ensure that the task was understood by participants. The

trial used (synthetic) vowels /i/ and /u/, based on cue values from Deterding (1997), using the same fundamental frequency taper (200 - 150 Hz) as generated for the actual test vowels. Once participants understood the task, the actual experiment was started. Participants were required to react to 24 stimulus sets: the teacher uttered a target vowel sound 24 times, and participants evaluated Danny and Donny's responses 24 times. After a decision had been made as to who imitated the teacher best, encouraging feedback (English) was provided through a voiceover. The interstimulus interval was set to 1500 milliseconds, to tap into the phonemic perception rather than phonetic or acoustic perception (Werker & Logan, 1985). The interval between each stimulus set was manually advanced by either the teacher or the child. It was left up to the experimenters (the teacher-trainees) to decide whether children could operate the computer. A protocol was created to ensure that all nine experimenters carried out the experiment in similar fashion (see Appendix B).

The experiments were carried out in schools using school computers. Experimenters were provided with a 2 Gigabyte Kingston USB memory stick that contained both discrimination tests and the PPVT-4 test. In addition, experimenters were given headphone splitter cables to plug two headsets into a single computer so that both participant and experimenter could listen to what was said. Headphones were present at most schools, but experimenters could take an extra headset along to be sure that the tests could be carried out. Upon completion of the discrimination experiment, experimenters would save the experiment data to the USB sticks that were provided.

6.1.3 PPVT-4

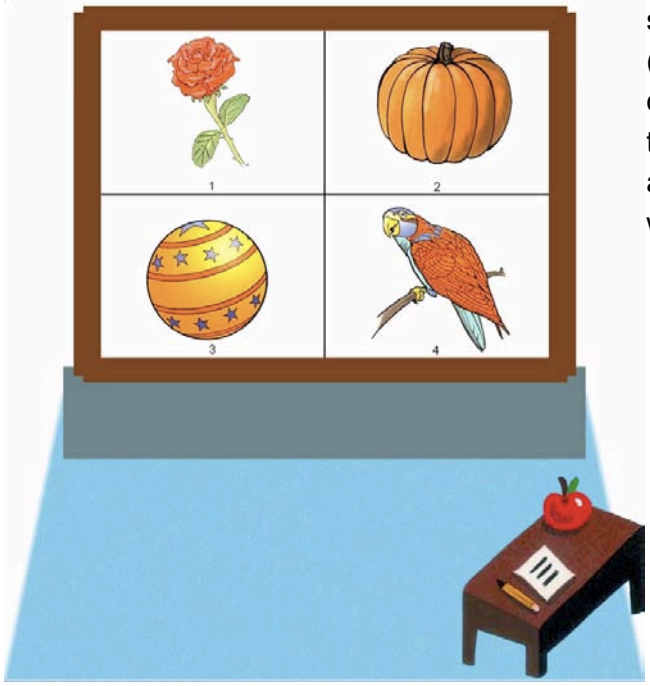
In addition to a perceptual discrimination task, measurement was made of participants' receptive vocabulary size using the Peabody Picture Vocabulary Test-4. For this study the PPVT-4 version of the test will be used in computerised form to monitor vocabulary development over a 3-month period (between 5-12 hours of instruction). Unsworth (2008, ms.) found that pupils aged 5 to 7 years (N=23) showed significant receptive vocabulary development across a similar timespan, using an English translation of a Dutch PPVTest. An actual English PPVT-4 test was used in this study, to ensure that the instrument's validity was not jeopardised through translation. Additionally, a native English PPVT-4 test provides a measure of English receptive vocabulary, which is essentially what early L2 English learners are in the process of acquiring. Testing receptive vocabulary development provides an indication of participants' language development (Dunn & Dunn, 2007) and Werker et. al. (2002 pg. 26) suggest that (for infants) the 'ability to attend to phonetic detail in words is one of the factors that leads to the increase in rate of vocabulary acquisition'.

6.1.3.1 Materials

Haaf et. al. (1999) compared a paper PPVT version to a digitalised version of the same test and found the different versions to be statistically equivalent. There is, however, no official digitalised format available so a paper PPVT-4 version was digitalised for the current study. This was done for various reasons:

- A PPVT-4 test is generally used in an L1 testing situation where inter-rater differences are much less likely in comparison to the current situation where teacher-trainees are both participant (see section 6.1.1) and experimenter;
- To prevent experimenter bias, and to guarantee similar testing conditions across schools;
- To ease the work of the experimenters involved.

To create a digital version of the PPVT-4 test, all images were scanned at 300 dpi and batch-resized to 583 x 430 pixels @ 72 dpi using Photoshop CS 4 on a MacBook 2.4 Ghz. A native SSBE speaker was asked to read aloud all PPVT target words in a quiet room. This was recorded at 441000 kHz, 32 bit using Audacity 1.3.6 and a Samson C03U USB condenser microphone. The recording was normalised to ensure equal volumes between words. Following this, the recording was cut into 228 target words and two seconds of silence was added to each word. Each sound file was saved as a .wav file to ensure cross-platform compatibility. The background that was used in the discrimination experiment was re-used and the blackboard now showed the four-picture quadrants. Keynote '08 was used to create a 228-page slideshow which was exported to an Adobe flash file, to make sure that school PC's could run the test (flash is a common browser plugin, and does not require costly software licenses to run). The Adobe Flash file was then added to the USB stick that experimenters were provided with.



6.1.3.2 Participants

Early English L2 learners from groups 1 to 4 (age: 4 - 7 years) are tested. A total of 122 early English L2 learners (57 male; 65 female) participated in the PPVT-4 task. Note that more children took part in the PPVT-4 task than the discrimination experiment (N = 111). This was due to technical difficulties encountered during the discrimination experiment.

6.1.3.3 Procedure

A protocol (see Appendix B) ensures that the same testing procedure is used between experimenters. The basic setup is the same as that used for the perceptual discrimination task; participants are tested at school using a school PC, and both experimenter and participant wear headphones throughout the task. The PPVT-4 test offers two training items to ensure that participants are familiar with the task of selecting the correct picture and pointing to it in the quadrant containing four images - after hearing the corresponding word.

The PPVT-4 test consists of 19 sets of 12 picture quadrants. Assuming that participants in the current study have little or no experience with English, it was decided to start with the first set of test items rather than the set that is appropriate for the age of a particular participant. The PPVT-4 manual (Dunn & Dunn, 2007) recommends this (non-standard) procedure for participants with very low ability. Thus, the basal set is set 1 for all participants in the current study. Participants progress through the sets until a total of 8 or more errors per set is noted. This is the ceiling set, and the task is stopped after completion of the entire 12 item set. During the task, experimenters note scores and participant data on a PPVT-4 standard form. Raw scores were calculated by subtracting the number of errors from the total number of completed test items.

6.2 Results

Both phoneme discrimination tasks and the PPVT-4 were carried out twice: the first session took place during the second and third week of March 2009 after an average of 2 hours of English L2 instruction, and the second session took place in the third week of June 2009 after an average of nearly 8 hours of English L2 instruction. For each task, descriptives for session one and two are first presented separately, after which they are contrasted to establish differences between the two testing sessions.

6.2.1 Results PPVT-4 task

A complete participant overview for this task is shown in section 6.1.3.2. A total of 122 children took part in the first test session, and 104[26] children were involved in the second test session. The 104 children taking part in the second test session took part in both test sessions. For the first PPVT-4 test session, the following independent variables were used for data collection:

[26]This decrease is the result of a) the difficulty of testing children near the end of a school-year when there are numerous other activities to be completed.

- Teacher-trainee
- Gender
- Participant age
- Class
- L2 input outside the classroom
- L1

The dependent variable is 'raw score'.[27] The resulting overall average PPVT-4 raw score (total number of test items completed - no. of errors) is 18.93. Data from the first test session was normally distributed (Kolmogorov-Smirnov $p > .05$). The tables below provides an overview of the raw scores grouped across the independent variables stated above:

Teacher-trainee	**Mean raw score**
1 - no. of pupils = 14	23.5
2 - no. of pupils = 13	14.92
3 - no. of pupils = 24	19.25
4 - no. of pupils = 15	16.73
5 - no. of pupils = 21	15
6 - no. of pupils = 5	16.2
7 - no. of pupils = 9	18
8 - no. of pupils = 10	20.5
9 - no. of pupils = 11	28.27

Table 26: Mean PPVT-4 raw scores across Teacher-trainee

The assumption of homogeneity of variance for an ANOVA was not met for the different teacher-trainee groups as a Levene's test for equality of variances revealed significant (Levene statistic = 3.332; df = 8; $p < .01$) variances between the different teacher-trainee groups. Consequently, a nonparametric Kruskal-Wallis test revealed a small significant difference between the teacher-trainee groups ($Chi^2 = 15.722$; df = 8; $p < .05$). A post-hoc analysis (Tamhane's T) showed that teacher-trainee group 9 differed significantly from teacher-trainee group 5 ($p < .05$). No other teacher-trainee groups differed significantly.

[27]A standard score is also calculated but not provided in the results, as the standard score is derived from an index particular to native English speakers. The standard score is therefore unlikely to provide an accurate indication for L2 learners.

Gender	Mean raw score
Female (n=65)	19.86
Male (n=57)	17.88

Table 27: Mean PPVT-4 raw scores across Gender

The assumption of homogeneity of variance for an ANOVA was met for the Male and Female groups (Levene statistic = 1.038; df = 1; $p > .05$), but no statistically significant differences for Gender was found ($F(1,122) = 1.038$; $p > .05$).

Age	Mean raw score
4 (n=20)	11.53
5 (n=50)	20.1
6 (n=39)	23
7 (n=11)	21.5
8 (n=2)	15

Table 28: Mean PPVT-4 raw scores across Age

Grouping the ages by year (i.e. age 4;3 is categorised as 4) the assumption of homogeneity of variance for an ANOVA was met for the different age groups as a Levene's test for equality of variances revealed no significant variances between the different age groups (Levene statistic = 1.208; df = 4; $p > .05$). An ANOVA revealed a significant effect for age ($F(4,117) = 3.909$; df = 4; $p < .01$; $\eta_p^2 = .118$). A Bonferroni post-hoc analysis showed that age-group 4 differed significantly from age-group 5 ($p < .05$), and age-group 4 also differed significantly from age-group 6 ($p < .01$). No other between group differences reached statistical significance.

Class	Mean raw score
group 1 (n=43)	17.67
group 2 (n=56)	19.75
group 3 (n=20)	19.55
group 4 (n=3)	17.67

Table 29: Mean PPVT-4 raw scores across Class

The assumption of homogeneity of variance for an ANOVA was met for the different classes (Levene statistic = .624; df = 3; $p > .05$). An ANOVA revealed no significant effect for Class ($F(3,118) = .404$; df = 4; $p < .05$).

L2 input via Multimedia	Mean raw score
No L2 input	18.51
Occasional L2 input (> 1 hr./week)	21.75

Table 30: Mean PPVT-4 raw scores across L2 input via multimedia

The assumption of homogeneity of variance for an ANOVA was met for the 'no multimedia' and 'occasional multimedia' groups (Levene statistic = .103; df = 1; p > .05), but an ANOVA revealed no significant effect for L2 input via Multimedia (F (1,120) = 1.528; df = 1; p < .05).

L1	Mean raw score
Dutch (n=117)	19.08
Berber (n=2)	14.5
Turkish (n=2)	12.5

Table 31: Mean PPVT-4 raw scores across L1

The assumption of homogeneity of variance for an ANOVA was met for the different L1 groups (Levene statistic = .715; df = 3; p > .05), but an ANOVA revealed no significant effect for L1 (F (3,118) = .515; df = 3; p < .05).

For the second PPVT-4 test session, the following independent variables were used in data collection:

- Teacher-trainee
- Gender
- Participant age
- Class
- L2 input inside the classroom
- L2 input outside the classroom
- L1

The dependent variable 'raw score'[28] represents the results obtained in the second test session (N = 104). The resulting overall average PPVT-4 raw score (total number of test items completed - no. of errors) is 23. Data from the second test session was normally distributed (Kolmogorov-Smirnov p > .05). The tables below provides an overview of the raw scores grouped across the independent variables stated above. Please note that differences

[28]A standard score is also calculated but not provided in the results, as the standard score is derived from an index particular to native English speakers. The standard score is therefore unlikely to provide an accurate indication for L2 learners.

between test sessions one and two will be provided following the presentation of the results from the second test session:

Teacher-trainee	Mean raw score
1 - no. of pupils = 11	21.82
2 - no. of pupils = 13	17.54
3 - no. of pupils = 24	22
4 - no. of pupils = 15	20.87
5 - no. of pupils = 19	19
7 - no. of pupils = 7	35.14
8 - no. of pupils = 10	28.8
9 - no. of pupils = 5	37.6

Table 32: Mean PPVT-4 raw scores across Teacher-trainee

The assumption of homogeneity of variance for an ANOVA for the different teacher-trainee groups was met (Levene statistic = .400; df = 7; $p < .05$). An ANOVA revealed a significant effect for teacher-trainee group ($F(7,96) = 4.696$; df = 7; $p < .001$; $\eta_p^2 = .225$). Post-hoc analysis (Bonferroni) showed that teacher-trainee group 9 differed significantly from teacher-trainee group 2 ($p < .01$), teacher-trainee group 4 ($p < .05$), and teacher-trainee group 5 ($p < .01$). Marginally significant was the difference between teacher-trainee group 9 and 3 ($p = .052$). No other teacher-trainee groups differed significantly.

Gender	Mean raw score
Female (n=57)	23.44
Male (n=47)	22.47

Table 33: Mean PPVT-4 raw scores across Gender

Homogeneity of variance can be assumed, based on Levene's test (Levene statistic = .09; df = 1; $p < .05$). An ANOVA revealed no significant effect for Gender ($F(1,102) = .196$; df = 1; $p > .05$).

