Intonational and durational contributions to the perception of foreign-accented Norwegian

An experimental phonetic investigation

Snefrid Holm

Thesis for the degree philosophiae doctor

Department of Language and Communication Studies

Norwegian University of Science and Technology

Table of contents

Acknowledgements

I am deeply indebted to my excellent advisor Prof. Wim van Dommelen. His high standards, unparalleled thoroughness and great enthusiasm have educated me both personally and professionally.

I am also grateful for the help I received from my second advisor, Prof. Ann Bradlow, who provided guidance regarding research and offered helpful suggestions about experimental design.

I wish to thank Prof. Francis Nolan for sharing ideas on analyses, Dr. Jane Setter for providing literature references, Dr. Jardar Eggesbø Abrahamsen and Cand. Philol. Jan Hognestad for help regarding dialectal variation, M.A. Egil Albertsen for discussions about perception models, and Marte Kristine Lindseth and Karin Hansen for practical assistance.

I would also like to say thank you to the 167 speakers and listeners who patiently participated in recordings and experiments. Thanks are also due to the teachers of Norwegian as a second language in our department, who provided me with their students' e-mail addresses, especially Lill Kristin Lund Kibakaya, Sissel Robbins, Eli Skarpnes and Sissel Jensen Nefzaoui.

A special thanks to my roommate Anne for sharing frustration and joy, seriousness and silliness through several years. Many thanks also to the rest of my departmental family, in particular Kaja, Heidi, Jørn, Petter, Rein Ove, Aleksander, and Jacques. Heartfelt thanks to retired phonetician Arne Kjell Foldvik for introducing me to phonetics and for always believing in me. His frequent inquiries regarding my work and well being have been greatly appreciated.

I would also like to thank Nancy Bazilchuk for correcting my English in an early version of this thesis. As parts of the thesis were later rewritten, any mistakes are entirely mine.

Hugs and kisses to my boyfriend Kjell for unfaltering support in rough times and cherished company in good times.

List of abbreviations

- L1= first language
- L2= second language
- N1= native Norwegian
- N2= Norwegian as a second language
- O= original stimuli
- COD= close-original duration stimuli
- COI= close-original intonation stimuli
- I= intonation manipulated stimuli
- D= duration manipulated stimuli
- ID= intonation-duration manipulated stimuli

1. Introduction

When a person over a certain age learns a second language (L2), he will normally pronounce the second language with a foreign accent (Flege, Munro & Mackay, 1995b; Major, 2001; Scovel, 1988; Patkowski, 1990). Native speakers can easily detect foreign accents in their own mother tongue (L1). For instance, Flege (1984) showed that speech samples as short as 30 ms were correctly identified as foreign-accented, and Munro, Derwing & Burgess (2003) showed that listeners could detect foreign-accentedness in a single word presented backwards. Despite our intuitive identifications of foreign accents, there is as yet no universally accepted definition of a foreign accent (Pennington, 1996; Gut, 2007). If we view the foreign accent as a perceptual concept, it can be defined as *"the extent to which an L2 learner's speech is perceived to differ from native speaker norms"* (Munro & Derwing, 1998). How various phonetic aspects contribute to listeners' perceptions of foreign accented speech is not known. The focus of the present thesis work is the relative contributions of intonational and durational aspects of non-native speech to native listeners' perceptions in terms of the perceived degree of foreign accent and in terms of intelligibility. Determining which aspects of speech are the most important in listener perceptions is useful not only in expanding our theoretic knowledge about the phenomenon of foreign accent, but also in helping teachers set priorities regarding which phonetic aspects to emphasize when teaching pronunciation to non-native speakers.

The following provides a brief overview of the most important lines of research regarding foreign accents, after which the focus will be narrowed to research on the relative contributions of various phonetic aspects to the degree of accent and intelligibility. Based on this literature review, the aim of the present investigation is further described and hypotheses about the outcomes of the investigation are offered. An outline of the structure of this thesis is provided at the end of the chapter.

1.1 Foreign accent research

L2 speech has been found to differ from L1 speech in a variety of different ways. Among the phonetic differences are deviant VOT duration (Flege, 1987; Schmidt & Flege, 1996), consonant articulation (McAllister, 2007), vowel articulation (Bohn & Flege, 1992; Flege, Bohn & Jang, 1997), liquid articulation (Major, 1986a; Aoyama, Flege, Guion, Akahane-Yamada & Yamada, 2004), stress placement (Archibald, 1994; Pater, 1997), and prosody (Aoyama & Guion, 2007; Guion, Flege, Liu & Yeni-Komshian, 2000). A great deal of foreign accent research has been devoted to accounting for the reasons why foreign accents occur, especially in terms of various types of interactions between the L1 and the L2. Early work in the 1950s by Robert Lado sparked this research approach. His Contrastive Analysis Hypothesis (Lado, 1957) claimed that all difficulties observed in L2 acquisition could be predicted from comparing the sound systems of the L1 and the L2. Similarities between the L1 and the L2 meant that the learner would successfully acquire the L2 structure (transfer), whereas differences were predictive of learner difficulties (interference). It was soon recognized that the Contrastive Analysis Hypothesis was too simple and could not account for all learner problems, which led the hypothesis to undergo several adjustments in the decades to follow. Contrastive Analysis and subsequent related approaches arose in a pedagogical setting that focused solely on speech, but researchers soon began to pay attention to processes that were internal to speakers themselves. Since the 1970s, researchers have focused a great deal of effort in the search for the perceptual basis for foreign accent (Strange, 1995; Strange, 2007). The relevance of non-native perception for non-native production is reflected in the term *perceptual foreign accent* coined by McAllister (1997). McAllister used this term to show that foreign accents relate to perception as well as to production. This research focuses on perceptual categorization differences between native and non-native speakers. But how is the native perception shaped in the first place? Many experimental investigations have studied native language development. For example, it has been found that at birth, infants can perceive the segmental contrasts of most of the world’s languages (Eimas, Siqueland, Jusczyk & Vigorito, 1971). Infants’ perceptions then evolve to become language-specific by 14 months of age, so that contrasts that are linguistically functional in the L1 continue to be well perceived, whereas the ability to discriminate (some but not all[1]) contrasts that are not functional in the L1 diminishes (Werker & Tees, 1999). The Native Language Magnet model (NLM, Kuhl & Iverson, 1995) builds on earlier work and seeks to explain how native speech perception is shaped. The model also shows how this shaping has implications for subsequent L2 perception. The NLM holds that, at birth, infants have an innate ability to perceptually distinguish between sounds belonging to different phonetic categories, and at the same time they perceive sounds that belong to the same phonetic category as similar. Then, over the first weeks and months of life, the infant accumulates experience with the native language surrounding it, and already at 6 months of age its perception begins to be shaped by the native language. The infant learns how to categorize the variability in the speech signal in terms of

[1] Not all contrasts are discriminated at birth. For instance, the English d - ð contrast is acquired late by English L1 speakers (Polka, Colantonio & Sundara, 2001).

phonetic categories specific to the native language. It develops perceptual prototypes which are typical tokens forming the centre of a category. A prototype exerts a magnet effect on similar sounds so that they cluster together perceptually. The perceptual space around a prototype has thus "shrunk" which entails that sounds in this area will not be discriminated. These native perceptual sound clusters make perception of the native language functionally robust, but can cause problems when a person hears sounds from a language with a different clustering of sounds. In other words, when the listener hears non-native sounds perceived as near the centre of a native magnet, the sounds will not be discriminated. This will be the case whether the sounds belong to the same or to different categories in the non-native language. In this way the listener has become less sensitive to phonetic distinctions that are not important in his native language.

The Native Language Magnet model has its main focus on the formation of native perceptual categories, but also shows how this formation influences the learning of an L2. Several other models focus on L2 perception problems long after the formation of the native language is completed. The most renowned models are the Speech Learning Model (Flege, 1995) and the Perceptual Assimilation Model (Best, 1995). The Speech Learning Model (SLM) makes the following claims about how an L2 learner's L1 influences the way in which he perceives sounds in the target language. The learner may perceive that an L2 sound is similar to a sound in his L1. If he fails to perceive the difference between them, then he will perceptually assimilate these sounds. The SLM claims that the L1 and L2 are represented in the same perceptual space, and so when an L2 sound is assimilated to an L1 sound, this perceptual category is expanded to comprise both. The result of such assimilation is an accented pronunciation. The learner may however perceive an L2 sound to be different from any sound in his L1. This causes the listener to form a new perceptual category. In order to keep the new L2 category sufficiently apart from the closest L1 category, the differences between these categories may be exaggerated both in perception and production. The SLM further claims that a higher degree of L1 / L2 interference will occur for older learners because they have more experience with their L1 phonological system. The Perceptual Assimilation Model (PAM), like the SLM and the NLM, describes the interferences that occur when listening to native versus non-native speech. This model meticulously outlines six different scenarios that can occur when discriminating between non-native sounds:

1) Excellent discrimination is predicted when the two non-native sounds are perceptually assimilated to two different native categories.
2) Moderate/very good discrimination is expected when the two non-native sounds are perceptually assimilated to the same native category, but one sound is perceived as a deviant exemplar of this category.
3) Poor discrimination is predicted when the two non-native sounds are perceptually assimilated to the same native category, and the sounds are perceived as equal regarding goodness of fit to this category.
4) Discrimination ranges from poor to very good when the sounds are not perceptually assimilated to any native category. Level of discrimination success depends upon the sounds' perceptual closeness to each other and to native categories.
5) Very good discrimination is predicted when one non-native sound is assimilated to a native category while the other non-native sound remains uncategorized.
6) Discrimination is expected to be good/very good when both non-native sounds are perceived as non-speech events (this is for instance the case for native English listeners' perceptions of isiZulu clicks, as reported in Best, McRoberts & Sithole, 1988).

Among the differences between the SLM and the PAM is that the SLM seeks to explain the perceptions of listeners who are in the process of learning an L2, whereas the PAM focuses on naïve listeners' perceptions of sounds from an unfamiliar language. However, this does not mean that the SLM and the PAM are in conflict, merely that the PAM describes perceptual phenomena at the very onset of learning a new language whereas the SLM describes phenomena at later stages when the listener has become familiar with the L2 phonetics and phonology and is actively engaging in acquiring it (Best & Tyler, 2005). Another interesting point is that the PAM addresses a scenario where some speech sounds are perceived as non-speech, and describes how this has special implications for the discrimination between non-native sounds. Perceptual models that seek to explain L2 listener problems do so in terms of phonetic similarity between L1 and L2 categories. The notion of phonetic similarity is incidentally a somewhat problematic concept that has been discussed by Strange (2007). Many studies have investigated non-native perception and production in order to evaluate the models referred to above. Support has been found for the Native Language Magnet model (Näätänen et al., 1997; Kuhl, 2000; Aaltonen, Eerola, Hellström, Uusipaikka & Lang, 1997), for the Speech Learning Model (Baker, Trofimovich, Mack & Flege, 2002; Flege & MacKay, 2004; MacKay, Meador & Flege, 2001), and also for the Perceptual Assimilation Model (Best & Avery, 1999; Morrison, 2003; Polka, Colantonio & Sundara, 2001). The three models referred to above seek to explain the non-natives' perception and production difficulties that arise from differences between the particular L1 and L2 sound systems. It can also be mentioned that non-native listeners' L2 comprehension is more negatively affected by adverse listening conditions than native listeners' perception. Non-natives have more difficulty in coping with whispered speech, background noise, poor telephone connections, radio signal interference, and simultaneous speech (Lane, 1963; Trudgill, 2005). These perception problems experienced by L2 listeners are probably due to the loss of redundancy caused by the poor listening conditions (Gaies, 1977).

There is much empirical evidence showing that as the age of immersion in the L2 increases, the level of ultimate L2 attainment decreases. Put more simply, children seem more apt than adults to learn a new language without a foreign accent. Exactly what constitutes this "age-limit" in L2 acquisition remains unclear. Many believe that there is a so-called "critical period" extending up to late childhood or early adolescence during which the human language learning capacity is at its most acute, and after which it is virtually impossible to learn a new language without a foreign accent. Neurophysiological research from the 1960s on (Lenneberg, 1967; Scovel, 1995) has lent some support to this critical period hypothesis (CPH), claiming that the immature brain undergoes a process of specialization of the hemispheres, called lateralization, which ends at some point in childhood or early adolescence. It was thought that once the lateralization process was completed, the brain no longer had the plasticity necessary in order to learn and master a new language in a native-like way. However, many researchers now refute the idea of a critical period, pointing out that some late learners in fact do perceive and produce L2 speech at native-like levels (Birdsong, 2007; Bongaerts, 1999). Markham (1997) argues that investigations of speakers' L2 levels has traditionally been averaged over many speakers, thus overlooking outliers in the form of speakers with native-like competence or speakers with very poor L2 competence. In his study, Markham shows that some L2 learners are in fact able to reach an extremely competent level of the L2. Although it is rare for adults to become indistinguishable from natives, it is therefore not impossible. It is becoming increasingly more common to explain adults' general difficulty in native-like L2 acquisition, not in terms of biological constraints, but in terms of perceptual interference between the L1 and the L2 (as described in the NLM, SLM and PAM models described in the above paragraphs). The SLM explicitly states that the language learning ability remains intact over the life-span. But if age in itself does not predict a foreign accent, which factors do? A factor that intuitively seems to be a strong predictor of L2 performance is length of residence in the L2 country. However, while many studies support this view (Flege, 1988; Purcell & Suter, 1980), there are also quite a few findings that go against it (Moyer, 1999; McAllister, 2001), showing that it is not uncommon for long-term residents to have relatively poor competence in the L2. For instance, Flege, Munro & MacKay (1995a) reported a foreign accent in the English of Italian immigrants who had lived in Canada for as long as 30 years or more. Single factors like age of immersion in the L2 and length of residence in the L2 country are not in themselves powerful enough to predict L2 learner success. Much stronger predictions can be done when several different factors are considered together. There seems to be interaction between maturational and socio-psychological factors such that while age of immersion undisputedly is the most important predictor for degree of foreign accent, factors such as the type of L2 instruction, motivation, aptitude, amount of L1 use and length of residence also play important roles (Piske, Flege & MacKay, 2001). There are even some studies indicating a possible relationship between musical ability and L2 intonation acquisition success. For instance, Gottfried (2007) found that university conservatory students were better at producing and perceiving unfamiliar linguistic Mandarin tones than subjects without musical training. However, the intriguing idea that musical ability should be a factor in L2 learning is at this point not very well founded in the literature.

The relatively large amount of research describing and accounting for learner problems may seem disheartening to the L2 teaching community. However, there is also research regarding the pedagogical challenges posed by teaching second languages. Some of this research regards how learners can be helped to achieve certain goals. For example, non-native listeners use other cues than native listeners when perceptually distinguishing between L2 sounds. Flege's SLM holds that production will eventually become aligned with perception. It should therefore be fruitful from a pedagogical perspective to work with changing the way that non-native listeners perceive the L2. Research has shown that it is possible to redirect listeners' attentions to the same cues as the native listeners use (Guion & Pederson, 2007). It has also been shown that improvement in perception leads to improvement in production, and that the improved production can have long-term effects (Akahane-Yamada, Tokhura, Bradlow & Pisoni, 1996). So called high variability training seems to be a particularly successful approach. In such training, listeners are presented with two contrasting sounds in the L2 embedded in many different words and read by many different speakers. This variability enables the learner to construct robust perceptual representations of the L2 contrast. Researchers also work with making such perceptual training more effective by for instance adding visual information or by intensity enhancement of important parts of the signal, like the formant transitions (Hazan, 2002).

L2 teaching may also profit from technological advances in the form of special computer programs referred to as computer-assisted language learning (CALL). These programs are interactive and allow the learner to explore differences between his L1 and the L2 both visually (e.g. looking at pitch contours) and auditorily (e.g. listening to native examples). A description of such a program can be found in for example Bonneau, Camus, Laprie & Colotte (2004). A further development of such programs includes a virtual language tutor (a talking head) with whom the learner interacts (Wik, 2004). The virtual tutor keeps track of the particular problems of the individual learner, and tailors the lessons so that they answer to the learner's needs. The main advantages with computer assisted language learning programs are firstly that they let the learner record his L2 pronunciation and subsequently let him hear his own pronunciation corrected, and secondly that these programs are available for use whenever the learner has the time and the desire to engage in L2 training.

Other pedagogical concerns relate to the various communicative implications of foreign accents. Compared to the amount of research that seeks to account for why foreign accents occur, studies regarding the communicative implications of foreign accents are scarcer but are on the rise (Munro & Derwing, 2005). Foreign accented pronunciation has various effects on the speaker, on the listener and on the interaction between them. Investigators have almost exclusively been concerned with the negative effects of foreign accents, but it should be noted that there are in fact also positive implications of a foreign accent. This is because a foreign accent signals to the interlocutors that the L2 speaker may need an adjusted speech input. The L2 speaker is thus provided with so called "foreigner speech" which alleviates the communication (Gass & Varonis, 1984; Varonis & Gass, 1982). Many studies have shown that speaking with a foreign accent can give rise to discrimination and various negative social evaluations (Brennan & Brennan, 1981; Fayer & Krasinski, 1987; Gynan, 1985; Beebe, 1988; Kalin & Rayko, 1978; Lippi-Green, 1997; Munro, 2003). For instance, Munro (2003) reviews cases where L2 speakers have been discriminated by their employers because of their foreign accent. It is also well known that a foreign accent can hinder intelligibility (James, 1998; Lane, 1963), and the intelligibility of foreign-accented speech suffers more from adverse listening conditions than native speech (Munro, 1998). Of course, many aspects of speech contribute to intelligibility, but studies that compare the effects of pronunciation with other aspects of speech in fact tend to find that pronunciation is the most important aspect (Jenkins, 2000; Rajadurai, 2007). Not only is a foreign accented speaker at risk of eliciting unwanted negative evaluations and causing a communication breakdown because of reduced intelligibility, but native listeners also require more time (Munro & Derwing, 1995b) and expend more effort (Derwing & Munro, 1997) to process non-native speech than they do native speech.

Communicative interaction involves a speaker, a listener and a context. Researchers sometimes focus exclusively on factors in the utterances. It is however important to be aware of the contributions of listener factors and context factors to the interaction. Munro (2008) discusses the relative contributions of so called *stimulus properties* (SP, i.e. aspects of the utterance), *listener factors* (LF), and *contextual factors* (CF). He presents a model (revised and extended on the basis of Varonis & Gass (1982) and Gass & Varonis (1984)) for non-native speech perception. In this model, two terms of L2 speech perception are used, namely comprehensibility and intelligibility. These terms will be discussed in Chapter 4. In short, comprehensibility is the rated ease with which a listener perceives non-native speech, and intelligibility is the degree to which a listener identifies the word forms in a non-native utterance. The model is illustrated below.

SCORE= SP+ LF+ CF+…+error
Where SCORE refers to one of accentedness (A), comprehensibility (C), or intelligibility (I) and SP= αSeg+ βPros+ γGram+ δFlue+ …
LF= εFTop+ ζFSpkr+ ηFAcc+ …
CF= θCtxt

In the model, SCORE refers to either a score of accentedness (ranging from low values meaning native-like to high values meaning very foreign-accented), a score of comprehensibility (ranging from low values meaning easy to understand to high values meaning hard to understand), or a score of intelligibility (ranging from low values meaning few word forms identified to high values meaning all word forms identified). A low value for

A, C or I therefore indicate more native-like speech. The model shows that the SP involve aspects like segmental, prosodic, grammatical and fluency deviances. The Greek letter coefficients show how much a particular deviance affects the SCORE. The model also lists listener factors like familiarity with topic, familiarity with a speaker, and familiarity with a particular accent. The model lastly shows that context influences the SCORE. Munro (2008) goes on to discuss the relative contributions of the SP versus the LF component in the model: If SP factors contribute most, for instance to comprehensibility, then one would expect that there would be strong agreement between listeners regarding comprehensibility ratings. In contrast, if LF factors contribute most, one would expect comprehensibility ratings to vary across different listeners. Munro points out that if the latter scenario were true, this would mean that pronunciation teaching would be of little help because the effect would be different for different listeners. Munro concludes that while there are as yet rather few studies that have examined the relative contributions of the SP versus LF components, the existing literature seem to suggest that the SP component outweighs the LF component.

A foreign accented speaker is potentially faced with a number of problems affecting his interaction and communication with native speakers. Moreover, L2 learners themselves have been found to consider speaking without an accent a desirable goal (Derwing & Rossiter, 2002). Do all these research findings thus suggest that L2 teaching should have as its goal the eradication of foreign accents? L2 teaching has evolved alongside L2 research (although often more in parallel than in dialectic symbiosis). Before the 1960s, the *nativeness principle* set the standards for L2 teaching. This principle stated that the goal of L2 teaching should be to eradicate foreign accents and to attain native-like competence (Levis, 2005). In the wake of research findings suggesting biological constraints on L2 attainment, and recognizing that native-like competence was unrealistic, L2 teachers redefined their goal from that of nativeness to that of intelligibility. The *intelligibility principle* stated that the goal of L2 instruction should be for learners to attain a functional level of intelligibility. A widely cited passage from Abercrombie (1949: 120) supports this view: *"language learners need no more than a comfortably intelligible pronunciation"*. The intelligibility principle has been further consolidated as a sensible teaching goal in light of research findings showing that even heavily accented speech can be highly intelligible (Munro & Derwing, 1995a; Derwing & Munro, 1997). This means that even though heavy accents can hinder intelligibility, this is not necessarily the case. This partial independence between the degree of foreign accent and intelligibility is a very robust finding that has been substantiated in several studies (Munro &

Derwing, 2005). The intelligibility principle implies that different pronunciation aspects have different effects on intelligibility (Levis, 2005). Field (2005) notes the following: *"For some 30 years, intelligibility has been recognized as an appropriate goal for pronunciation instruction, yet remarkably little is known about the factors that make a language learner's speech intelligible"*.

The preceding paragraphs have provided a brief overview of the last several decades of research regarding foreign accents. In the following, the role of particular pronunciation deviations in foreign accented speech is discussed, both regarding the perceived degree of foreign accent and regarding the intelligibility of foreign accented speech. The roles of durational and intonational factors are discussed in light of this literature. Note that the term comprehensibility will be used to refer to a methodology where listeners rate how intelligible they feel the L2 speech to be. The term intelligibility will refer to methodologies of transcription, word identification tasks, paraphrasing of text content and the like. Note also that the following short presentations of different investigations will comprise only the information judged relevant for the present investigation. For instance, if an investigation has measured aspects of L2 speech such as grammatical correctness as well as degree of foreign accent and intelligibility, only the information regarding degree of foreign accent and intelligibility will be extracted for the short presentations here.

1.2 The relative importance of pronunciation deviations

Some studies have investigated the impact of *one single* pronunciation aspect upon the degree of accent and/or intelligibility. While such studies do not show the relative importance of pronunciation aspects, they do show if particular aspects are relevant for the perception of L2 speech. We will therefore first have a brief look at some of these studies.

Tajima, Port & Dalby (1997) investigated the effects of durational corrections on the intelligibility of Chinese-accented English. They recorded a native Chinese speaker and a native English speaker reading the same English utterances. Utterances were manipulated so that the non-native utterances' segments were given native segment durations. Also, the native English utterances were manipulated so that their segment durations matched the Chinese speaker's segment durations. 36 listeners were used. Intelligibility was measured through forced-choice identification tests (the correct utterance plus three similar utterances). They found that intelligibility had been affected by the durational adjustments.

Munro (1995) investigated low-pass filtered Mandarin-accented English in which all segmental information was removed. The 20 native English listeners were still able to distinguish the foreign-accented speech from the native speech. This is interpreted to mean that listeners make judgements based on prosodic characteristics such as intonation and speaking rate.