Age	Mean raw score
4 (n=15)	15.87
5 (n=45)	24.93
6 (n=34)	29.25
7 (n=10)	18.75

Table 34: Mean PPVT-4 raw scores across Age

For the different age groups presented here, homogeneity of variance may be assumed

(Levene statistic = 1.804; df = 3; p < .05). An ANOVA showed a marginally significant effect for Age (F (3,100) = 2.526; df = 3; p = .062; η_p^2 = .07). Post-hoc analysis (Bonferroni) showed a significant difference between age group 4 and age group 5 (p < .05), and age group 4 and age group 6 (p < .05).

Class	**Mean raw score**
group 1 (n=34	19.76
group 2 (n=52	25.29
group 3 (n=18)	22.5

Table 35: Mean PPVT-4 raw scores across Class

Homogeneity of variance may be assumed across classes (Levene's statistic = .159; df = 2; p > .05). An ANOVA revealed no significant effect for class (F (2,101) = 2.653; df = 2; p = .075; η_p^2 = .05). Post-hoc analysis (Bonferroni) did not reveal any significant group differences.

Hrs. L2 input classroom	**Mean raw score**
5 (n=26)	21.27
6 (n=61)	21.39
7 (n=10)	28.8
12 (n=7)	35.14

Table 36: Mean PPVT-4 raw scores across hrs. of L2 classroom input

A Levene's test revealed significant homogeneity of variance (Levene's statistic = .067; df = 3; p > .05), and an ANOVA revealed a significant effect for hours of classroom L2 input (F (3,100) = 4.828; df = 3; p = .004; η_p^2 = .127). Post-hoc analysis showed significant differences between 5 and 12 hours of classroom input (p < .05) and between 6 and 12 hours of classroom input (p < .01).

L2 input via multimedia	**Mean raw score**
No L2 input (n=89)	22.71
Occasional L2 input (n=15)	24.73

Table 37: Mean PPVT-4 raw scores across L2 input via multimedia

A Levene's test shows significant homogeneity of variance (Levene's statistic = .01; df = 1; p > .05), but no effect for L2 input via multimedia was found using an ANOVA (F (1,102) = .425; df = 1; p > .05).

L1	Mean raw score
Dutch (n=99)	22.67
Berber (n=2)	26.5
Turkish (n=2)	25

Table 38: Mean PPVT-4 raw scores across L1

Homogeneity of variance can also be assumed for the groups in L1 (Levene's statistic = .852; df = 2; p > .05), but an ANOVA revealed no effect for L1 (F (3,100) = 1.447; df = 2; p > .05).

Having detailed the PPVT-4 results from test sessions one and two, we will now turn to the comparative analysis. The comparative analysis involves only data from children who took part in both test sessions (N=104) through listwise exclusion. Details are shown in Table 26:

	N	Mean	Std. Deviation
Raw score 1st test	104	18.62	9.59
Raw score 2nd test	104	23	11.09

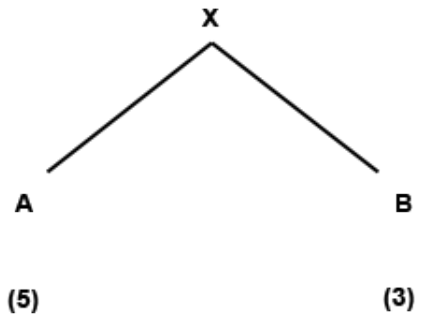
Table 39: Descriptive statistics PPVT-4 task raw scores

To establish whether or not the difference between results from test session one and test session two are significant, a paired samples t-test was carried out. The difference between the two sessions is statistically significant (T = -4.204; df = 103; p < .001). To establish a measure of effect size, Cohen's D was found to be -0.42, which may be regarded as a moderate effect size (Cohen, 1977). These findings will be elaborated in the discussion in the following chapter. We will now turn the the results from the phoneme discrimination tasks.

6.2.2 Results Categorical discrimination tasks

To ensure that only data from attentive subjects who understood the task was included in their analysis, Nittrouer (pg. 280, 1996) selected data from participants with a minimum of 80% accuracy (at the end-points of the continuum) only, and Brasileiro (2009) used a 75% accuracy criterion. Brasileiro (2006) shows that all adults met the set criterion, while leaving 40 of the 130 children out of the data analysis. The current study will remove no data from the data set during analysis, as the teacher-trainees were instructed to abort the task if they

found a child to be consistently unattentive.[29] It is hoped that in doing so the risk of obtaining less consistent results is minimised while including as many tokens in the study as possible.



Each of the four data sets (/ɛ/ - /æ/ discrimination task March & June, /i/ - i:/ discrimination task March& June) contained 'A' or 'B' responses to a token 'X'. A count was made of the total number of responses to each 'X' token, resulting in two scores per token X (see illustration). In order to make a comparison between the different 'X' tokens in the phoneme continuum, the scores obtained for A and B were converted using the following formula:

$$\text{Proportion X} = ((A / (A + B)) - (B / (A + B))) + 5$$ [30]

Adding the values from the diagram above results in a value of 5.25 which indicates an above chance bias toward an /ɛ/ response or /i:/ response, depending on the continuum that is investigated. The calculated value ranges between 4 and 6 for the entire dataset, where 4 represents only /æ/ or /i/ responses to token X and 6 represents only /ɛ/ or /i:/ responses to token X, depending on the phoneme continuum. An equal number of A and B responses to token X (chance level performance) is represented by 5[27].

Based on previous studies by Flege (1988), Weber (2004) and others, it is expected that participants will be less accurate in identifying an /æ/, when compared to identifying an /ɛ/ token. This is a consequence of the fact that /ɛ/ but not /æ/ is part of the Dutch phoneme inventory. Extending this distinction to the /i/ and /i:/ continuum, it is expected that /i/ will be more accurately identified in contrast to the /i:/ phoneme while it is likely that the distinction between phonemes in this continuum is possibly more difficult due to the fact that spectral differences are small, leaving duration as the main cue to distinguish between the tokens.

6.2.2.1 Results Categorical discrimination task /æ/ - /ɛ/

The results for the categorical /æ/ - /ɛ/ discrimination tasks recorded during both test sessions are presented here consecutively. Descriptive data analysis is shown first, followed by an analysis of the effects of the independent variables on the data obtained. This is followed

[29] If participants consistently failed to pay attention to the experiment after repeated encouragement by teacher-trainees, the experiment was aborted for that participant.
[30] The range 4 to 6 was used in data analysis to prevent 0 value scores, which may be interpreted as missing values by SPSS/PASW

by a contrastive analysis to establish differences between the two test sessions and whether or not there is a shift in perception regarding the stimuli in the phoneme continuum.

The descriptive data in the table shows collective child responses from the first test session to the categorical /æ/ - /ɛ/ discrimination task. Note that calculated mean values range between 4 and 6, where 4 represents and /æ/ response and 6 represents an /ɛ/ response. Thus, chance level performance is represented by the value 5.[30]

Session 1	Stimulus	F1	F2	duration	N[31]	Mean Score[32]
/ɛ/	1	744	1823	0.132	78	5.7004
	2	744	1823	0.167	71	5.6714
	3	770.14	1746.92	0.143	76	5.539
	4	796.8	1673.28	0.154	78	4.7402
	5	824	1602	0.132	78	5.0521
/æ/	6	824	1602	0.167	75	4.5356

Table 40: Descriptive statistics discrimination task /æ/ - /ɛ/ first test session

The descriptives shown above reveal mean scores that generally replicate the phoneme continuum (stimuli 1-6) where stimulus 1 is identified as /ɛ/ by showing the highest score (5.7004), and stimulus 6 as /æ/ by showing the lowest score (4.5356). Stimuli 2 through 5 are interpreted less accurately than the end-points of the continuum, with stimulus 5 performing at near-chance level. These results are represented graphically in Appendix D.

To establish statistical significance of above chance level performance, t-tests were carried out using 5 as the test value. As expected, performance on stimulus 5 does not reach significant above-chance performance ($p > .05$), but all other stimuli reveal significant above-chance performance in phoneme identification.

	t	df	Sig. (2-tailed)	Mean Difference
Stimulus 1	11.542	77	0.000	0.70043
Stimulus 2	10.248	70	0.000	0.67136
Stimulus 3	7.983	75	0.000	0.53904
Stimulus 4	-2.78	77	0.007	-0.25983
Stimulus 5	0.553	77	0.582	0.05214
Stimulus 6	-5.356	74	0.000	-0.46444

Table 41: T-test values establishing above-chance test performance across the stimuli test session 1

[31] Note that the number of cases across the stimuli varies as a consequence of the random nature of the experiment.
[32] Calculated mean values range between 4 and 6, where 4 represents an /æ/ response and 6 represents an /ɛ/ response. Thus, chance level performance is represented by the value 5.

The findings are also analysed across the independent variables as detailed in the PPVT-4 results section. These are repeated here:

- Teacher-trainee
- Gender
- Participant age
- Class
- L2 input outside the classroom
- L1

A Levene's test for homogeneity of variance across six teacher-trainee groups showed significant differences between some groups (Levene statistic = 1.675; df = 5; p > .05). A Kruskal-Wallis (nonparametric) test was carried out as a consequence:

	Stimulus 1	**Stimulus 2**	**Stimulus 3**	**Stimulus 4**	**Stimulus 5**	**Stimulus 6**
Chi-square	10.722	9.493	16.162	19.554	40.021	19.584
df	5	5	5	5	5	5
Asymp. Sig.	0.057	0.091	0.006	0.002	0.000	0.001
a. Kruskal Wallis Test; b. Grouping Variable: Teacher-trainee (6 Teacher groups)						

Table 42: Performance differences per stimulus between teacher-trainee groups test session 1

Teacher-trainee group performances do not differ significantly for stimulus 1 and 2. Performance on the other stimuli is significantly different however.

Post-hoc analysis (Tamhane's T) results are shown in the table below:

Stimulus	Teacher-trainee	Teacher-trainee	Mean Difference	Std. Error	Sig.
Stimulus 3	Teacher-trainee 2	Teacher-trainee 5	.58694*	0.13549	0.004
Stimulus 4	Teacher-trainee 2	Teacher-trainee 3	1.06154*	0.23428	0.002
		Teacher-trainee 5	.90696*	0.24162	0.011
		Teacher-trainee 8	1.34744*	0.32133	0.020
		Teacher-trainee 2	-1.06154*	0.23428	0.002
Stimulus 5	Teacher-trainee 1	Teacher-trainee 3	1.26319*	0.17058	0.000
		Teacher-trainee 5	.98016*	0.20527	0.001
		Teacher-trainee 8	1.34524*	0.16057	0.000
	Teacher-trainee 2	Teacher-trainee 3	1.35385*	0.25518	0.002
		Teacher-trainee 5	1.07082*	0.27956	0.018
		Teacher-trainee 8	1.43590*	0.2486	0.001
	Teacher-trainee 3	Teacher-trainee 1	-1.26319*	0.17058	0.000
		Teacher-trainee 2	-1.35385*	0.25518	0.002
		Teacher-trainee 7	-1.19573*	0.23618	0.010
Stimulus 6	Teacher-trainee 1	Teacher-trainee 3	.89530*	0.20218	0.011
		Teacher-trainee 8	.83974*	0.21691	0.021
	Teacher-trainee 2	Teacher-trainee 3	1.10043*	0.27756	0.027
		Teacher-trainee 8	1.04487*	0.28846	0.040
	Teacher-trainee 3	Teacher-trainee 1	-.89530*	0.20218	0.011
		Teacher-trainee 2	-1.10043*	0.27756	0.027
*. The mean difference is significant at the 0.05 level.					

Table 43: Post-hoc analysis differences between teacher-trainee groups

The table shows consistent differences between teacher-trainee 2 and teacher-trainees 3, 5 and 8 for stimulus 4, 5 and 6. It also shows a significant difference between teacher-trainee 1 and teacher-trainees 3, 5, and 8 for stimulus 5. Dividing the data into male and female groups, Levene's test showed no significant differences in variance for any of the stimuli ($p > .05$). An ANOVA revealed no significant effect for Gender across stimulus 1 through 6 ($p > .05$). To establish an effect for participant age, a Levene's test was used as a precondition for carrying out an ANOVA. Homogeneity of variance could not be established across all 4 age groups for stimulus 5 (Levene statistic = 6.752; df = 3; $p < .05$) and stimulus 6 (Levene statistic = 5.079; df = 3; $p < .05$). A Kruskal-Wallis test was therefore used for these stimuli, and an ANOVA was used to establish the effect for participant age in stimulus 1 to stimulus 4. The effect of age for stimulus 1 ($F(3,77) = 1.421$; df = 3; $p = .243$), stimulus 2 ($F(3,70) = 0.91$; df = 3; $p = .441$), and stimulus 3 ($F(3, 75) = 0.456$; df = 3; $p = .714$) is not significant. It does reach significance for stimulus 4 ($F(3,77) = 3.004$; df = 3; $p = .036$). The Kruskal-Wallis

test revealed a significant age effect for stimulus 5 (Chi^2 = 22.440; df = 3; p = .000) and 6 (Chi^2 = 7.844; df = 3; p < .049). The table below illustrates findings using a post-hoc analysis (Tamhane's T):

	Age (yrs.)	**Age (yrs.)**	**Mean Difference**	**Std. Error**	**Sig.**
Stimulus 4	5	7	1.10351*	0.20863	0.0210
Stimulus 5	4	5	.58377*	0.20292	0.0430
		6	1.17833*	0.19441	0.0000
		7	1.45833*	0.14953	0.0000
	5	4	-.58377*	0.20292	0.0430
		6	.59456*	0.18508	0.0130
		7	.87456*	0.13717	0.0000
Stimulus 6	4	7	1.00000*	0.29814	0.0500
	5	7	.61404*	0.12287	0.0000
	6	4	-0.71528	0.32082	0.2450
*. The mean difference is significant at the 0.05 level.					