Bond, Stockmal & Markus (2003) studied the impact of sentence durations on degree of foreign accent in Latvian. They recorded native Latvian speakers as well as long-term residents of Latvia with Russian as their L1. Three listener groups listened to the native and Russian-accented Latvian speech. These groups were a) native Latvian, b) long-term residents of Latvia with Russian as their L1, and b) American with no knowledge of either Latvian or Russian. The listeners judged whether the speech they heard was native or non-native Latvian. The native Latvian listeners were very good at identifying the foreign accented speech (88 % correct), the residents of Latvia for whom Latvian was an L2 were slightly less correct (83 %), but most surprisingly, the American listeners also scored above chance (62 %). The researchers found that sentence durations significantly correlated with the degree of accent ratings, but only for the American listeners.

Munro & Derwing (1998) investigated the effects of natural as well as manipulated variation in speaking rate on the degree of foreign accent in the foreign-accented English of 10 native speakers of Mandarin. The non-native speech was produced at slower rates when compared with the speech of native Canadian English speakers. 10 native Canadian English listeners performed foreign-accent ratings. The non-native speech became more foreign-accented when the speech was slowed down, while it became less foreign-accented when the speech was speeded up. The optimal speaking rate for the non-native speech was however slower than the native rate.

In a follow-up study to their 1998 investigation, Munro & Derwing (2001) used 48 non-native speakers of Canadian English from various L1s in the first part of the investigation and 10 non-native speakers of Canadian English as produced by native Mandarin speakers in the second part. A total of 55 native Canadian English listeners assigned a stronger foreign accent to non-native speech that was produced at slower rates. The results also consolidate the finding from their 1998 study that the optimal speaking rate for non-native speech was slower than the rate for native speech.

Flege (1988) investigated the durational aspect of pauses. He removed pauses from 47 Mandarin and Taiwanese speakers' foreign-accented English. Native English listeners as well as Taiwanese and Mandarin L2 listeners judged the degree of foreign accent in the L2 utterances. He observed no effect of pause-removal on the degree of accent.

This selection of studies shows that both durational and intonational aspects *do* affect the degree of foreign accent and intelligibility, although they *do not* show their relative importance. Moreover, the last investigation in the selection (Flege, 1988) exemplifies that not all L2 pronunciation deviation is always found to be relevant for the perception of L2 speech. The following examines studies exploring the relative impacts of various pronunciation aspects on the degree of accent and on intelligibility. We begin by looking at studies that gauge the impacts of *prosodic* versus *segmental* deviations in L2 speech.

1.2.1 Prosody versus segmentals

It has long been debated whether segmental or prosodic deviations are more important for the perception of L2 speech. Moreover, as explained earlier, this may very well differ according to the perceptual dimension investigated, for instance the degree of foreign accent or the intelligibility. The results from the following studies contribute to the discussion of the relative importance of prosodic versus segmental aspects.

Boula de Mareüil & Vieru-Dimulescu (2006)[2] recorded Italian and Spanish speech produced by 1 Italian, 1 Spanish and 3 Italian/Spanish bilingual speakers. In the recordings they crossed segment durations and intonation between the utterances. The intonation and duration were manipulated as one compound prosodic parameter. Their aim was to find the relative importance of this compound prosodic parameter as compared to segmental information (e.g. articulation of segments). 20 native Spanish and 20 native Italian listeners were asked to identify the speech as either Italian, Spanish-accented Italian, Spanish or Italian-accented Spanish. The perception tests were conducted through the Internet. The results were interpreted such that degree of foreign accent was equally influenced by the prosodic and the

[2] Their paper reported two experiments, but because the first yielded unclear results and had a methodology that was subsequently questioned by the researchers themselves, it was chosen to only discuss their second experiment.

segmental information. Of course, the fact that this perception test was done over the Internet with non-optimal listening conditions, and probably varying listening conditions for different listeners (different headsets and surroundings), makes the results from this investigation somewhat unreliable.

Anderson-Hsieh, Johnson & Koehler (1992) investigated perceptions of English utterances produced by 60 speakers from many different L1s (Arabic, Armenian, Assamese, Chinese, Farsi, German, Greek, Hindi, Indonesian, Kannada, Korean, Malayam, Punjabi, Serbo-Croatian, Spanish and Tamil). Three native English listeners rated the degree of accent as well as the perceived intelligibility (ratings of how intelligible they felt that the utterances were). The researchers then tried to relate these ratings to analyses of deviances regarding prosody, segmentals and syllable structure. They found that all types of errors correlated with both the degree of foreign accent and with the perceived intelligibility, but that prosody was more important than segmentals, and segmentals were more important than syllable structure in determining the degree of foreign accent. This investigation therefore suggests that prosodic information is of superior importance in terms of degree of foreign accent.

Derwing & Rossiter (2003) investigated the effects of prosodic versus segmental pronunciation training on 48 learners of English from a variety of L1s. Six ESL teachers rated the speakers' degree of foreign accent, fluency and perceived comprehensibility (ratings of how comprehensible they felt that the speakers were). Improvement in terms of higher ratings for comprehensibility and fluency were shown only for the learners that had received prosodic training. The results from this investigation support the results from Anderson-Hsieh, Johnson & Koehler (1992) above, in that prosody was found to be of greater importance. The main difference regarding the outcomes of these investigations is that prosody was found to be the most important factor for *degree of accent* in Anderson-Hsieh, Johnson & Koehler (1992) whereas it was the most important factor for *perceived intelligibility* in Derwing & Rossiter (2003).

Derwing, Munro & Wiebe (1998) conducted a study similar to Derwing & Rossiter (2003), in that they too compared the effects of segmental versus prosodic pronunciation training. They investigated the non-native English of 48 learners from the L1s Arabic, Cantonese, Mandarin, Japanese, Persian, Polish, Russian, Serbo-Croatian, Spanish, Turkish, Ukrainian and Vietnamese. 48 native English listeners rated the degree of accent as well as comprehensibility (how intelligible they felt that the utterances were) in non-native sentences before and after training. 6 native English listeners also rated the degree of foreign accent, the comprehensibility and the fluency in extemporaneously produced narratives. While both speaker groups showed improvement as for degree of accent and comprehensibility, only the group which had received prosodic training also showed improvement regarding fluency in their narratives. This investigation adds to the impression that while segmentals are important, prosody may be even more important for the perception of L2 speech.

The investigations referred to above indicate that prosody is more important than segmentals in L2 perception. In contrast, Boyd, Abelin & Dorriots (1999) came to a different conclusion when they investigated how segmental, prosodic and phonotactic factors affected degree of foreign accent in their material. They investigated the Swedish productions of 5 speakers with the 4 L1s Hungarian, Spanish, Persian and Russian. 54 judges rated the degree of foreign accent in their Swedish L2 speech. The researchers then analyzed the L2 productions auditively and concluded that the various deviations had contributed *equally* to the degree of foreign accent.

In spite of conflicting results across studies, it is now generally believed in both teaching and research communities that prosodic aspects may be more important than segmental aspects (Anderson-Hsieh, Johnson & Koehler, 1992; Anderson-Hsieh & Koehler, 1988; Hahn, 2004; Derwing & Munro, 1997; Munro & Derwing, 2005; Jilka, 2000). However, as Munro & Derwing (2005) pointed out, there are as yet rather few studies on which to base this belief. Munro & Derwing (1995b) suggested that segmentals may be more important in determining the degree of foreign accent while less important for intelligibility.

Field (2005) pointed out that many studies have treated prosody as a unitary aspect of speech, and went on to advocate that the impacts of the various constituents of prosody be determined. The remainder of this chapter looks at how particular pronunciation deviations, especially those that can be viewed as prosodic, influence listeners' perceptions, first in terms of the degree of foreign accent and second in terms of intelligibility.

1.2.2 Degree of foreign accent

It seems fairly well established that the perceived degree of foreign accent correlates with simple error counts in non-native speech, such that more errors give the impression of a stronger foreign accent (Anderson-Hsieh, Johnson & Koehler, 1992; Boyd, Abelin & Dorriots, 1999; Brennan & Brennan, 1981; Munro & Derwing, 1999). At the same time, however, several studies have shown that listeners assign different perceptual weightings to different pronunciation aspects such that certain aspects have a greater impact than others (Anderson-Hsieh, Johnson & Koehler, 1992; Munro & Derwing, 1995a). Several studies have investigated the relative impact of different pronunciation aspects on the degree of foreign accent. This section gives an overview of such studies.

Boyd, Abelin & Dorriots (1999) investigated the Swedish L2 speech of 5 speakers from the L1s Hungarian, Spanish, Farsi, Persian and Russian. 54 native Swedish listeners rated the degree of foreign accent in their utterances. The researchers analyzed the L2 utterances auditively in terms of segmental, prosodic and phonotactic deviances. They concluded that all types of deviations had contributed equally to the degree of foreign accent. This study did therefore not find any evidence for different weightings of different pronunciation aspects.

Wayland (1997) investigated foreign accented Thai. He recorded 3 native Thai and 6 native English speakers reading Thai. The native and non-native productions were analyzed and found to differ more spectrally (F0 and formant frequencies) than temporally (VOT and vowel durations). 3 native Thai listeners then rated the degree of foreign accent. Regression analyses between the production data and the rating data showed that the deviant production of Thai tone significantly contributed to the degree of foreign accent in his material. Wayland's investigation thus indicated that intonational aspects were of superior importance for degree of accent in his material.

Magen (1998) recorded two native Spanish speakers' productions of English. She manipulated a range of aspects of foreign accented speech in terms of syllable structure (epenthetic schwa), vowel quality (reduction, tense/lax), consonant articulation (articulation manner, deletions), fricative voicing, stop voicing and stress (lexical, phrasal). 10 native English listeners judged the degree of foreign accent of the utterances. It was found that syllable structure, consonants and stress affected the degree of foreign accent more than voicing.

Munro & Derwing (1995a) investigated perceptions of foreign-accented English as spoken by 10 native speakers of Mandarin. 18 native English listeners rated the degree of accent in the utterances. The researchers found that the degree of accent correlated with phonetic, phonemic and grammatical errors and with goodness of intonation ratings.

Gut (2007) investigated foreign-accented German as produced by 55 speakers from 24 different L1s and foreign-accented English by 46 speakers from 17 different L1s (the particular L1s were not specified). She performed production analyses exploring the differences between the native and non-native speech. She then conducted perception experiments. 7 native German listeners rated the degree of foreign accent for each of the 55 German L2 speakers, and 4 native British English listeners rated the degree of foreign accent for each of those English L2 speakers who aimed at a British English pronunciation (number not specified). The speaker ratings were investigated for correlation with durational (rate and reduction) and intonational (range and movement) aspects of their non-native productions. In general, Gut found that speaking rate was the most important factor affecting the degree of foreign accent.

Kamiyama (2004) investigated intonational and durational contributions to degree of foreign accent in Japanese-accented French. Kamiyama used both synthesized speech and manipulated natural speech to investigate the roles of intonation and duration. The speech was based on 11 Japanese L2 French speakers and 4 native French speakers. 17 native French listeners judged the degree of foreign accent in the utterances. The main finding was that intonation affected the foreign accent more than durations. One utterance had also been selected in which speaking rate and pauses were manipulated. Kamiyama reported that neither speaking rate nor pauses affected the degree of accent in this material. The finding that speaking rate does not affect degree of foreign accent is in conflict with the findings in Gut (2007) above, and the findings in Munro & Derwing (1998) and Munro & Derwing (2001) in the beginning of section 2.1, which showed that speaking rate *did* affect the accent. The finding in Kamiyama (2004) may be less reliable than the findings in the other studies because Kamiyama's speech material was short fragments of sentences lacking verbs. A listener may need a somewhat longer stretch of speech to get a clear impression of the speaking rate. In contrast, Gut (2007) reported investigating story retellings and read passages, Munro & Derwing (1998) investigated read passages, and Munro & Derwing (2001) investigated complete sentences.

Trofimovich & Baker (2006) investigated 30 speakers' Korean-accented English. 10 native English listeners rated the degree of foreign accent in the Korean-accented utterances. The researchers then analyzed the accented utterances for deviations regarding stress timing, peak alignment, speaking rate, pause length and pause frequency. Correlation analyses between the particular deviances and the rated degrees of foreign accent showed that all these types of deviances had affected the perception of accent, but the analyses also showed that pause duration and speaking rate had a greater effect than stress timing and peak alignment. This study indicates that durational factors like speaking rate are of great importance in the perception of degree of accent.

Lastly, we consider two investigations in which the target languages were Scandinavian, namely Almberg & Husby (2000) and Bannert (1995).

Almberg & Husby (2000) compared the effects of manipulating durational and intonational aspects of one native Russian speaker's Norwegian speech. 16 native Norwegian listeners rated the utterances for degree of foreign accent. Durational aspects were found to affect degree of foreign accent more than intonational aspects.

Bannert (1995) investigated foreign accent in Swedish. His two speakers' native languages were Punjabi and Persian. He manipulated durational and intonational aspects of their Swedish utterances. 20 native Swedish listeners rated the acceptability of the utterances (acceptability was equated with degree of foreign accent in this study). Bannert concluded that intonational aspects affected degree of accent more than durational aspects. This investigation had however used only fragments of sentences, which could have influenced listeners' perceptions of these prosodic aspects.

Although there were different findings across many of the investigations presented above, the investigations that were methodically more reliable in terms of the largest number of subjects (Gut, 2007; Trofimovich & Baker, 2006) suggest that durational aspects affect the degree of foreign accent more than intonational aspects. Among durational aspects, speaking rate seems to be particularly important, with a faster rate found to reduce foreign accent. The two investigations with closely related Scandinavian target languages (Almberg & Husby, 2000; Bannert, 1995) should intuitively show similar results, and the fact that there were instead conflicting results across these two studies was perhaps due to few subjects and different L1 groups. However, the investigation with Norwegian as the target language (Almberg & Husby, 2000) supports the finding from Gut (2007) and Trofimovich & Baker (2006) in showing that duration seems to be more important for degree of foreign accent than intonation.

1.2.3 Intelligibility

The previous section showed that intonation and duration contribute to the degree of foreign accent, and suggest that durational aspects, particularly speaking rate, may be more important than intonational aspects. This section presents investigations of the intonational and durational contributions to intelligibility. Note that several studies have investigated both degree of foreign accent and intelligibility. Because of this, some of the studies that were referred to in the previous section are also discussed here.

Intelligibility (and other dimensions relevant for understanding) has been investigated in different ways and using different terms across different investigations. A widely used approach is to let listeners write down the words that they perceive. This will be called *intelligibility* regardless of the term used in the particular investigations. Another widely used method is to let listeners rate how well they feel that they understand the speech (a method that some researchers, e.g. Munro & Derwing, 1995a, believe show the perceptual processing load rather than show how much the listener can actually perceive). This will be referred to as *perceived comprehensibility*. A third common method is to ask listeners about the content or message of what they hear. This will be referred to as *comprehensibility*. Methodologies and terms differ in the field of foreign accent research, but the choice to use the terms intelligibility, perceived comprehensibility and comprehensibility as explained above is in line with for instance Smith & Nelson (1985) and Munro & Derwing (e.g. 1995a). It is important to keep methodologies and terms apart because differences in this respect can yield different results, as will be discussed in the beginning of Chapter 4. In that chapter, it will also be explained that the present investigation will investigate intelligibility through listener transcriptions of the words that they perceive. As a background for this investigation, studies that investigate intelligibility through listener transcriptions, i.e. *intelligibility* studies, are therefore of particular interest.

Huckvale (2006) investigated one English speaker's English-accented Japanese. The speaker read Japanese words from a list. The segmental quality, pitch and timing were manipulated to match native Japanese pronunciation. Intelligibility was measured by letting eight Japanese listeners write down the words that they perceived. The results showed that pitch was the only significant aspect that affected intelligibility. This study therefore suggests that pitch is of great importance for intelligibility. The speaker in this study was however unfamiliar with Japanese (and therefore read the Japanese words in Romanised re-spelling), which may have affected the results.

Anderson-Hsieh & Koehler (1988) investigated speaking rate in 3 native Chinese speakers' English. The investigation used 224 native English listeners. Their comprehensibility scores were measured through questions about text content. When the L2 speakers spoke faster, the degree of comprehensibility dropped. This study used naturally produced different speaking rates. The results may therefore have been affected by the likelihood that the speakers produced more errors when speaking faster than normal. It is therefore uncertain if this study has actually investigated the effect of speaking rate per se.

Munro & Derwing (1995a) investigated Mandarin-accented English. Just a few listeners showed a correlation between intelligibility and measurements of phonetic, phonemic and prosodic deviations and intonation goodness ratings. In other words, these deviations were not found to affect intelligibility across listeners. The results from this investigation are therefore unclear.

In a follow-up study to their 1995a investigation, Derwing & Munro (1997) investigated more L1 groups. They investigated the foreign-accented English of 12 speakers from each of the L1s Cantonese, Japanese, Polish, and Spanish. 26 native English listeners rated the perceived comprehensibility and also provided transcripts of the utterances that they heard (intelligibility). The perceived comprehensibility was affected by speaking rate whereas the intelligibility remained unaffected. This study therefore points to the importance of speaking rate for perceived comprehensibility, but not for intelligibility.

Munro & Derwing (1998) investigated speaking rate in the non-native Canadian English of 10 Mandarin speakers. The speaking rate was both varied naturally and by means of manipulation. 10 native Canadian listeners rated the perceived comprehensibility. The perceived comprehensibility dropped when the speaking rate became slower. When the speaking rate was *slightly* speeded, the perceived comprehensibility increased. However, if the non-native speaking rate was speeded to the extent that it became as fast as the *native* speaking rate, the perceived comprehensibility dropped again. A moderate acceleration was thus optimal.

Munro & Derwing (2001) was a follow-up study to their 1998 study. They used 48 non-native speakers of Canadian English from various L1s (Arabic, Cantonese, Mandarin, Japanese, Persian, Polish, Russian, Serbo-Croatian, Spanish, Turkish, Ukrainian and Vietnamese). A total of 55 native Canadian English listeners rated the perceived comprehensibility. Slower speaking rates caused poorer perceived comprehensibility. The results were in accordance with the results from Munro & Derwing (1998) also in that faster non-native speaking rates were beneficial for perceived comprehensibility as long as the rate of the non-native speech did not became as fast as the native speaking rate. These two studies taken together (Munro & Derwing, 1998 and 2001) therefore suggest that perceived comprehensibility is affected by speaking rate, such that faster rate is beneficial, but speaking rates as fast as the native rate is detrimental.

Almberg & Husby (2000) manipulated durational and intonational aspects of one native Russian speaker's Norwegian speech. 16 listeners participated. They rated the perceived comprehensibility of the utterances. The results showed that durational aspects were more important than intonational aspects for the perceived comprehensibility.

Bannert (1995) manipulated intonational and durational aspects of Swedish L2 speech as produced by two native speakers of Punjabi and Persian. There were 20 native Swedish listeners who rated the perceived comprehensibility of the utterances. The results showed that the intonational corrections affected the perceived comprehensibility more than the durational corrections.

The results from the investigations presented in this section suggest that durational aspects, particularly speaking rate, affects the intelligibility of foreign accented speech. In the previous section it was concluded that a faster speaking rate seems to be desirable in order to reduce the degree of foreign accent. It seems that the effect of speaking rate on *intelligibility* may be such that a rate as fast as the native rate impedes intelligibility. This may be explained in light of the finding that native speakers need more time to process foreign accented speech (Munro & Derwing, 1995b). Therefore, as the speaking rate accelerates, the listener has less time to process the speech, and if the speaking rate is accelerated too much, the intelligibility and the perceived comprehensibility drop. A *moderate* acceleration of the speaking rate therefore seems to be desirable for the purpose of enhanced intelligibility. The two investigations with Scandinavian target languages (Almberg & Husby, 2000; Bannert, 1995) show opposite results. The investigation of Norwegian N2 speech (Almberg & Husby, 2000) however supports the general impression from the literature in that duration was found to be more important than intonation for the purpose of L2 intelligibility.

1.3 Aim and hypotheses

1.3.1 Aim of study

The present investigation is a phonetic exploration of the relative impacts of durational and intonational aspects on both the degree of foreign accent and the intelligibility of foreign accented speech. Relatively few studies have investigated the relative impacts of different pronunciation aspects. This is especially the case for intelligibility. Because improved intelligibility has become the aim in L2 teaching, studies of this type are important in establishing guidelines for teachers' priorities in second language pronunciation teaching. Note that intonation is here narrowly defined as "the ensemble of pitch variations in the course of an utterance" ('t Hart, Collier & Cohen, 1990: 10).

1.3.2 Hypotheses

Formulating hypotheses about foreign accented speech based on existing literature is not a straightforward proposition because, as Bent, Bradlow & Smith (2007) pointed out, it is difficult to make clear comparisons across studies due to their widely different methodologies (which incidentally led Kent, Weismer, Kent & Rosenbek (1989) to comment that intelligibility studies cannot be compared without considering the specific conditions under which the data were collected). Moreover, the target language in the present thesis work is Norwegian, whereas the literature is heavily dominated by studies with English as the target language. Almberg & Husby (2000) referred to above had Norwegian as the target language. This investigation found that durational aspects were more important than intonational aspects in determining the degree of foreign accent and for perceived comprehensibility. This finding is thus in accordance with the general impression from the literature, which in points towards the superior role of duration, particularly speaking rate, both for the degree of accent and for intelligibility.

In light of the existing literature, the following hypotheses have been formulated:

A. Both intonational and durational aspects will affect the degree of foreign accent.
B. Durational aspects will affect the degree of foreign accent more than intonational aspects.
C. Both intonational and durational aspects will affect intelligibility.
D. Durational aspects will affect intelligibility more than intonational aspects.

1.4 Norwegian prosody

This thesis investigates the perceptual roles of prosodic aspects in non-native speech. As a general background for this investigation, this section provides an overview of the phonetics and phonology of Norwegian prosody. Note that the present thesis is a phonetic work, and the remainder of the thesis will therefore use phonetic analysis approaches and mostly phonetic terminology.

1.4.1 The syllable

In a traditional approach, the Norwegian syllable can have complex onsets and complex codas. The nucleus does not have to be a vowel. It can also be a liquid or a nasal.

1.4.2 Stress

A light syllable has only short segments in the rhyme, whereas a heavy syllable has one long segment in the rhyme. In Norwegian, all heavy syllables are stressed. There are long segments only in syllables with (primary or secondary) stress in Norwegian. In virtually all Norwegian dialects, all stressed syllables are heavy. Syllables with primary stress display certain F0 patterns, which will be discussed in section 1.4.4.

1.4.3 Quantity

Norwegian has a two-way phonological vowel length opposition between long and short vowels. Phonologically long vowels occur only in syllables with (primary or secondary) stress, and most often in stressed syllables with no more than one consonant in the coda. The perceived quantity of a vowel is not only caused by the duration of the vowel itself. The duration of the following consonant is also important. In general, phonologically short vowels are followed by a (phonetically) longer consonant, forming a VC: pattern, and phonologically long vowels are followed by a (phonetically) shorter consonant in a V:C pattern. Figures 1.1 and 1.2 exemplify the durational aspect of the Norwegian vowel quantity distinction.

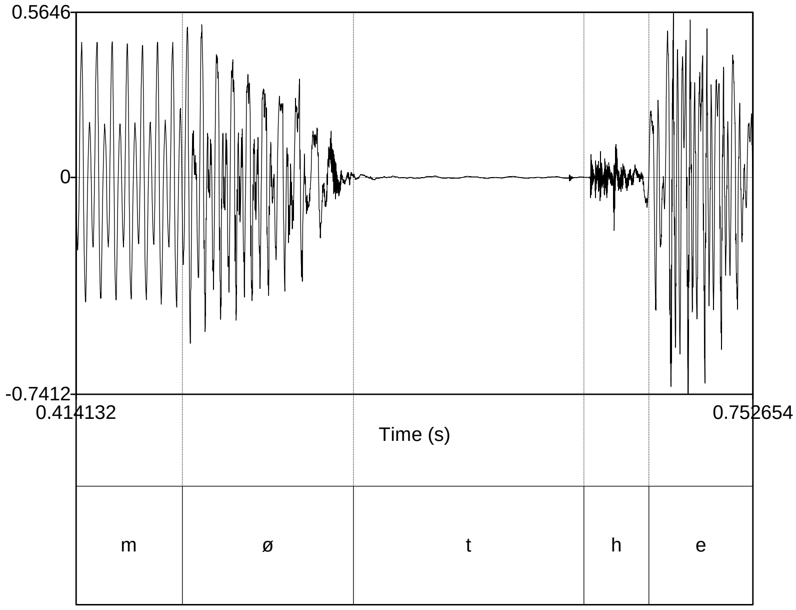

Figure 1.1: Phonologically short vowel in the word “møtte” (*met).* The illustration shows that the vowel /ø/ is followed by a phonetically longer consonant /t/.