Table 44: Post-hoc analysis differences between age-groups

To establish an effect for Class, a Levene's test pointed out a similar pattern of variance compared to Age, where stimuli 1 through 4 showed homogeneity of variance across different classes (p > .05), with stimulus 5 (Levene statistic = 11.057; df = 2; p < .05) and 6 (Levene statistic = 10.777; df = 2; p < .05). A Kruskal-Wallis is used for the latter stimuli while an ANOVA is used to establish an effect of Class for stimulus 1 to 4. There is no effect of Class for stimulus 1 (F (2,77) = 1.801; df = 2; p = .172), stimulus 2 (F (2,70) = .379; df = 2; p = .686), stimulus 3 (F (2,75) = .112; df = 2; p = .894) and stimulus 4 (F (2, 77) = 1.983; df = 2; p = .145). The Kruskal-Wallis test revealed a significant effect of Class for stimulus 5 (Chi^2 = 21.766; df = 2; p = .000) and 6 (Chi^2 = 6.010; df = 2; p = .05). Post-hoc analysis revealed significant differences between class-groups 1, 2 and 3 (p < .05) for stimulus 5 and 6. Turning to the effect of L2 input via multimedia, a Levene's test showed homogeneity of variance for the groups 'No L2 input via Multimedia' and 'Occasional L2 input' (p < .05). An ANOVA revealed no significant effect of L2 input via multimedia on any of the stimuli in the phoneme continuum. Lastly, the effect L1 on stimuli scores was determined: A Levene's test showed homogeneity of variance across the 'Dutch', 'Turkish' and 'Berber' L1 groups (p < .05). An ANOVA revealed a significant effect of L1 for stimulus 1 (F (2,77) = 4.145; df = 2; p = .009), while other stimuli did not reach significance. These results will be discussed in the final chapter. We will now turn to data obtained in the second test session.

Descriptives for the second /æ/ - /ɛ/ phoneme discrimination test session are shown in the table below:

Session 2	Stimulus	F1	F2	duration	N[33]	Mean Score[34]
/ɛ/	1	744	1823	0.132	76	5.6633
	2	744	1823	0.167	75	5.6542
	3	770.14	1746.92	0.143	74	5.6032
	4	796.8	1673.28	0.154	75	4.9333
	5	824	1602	0.132	75	4.5178
/æ/	6	824	1602	0.167	73	4.2922

Table 45: Descriptive statistics discrimination task /æ/ - /ɛ/ second test session

The descriptives in the table above reveal mean scores that generally replicate the phoneme continuum (stimuli 1-6) where stimulus 1 is identified as /ɛ/ by showing the highest score (5.6633), and stimulus 6 as /æ/ by showing the lowest score (4.2922). The results Shown here are represented graphically in Appendix D.

To establish statistical significance of above chance level performance, t-tests were carried out using 5 as the test value. Scores for all stimuli are significantly above-chance level performance. Note however, that results from stimulus 4 are only marginally significant.

	t	df	Sig. (2-tailed)	Mean Difference
Stimulus 1	10.565	75	0.000	0.66335
Stimulus 2	10.477	74	0.000	0.65422
Stimulus 3	8.873	73	0.000	0.60315
Stimulus 4	-0.704	74	0.048	-0.06667
Stimulus 5	-6.031	74	0.000	-0.48222
Stimulus 6	-11.321	72	0.000	-0.70776

Table 46: T-test values establishing above-chance test performance across the stimuli test session 2

These findings are analysed across the independent variables as detailed in the previous section, adding the variable ‘L2 input inside the classroom’. To establish the effect of Teacher-trainee on performance across the phoneme continuum, a Levene’s test revealed significant differences in variance between the six teacher-trainee groups across all stimuli.

[33]Note that the number of cases across the stimuli varies as a consequence of the random nature of the experiment.
[34]Calculated mean values range between 4 and 6, where 4 represents an /æ/ response and 6 represents an /ɛ/ response. Thus, chance level performance is represented by the value 5.

Consequently, a Kruskal-Wallis (nonparametric) test was used to establish the effect of Teacher-trainee:

	Stimulus 1	**Stimulus 2**	**Stimulus 3**	**Stimulus 4**	**Stimulus 5**	**Stimulus 6**
Chi-square	5.453	6.959	6.316	5.993	24.661	20.681
df	5	5	5	5	5	5
Asymp. Sig.	0.363	0.224	0.277	0.307	0.000	0.001
a. Kruskal Wallis Test; b. Grouping Variable: Teacher-trainee (6 Teacher groups)						

Table 47: Performance differences per stimulus between teacher-trainee groups test session 2

There is no effect of Teacher-trainee for stimulus 1 through stimulus 4. Only stimulus 5 and 6 appear to be affected by the independent variable Teacher-trainee. A post-hoc (Tamhane's T) revealed that subject 7 differed significantly ($p < .05$) from teacher-trainees 2, 3 and 5 across stimulus 5 and stimulus 6 as shown in the table below:

	Teacher	**Teacher**	**Mean Difference**	**Std. Error**	**Sig.**
Stimulus 5	Teacher-trainee 2	Teacher-trainee 7	-1.37897*	0.21539	0.001
	Teacher-trainee 3	Teacher-trainee 7	-1.44147*	0.19914	0.001
	Teacher-trainee 5	Teacher-trainee 7	-1.16917*	0.22699	0.002
Stimulus 6	Teacher-trainee 2	Teacher-trainee 5	-.61111*	0.15481	0.015
	Teacher-trainee 3	Teacher-trainee 5	-.53472*	0.16054	0.05
*. The mean difference is significant at the 0.05 level.					

Table 48: Post-hoc analysis differences between teacher-trainee groups

Turning to Gender, a Levene's test showed no significant differences in variance across any of the stimuli ($p > .05$) while an ANOVA revealed no significant effect for Gender across stimulus 1 through 6 ($p > .05$). A Levene's test was used as a precondition for carrying out an ANOVA to establish an effect for Participant age. Homogeneity of variance could not be established across all 4 age groups for stimulus 2 (Levene statistic = 3.462; df = 3; p = .021), stimulus 3 (Levene statistic = 3.836; df = 3; p = .013), stimulus 5 (Levene statistic = 4.242; df = 3; p = .008) and stimulus 6 (Levene statistic = 3.260; df = 3; p = .027). A Kruskal-Wallis test was therefore used to establish an effect for Participant age:

	Stimulus 1	**Stimulus 2**	**Stimulus 3**	**Stimulus 4**	**Stimulus 5**	**Stimulus 6**
Chi-square	2.481	7.365	2.851	1.148	7.084	3.252
df	3	3	3	3	3	3
Asymp. Sig.	0.479	0.061	0.415	0.765	0.069	0.354
a. Kruskal Wallis Test; b. Grouping Variable: Age						

Table 49: Performance differences per stimulus between Participant age groups test session 2

No effect for Participant age was found across the stimuli. To establish an effect for Class, a Levene's test pointed out a similar pattern of variance compared to Age, where stimulus 2 (Levene statistic = 6.327; df = 2; p = .003), stimulus 5 (Levene statistic = 15.025; df = 2; p = .000), and stimulus 6 (Levene statistic = 11.325; df = 2; p = .000) failed to meet the homogeneity of variance criterion. A Kruskal-Wallis test revealed the following results:

	Stimulus 1	**Stimulus 2**	**Stimulus 3**	**Stimulus 4**	**Stimulus 5**	**Stimulus 6**
Chi-square	2.907	9.269	3.491	.280	7.923	4.648
df	2	2	2	2	2	2
Asymp. Sig.	.234	.010	.175	.869	.019	.098
a. Kruskal Wallis Test; b. Grouping Variable: Class						

Table 50: Performance differences per stimulus between Class groups test session 2

The Class effect on stimulus 2 and stimulus 5 is statistically significant. No other differences are significant however. A Tamhane's T post-hoc analysis showed that for stimulus 2, class-groups 1 and 3 differed significantly while for stimulus 5 all class groups differed significantly in their phoneme discrimination performance ($p < .05$). Turning to the effect of L2 input via multimedia, a Levene's test showed homogeneity of variance for the groups 'No L2 input via Multimedia' and 'Occasional L2 input' ($p < .05$). An ANOVA revealed no significant effect of L2 input via multimedia on any of the stimuli in the phoneme continuum ($p > .05$). The effect of L1 on stimuli scores was also determined: A Levene's test showed homogeneity of variance across the 'Dutch', 'Turkish' and 'Berber' L1 groups ($p < .05$). An ANOVA revealed no significant effect of L1 on any of the stimuli ($p > .05$). A Levene's test revealed significant differences in variance across three groups receiving different amounts of L2 input in the classroom, and a concomitant ANOVA showed no significant effect for L2 input in the classroom ($p > .05$). Again, a Kruskal-Wallis test is used to show the effect of L2 input inside the classroom. The table below shows that this independent variable only affects stimulus 5:

	Stimulus 1	**Stimulus 2**	**Stimulus 3**	**Stimulus 4**	**Stimulus 5**	**Stimulus 6**
Chi-square	3.171	3.561	1.769	4.577	22.505	4.472
df	3	3	3	3	3	3
Asymp. Sig.	.366	.313	.622	.206	.000	.215
a. Kruskal Wallis Stimulus; b. Grouping Variable: Exposure to classroom English (hours)						

Table 51: Performance differences per stimulus between L2 input in classroom groups test session 2

Post-hoc analysis showed that discrimination performance of stimulus 5 based on 6hrs. of classroom input differed significantly from 12 hrs. of classroom input ($p < .01$).
Following the presentation of findings from both test sessions of the /æ/ - /ɛ/ phoneme discrimination experiment, a comparison is made between results from test session one and

two. This is done using a T-test[35]:

	Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
Stimulus 1	0.04696	0.60793	0.0754	0.623	64	0.536
Stimulus 2	0.14576	0.64813	0.08438	1.727	58	0.089
Stimulus 3	0.02646	0.62425	0.07865	0.336	62	0.738
Stimulus 4	0.16771	0.95546	0.11943	1.404	63	0.165
Stimulus 5	0.0349	0.95903	0.11988	0.291	63	0.772
Stimulus 6	0.13388	0.87603	0.11216	1.194	60	0.237

Table 52: Comparison Scores Test session 1 - Test session 2

As is shown in the Table 36, there are no significant differences in scores across the phoneme continuum between test session one and test session two. While differences are clearly not statistically significant, there are differences between the data obtained across the stimuli. These are shown in the table below:

	Mean	Std. Error	Std. Deviation
Stimulus 1 - ae	0.047	0.0754	0.60793
Stimulus 2 - ae	0.1458	0.08438	0.64813
Stimulus 3 - ae	0.0265	0.07865	0.62425
Stimulus 4 - ae	0.1677	0.11943	0.95546
Stimulus 5 - ae	0.0349	0.11988	0.95903
Stimulus 6 - ae	0.1339	0.11216	0.87603

Table 53: Mean differences per stimulus

These findings are discussed in more detail in the final chapter. We will now turn to the results from the second phoneme discrimination experiment.

6.2.2.2 Results Categorical discrimination task /i/ - /i:/

The result obtained in the /i/ - /i:/ phoneme discrimination experiments are detailed in the current section. Descriptive data analysis is shown first, followed by an analysis of the effects of the independent variables on the data obtained. This is followed by a contrastive analysis to establish differences between the two test sessions and whether or not there is a shift in perception regarding the stimuli in the phoneme continuum. The descriptives shown in the table below reveal mean scores that do not differ greatly from chance-level performance. These results are also represented graphically in Appendix D.

[35]The comparative analysis involves only data from children who took part in both test sessions through listwise exclusion

Test session 1	stimulus	F1	F2	duration	N[36]	Mean Score[37]
/i:/	1	270	2410	0.125	68	5.0368
	2	270	2410	0.092	68	4.975
	3	277.92	2447.52	0.113	69	5.0401
	4	285.92	2485.52	0.102	64	4.8703
	5	294	2524	0.125	64	4.863
/i/	6	294	2524	0.092	64	5.0245

Table 54: Descriptive statistics discrimination task /i/ - /i:/ first test session

To establish statistical significance of above chance level performance, t-tests were carried out using 5 as the test value. None of the stimuli elicit a statistically significant above-chance performance:

Stimulus	t	df	Sig. (2-tailed)	Mean Difference
Stimulus 1	0.528	67	0.60	0.03676
Stimulus 2	-0.307	67	0.76	-0.025
Stimulus 3	0.537	68	0.593	0.0401
Stimulus 4	-1.499	63	0.139	-0.12969
Stimulus 5	-1.704	63	0.093	-0.13698
Stimulus 6	0.285	63	0.777	0.02448

Table 55: T-test values establishing above-chance test performance across the stimuli test session 1

Findings are analysed across the independent variables as detailed in the previous section. To establish the effect of Teacher-trainee on performance across the phoneme continuum, a Levene's test revealed no significant differences in variance between the six teacher-trainee groups across all stimuli ($p > .05$). An ANOVA revealed no significant effect of Teacher-trainee for any of the six stimuli in the phoneme continuum ($p > .05$). For the independent variable Gender, homogeneity of variance may be assumed ($p > .05$), and an ANOVA showed no significant effect for any of the six stimuli ($p > .05$). Homogeneity of variance may also be assumed for the independent variable Participant age ($p > .05$). An ANOVA showed showed no effect of Participant age for any of the six stimuli ($p > .05$). The independent variable Class showed a similar pattern, with equality of variance assumed ($p > .05$) while an ANOVA revealed no significant effect for this independent variable. Turning to L2 input via Multimedia, the findings in previous independent variables were again replicated, showing no significant effect ($p > .05$) using an ANOVA. For the independent variable

[36] Note that the number of cases across the stimuli varies as a consequence of the random nature of the experiment.
[37] Calculated mean values range between 4 and 6, where 4 represents an /i/ response and 6 represents an /i:/ response. Thus, chance level performance is represented by the value 5.

L1, no significant effect was found using an ANOVA (p > .05).