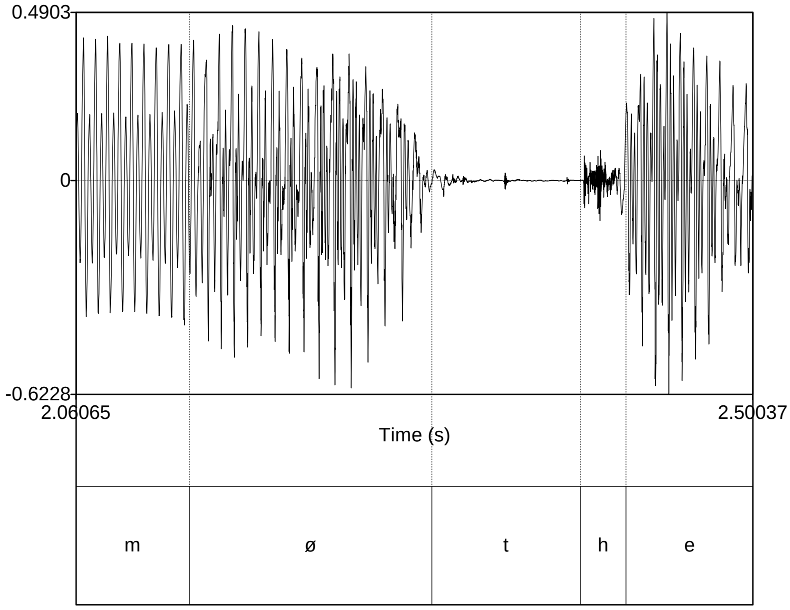

Figure 1.2: Phonologically long vowel in the word “møte” (*meeting*). The illustration shows that the vowel /ø/ is followed by a phonetically shorter consonant /t/.

Figure 1.1 shows that the phonologically short vowel is followed by a phonetically longer consonant, and Figure 1.2 shows that the phonologically long vowel is followed by a phonetically shorter consonant. In addition to this durational ratio between the vowel and the consonant, listeners also use spectral cues in the vowel in order to determine the quantity. Often, the phonologically short vowels have more lax articulations than their long counterparts, causing the formants to centralize. The relative importance of the durational V/C cue and the spectral vowel cue has not been established for Norwegian. However, in Swedish, which is a language very closely related to Norwegian, it has been found that listeners use both the V/C durational ratio cue and spectral vowel cues when determining the phonological vowel length of Swedish vowels, but that the relative importance of these two cues varies according to the particular vowel phoneme (Thorén, 2003). Because Norwegian and Swedish are very similar and closely related languages, this finding may also apply to how Norwegian listeners perceive Norwegian vowel quantity.

1.4.4 Word accents

Norwegian has two tonal accents that are lexically determined, and that are generally referred to as accent 1 and accent 2. An accent contour stretches over a stressed syllable and at least one following unstressed syllable. Quite a few minimal pairs are distinguished only on the basis of the accents. The accents are however not only realized within isolated words. In continuous speech, a stressed syllable followed by at least one unstressed syllable always initiates one of the two accent contours. The domain of an accent contour is called an accent phrase.

In Norwegian dialectology, a dichotomy exists between the dialects where accent 1 is realized with a *low* tone on the stressed syllable (so called low tone dialects) and the dialects where accent 1 is realized with a *high* tone on the stressed syllable (so called high tone dialects). The following figures show the difference between the accent 1 realization in a low tone dialect (Figure 1.3) and the accent 1 realization in a high tone dialect (Figure 1.4). The same word is used in both examples.

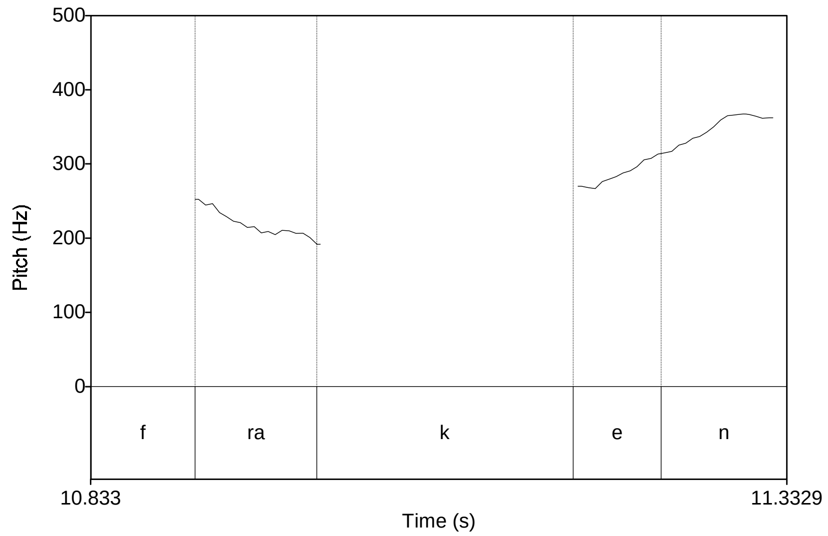

Figure 1.3: The accent 1word "frakken" (*the coat*) as spoken by a speaker with a low tone dialect (Southeast Norwegian). There is a low tone associated with the stressed syllable.

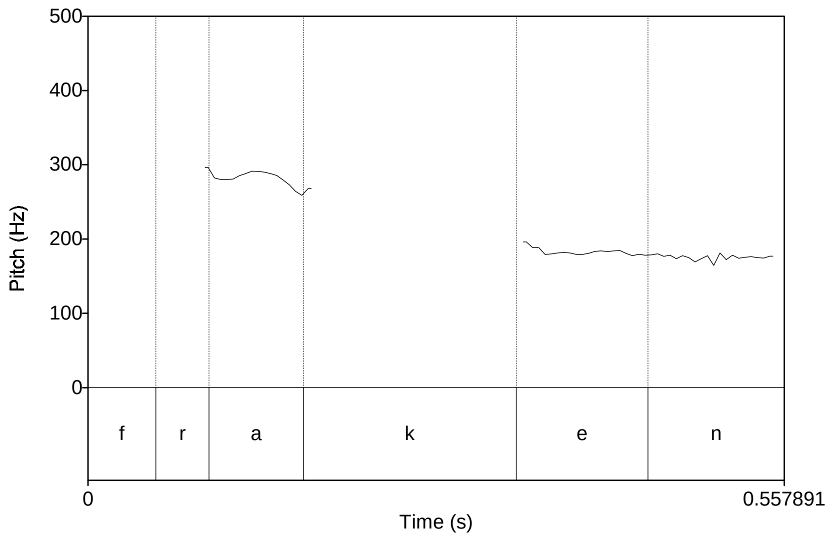

Figure 1.4: The accent 1word "frakken" (*the coat*) as spoken by a speaker with a high tone dialect (North Norwegian). There is a high tone associated with the stressed syllable.

The examples show that accent 1 is realized with a low tone on the stressed syllable in the low tone dialect (Figure 1.3) and with a high tone on the stressed syllable in the high tone dialect (Figure 1.4).

The contrast between accent 1 and accent 2 is illustrated in the following figures which show the difference between the accents in a low tone dialect (Southeast Norwegian). Figure 1.5 shows accent 1 and Figure 1.6 shows accent 2.

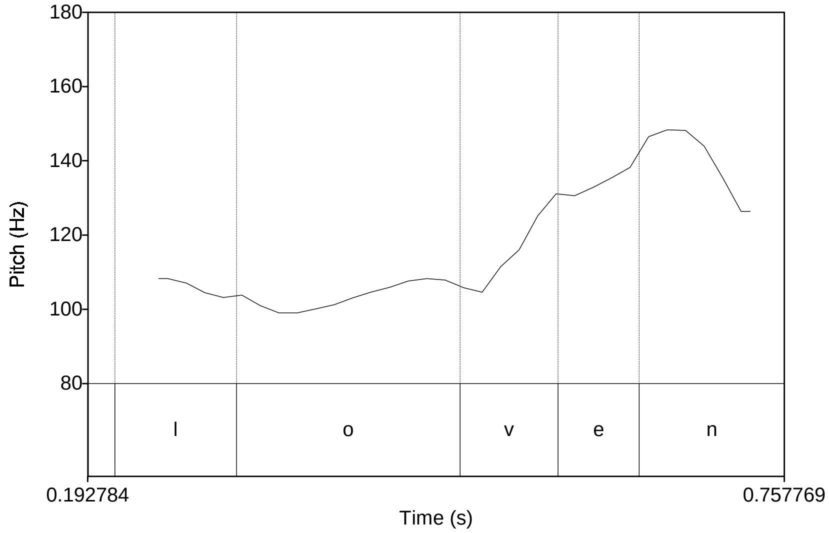

Figure 1.5: The accent 1 word “loven” (the law) as spoken by a speaker with a low tone dialect (Southeast Norway). There is a low tone associated with the stressed syllable.

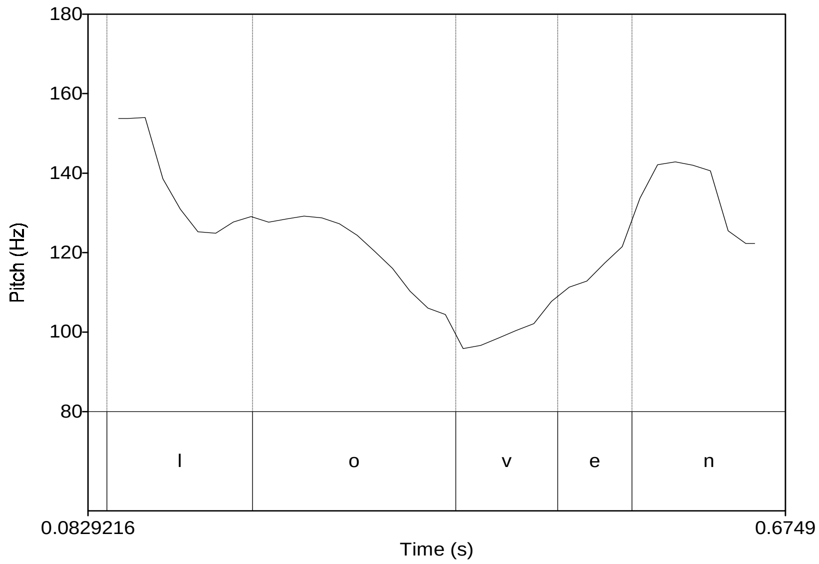

Figure 1.6: The accent 2 word “låven” (the barn) as spoken by a speaker with a low tone dialect (Southeast Norway). There is a movement from high to low associated with the stressed syllable.

The figures show a low tone on the stressed syllable in the accent 1 word “loven” (Figure 1.5) and a movement from high to low on the same syllable in the accent 2 word “låven” (Figure 1.6).

High tone dialects are found in the North, West and South of Norway, whereas low tone dialects are in the South East and Trønder (middle part of Norway) areas. As explained earlier, an accent contour stretches over a stressed syllable and at least one following unstressed syllable. However, there are a few Norwegian dialects in which accent 2 contours are realized over just *one syllable*, and they are referred to as circumflex tones. The circumflex tone is often explained with reference to a diachronic phenomenon whereby some accent 2 words have lost their final syllable yet retained the accent 2 contour. The circumflex tone exists in just a few Norwegian dialects around the Trønder area and in some mid-Northern parts of the country. It should lastly be mentioned that there are some Norwegian dialects which do not have accents. These are dialects in Nord-Troms, Finnmark, parts of Helgeland and areas around Bergen. For information about the Norwegian accents, see for instance Fintoft, Mjåvatn, Møllegård & Ulseth (1978), Kristoffersen (2000) and Hognestad (1997).