Descriptives for the second /i/ - /i:/ phoneme discrimination test session are shown in the table below:

Session 2	stimulus	F1	F2	duration	N[38]	Mean Score[39]
/i:/	1	270	2410	0.125	72	5.0324
	2	270	2410	0.092	73	4.8809
	3	277.92	2447.52	0.113	71	5.0235
	4	285.92	2485.52	0.102	72	4.9505
	5	294	2524	0.125	74	4.8779
/i/	6	294	2524	0.092	71	4.8822

Table 56: Descriptive statistics discrimination task /i/ - /i:/ second test session

The descriptives shown in the table below reveal mean scores that do not differ greatly from chance-level performance. The results shown here are represented graphically in Appendix D. To establish statistical significance of above chance level performance, t-tests were carried out using 5 as the test value. Performance does not reach significantly extend beyond chance levels across all six stimuli (p > .05):

	t	df	Sig. (2-tailed)	Mean Difference
Stimulus 1	0.434	71	0.665	0.03241
Stimulus 2	-1.692	72	0.095	-0.11911
Stimulus 3	0.322	70	0.749	0.02347
Stimulus 4	-0.588	71	0.558	-0.04954
Stimulus 5	-1.547	73	0.126	-0.12207
Stimulus 6	-1.38	70	0.172	-0.11784

Table 57: T-test values establishing above-chance test performance across the stimuli test session 2

These data are analysed across the independent variables as detailed in the previous sections, with the addition of the independent variable ‘L2 input inside the classroom’.
To establish the effect of Teacher-trainee on performance across the phoneme continuum, a Levene’s test revealed no significant differences in variance between the six teacher-trainee groups across all stimuli (p > .05). An ANOVA revealed no significant effect of Teacher-trainee for stimulus 1, 2, 3, 5, and 6. Only stimulus 4 showed a small effect for Teacher-trainee (F (5,71) = 2.428; df = 5; p = .044). For the independent variable Gender, homogeneity of

[38] Note that the number of cases across the stimuli varies as a consequence of the random nature of the experiment.
[39] Calculated mean values range between 4 and 6, where 4 represents an /i/ response and 6 represents an /i:/ response. Thus, chance level performance is represented by the value 5.

variance may be assumed across all stimuli (p > .05), and an ANOVA showed no significant effect for any of the six stimuli (p > .05). Homogeneity of variance may also be assumed for the independent variable Participant age for all stimuli (p > .05). An ANOVA showed no effect of Participant age for any of the six stimuli (p > .05). The independent variable Class showed a similar pattern, with equality of variance assumed for all stimuli (p > .05) While an ANOVA revealed no significant effect for this independent variable. Turning to L2 input via Multimedia, a Levene's test revealed no significant differences in variance between the six teacher-trainee groups across all stimuli (p > .05). An ANOVA revealed no significant effect of L2 input via multimedia for stimulus 1, 2, 3, 5, and 6. Stimulus 4 showed a significant effect of L2 input via multimedia (F (1,71) = 5.411; df = 2; p = .023). For the independent variable L1, no significant effect was found using an ANOVA (p > .05).
Lastly, for the independent variable L2 input inside the classroom, a Levene's test for homogeneity of variance revealed no significant differences in variance for stimulus 1, 2, 3, and 5. The assumption was however, not met for stimulus 4 and 6. A Kruskal-Wallis test revealed significant differences for stimulus 4 and stimulus 6, as shown in the table below:

	Stimulus 1	**Stimulus 2**	**Stimulus 3**	**Stimulus 4**	**Stimulus 5**	**Stimulus 6**
Chi-square	1.248	1.093	4.571	10.451	3.543	8.752
df	3	3	3	3	3	3
Asymp. Sig.	.741	.779	.206	.015	.315	.033
a. Kruskal Wallis Test; b. Grouping Variable: Exposure to classroom English (hours)						

Table 58: Performance differences per stimulus between L2 input in classroom groups test session 2

Following the presentation of findings from both test sessions of the /i/ - /i:/ phoneme discrimination experiment, a comparison is made between results from test session one and two. The comparative analysis involves only data from children who took part in both test sessions through listwise exclusion:

	Mean	**Std. Deviation**	**Std. Error Mean**	**t**	**df**	**Sig. (2-tailed)**
Stimulus 1	-0.03051	0.82885	0.10791	-0.283	58	0.778
Stimulus 2	0.18612	0.84517	0.11098	1.677	57	0.099
Stimulus 3	0.07368	0.68181	0.09031	0.816	56	0.418
Stimulus 4	-0.15597	0.94636	0.12999	-1.2	52	0.236
Stimulus 5	-0.07152	0.79573	0.1073	-0.667	54	0.508
Stimulus 6	0.11879	0.99796	0.13456	0.883	54	0.381

Table 59: Comparison Scores Test session 1 - Test session 2

As is shown in the table, there are no significant differences in scores across the phoneme continuum between test session one and test session two. As is shown in Table 43, there are no significant differences in scores across the phoneme continuum between

test session one and test session two. While differences are clearly not statistically significant, there are differences between the data obtained across the stimuli. These are shown in the table below:

	Mean	**Std. Error**	**Std. Deviation**
Stimulus 1 - i	-0.0305	0.10791	0.82885
Stimulus 2 - i	0.1861	0.11098	0.84517
Stimulus 3 - i	0.0737	0.09031	0.68181
Stimulus 4 - i	-0.156	0.12999	0.94636
Stimulus 5 - i	-0.0715	0.1073	0.79573
Stimulus 6 - i	0.1188	0.13456	0.99796

Table 60: Mean differences per stimulus

These findings are discussed in more detail in the final chapter.

7. Discussion

The results as detailed in the previous chapter are discussed in three separate sections. Section 7.1 addresses the PPVT-4 results, section 7.2 reviews the native and non-native vowel realisations, section 7.3 discusses the results of the two phoneme discrimination tasks. The conclusion, limitations of the present study and recommendations for further research are found section 7.4 of this final chapter.

7.1 PPVT-4

The PPVT-4 task is included in this study to establish a measure of early L2 learners' receptive vocabulary. The current section reviews the findings from both test sessions and discusses differences found between the two sessions. Results are linked to findings from previous research.

A total of 122 children (age: 4;0-8;0 yrs.) took part in the first PPVT-4 test session and 104 children (age: 4;0-7;7) took part in the second test session. The decrease in number of participants is the result of time pressure at primary schools near the end of a school-year when there are numerous other activities to be completed. The average raw scores for test session one (18.93) and test session two (23.00) differed significantly from each other ($T = -4.204$; $df = 103$; $p < .001$; cohen's $d = -0.42$) which amounts to a small to moderate gain (Cohen, 1977) in receptive vocabulary score. The effect of teacher-trainee for the first test session actually revealed differences between different classes in different schools and a Kruskal-Wallis test revealed a small significant difference between these groups. Post-hoc analysis showed that teacher-trainee group 9 differed significantly from teacher-trainee group 5, but that other teacher-trainee groups did not differ significantly. Differences between teacher-trainee groups were more clear-cut after a period of three months: A significant effect for teacher-trainee group was found, where post-hoc analysis showed that teacher-trainee group 9 raw scores differed significantly from teacher-trainee group 2, teacher-trainee group 4, and teacher-trainee group 5. Marginally significant was the difference between teacher-trainee group 9 and 3 ($p = .052$). No other teacher-trainee groups differed significantly. It is clear that teacher trainee 9 (proficiency level: B2) was most successful in terms of her pupils' receptive vocabulary raw score growth, while this teacher-trainee belonged to the lower proficiency level group and also showed greatest deviance from other teacher-trainees in the vowel realisation measurements.

No statistically significant differences for Gender was found across both testing periods. This suggests an even spread of raw scores for males and females, and it would seem that gender does not affect receptive vocabulary scores. A small but robust (i.e. across both test sessions) effect for age was found, and post-hoc analysis showed that age-group 4 differed significantly from age-group 5, and age-group 4 also differed significantly from age-group 6 ($p < .01$), where the older L2 learners showed higher scores than the younger learners. Differences could be attributable to general age-related changes in cognitive functioning (Werker & Tees, 2002), which leads to advantages in test-taking in older children. No other

between group differences reached statistical significance, possibly due to the fact that there were fewer 7- and 8-year-olds involved in the study. These results therefore suggest a small effect of age where 6-year-olds score significantly better than 4-year-olds. Raw scores did not differ significantly across different classes, and while this finding may be due to the relatively small number of participants involved in this study, this does allow us to state that a class contains a very heterogeneous group of children and that raw scores cannot be attributed to being in a certain group. Input via multimedia did not reveal any effect, which may be due to the lack of definition of the variable or the lack of accurate information collected. It is also possible that the variable simply shows no effect at an early age due to comparatively low amount of L2 English input via multimedia.[40] Participant's L1 was not found to significantly affect raw scores, but it should be noted that there were very few non-native Dutch speakers involved in this study, and that an effect was unlikely to surface as a consequence. A small but statistically significant effect of hours of classroom input was found, where more input yields higher raw scores.

In sum, the PPVT-4 test showed a significant raw score increase across a three-month period with 5 to 12 hours of L2 input during that timespan. It was shown that 4-year-olds score significantly lower than 6-year-olds on the PPVT-4 test, and that there was a limited effect for teacher-trainee, showing highest raw scores for a teacher-trainee with B2 proficiency with 6 hrs. of classroom English L2 (participant age: 5;5 yrs.). These findings replicate findings from Unsworth (2008, ms.) in showing increased PPVT scores over a limited amount of L2 input. A small effect for amount of L2 input was also found, showing higher PPVT-4 raw scores with an increased amount of L2 input. Note that i. a greater number of participants from a more diverse (and nationally representative) L1 background across a longer period of time could provide more accurate insight into the factors affecting L2 receptive vocabulary scores in Dutch primary schools.

7.2 Native and non-Native English vowel realisations

Previous research suggested children's non-adult like perceptual weighting (Nittrouer, 1996, Mayo, 1999, Gerrits 2001, Heeren, 2004) where adults are consistently more sensitive to smaller differences in perceptual cues than children. The current study investigated differences in the phonemic realisation of teachers and its effect on Early English language learners. In testing the phonemic realisations of NNS and NS of English, it was found that there were significant differences between speakers (i.e. NNS teacher-trainees and the NS English speaker) across spectral and durational values for phoneme /æ/ and across the spectral but not the durational values for phoneme /i:/. The distinction NS - NNS did not reveal significant differences in spectral and durational cue values for either phoneme. Interestingly, in post-hoc analysis a trend was found between proficiency level and /æ/ cue values; NNS with a lower proficiency level (participants 4 and 9) showed significantly

[40] See Nuhaan (2009, ms.) for data relating to L2 input via multimedia.

different F1 cue values compared to other NNS and NS participants. This was not replicated for the /i:/ phoneme, where differences were clearly not related to differences in proficiency levels. This suggests that non-native phoneme realisations may provide an indication of a speaker's general proficiency level. It should be pointed out that the data collected did not reveal highly significant differences, and generalisations should be treated with caution.

A preliminary study carried out by Schutter & Chevalking (2009) revealed significant differences between native SBE speakers and non-native speakers at B1 level (CEFR) for the F1 formant value ($p < .001$), while both F2 and durational cues value differences failed to reach significance. The results obtained in the current study differ where the F1 formant is concerned. This may be due to two factors: i. greater proficiency level differences between NNS and NS in the Schutter & Chevalking (2009) study, and/or ii. differences in the way in which tokens were obtained (CVC imitation versus reading text recording).

7.3 Perceptual discrimination task

A total of 111 participants took part in the first /æ/ - /ɛ/ phoneme discrimination test session, in contrast to 78 participants for the second test session. The /i/ - /i:/ phoneme discrimination task was consistently more difficult,[41] leading to fewer participants completing the phoneme discrimination task; 69 participants successfully completed the first /i/ - /i:/ phoneme discrimination test session, while 75 participants successfuly completed the second /i/ - /i:/ phoneme discrimination test session. Participants found it very difficult to discern between the /i/ - /i:/ endpoints, and a number of children did not complete the task as the result of lack of concentration and motivation. The section below reviews the main findings presented in the previous chapter, and discusses these in relation to the research questions outlined in chapter 5. Recall that the aim of these experiments was to establish i. suitability of the perceptual discrimination task for early learners aged 4 - 8 years, and ii. test whether there is a relationship between quality of phonological input and a perceptual shift in early L2 learners.

Results for the categorical /æ/ - /ɛ/ discrimination task showed that participants scored significantly above chance-levels for all stimuli except stimulus 5 in the first test session. The end-points were consistently recognised as instances of the phonemes /æ/ - /ɛ/, suggesting that participants (even four-year-old participants) understood the task as was instructed. Analysing the data across the independent variables showed no effect for Gender, L2 input via multimedia and L1 across both test sessions. An effect of teacher-trainee was found for stimulus 3 to 6 in the first test session, but the second test session replicated this effect only for stimulus 5 and 6. It would appear that the teacher-trainee affects the way in which the end-point /æ/ is discriminated by L2 learners, suggesting an influence of teacher input on L2 phoneme acquisition that are new to early L2 learners. Post-hoc analysis does not show differences between teacher-trainees that is paralleled by proficiency levels (see table 48). It is, however, unfortunate that the lower level teacher-trainees were unable to take part in the

[41] Teacher-trainees reported that children had difficulty in distinguishing between phones.

phoneme discrimination experiments as this could have provided better insight into the effect of teacher-trainee proficiency on phoneme perception in early L2 learners. It should be noted that the effects were found using a nonparametric test, and that these findings are less than robust as a consequence. They should be interpreted as being indicative rather than conclusive.

The results suggest a small effect of Age on stimulus 4, 5 and 6 during the first session, but this is not sustained in the data from the second test session. Age is therefore not a factor of significance in this experiment. The independent variable Class shows a sustained effect on stimulus 2 (F1=744; F2=1823; duration=.167) and 5 (F1=824; F2=1602; duration = .132). These stimuli differ maximally from each other in terms of both spectral and durational cues which suggests that participants from different classes use maximum spectral and durational cues to differentiate between these phonemes. There was no sustained effect of L1 on discrimination of any of the stimuli, and L2 input in the classroom showed an effect on stimulus 5 only. Again, the differences shown here are based on nonparametric analysis, and these do not reveal clear patterns. It is however, clear that participants understood the task, and that the teacher-trainee exerts the greatest effect on phoneme discrimination at the /æ/ end of the phoneme continuum; Recall that post-hoc analysis revealed that subject 2 differed significantly from subject 3, 5, and 8 across stimulus 4, stimulus 5 and stimulus 6. Subject 2 showed B2 proficiency levels, and subject 3 was also at the same proficiency level. Subjects 5 and 8 however, were more proficient (C1) English L2 speakers.