1.4.5 Rhythm

Different languages give the impression of different speech rhythms. Exactly what constitutes this perceived rhythm remains unclear. A traditional approach holds that the impression of rhythm is caused by the isochronony (= time constancy) of some unit in the signal. In so called *stress timed* languages, stressed syllables are thought to recur at equal durations. The isochrony in this type of rhythm therefore comes from the regular intervals between stress beats. In *syllable timed* languages, the syllables are supposedly of equal length (Pike, 1945; Abercrombie, 1967), and in *mora timed* languages, the successive morae are thought to have equal durations (Han, 1962; Bloch, 1950). Norwegian has been classified as a stress timed language. French is often mentioned as an example of a syllable timed language, as Japanese often exemplifies a mora timed language. However, phonetic endeavours to find the acoustic correlates of rhythm find very little support for the existence of such isochrony-based rhythm categories (Beckman, 1992; Laver, 1994). There have been many other approaches to try and measure speech rhythm. A current and fairly widespread approach is the so called Pairwise Variability Index (Grabe & Low, 2002) which measures the level of variability in vocalic and inter-vocalic intervals. The search for the acoustic correlates of speech rhythm is very much ongoing, and at this point there is therefore no universally standardized method of measurement. In general, phonetic approaches to speech rhythm often measure the ratios or intervals between successive units in the time domain.

The phonetic manipulations that will be carried out in this investigation will affect perceptions of stress, quantity, tone and rhythm. The relevance of these features for degree of foreign accent will be investigated in section 3.6, Chapter 3. The features' relevance for intelligibility will be investigated in section 4.5, Chapter 4.

1.5 Outline of study

In Chapter 2, the design and recording of the speech material, comprising both native Norwegian and foreign accented Norwegian, will be described. The same chapter explains the methods of speech manipulation for the generation of speech stimuli to be used in perception experiments. Chapter 3 describes the perception experiment used to investigate the degree of foreign accent, and Chapter 4 describes the perception experiment used for intelligibility. Towards the ends of Chapters 3 and 4, production analyses will be presented that relate specific details in the manipulations to the perceptual effects of the same manipulations to determine exactly which detailed changes of the foreign accented speech have caused the observed perceptual effects. In Chapter 5, the results and conclusions of the investigation are discussed. Information regarding statistical tests can be found in Appendices A, B, C and D. Appendix E provides a list of the recorded sentences from which the sentences in the experiments have been selected.

2. Speech corpus and manipulation methods

Towards the end of the previous chapter it was explained that the purpose of the present investigation is to measure the impacts of durational and intonational aspects on the degree of foreign accent and intelligibility of N2 speech. In order to investigate perceptions of non-native speech, it was necessary to first compile a speech corpus. In section 2.1, the design and recording of a speech corpus and the subsequent selection of speakers for the present investigation will be described. The recorded N2 speech was further subjected to digital speech manipulations. The manipulations adjusted durational and intonational aspects of the N2 speech. The speech manipulations will be described in section 2.2.

2.1 Speech corpus

The speech corpus consists of both N2 and N1 speech and comprised speakers from many different L1s. It will be explained how a fairly large speech corpus was recorded and how a smaller part of this corpus was subsequently selected for the present investigation.

2.1.1 Speech corpus design

Different types of speech were recorded. The recordings consisted of a spontaneous interview, a read text and read sentences. It was ultimately decided to use only the read sentences. It has long since been established that there are acoustic differences between read and spontaneous speech (Blaauw, 1995; Koopmans- van Beinum, 1980; Caldognetto, 1997; Holm, 2003), but it has also been found that the perceived degree of foreign accent does not differ between read and spontaneous speech (Munro & Derwing, 1995a). It is therefore unlikely that the choice of speaking style will affect the results of the investigation.

There are advantages with the choice of read speech over spontaneous speech, in particular regarding the level of control the experimenter has with the linguistic content of the speech and the length of sentences. The most important advantage is that using read sentences enables the comparison of the same sentence as uttered by an N1 speaker and an N2 speaker. As will be explained in the next chapter, comparisons across N1 and N2 utterances of the same sentence are at the core of the experimental design of the present study.

Sixty sentences were constructed, each consisting of 5 to 11 words. The sentences were designed so as to capture all segmental phonemic variation in Norwegian. This was done by letting each Norwegian consonant phoneme occur at least 4 times in word *initial* position, at least 4 times in word *medial* position and at least 4 times in word *final* position across the sixty sentences. All positions were however not possible for all consonants because of Norwegian phonotactic constraints. As for the vowels, the phonologically *long* vowels occurred at least 4 times, and the phonologically *short* vowels also occurred at least 4 times across the sixty sentences. Each of the Norwegian *diphthongs* occurred at least 4 times. Note that this approach assured that there was a *minimum* of 4 occurrences for all the mentioned segments, but that many of them occurred *more* than 4 times across the 60 sentences. This sentence design, which assured segmental variability, was chosen in the early stages of the investigation because the experimenter wanted to have the possibility to investigate segmental aspects of non native speech. However, it was later decided to only investigate prosody in this work.

The use of different sentences could potentially affect the results in this type of investigation. However, as will be explained later, the degree of accent experiment (Chapter 3) uses only *one sentence*. In that experiment therefore, differences between different sentences does not affect the results. In the intelligibility experiment on the other hand (Chapter 4), many different sentences are used. However, that experiment seeks to make statements about the non native speech of different L1 groups, and each L1 group will be represented with as many as 6 different sentences. The sentence differences between the L1 groups will therefore presumably be evened out in the wash across the 6 sentences for each L1 group. It should further be mentioned that the intelligibility investigation will be based on the comparison of various original and manipulated versions of the *same sentence*. This minimises the effect of differences between the individual sentences regarding for instance relative predictability and difficulty. A list of all the sentences can be found in Appendix E.

2.1.2 Representation of L1 groups

Before N1 speakers could be recruited, it was first decided which L1 groups that these speakers should represent. The criteria were as follows: a) The groups should be strongly represented in Norway so that the investigation was as relevant as possible for the particular situation for Norwegian as a second language, b) the selected L1 groups should represent linguistic diversity and be selected from several different language families, and finally, c) many L1s rather than just one or two should be represented. These criteria ensured that the speech corpus would represent a broad range of N2 speech.

Finding speakers was very time consuming. In particular, some L1 groups were very hard to recruit. For example, it was not possible to find more than one native speaker of Urdu, which may be partly because most Urdu speakers live in the Oslo area and not in the Trønder area where the recordings took place. Most of the speakers were ultimately recruited from the Norwegian as a second language courses at the Norwegian University of Science and Technology.

2.1.3 Recording of a speech corpus

The texts were sent by e-mail to the participants, and they were encouraged to read through the texts in advance and take notes of any unknown words or other difficulties. When they came to the studio they were asked if they had any problems with the texts. The speakers were also given the opportunity to read through the texts in the studio before the recording. The recordings were conducted in a soundproof recording studio using high-quality equipment. Each read sentence was subsequently excised and LP filtered with 75 Hz to remove low frequency noise. The speakers were paid for their participation.

A total of 41 N2 speakers and 6 N1 speakers were recorded. Such a relatively large corpus was recorded because the present work was part of a larger project involving several researchers and PhD students. The large corpus was recorded so that the other participants in the project, and subsequently any other interested researchers, would have the opportunity to use the corpus for their research.

2.1.4 Selection of speakers from the corpus

From the speech corpus described above, a selection was made for the present investigation. It was decided to use one Norwegian speaker which would serve as an N1 template as well as two N2 speakers from each of 7 L1s.

2.1.4.1 Selection of N1 template

One of the recorded N1 speakers was selected for the present investigation. This was a male speaker from the Southeast area. This N1 speaker was used as a native Norwegian template (as will be explained in section 2.2 on speech manipulation). It was decided to use an N1 speaker from the Southeast area for the following reasons. Although there is no officially recognized spoken norm in Norwegian, the Southeast dialect has traditionally represented the "unmarked" version of Norwegian pronunciation (Kristoffersen, 2000). It is likely that N2 learners aspire to attain this type of pronunciation both because their Norwegian language teachers tend to approximate this type of pronunciation in their teaching and interaction with students, and because many foreigners do not share Norwegians' exceptionally positive view of dialectal variation. Also, a study investigating which variety American immigrants choose as their phonetic goal shows that they tend to choose a standard variety rather than the dialect of the region in which they live (Fox & McGory, 2005). An additional reason to choose an N1 speaker with a Southeast dialect was that it is the author's dialect. In the process of digitally manipulating the recordings of the N2 speakers' utterances to make them were more similar to the N1 speaker's utterance (described in section 2.2), it was an advantage to be able to judge the degree of success of these adjustments.

2.1.4.2 Selection of N2 speakers

Vanishingly few investigations have investigated foreign-accented speech with *Norwegian* as the target language (the literature is dominated by investigations of foreign-accented *English*). Consequently, very little is known about foreign-accented Norwegian. The aim of the present investigation is therefore to provide a first *broad overview* of foreign-accented Norwegian as spoken by speakers from many different L1s, so that later investigations may have the possibility to build on the results from the present investigation when exploring a particular L1 group more in depth. The broad approach of the present investigation makes it possible to discover similarities that can be generalized across L1s, and also makes it possible to discover differences between L1 groups.

It was decided to use two speakers from each of 7 L1 groups. There were purely pragmatic reasons for selecting few speakers from each of the 7 L1s. There would simply not have been enough time, within the frames of this project, to both investigate *many* L1 groups and investigate *many* speakers from each L1.

It was decided to use recordings without disturbances that could annoy or distract listeners. For this reason, recordings where speakers read very hesitantly, re-read words or syllables within a sentence, paused very long, mumbled, breathed heavily into the microphone, made noise (e.g. scratched their face or touched the paper from which they read) and so on, were *not* selected. Some speakers were not included because their voice qualities were too breathy, creaky or otherwise deviant. These speakers' recordings were deemed unsuitable for the present investigation because their deviant voice qualities could interfere with analyses and manipulations. For instance, a breathy voice could be a problem for reliable F0 analysis.

Because only those speakers were selected who met the criteria defined above, two of the L1s (Persian and German) were represented by speakers from just one gender, whereas the remaining L1 groups were represented by one male and one female speaker each. Because it had been decided to use *two* speakers from each L1, speakers were selected from the L1 groups from which *at least two* speakers had been recorded. For instance, only one Urdu speaker had been recorded, and therefore Urdu was not possible to select as an L1 group. Apart from the criteria described in this paragraph, the selection of L1 groups and speakers was random.

The 14 N2 speakers had the 7 following L1s: English, French, German, Chinese (Mandarin), Russian, Persian and Tamil. In the large speech corpus, each speaker has been given a label. For instance, the five French speakers were labelled Fr1, Fr2, Fr3, Fr4 and Fr5. In the selection of speakers for the present investigation, the labels for each speaker have been preserved so that it will be possible for other researchers to relate the results from this investigation to the individual speakers in the speech corpus. This explains the otherwise peculiar labels used for the speakers throughout the present investigation. For instance, the labels Fr2 and Fr3 were used for the two French speakers instead of the more intuitive Fr1 and Fr2.

All speakers were adults between 21 and 61 years of age. Almost everyone had a high level of education (many were PhD students or researchers). The amount of N2 use varied greatly. Most of the speakers had been recruited from the Norwegian as a second language courses at the Norwegian University of Science and Technology. These courses have four levels. Most of the speakers recorded in the large speech corpus were on levels 2 and 3. The selection of speakers that was made for the investigation includes speakers from course levels 1, 2 and 3. Table 2.1 provides background information about each of the selected speakers.

Table 2.1: Background information about each of the speakers.