Differences between test session one and two were not statistically significant. A significant shift across any of the stimuli in the phoneme continuum is not noted in this experiment. Mean differences per stimulus show that there is a positive gain across stimuli, specifically the stimuli at the /æ/ end-point of the phoneme continuum; The shift indicates a general trend toward more test tokens being (correctly) identified as /æ/. It is also shown that the shift is greatest for stimulus 4, 2 and 6 respectively. Based on this no conclusion can be drawn as to which cue is used more in identifying a phone; i.e. both spectral and durational cues appear to be used in token discrimination. A greater number of participants and more participant L2 contact could possibly reveal more robust findings.

Turning to the second categorical discrimination experiment where the task was to discriminate between phones in an /i:/ - /i/ continuum. Participants did not score significantly above chance-level in either test session, which shows that findings from the previous experiment are indeed attributable to the experimental task and not a result of chance. For the second categorical discrimination experiment no effects were discovered across any of the independent variables for the first test session, and the effects found in the second test sessions affected only the fourth stimulus in the phoneme continuum. There were no significant differences between both test sessions, but examination of the differences obtained show a similar pattern of shift in scores as the previous discrimination experiment where stimulus 2, 4 and 6 show the greatest shift across test sessions. This shift shows that participants use

both spectral and durational cues to identify stimuli in both phoneme continua.

7.4 Conclusion

The aim of these experiments was to establish the suitability of the perceptual discrimination task for early learners aged 4 - 8 years, to test whether there is a relationship between quality of phonological input and a perceptual shift in early L2 learners and to monitor receptive vocabulary development, given a limited amount of L2 input.

The native and non-native vowel realisation measurement provided insight into differences between teacher-trainees and the single SSBE participant. It has been shown that non-native phoneme realisations may provide an indication of a speaker's general proficiency level. It was unfortunate however, that teacher-trainees 4 and 9 were unable to participate in the phoneme discrimination experiments as this could have provided information regarding the relationship between phonological input and a perceptual shift in early L2 learners.

The /æ/ - /ɛ/ phoneme discrimination experiment showed a promising outcome through a shift towards more correct /æ/ identifications at the /æ/ end-point of the phoneme continuum. An small effect of teacher-trainee on identification of the /æ/ phoneme for stimulus 5 and 6 has been shown. This effect cannot, however, be attributed to teacher-trainee proficiency levels. In other words, a relationship between quality of phonological input and a perceptual shift in early L2 learners could not be established in this study. The /i:/ - /i/ phoneme discrimination experiment showed chance-level performance, which should be attributed to the small cue differences in that experiment, given the slightly better performance on the /æ/ - /ɛ/ phoneme discrimination experiment. This is in line with findings from Heeren (2006), who showed that children have difficulty attending to small acoustic differences.

Although there was no significant perceptual shift across 5 to 12 hours of English L2 input, it was found that children did increase their number of correct /æ/ phoneme identifications following a greater amount of L2 input, which underlines the importance of amount of English L2 input in classroom L2 learning situations. Unfortunately, no conclusion can be drawn on the basis of current findings.

The PPVT-4 has shown to be an effective instrument in recording a (small) difference across length of exposure and age. The PPVT-4 results show the greatest increase in raw score with a B2 proficiency-level teacher-trainee, which suggests that more native-like proficiency and phoneme realisation do not contribute to increased PPVT-4 raw scores. Increased age and hours of exposure are shown to lead to increased raw scores, which corroborates the trend noted for the /æ/ phoneme discrimination experiment. In sum, the findings in the present study do not contradict previous findings, but neither do they provide robust evidence to confirm results found in previous studies as detailed in chapter 5.

Limitations of the present study

This study has involved teacher-trainees as both participants and class teachers. While care was taken to ensure that teacher-trainees did not train classes to 'do well' on any of the experiments by explaining as little as possible about the aim of the experiment, little could be done to actively control for such bias. And, time-pressure on teacher-trainees led to fewer early learners being able to take part in this study, which limited the statistical power of the current study. Involving 11 teacher-trainees in collecting data as consistently as possible has proved to be no small task, and although this has been a rich educational experience for all parties involved, care should be taken in planning and timetabling so that teacher-trainees have sufficient time and resources to carry out these tests. Another limitation in this study was the short period between the first and second test sessions. This was imposed by the 3-month internship period, and nothing could be done to extend this period of time. Greater differences in proficiency levels and phoneme realisations could also have provided more insight into effects of L2 quality on early L2 learners. Additionally, video observation could have provided a cross-check to establish the actual input provided during an English lesson.

This study has taken place in everyday primary school teaching practise, which allows results to be easily extended to similar contexts, but it is important to realise that there are limitations imposed on research by primary school logistics. Technical issues in several schools also limited the number of early learners being able to take part in the current study.

Implications for future studies

Given the previous section, future research should focus on establishing better vowel measurement leading to greater within-speaker cue consistency, and accurate proficiency measurement to be able to establish accurate between-speaker differences. When doing research inside the primary school classroom, researchers should use their own (tested) hardware rather than reliance on school hardware, and tests should not be carried out at the end of the school-year as this leads to too many distractions in the school timetable. The results from the /i/ - /i:/ phoneme discrimination experiment suggest that early L2 learners are indeed insensitive to very small acoustic differences, and future research in this vein should focus on phonemes that are less similar to the native phoneme inventory as compared to /i/ - /i:/ in this study. It is clear however, that the instruments used in the current study are suited to research invovling children. These instruments could be refined further to aid and simplify data collection. If these considerations are taken into account the tentative findings from this study are likely to be improved in future studies.

8. Bibliography

- Arva, V. & Medgyes, P. (2000) Native and non-native teachers in the classroom. System, 28(3), 355-372.
- Barik, H.C., Swain, M. (1976) A Longitudinal Study of Bilingual and Cognitive Development, International Journal of Psychology, 11: 4, 251 — 263
- Bialystok, E. (1997). The structure of age: In search of barriers to second-language acquisition. Second Language Research, 13, 116-137.
- Beddor, P.S., Gottfried, T.L., (1995). Methodological issues in cross-language speech perception research with adults. In W. Strange (ed.), Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research, 207-232. Timonium, MD: York Press.
- Blondin, CH., Candelier, M., Edelenbos, P., Johnstone, R., Kubanek-German, A., Taeschner, T. (1998). Foreign languages in primary and pre-school education: contexts and outcomes. London: CILT.
- Broekhof, K. (2007). Engels als derde taal. In: Philipsen, K., E. Deelder, M. Bodde-Alderlieste (red.). Early English: a good start! Conferentiebundel. Europees Platform voor het Nederlandse Onderwijs, Den Haag, p. 8593
- Boersma, P., P. Escudero & R. Hayes (2003). Learning abstract phonological from auditory phonetic categories: an integrated model for the acquisition of language-specific sound categories. In M. J. Sole, D. Recansens & J. Romero (eds.), Proceedings of the 15th International Con- gress of Phonetic Sciences, 1013-1016. Barcelona: Causal Productions.
- Carroll, J. B. (1975). The teaching of French as a foreign language in eight countries. New York: Wiley.
- Clyne, M., Jenkins, C., Chen, I., Tsokalidou, R., and Wallner, T., Developing Second Language From Primary School. NLLIA., Canberra, 1995.
- Cohen, J. (1977). Statistical power analysis for behavioral sciences (revised ed.). New York: Academic Press.
- Deelder, E., Maljers, A., (2007) Jong geleerd is oud gedaan, in 'Early English a good start', Europees Platform voor het Nederlandse Onderwijs.
- Doyé, P. & Hurrell, A., (1997). Foreign Language learning in primary schools (age 5-6 to 10-11). Strasbourg: Council of Europe.
- Driscoll P, Jones J, Martin C, Graham-Matheson L, Dismore H, Sykes R (2004). A systematic review of the characteristics of effective foreign language teaching to pupils between the ages 7 and 11. In: Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Dunn, L.M., Dunn, D.M., (2007) Peabody Picture Vocabulary Test, Fourth Edition - Manual, Pearson Assessments.
- Edelenbos, P., Schoot, F. van der, Verstralen, H. (1996). Balans van het Engels aan het einde van de basisschool 2. Uitkomsten van de tweede peiling in 1996. Arnhem: Citogroep (1996).
- Edelenbos, P., Jong, J.H.A.L., de (2004). Vreemdetalenonderwijs in Nederland - Een analyse in opdracht van OCenW uitgevoerd door het Nationaal Bureau Moderne Vreemde Talen (NaB-MVT), Enschede.
- Edelenbos, P., Johnstone, R. & Kubanek, A. (2006). The main pedagogical principles underlying the teaching of young learners. [EAC 89/04]. Key study for the

European Commission. Brussels. German and French translation 2007.

- Eimas, P.D., Siqueland, E.R., Jusczyk, P.W., & Vigorito, J. (1971). Speech perception in infants. Science, 171: 303-306.
- Eimas, P.D., Corbit, J.D. (1973). Selective adaptation of linguistic feature detectors. Cognitive Psychology, 4, 99-109.
- Escudero, P.R. (2005). Linguistic perception and second language acquisition, Doctoral dissertation, Utrecht University.
- Eurydice Studies (2001). Foreign Language Teaching in Schools in Europe. Brussels, Belgium: European Commission.
- Eurydice Studies (2005). Key data on teaching Languages at School in Europe. Brussels, Belgium: European Commission.
- Flege, J.E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. Journal of Phonetics, 15, 47–65
- J. Flege (1988), The production and perception of speech sounds in a foreign language, in Human Communication and Its Disorders, A Review - 1988, ed. by H. Winitz (Ablex, Norwood, NJ), pp. 224-401.
- Flege, J., Munro, M., & MacKay, I. (1995). Factors affecting degree of perceived foreign accent in a second language. Journal of the Acoustical Society of America, 97, 3125-3134.
- Flege, J., Munro, M. & MacKay, I. (1995). The effect of age of second language learning on the production of English consonants. Speech Communication, 16, 1-26.
- Flege, J., Frieda, A., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. Journal of Phonetics, 25, 169-186.
- Flege, J., Yeni-Komshian, G., & Liu, S. (1999). Age constraints on second language learning. Journal of Memory and Language, 41, 78-104.
- Gerrits, P.A.M. (2001). The categorisation of speech sounds by adults and children,
- Doctoral dissertation, Utrecht University.
- Gilzow, D. & Rhodes, N. (2000). Establishing High-Quality Foreign Language Programs in Elementary Schools. In: Perspectives on Policy and Practise. Northeast and Islands Regional Educational Laboratory at Brown University/LAB, December 2000.
- Heeren, W.F.L. Perceptual Development of Phoneme Contrasts in Adults and Children, Doctoral Thesis, Utrecht University, 2006.
- Heesters, K., Feddema, M., Schoot, F. van der, Hemker, B., Balans van het Engels aan het einde van de basisschool 3. Uitkomsten van de derde peiling in 2006. Arnhem: Citogroep (2008).
- Herder & de Bot (2005). Vroeg vreemdetalenonderwijs in internationaal perspectief literatuurstudie. Expertisecentrum voor taal, onderwijs en communicatie, Rijksuniversiteit Groningen.
- Hunt, M., A. Barnes, B. Powell, G. Lindsay & D. Muijs (2005). Primary modern foreign languages: an overview of recent research, key issues and challenges for educational policy and practice. In: Research Papers in Education, vol. 20, no. 4, p. 371-390
- Hyltenstam, K. (1992). Non-native features of near-native speakers: On the ultimate attainment of childhood L2 learners. In R. J. Harris (Ed.), Cognitive

processing bilinguals. Amsterdam: Elsevier Science, 351-368.

- Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), Handbook of second language acquisition (pp. 539-588). London: Blackwell.
- Ioup, G., Boustagui, E., Tigi, M. E., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. Studies in Second Language Acquisition, 16, 73-98.
- Johnson, J. S. & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. Cognitive psychology, 21, 60-99.
- Johnstone, R. (2003). Reviews research on languages teaching, learning and policy published in 2002. Language Teaching, 36 3, 165 - 189.
- Jusczyk, P., (1992). Developing phonological categories from the speech signal In: C.A. Ferguson, L. Menn, & C. Stoel-Gammon (eds.), Phonological Development: Models, Research, Implications. Maryland: Timonium. pp. 17-64.
- Jusczyk, P.W. (1993). From general to language-specific capabilities: The WRAPSA model of how speech perception develops. Journal of Phonetics, 21, 3-28.
- Kubanek-German, A., (1998). Primary foreign language teaching in Europe - trends and issues, Cambridge University Press.
- Larson-Hall, J. (2008) Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. Second Language Research 24, 1, pp. 35-63.
- Leather, J. (1999) Second Language Speech Research: An Introduction, Language Learning 49, S1: 1-56.
- Leather, J. (2003) Phonological Acquisition in Multilingualism. In Garcia Mayo, M. and Garcia Lecumberri, M., editors, Age and the acquisition of English as a foreign language. Clevedon: Multilingual Matters, 23-58.
- Lecumberri, G., Gallardo, F. 2003: English FL sounds in school learners of different ages. In Garcia Mayo, M. and Garcia Lecumberri, M., editors, Age and the acquisition of English as a foreign language. Clevedon: Multilingual Matters, 115–35.
- Lenneberg, E. H. (1967). Biological Foundations of Language. Wiley
- Lightbrown, P.M., (2000). Classroom SLA Research and Second Language Teaching, Applied Linguistics, Oxford University Press, 2000.
- Marinova-Todd, S. H., Marshall, B., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. TESOL Quarterly, 34, 9-34.
- Mayo, G. 2003: Age, length of exposure and grammaticality judgements in the acquisition of English as a foreign language. In Garcia Mayo, M. and Garcia Lecumberri, M., editors, Age and the acquisition of English as a foreign language. Clevedon: Multilingual Matters, 94–114.
- Medgyes, P. (1994). The non-native teacher. London: Macmillan publisher.
- Moussu, L., Llurda, E., (2008). Non-native English-speaking English language teachers: History and research. Language Teaching, 41 , pp 315-348
- Moyer, A. (1999). Ultimate attainment in L2 phonology. Studies in Second Language Acquisition, 21, 81-108.
- Munoz, C., (2008) Symmetries and Asymmetries of Age Effects in Naturalistic and Instructed L2 Learning, Applied linguistics, pg. 578-596, Oxford University

Press.
- Pisoni, D.B. (1975). Auditory short-term memory and vowel perception. Memory & Cognition, 3: 7-18.
- Riney, T. & Flege, J. (1998). Changes over time in global foreign accent and liquid identifiability and accuracy. Studies in Second Language Acquisition, 20, 213-244.
- Rochet, B. 1995. Perception and production of second-language speech sounds by adults. In W. Strange (ed.), Speech Perception and Linguistic Experience: Issues in Cross- Language Speech Research. 379-410. Timonium, MD: York Press.
- Schoot, F. van der, (2008). Onderwijs op peil? Een samenvattend overzicht van 20 jaar PPON. Arnhem, Cito Instituut voor Toetsontwikkeling.
- Schouten, M.E.H., Hessen, van , A.J. (1992). Modeling phoneme perception.I: Categorical perception. Journal of the Acoustical Society of America, 92: 1841-1855.
- Schutter, J.S., Chevalking, J.W., Perception and production (a)symmetry in /æ/ ~ /ɛ/ contrasts among intermediate Dutch learners of English?, unpublished, 2009.
- Serrano, R.S., (2007) Time distribution and the acquisition of English as a foreign language, Doctoral Thesis University of Barcelona, September 2007
- Singleton, D. (1997). Second language in primary school: The age dimension. The Irish Yearbook of Applied Linguistics, 15, 155-1 66.
- Singleton, D. (2003) Critical Period of General Age Factor(s)? In Mayo, M., Lecumberri, M.L.G. (Eds.), Age and the Acquisition of English as a Foreign Language (pp. 3-22)
- Snow, C. and M. Hoefnagel-Höhle. 1978. The critical period for language acquisition: Evidence from second language learning, Child Development 49: 1114–28.
- Strange, W., 1995. Cross-language studies of speech perception: A historical review, in Speech Perception and Linguistic Experience, edited by W. Strange York, Timonium, MD, pp. 3–45.
- Sussman, J.E. (1993). Auditory processing in children's speech perception: results of selective adaptation and discrimination tasks. Journal of Speech and Hearing Research, 36: 380-395.
- Sussman, J.E. (2001). Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions? Journal of the Acoustical Society of America, 109: 1171-1180.
- Swain, M. (1981). Time and timing in bilingual education, Language Learning, 31, 1-15.
- Turnbull, M. & Arnett, K. (2002). Teachers' uses of the target and first languages in second and foreign language classrooms. Annual Review of Applied Linguistics, 22, 204–18.
- Walley, A.C., & Flege, J.E. (1999). Effect of lexical status on children's and adults' perception of native and non-native vowels. Journal of Phonetics, 27: 307-332.
- Weber, A. & A. Cutler (2004). Lexical competition in non-native spoken-word recognition. In Journal of Memory and Language 50, pp. 1–25.
- Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. Child Development, 52, 349–355.

- Werker, J. E, & Tees, R. C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development, 7, 49-63.
- Werker, J. E, & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. Journal of the Acoustical Society ofAmerica, 75, 1866-1878.
- Werker , J. F., J. S. Logan (1985). Cross-language evidence for three factors in speech perception. Perception & Psychophysics 37 (1), pp. 35-44.
- Werker, J.F., Lalonde, C. (1988). Cross-language speech perception: Initial capabilities and developmental change. Developmental Psychology, 24, 672–683.
- Werker, J.F., Fennell, C.T., Corcoran, K.M., Stager, C.L., (2002). Infants' Ability to Learn Phonetically Similar Words: Effects of Age and Vocabulary Size. Infancy, 3(1), 1–30

9. Appendix A - Common European Framework Levels

Level	Description
A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.
A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.

Table 61: CEF Level description (http://www.coe.int/t/dg4/Linguistic/Source/Framework_EN.pdf)

10. Appendix B - Protocol dataverzameling

1. Inleiding

Dit protocol biedt een handreiking ter verzameling van gegevens rondom 'Vroeg Engels'. Gegevens worden via drie experimenten verzameld:

- linguistic_experiment_set_a (21 test items)
- linguistic_experiment_set_a (21 test items)
- ppvt set 1-6 (max. 72 test items)

Om een meting van het effect van kwaliteit en kwantiteit van tweede taal input bij kinderen in groep 1-3 te maken, is het van belang de instructies nauwkeurig op te volgen. Mocht de onderstaande informatie onvoldoende duidelijkheid bieden, of wilt u meer achtergrondinformatie ontvangen, óf ontstaan er problemen tijdens het testen, dan kunt u altijd contact opnemen met Jan Willem Chevalking via:

jwchevalking@che.nl / chevalking@mac.com / 0318-510409 / 0318-696300

2. Systeemvereisten

Het testmateriaal vereist een PC met ten minste Windows XP of Mac OS 10.3.9. Tevens dient een Flash Player op de computer geïnstalleerd te zijn. Als dit niet het geval is (dus: het doorlopen van de stappen onder "werken met..." resulteert niet in het opstarten van de .swf bestanden) kunt u deze downloaden vanaf:

- http://get.adobe.com/flashplayer/

3. Benodigde bestanden

U dient over een USB-stick te beschikken met daarop de bestanden zoals vermeld in Bijlage 1. Controleer of alle bestanden aanwezig zijn. BELANGRIJK: De bestanden dienen altijd in de vermelde hierarchie te blijven staan omdat de tests anders niet werken. Kopieer/knip dus nooit een enkel bestand van USB-stick naar computer! Indien u niet beschikt over de juiste bestanden kunt u deze downloaden (.zip bestand) vanaf de volgende locatie (voor Early English minor studenten):

- http://hbo-english.nl/file.php/27/language%20tests/060309_experiment_files.zip

of (voor externen):

- http://homepage.mac.com/chevalking/.Public/linguistics_experiment/experiment_files.zip

U kunt het gedownloade bestand 'uitpakken' met o.a. Winzip. Vervolgens kunt u de uitgepakte bestanden als geheel naar uw USB-stick kopiëren. Voor nadere instructies rondom downloaden/installeren van de bestanden kunt u Bijlage 3 raadplegen.

4. Werken met 'linguistic_experiment_set_a.swf'

Als alle bestanden op uw USB-stick staan opent u de windows verkenner en zoekt u de USB-drive. Vervolgens opent u de map "linguistic_experiment_set_a", en dubbelklikt u op het bestand "linguistic_experiment_set_a.swf". Als alles naar behoren werkt, opent zich een scherm dat lijkt op de onderstaande illustratie.

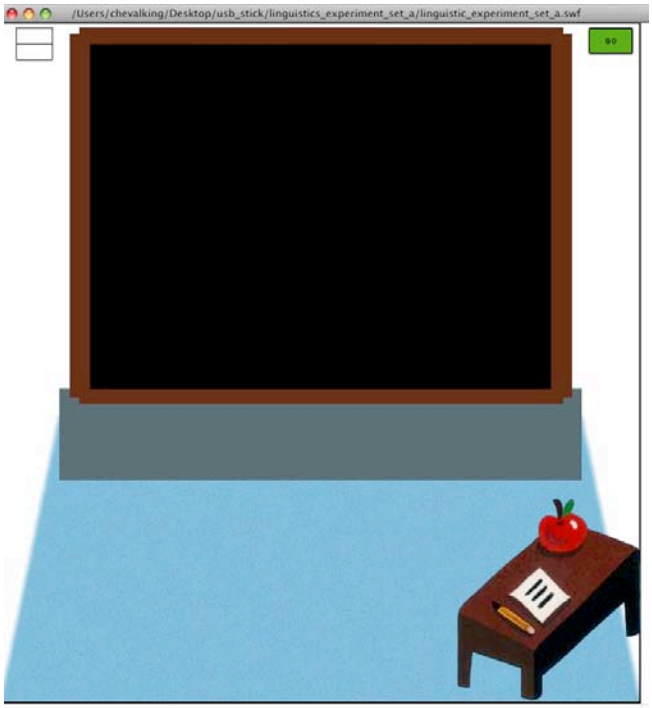


Table 62: opstartscherm

4.1 Voorbeeldronde

Doorloop de volgende stappen om bekend te raken met het eerste experiment:
Klik op de groende knop rechtsboven om het experiment te starten
Een kort introductie-muziekje en de 'spelers' genaamd 'Teacher', 'Danny' en 'Donny' verschijnen, alsook de onderstaande voiceover:

"*This is the teacher. This is Danny, and this is Donny."*

Hierna tikt 'Teacher' op het bord en spreekt een klank. Daarna vervolgt de voiceover met:

"Danny and Donny are trying to copy their teachers. Who said what the teacher said?"

Table 63: voorbeeldronde

De namen boven de poppetjes 'Danny' en 'Donny' lichten vervolgens groen op. Er kan nu een keus gemaakt worden door op de naam 'Danny' of 'Donny' te klikken. De keus wordt (zoals de voiceover ook meldt) gemaakt voor degene die de 'Teacher' het best nazegt. Als voor leerlingen niet helemaal duidelijk is wat de bedoeling is, kan de 'voorbeeldronde' herhaald worden door op de 'example' knop linksboven te drukken. Als de opdracht duidelijk is dan kan er aan de testronde begonnen worden. LET OP: Begin niet aan de testronde als voor de leerling niet geheel duidelijk is wat de bedoeling is - herhaal de voorbeeldronde zolang als nodig is!

4.2 Testrondes

U klikt op de groene knop rechtsboven om door te gaan naar de testronde. Dit is het gedeelte in de test waarbinnen gegevens verzameld worden. Het testgedeelte bestaat uit 24 rondes. Aan het begin van iedere ronde tikt 'Teacher' op het bord, en volgt er een klank die 'Danny' en 'Donny' nadoen. De opdracht is wederom te klikken op de naam van degene die de 'Teacher' het best nazegt. Na het maken van een keus volgt altijd een compliment als feedback. Daaropvolgend klikt u op de groene knop rechtsbovenaan om naar de volgende ronde te gaan.

DE TOETS KAN VROEGTIJDIG WORDEN AFGEROND DOOR LINKSBOVEN OP 'OUTPUT' TE KLIKKEN.

4.3 Resultaten

Na de 24e ronde volgt een scherm met een weergave van de behaalde resultaten. Deze resultaten dienen opgeslagen te worden in een tekstbestand. Doorloop de onderstaande stappen om de resultaten om te slaan:

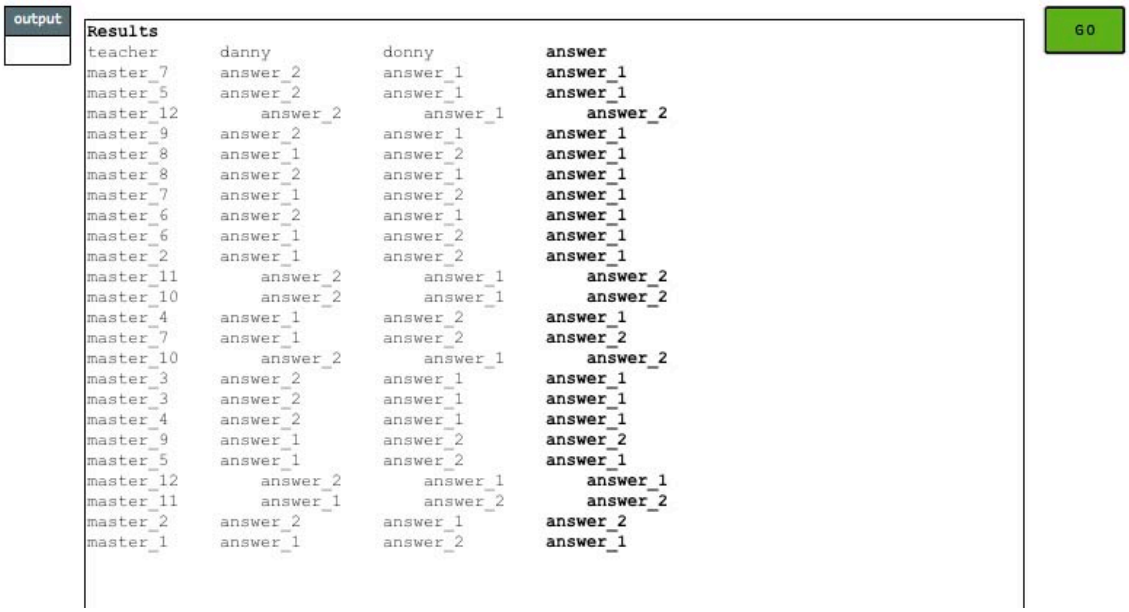


Table 64: resultatenscherm

- Klik met je rechtermuisknop in het witte tekstvak en selecteer alle tekst
- Klik weer met je rechtermuisknop op de geselecteerde tekst en kopieër de tekst
- Open nu een tekstverwerkingsprogramma (kladblok, MS Word) en plak de tekst in een nieuw document

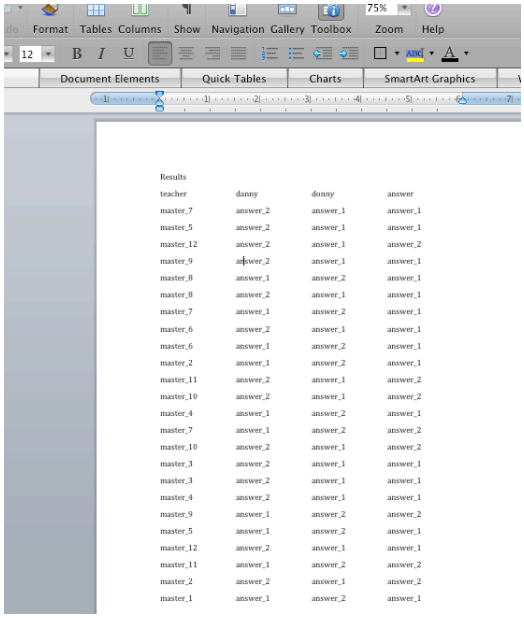


Table 65: tekstverwerker

- Sla het bestand op als leerlingnaam_set_a.txt, leerlingnummer_set_a.txt, leerlingnaam_set_a.doc of leerlingnummer_set_a.doc
- Plaats (indien u dit nog niet had gedaan) de opgeslagen bestanden op uw USB-stick in de map "put_your_data_here"