L1	Speaker	Sex	Age	N2 course level	Length of residence	Amount of N2 use 1 (rarely) – 4 (often)
English	En2	Female	26	On level 3	1.5 years	4
	En3	Male	61	Finished level 1	22 years	2
French	Fr2	Male	21	On level 2	3 months	2
	Fr3	Female	21	On level 2	1 year	3
Tamil	Ta1	Female	24	On level 2	8 months	4
	Ta2	Male	23	On level 1	7 months	1
Chinese (Mandarin)	Chi6	Female	26	On level 2	1 year, 2 months	2
	Chi7	Male	38	Finished level 2	5 years	1
Russian	Ru1	Male	26	Finished level 2	5.5 years	4
	Ru4	Female	26	Finished level 1	1 year	4
German	Ge2	Male	33	On level 2	5 years	4
	Ge3	Female	32	On level 3	8 months	3
Persian	Pe2	Female	23	Finished level 2	2.5 years	4
	Pe3	Female	30	Finished level 3	5 years	4

The table shows that three speakers were on or had finished level 1, eight were on/had finished level 2, and three were on/had finished level 3. There was variation both regarding length of residence in Norway and amount of N2 use. Learners who were on a higher course level, or who had lived for a long time in Norway, did not necessarily use Norwegian extensively. For instance, the oldest speaker, En3, who had lived in Norway for as long as 22 years, reported using Norwegian only to a moderate degree. Then again, this speaker had only completed a level 1 course. In contrast, Ru4, who had also completed level 1, but had lived in Norway for only one year, used Norwegian extensively. Later in this thesis (section 3.7, Chapter 3 and section 4.6, Chapter 4), the degree of similarity between the two individuals who represent the same L1 will be investigated.

Table 2.2 summarizes the number of speakers and their L1s.

Table 2.2: The speakers selected for the present investigation.

L1	Men	Women	Sum
Chinese	1	1	2
English	1	1	2
French	1	1	2
Persian	-	2	2
Russian	1	1	2
German	2	-	2
Tamil	1	1	2
Sum	*7*	*7*	*14*

The table shows the 7 L1 groups from which the speakers have been selected. There were two speakers from each of these L1 groups, yielding a total of 14 N2 speakers.

2.2 Manipulation methods

The previous sections described the speech material selected for the present investigation. As explained earlier, the aim of the investigation is to measure the impacts of durational and intonational aspects on perceptions of N2 speech. In order to perform this investigation, the recorded utterances have been subjected to adjustments in the form of digital speech manipulation. The manipulations involved both the adjustment of intonation and the adjustment of duration. The manipulated and original utterances functioned as speech stimuli in two perception experiments described in Chapters 3 (investigating degree of foreign accent) and 4 (investigating N2 intelligibility) respectively.

The selected speech material consisted of N2 utterances from 7 different L1 groups and N1 utterances from one native Norwegian speaker (see section 2.1.3 above). The N1 utterances and the N2 utterances were readings of the same sentences. The speech manipulations involved the adjustment of two phonetic aspects of speech, namely duration and intonation. Each N2 utterance was manipulated so that the durations of every phoneme equalled the duration of each phoneme in the N1 utterance of the same sentence. Also, each N2 utterance was manipulated so that the global utterance intonation equalled that of the N1 utterance of the same sentence. Stimuli were also generated in which both of these aspects were manipulated. In this way three manipulated utterances were created on the basis of each original N2 utterance; these were one duration manipulated, one intonation manipulated and one intonation-duration manipulated utterance. These speech stimuli were used to test listeners' perceptions in terms of the degree of foreign accent (Chapter 3) and intelligibility (Chapter 4). In the following, the methods of the speech manipulations are described and discussed. All manipulations were performed with the Praat program (Boersma & Weenink, 2004) and the manipulated files were resynthesized into wav files.

The following describes the manipulation methods in detail, first for duration manipulation and subsequently for intonation manipulation.

2.2.1 Duration manipulation

This section describes the method for manipulating the durations of the N2 utterances. For each sentence, the durations of the phonemes in the N2 utterances were manipulated so that they matched the durations of the phonemes in the N1 utterance. For this purpose, it was first necessary to segment and measure the duration of each phoneme in the N2 utterance and each phoneme in the corresponding N1 utterance.

2.2.1.1 Segmentation

Segmentation was guided by visual impressions from waveforms and spectrograms, coupled with the author's auditory impression. Segmentation was easier when the consonants were articulated with full closure (plosives, nasals, laterals and to a certain extent taps) or friction (fricatives) than when the articulation was approximantic, especially when the formant structure showed smooth transitions rather than abrupt changes. In these cases it was necessary to rely more heavily on auditory impressions. In order to determine the boundaries between vowels and plosives it was necessary to decide how to treat portions of aspiration. Post-aspiration (following the plosive and preceding the vowel) was treated as a separate segment, whereas pre-aspiration (following the vowel and preceding the plosive) was treated as part of the vowel. This approach was chosen because post-aspiration is a feature that occurs in regularly across dialects, whereas pre-aspiration occurs only in particular dialects and is often facultative[3] (Helgason, Stölten & Engstrand, 2003). Vowels at the very end of utterances were left unadjusted because it was impossible to decide where the vowel ended and the exhalation started.

2.2.1.2 Manipulation

As previously described, the segmentation of each phoneme in the N1 and N2 utterances of the same sentence provided each phoneme's duration. The following explains how the N2

[3] Traditionally, it has been believed that pre-aspiration occurs in only a few Norwegian dialects but research shows that pre-aspiration may be much more common than previously assumed. The view on pre-aspiration in Norwegian is currently changing also due to recent investigations into its linguistic function (van Dommelen, 1998; van Dommelen, 2000).

utterance's phoneme durations were adjusted to match the corresponding N1 utterance's phoneme durations. The duration of each N2 phoneme was divided by the duration of the corresponding N1 phoneme. This yielded a factor number with which each N2 phoneme was multiplied. The result was that each N2 phoneme ultimately matched the duration of each corresponding N1 phoneme. The following illustration shows an excised word from an N2 utterance in its original form and in the duration manipulated version.

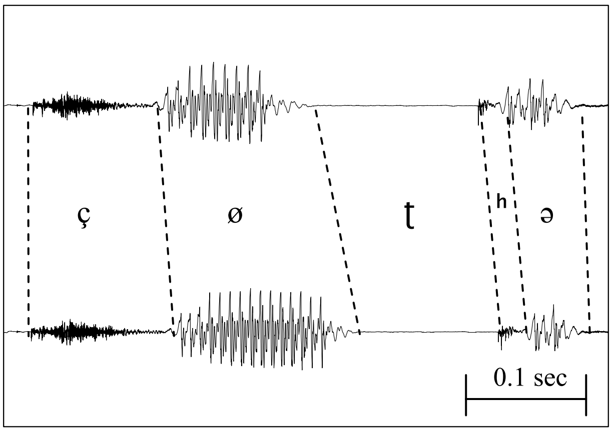

Figure 2.1: The word "kjørte" (= *drove*) as spoken by a Russian N2 speaker. Original N2 utterance above and duration manipulated N2 utterance below. The Southeast pronunciation of the sequence "kj" is pronounced as a palatal [ç] and the sequence "rt" is pronounced as a retroflex [ʈ].

The example shows that there are durational differences between the N2 original and the N2 duration manipulated utterances. The most prominent difference is that the ratio between the vowel [ø] and the following plosive [ʈ] has been altered. In the N2 original, the VC ratio is positive (i.e. C longer than V) whereas it is negative (i.e. V longer than C) in the manipulated version. The VC ratio is important in Norwegian because the language has phonological opposition between long and short vowels. This opposition is realized as a durational trade-off between the vowel and subsequent consonant. There are many Norwegian word pairs that differ only in the VC ratio. For instance, the (main) difference between the words "sette" (= *to put*) and "sete" (= *seat*) is that the former is pronounced with a VC: syllable (short vowel and long consonant) and the latter with a V:C syllable (long vowel and short consonant). Although the word "kjørte" is not among the words that change into a different word if the VC: syllable is instead pronounced as a V:C syllable, the pronunciation of the word becomes foreign accented nevertheless.

2.2.1.3 Problems

The previous section explained that the duration manipulation was performed by changing each N2 phoneme's duration so that it matched the corresponding N1 phoneme's duration. This procedure posed difficulties for several reasons.

In some cases the segment inventories were not identical across the N2 and N1 utterances. One reason for this was epenthesis (the insertion of sounds) in the N2 utterances. Epenthesis is typically a strategy that non-native speakers use when coping with a phonotactic pattern different from that found in their L1 (Husby & Kløve, 1998). Epentheses were left unaltered as a rule, but if the insertion made the duration manipulated utterance sound unnatural (which could happen if the surrounding segments were considerably shortened) the insertion was shortened just enough to restore the naturalness of the utterance. A second problem regarding discrepancies between the N2 and N1 utterances was that phonemes found in the N1 utterance were sometimes not realized in the N2 utterance. For instance, the word "bilen" (= *the car*) was sometimes pronounced without the final nasal. Such deletions did not affect the manipulation procedure. In the example with the word "bilen", the / e / would then simply be manipulated to match the duration of the corresponding N1 / e /. Another discrepancy between the N2 and N1 utterances was that the N2 utterances sometimes had pauses in them. Pauses were left unaltered except in a few cases where the duration manipulation made the pause sound unnatural in the modified surroundings. In these rare cases the pauses were shortened enough to remove this effect of unnaturalness.

The reason why epentheses, deletions and pauses were left unadjusted (as a rule) was that the focus of this investigation was on the durational pattern of the segments found in the utterances. Therefore, while the experimenter recognizes the potential interesting contributions of epentheses, deletions and pauses in perceptions of non-native speech, disfluencies of this kind lie outside the scope of this investigation.

In addition to problems arising from discrepancies between the N2 and N1 realizations of the same sentence, there were also some inherent problematic issues regarding the type of duration manipulation itself. Firstly, the manipulation affected not only the internal durational organization of the utterance, but also the duration of the entire utterances. This is because for a particular utterance, the utterance duration equals the sum of each phoneme's duration. For instance, if most of the phonemes in an utterance were shortened, the whole utterance was made shorter. This affects the impression of speaking rate. (The effects of altered speaking rate will be investigated). Secondly, the manipulation of duration inadvertently also affected the intonation. This is because when the duration of a certain portion of the signal was altered, the steepness of the intonation slope was also changed. The three illustrations below show how the slope of the intonation contour changes when a segment is lengthened and shortened.

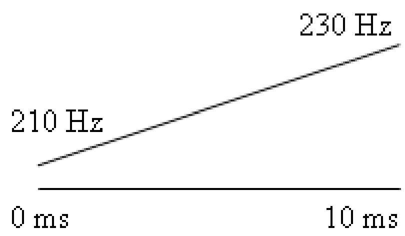

Fig. 2.2: Original segment duration.

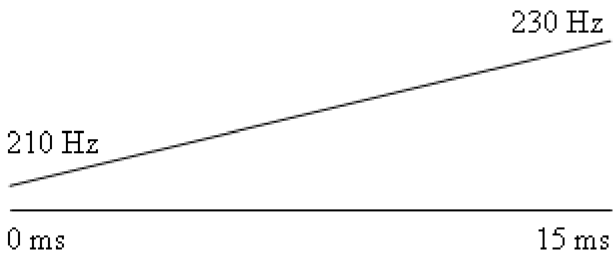

Fig. 2.3: Lengthened segment duration.

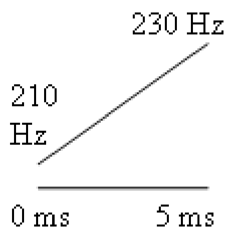

Fig. 2.4 Shortened segment duration.

Figure 2.2 shows a segment of 10 ms duration and an intonation contour that rises from 210 to 230 Hz. When the slope is calculated as the difference in Hertz divided by the difference in milliseconds, the slope is 20 Hz/ms. In Figure 2.3 the segment has been lengthened to 15 ms. The intonation still rises from 210 to 230 Hz, but the slope is now clearly less steep, only 1.33 Hz/ms. In Figure 2.4 the segment has instead been shortened. The intonation contour, which still rises from 210 to 230 Hz, now has a steeper slope of 4 Hz/ms.