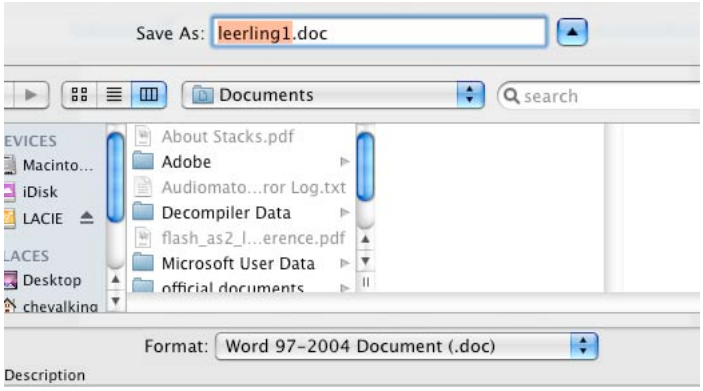


Table 66: bestandsnaam

5. Werken met 'linguistic_experiment_set_b.swf'

Dit tweede experiment heeft eenzelfde opzet als 'linguistic_experiment_set_a.swf'. U dient m.b.v. de windows verkenner de map "linguistic_experiment_set_b" te openen en te dubbelklikken op het bestand "linguistic_experiment_set_b". Eenzelfde scherm opent als bij het voorgaande experiment. Verder is ook de gehele procedure is identiek aan het vorige experiment. Wat verschilt is het klankspectrum dat wordt voorgelegd. Let wel op dat u de resultaten opslaat als leerlingnaam_set_b.txt, leerlingnummer_set_b.txt, leerlingnaam_set_b.doc of leerlingnummer_set_b.doc om overschrijven van resultaten te voorkomen.

6. Werken met 'ppvt set 1-6.swf'

De ppvt test is anders van opzet dan de voorgaande twee experimenten omdat gegevens niet door het programma, maar door u worden verzameld middels het invulformulier (zie Bijlage 2). Deze bijlage bestaat uit twee bladzijden waarvan de eerste gebruikt wordt voor het invullen van enkele (persoonlijke) gegevens.

LET OP: mocht uw school prijs stellen op anonimiteit van haar leerlingen, kunt u een nummer i.p.v. een naam opgeven op basis van bijvoorbeeld alfabetische volgorde, zodat gegevens na een tweede test met de oude gegevens vergeleken kunnen worden.

De tweede bladzijde bestaat uit een lijst van 6 sets met ieder 12 woorden. Iedere set komt overeen met een vastgestelde receptieve woordenschat. Het is dan ook absoluut niet de bedoeling dat u woorden uit deze lijst expliciet aan de orde stelt tijdens uw lessen om uw studenten een 'voorsprong' te geven!
U werkt altijd een volledige set door voordat u de toets afbreekt. Het is de bedoeling dat u bijhoudt welk antwoord (optie 1, 2, 3 of 4) door de leerling gegeven wordt, en of deze juist is. Omcirkel op de lijst het cijfer van het door de leerling aangewezen antwoord, en geef in de ruimte daarnaast aan of het antwoord goed is. Als een leerling in een set 8 (of meer) foute

antwoorden geeft maakt, dan breekt u na beëindiging van de set de toets af.
Open opnieuw de windows verkenner en zoek de USB-drive, en ga naar de map "ppvt set 1-6". Dubbelklik dan op "ppvt set 1-6.swf". Het volgende scherm verschijnt:

Table 67: ppvt training set

6.1 Trainingset

De eerste twee slides zijn bedoeld als oefening, om de opdracht duidelijk te maken. Na het openen van het bestand klinkt de volgende voiceover:

"Look at the pictures on this page. Put you finger on ... ...boy"

Als de leerling het juiste plaatje aanwijst (plaatje 3) kunt u doorgaan naar de volgende slide door met uw linkermuisknop te klikken. De volgende slide ziet er hetzelfde uit, met als verschil de voiceover:

"Now look at the pictures on this page. Put your finger on ... ...chair"

6.2 PPVT test-items

Als de leerling ook dit plaatje aanwijst (plaatje 2), kunt u beginnen met het formele gedeelte waarbij u de score van de leerling bijhoudt.

BELANGRIJK: Mocht u een slide nogmaals willen afspelen klikt u met uw rechtermuisknop en kiest u 'rewind' of 'terugspoelen'.

7. Stappenplan dataverzameling

Het stappenplan dient ter voorbereiding en handleiding tijdens het uitvoeren van de experimenten. Uitgangspunt is dat u het technische gedeelte voldoende beheerst vanuit de hierboven gegeven uitleg.

7.1 Voorbereiding

- Controleer de volgende punten zorgvuldig vóórdat de proefpersoon (leerling) achter de computer plaatsneemt:
- Zorg voor een blanco invulformulier (zie Bijlage 2)
- Verzeker u ervan dat de computer naar behoren werkt, en dat er 2 koptelefoons m.b.v. de meegeleverde 'splitter' op de PC zijn aangesloten
- Controleer het geluidsniveau van beide koptelefoons, en zorg voor een adequaat volume
- Verzeker u ervan dat de USB-stick in de computer zit en dat de testbestanden naar behoren werken.

7.2 Handleiding

- Als je de leerling niet kent stel je jezelf voor in het Engels
- Vertel dat je drie spelletjes gaat spelen en dat je erbij bent om de leerling te helpen/ mee te 'spelen'
- Vraag vervolgens naar de gegevens zoals vermeld in Bijlage 2 (1e bladzijde)
- Plaats de koptelefoon op het hoofd en zorg dat deze goed past
- Start 'linguistic_experiment_set_a.swf'
- Leg de bedoeling van het spelletje uit als dit niet duidelijk blijkt, en probeer dit in het Engels m.b.v. herhalen voorbeeldronde. Blijf (uiteraard) positieve feedback geven
- Als het doel eenmaal duidelijk is kunt u verder naar de testrondes.
- Eenmaal aangekomen bij het resultaatscherm kunt u de leerling b.v. een stickertje geven ter beloning - waardoor u tijd heeft de resultaten op de USB-stick op te slaan (denk aan bestandsnaam = leerlingnaam / nummer)
- Ga daarna verder met 'linguistic_experiment_set_b.swf', en herhaal de bovenstaande stappen
- Start dan de ppvt test (ppvt set 1-6.swf), en leg uit dat dit speletje weer anders is.
- Leg uit dat de leerling slechts hoeft te wijzen naar het plaatje dat het denkt gehoord te hebben. Herhaal de 'training set' indien nodig om ervan verzekerd te zijn dat de opdracht duidelijk is
- Als de leerling een woord niet heeft verstaan kunt u de slide herhalen door met uw rechtermuisknop te klikken en 'replay/rewind' te selecteren
- Stop aan het eind van een set als er 8 of meer 'foute' aanwijzingen geweest zijn (dus: aangewezen plaatje correspondeert niet met gesproken woord)
- Bedank de leerling!

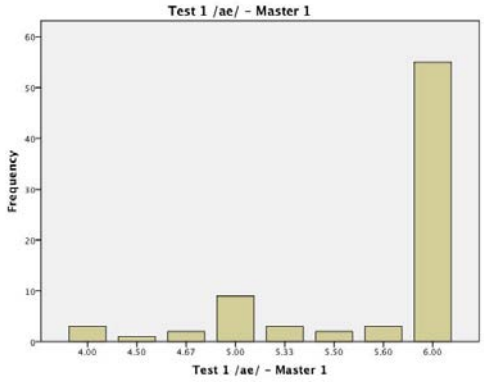
11. Appendix C - Praat script used to create stimuli (adapted from I. Brasileiro Reis Pereira)

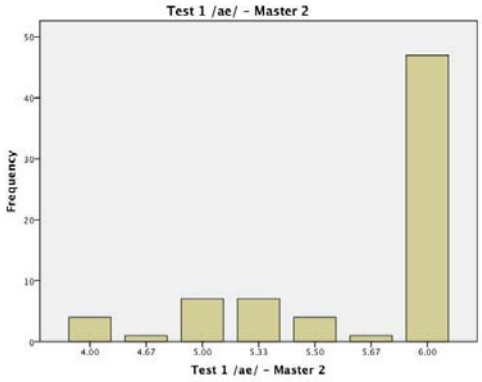
```
# Generate synthetic vowels with duration, F1 and F2 steps
# Stores resulting sounds in specified directory

form Generate vowels (cascade mode) with duration, F1 and F2 steps
    positive Initial_F0_(Hz) 150
    positive Final_F0_(Hz) 100
    sentence Directory_to_write_to ~\Desktop\vowels
    positive Minimum_duration_(ms) 96
    positive Maximum_duration_(ms) 203
    positive Number_of_duration_values 4
    positive Minimum_F1_(Hz) 698
    positive Maximum_F1_(Hz) 830
    positive Number_of_F1_values 4
    comment If F1 values are equal to or higher than F2 values the sounds are
    comment marked as "1" in the column "rep". The marked sounds are not
generated!
    comment
    positive Minimum_F2_(Hz) 1070
    positive Maximum_F2_(Hz) 1336
    positive Number_of_F2_values 4
endform

# calculate duration steps
logrange = log10(maximum_duration / minimum_duration)
logstep = logrange / (number_of_duration_values - 1)

for i to number_of_duration_values
    d'i' = minimum_duration * 10^((i-1)*logstep)
endfor

# calculate F1 values
if number_of_F1_values > 1
    maxmel = hertzToMel(maximum_F1)
    minmel = hertzToMel(minimum_F1)
    melrange = maxmel - minmel
    melstep = melrange / (number_of_F1_values - 1)
    for i to number_of_F1_values
        melvalue = minmel + (i-1) * melstep
        first'i' = melToHertz(melvalue)
    endfor
else

    first1 = minimum_F1
endif
```

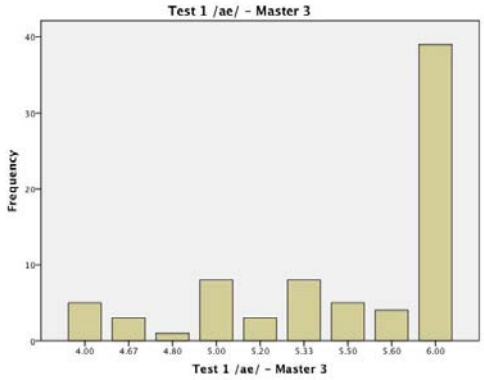
```
# calculate F2 values
if number_of_F2_values > 1
    maxmel = hertzToMel(maximum_F2)
    minmel = hertzToMel(minimum_F2)
    melrange = maxmel - minmel
    melstep = melrange / (number_of_F2_values - 1)
    for i to number_of_F2_values
        melvalue = minmel + (i-1) * melstep
        second'i' = melToHertz(melvalue)
    endfor
else
    second1 = minimum_F2
endif

# initialize duration and formants table
numsounds = number_of_duration_values * number_of_F1_values *
number_of_F2_values
Create TableOfReal... params numsounds 4
Set column label (index)... 1 rep
Set column label (index)... 2 f1
Set column label (index)... 3 f2
Set column label (index)... 4 dur

# generate sounds & update table
row = 0

for d to number_of_duration_values
    dur = d'd'/1000
    for second to number_of_F2_values
        f2 = second'second'
        for first to number_of_F1_values
            rep = 0
            f1 = first'first'
            if f1 >= f2 - 100
                rep = 1
            endif
            select TableOfReal params
            row += 1
            Set row label (index)... 'row' 'first'_'second'_'d'
            Set value... row 2 f1
            Set value... row 3 f2
            Set value... row 4 dur
            if rep = 1
               Set value... row 1 rep
            endif
            call generate
            if rep = 0
```

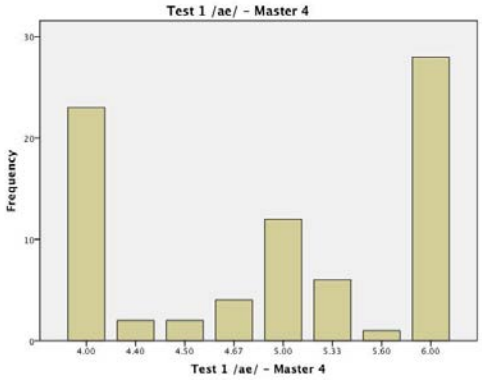
```
            Write to WAV file... 'directory_to_write_to$'\'first'_'second'_'d'.wav
        endif
        Remove
# pause 'f1' 'f2' 'dur'
    endfor
  endfor
endfor
select TableOfReal params
Write to binary file... 'directory_to_write_to$'\vowelparams.TableOfReal
Write to headerless spreadsheet file... 'directory_to_write_to$'\vowelparams.txt

procedure generate

# Create voice source signal
Create PitchTier... sweep 0.0 dur
Add point... 0 initial_F0
Add point... dur final_F0
To PointProcess
Remove points between... 'dur'-0.005 'dur'
To Sound (phonation)... 44100 1 0.01 0.7 0.01 3 4

# Add some extra formants to get a flatter spectrum.
f3 = max (2500, f2 + 1000)
f4 = max (3500, f3 + 400)
f5 = max (4000, f4 + 600)
f6 = f5 + 1000
f7 = f6 + 1000
f8 = f7 + 1000
f9 = f8 + 1000
f10 = f9 + 1000
for i to 10
  Filter with one formant (in-line)... f'i' sqrt(80^2+(f'i'/20)^2)
endfor

# clear up
select PitchTier sweep
plus PointProcess sweep
Remove

select Sound sweep
Scale... 0.99

endproc
```

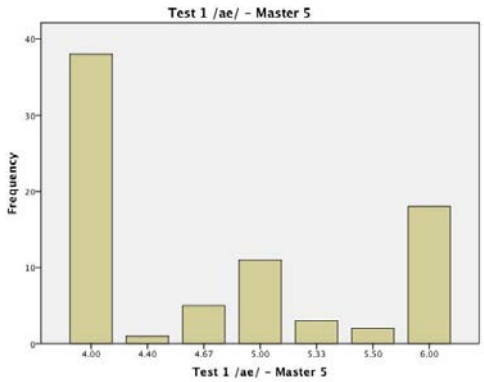
12. Participant responses phoneme discrimination task

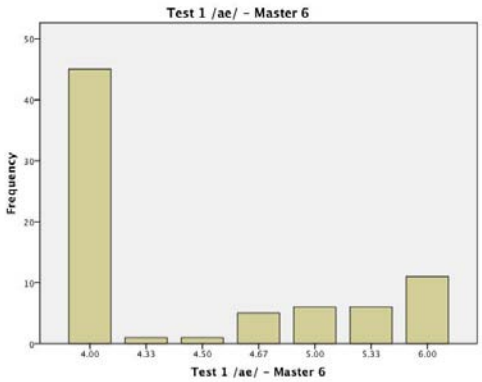


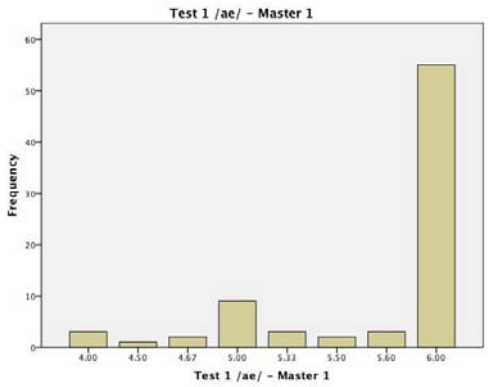


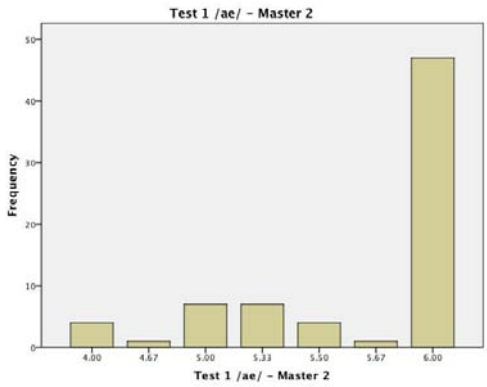


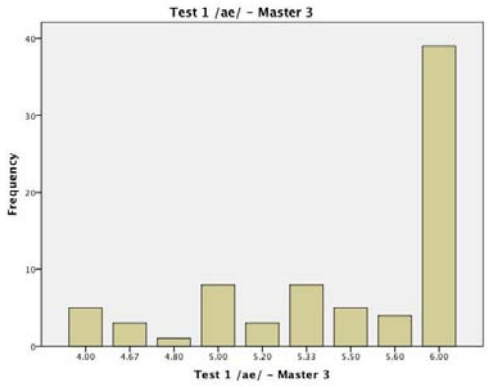


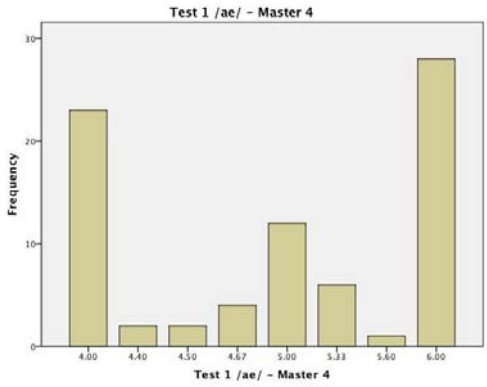


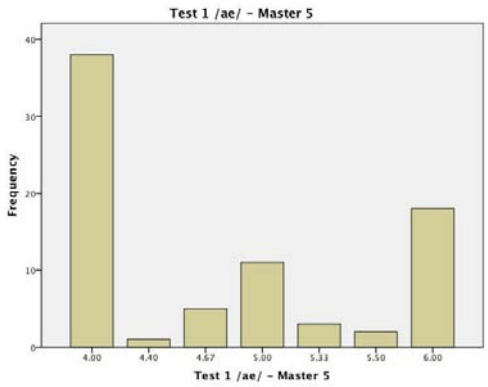


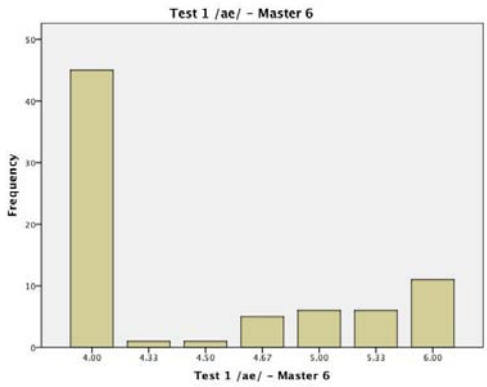

Test 1 /ae/ - Master 1
Frequency
Test 1 /ae/ - Master 1

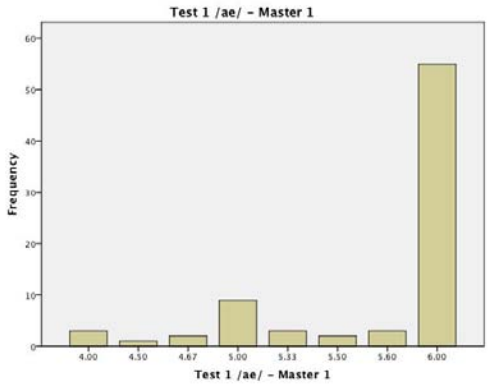

Test 1 /ae/ - Master 2
Frequency
Test 1 /ae/ - Master 2

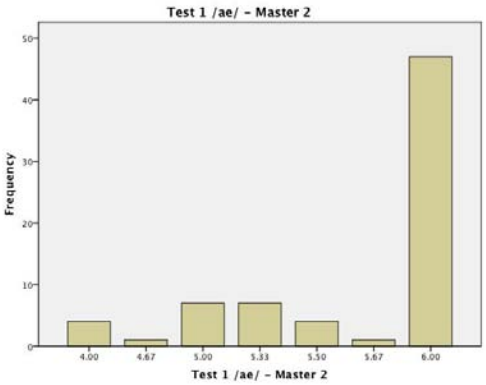

Test 1 /ae/ - Master 3
Frequency
Test 1 /ae/ - Master 3

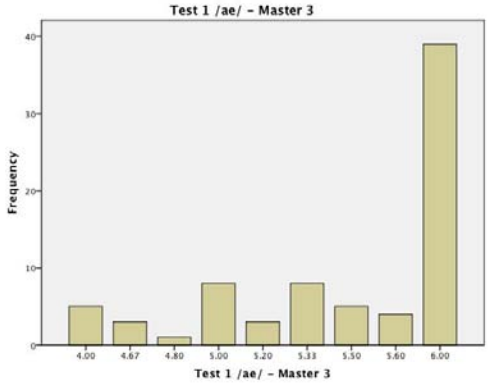

Test 1 /ae/ - Master 4
Frequency
Test 1 /ae/ - Master 4

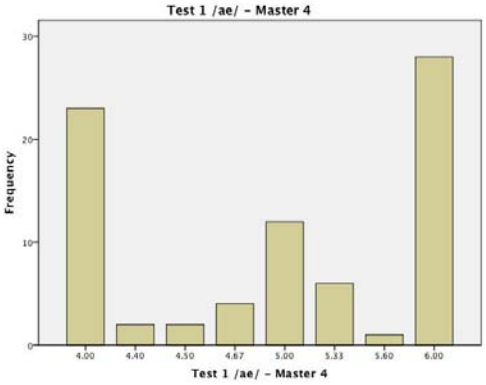

Test 1 /ae/ - Master 5
Frequency
Test 1 /ae/ - Master 5

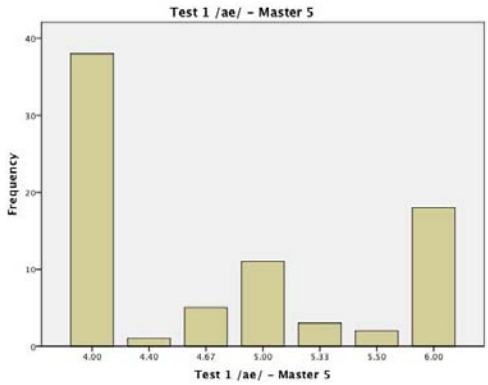

Test 1 /ae/ - Master 6
Frequency
Test 1 /ae/ - Master 6

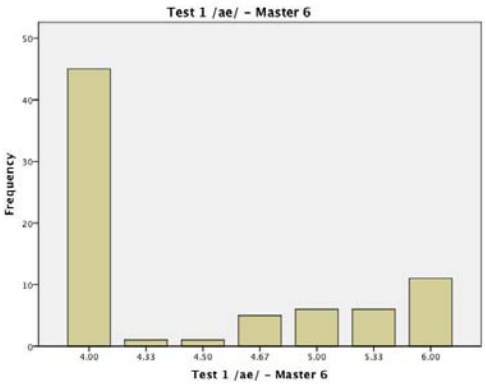

Test 1 /ae/ - Master 1
Frequency
Test 1 /ae/ - Master 1

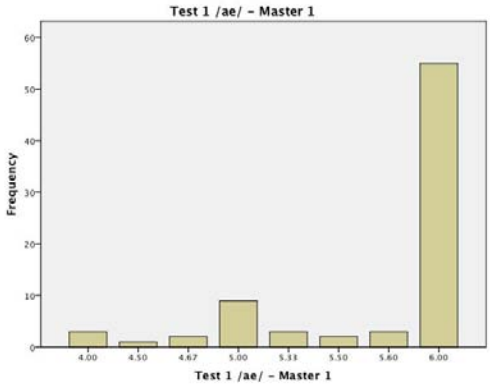

Test 1 /ae/ - Master 2
Frequency
Test 1 /ae/ - Master 2

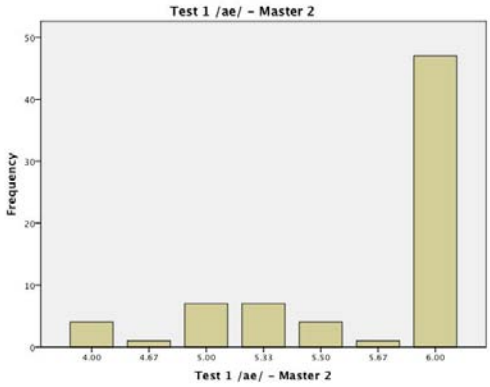

Test 1 /ae/ - Master 3
Frequency
Test 1 /ae/ - Master 3

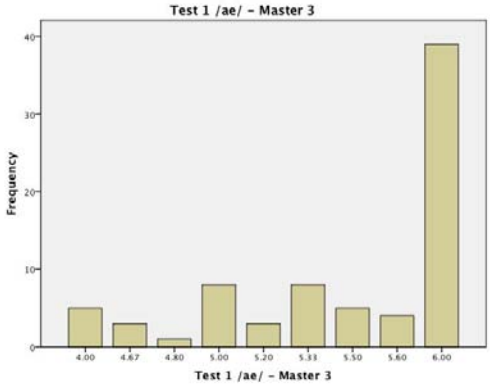

Test 1 /ae/ - Master 4
Frequency
Test 1 /ae/ - Master 4

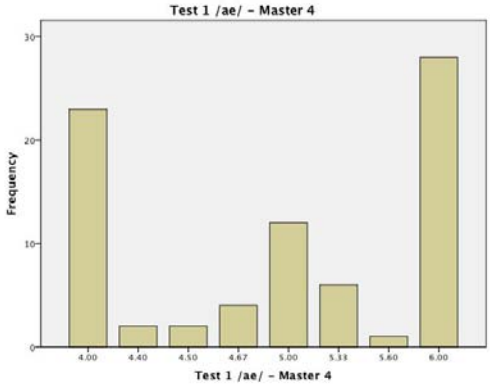

Test 1 /ae/ - Master 5
Frequency
Test 1 /ae/ - Master 5

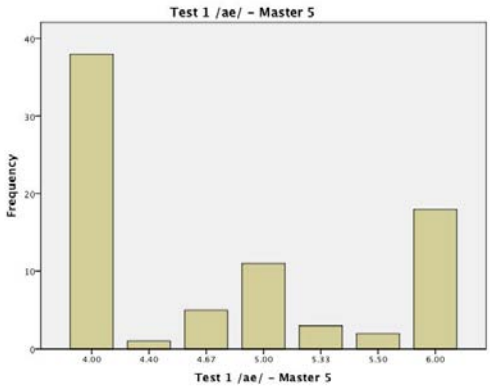

Test 1 /ae/ - Master 6
Frequency
Test 1 /ae/ - Master 6

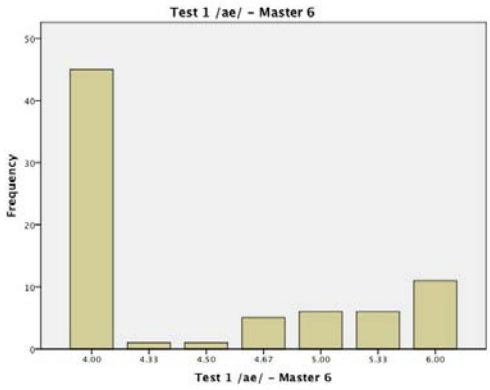

Test 1 /ae/ - Master 1
Frequency
Test 1 /ae/ - Master 1


Test 1 /ae/ - Master 2
Frequency
Test 1 /ae/ - Master 2


Test 1 /ae/ - Master 3
Frequency
Test 1 /ae/ - Master 3


Test 1 /ae/ - Master 4
Frequency
Test 1 /ae/ - Master 4


Test 1 /ae/ - Master 5
Frequency
Test 1 /ae/ - Master 5


Test 1 /ae/ - Master 6
Frequency
Test 1 /ae/ - Master 6