In other words, if a portion of the signal was shortened, then the intonation slope of this portion automatically became steeper and vice versa, when a portion was lengthened the intonation slope became less steep. However, the duration manipulation affected the intonation slopes only to a very moderate degree and was regarded as having a negligible effect, because an effect could not be detected when the author, a trained phonetician, listened carefully to the stimuli.

2.2.2 Intonation manipulation

So far, only the method for duration manipulation has been described. As explained earlier, the aim of the investigation was not only to study the role of duration, but also to study the effects of intonation. The N2 utterances have therefore also been subjected to intonation manipulation. The intonation manipulation involved analyzing the N1 utterance's global intonation contour and applying it to the corresponding N2 utterance.

The same difference in Hertz is perceived (often measured in Just Noticeable Difference, abbreviated JND) as larger in the lower than in the higher F0 regions. In other words, the human ear is more sensitive to F0 changes in lower frequency ranges than in higher frequency ranges. Therefore, if the N1 intonation contour of a male speaker in Hertz were to be superimposed onto an N2 utterance produced by a speaker with a very different F0 range, for instance a female speaker, then the same difference would have different perceptual effects across the N1 speaker and the N2 speaker. The semitone scale normalizes this difference. The same difference in semitones is perceived as similar regardless of the F0 range. In order for the analyses and manipulations to be as perceptually accurate as possible, semitones were used instead of Hertz.

The following describes the method used for manipulating intonation.

2.2.2.1 Stylization

The first step in the manipulation process was the stylization of the N2 utterances' intonation contours. Stylization means that the intonation curve was represented by a limited number of coordinates, representing only the important turning points of the curve. The curves were stylized such that only turning points of at least 2 semitones were represented. Figure 2.5 gives an example of an intonation contour in the natural and the stylized versions.

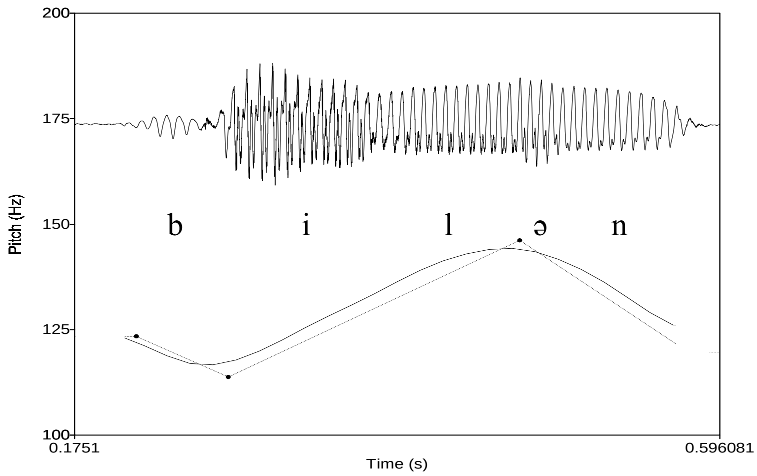

Figure 2.5: The word "bilen" (= *the car*) with natural intonation curve and stylized intonation curve. (In this figure the curve is analyzed in Hertz rather than in semitones due to technical constraints in the program).

The illustration in Figure 2.5 shows that for this particular token of the word "bilen" (*the car*), the stylized curve represents the signal with two turning points. Stylization left the utterances sounding fairly natural in the ears of the phonetically trained author. In his PhD thesis, Werner (2000) also remarked that stylization did not affect the impression of his speech material.

2.2.2.2 Manipulation

As explained in the previous section, stylization was applied only to the N1 utterances. The next step in the manipulation process was to replace each N2 utterance's intonation curve with the stylized N1 curve of the same sentence. This step was carried out by copying the N1 contour and superimposing it onto the N2 utterance. However, because there are durational differences between the N1 and N2 utterances, the N1 intonation curve did not automatically fit the corresponding N2 utterance. The maxima and minima of the superimposed curve occurred at the wrong places relative to the segmental inventory of the utterance. Therefore, the curves subsequently had to be manually adjusted in the time domain. These manual adjustments were greatly facilitated, if not made possible altogether, by the fact that the superimposed N1 curve was stylized. Each turning point was now simply "dragged" horizontally in the time domain so that the F0 excursions of the curve coincided with the same segmental phenomena across the N2 and N1 utterances.

When the N1 contour was superimposed on an N2 utterance, the contour also had to be shifted up or down to fit the particular speaker's voice. For instance, because the N1 template was a rather low pitched male voice, when it was applied to a female voice it had to be shifted upwards to suit the particular speaker's F0 range.

After the N2 intonation curves were replaced with the stylized N1 curves, the manipulated utterances were resynthesized. Resynthization is a procedure in the Praat program which smoothes the curve so that the turning points became less abrupt. This makes the curve more natural. Figure 2.6 shows the intonation curve for an N2 word in the original and the intonation manipulated versions.

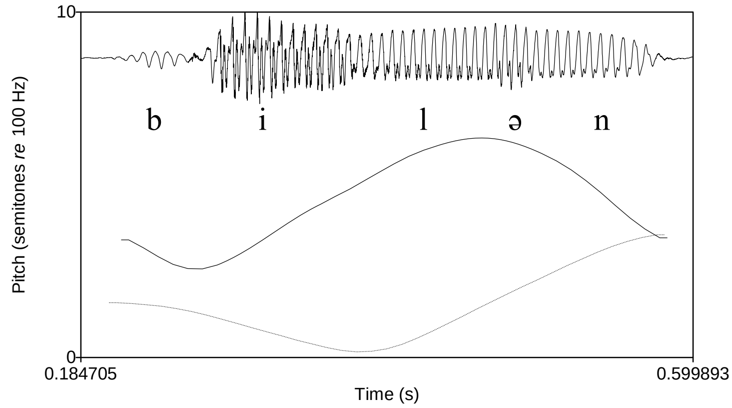

Figure 2.6: Intonation curves in an N2 speaker's utterance of the word "bilen" (=*the car*). The rising-falling contour is the N2 original utterance. The falling-rising contour is the N2 intonation manipulated utterance.

Figure 2.6 shows the different intonation contours of the N2 original utterance (rising-falling) and the intonation manipulated N2 utterance where the N2 contour has been replaced by the N1 template contour (falling-rising). In this example, the N2 original curve rises and then falls again just before the schwa in the second syllable, whereas the N2 intonation manipulated curve has an initial fall and then a rise which coincides with the onset of the lateral.

2.2.3 Intonation-duration manipulation

The previous sections describe how the N2 utterances have been manipulated. The utterances were both duration manipulated and intonation manipulated, but in separate steps, which generated one duration manipulated utterance and one intonation manipulated utterances. In addition, stimuli were generated in which both duration and intonation were manipulated. These intonation-duration manipulated utterances were generated by superimposing the stylized N1 intonation curve onto the corresponding N2 duration manipulated utterance.

Because the duration manipulated utterance had phoneme durations equal to the N1 utterance, the superimposed N1 curve fit the N2 utterance fairly well, but not always perfectly because of the discrepancies in terms of pauses, epentheses and deletions described earlier. As in the intonation manipulation process described above, the curve had to be somewhat manually adjusted in the time domain in order to appropriately align the events across the F0 and time domains.

In section 2.1, the speech corpus selected for the present manipulations was described. There were 14 speakers from 7 different L1 groups. These speakers' N2 utterances were manipulated as explained in section 2.2. The two following two chapters describe experiments in which native Norwegian listeners were presented with the original and manipulated N2 utterances. In Chapter 3, the manipulations' impacts on listeners' perceptions in terms of the perceived degree of foreign accent will be investigated, and the same manipulations' role for N2 intelligibility will be investigated in Chapter 4.

3. Degree of foreign accent in N2 speech

The previous chapter has described the recording of the speech corpus and how the material for the present investigation was selected from that corpus. The same chapter also described the manipulation of N2 utterances generating the speech stimuli. In this chapter the experiment investigating degree of foreign accent is described and discussed. As explained in the introductory chapter, the aim of this experiment was to investigate the relative impacts of durational and intonational aspects upon the degree of foreign accent in N2 speech. However, before describing this experiment it is necessary to first discuss some methodological issues.

3.1 Method of measurement

It is common to determine the degree of foreign accent by means of listener ratings. Jesney (2004) gives an overview of foreign accent rating studies and has identified three different methods that have been used to obtain such ratings: Likert scales, sliding scales and Direct Magnitude Estimation. Likert scales feature from three to ten gradients (e.g. Munro & Derwing, 1994; Anderson-Hsieh & Koehler, 1988; Magen, 1998). The listener determines the degree of foreign accent in the stimulus by placing it at a point on the graded scale. Sliding scales have no gradients, only endpoints (Major, 1986b; Flege & Fletcher, 1992). The listener will judge the degree of foreign accent by placing it on the scale, and for this point a number is subsequently calculated. Because sliding scales have no gradients they provide finer distinctions than Likert scales. The third method identified by Jesney is Direct Magnitude Estimation, where raters assign a score to the first stimulus they hear and then assign scores to the following stimuli to show if they perceived them as more or less accented in relation to the first stimulus. Jesney points out that the method of Direct Magnitude Estimation focuses on the relationship between scores rather than on the raw scores.

The three methods discussed above are the most commonly used in determining the degree of foreign accent. However, when using these methods, speaker factors such as age, gender, L1 or voice quality could influence listeners' judgments. Such speaker factors could thus obscure measurement of the foreign accent. For the present experiment, a method was therefore used that eliminates these speaker factors. This method is similar to Direct Magnitude Estimation in that it does not assign scores to individual stimuli, but instead gives a score to the distance between stimuli, and resembles Likert scale ratings in that the listener chooses among a closed set of gradients to express his assessment. With this method, couples of stimuli from the same speaker are compared between themselves. The two stimuli are always based on one particular recording that has been manipulated to produce different versions. For instance, the two stimuli compared could be the original and the duration manipulated stimuli for one particular speaker's utterance, or the two stimuli could be another particular speaker's utterance in the duration manipulated and intonation manipulated versions. The advantage of this method is that speaker factors are eliminated because manipulated versions of the same utterance from the same speaker are compared internally. A disadvantage of this method is however that the perceptual data are purely relational, showing *differences* in degree of accent between stimuli, and therefore do not give information about the degree of foreign accent of the particular stimuli. It could be useful to know the degree of foreign accent of each stimulus because this would indicate whether the two speakers from the same L1 are indeed similar. In this experiment, the analyses of the manipulation effects will be carried out with data pooled across the two speakers because it is likely that speakers from the same L1 will have the greatest benefit from the same manipulation. At the end of the chapter (section 3.7), the similarity between the listeners, in particular the two listeners from the same L1, will be investigated.

3.2 Pilot experiment

Before launching the experiment, a small pilot experiment was carried out in order to test whether the listeners should be presented with whole sentences or just sentence fragments. In his study, Bannert (1995) used sentence fragments, some of which were as short as one word. In the pilot experiment, six listeners were presented with both a whole sentence as well as an excised two-word sentence fragment in the form of a subject and the following verb. The listeners' reactions were that a sentence fragment was too short to evaluate the degree of foreign accent. For this reason it was decided to use whole sentences for the present experiment. Furthermore, it is most common in investigations with foreign accent ratings to use whole sentences (Jesney, 2004). The pilot experiment was small and unstructured and will therefore not be discussed in more detail.

3.3 Stimuli

As explained in Chapter 2, there were 14 speakers, 2 from each of 7 L1 groups. In the present experiment one and the same read sentence was used from each of these speakers. The sentence was "Bilen kjørte forbi huset vårt" (= *The car drove past our house*). Apart from the need to find a sentence that all the speakers had read without any disturbances or noise, the choice of this particular sentence was arbitrary. For each of the 14 utterances there was an original, a duration manipulated, an intonation manipulated and an intonation-duration manipulated version. All utterance versions, including the original utterance, are referred to as stimuli.

In the present experiment, stimuli from the same speaker were paired in files with a two second pause in between. The pairs are called stimulus pairs. A stimulus pair thus consisted of two single stimuli. In their investigation, Munro & Derwing (1994) found that whichever utterance the listeners heard *second* was judged as more foreign-accented. Because of the possibility that the order of the stimuli in the stimulus pair might affect listeners' judgements, each stimulus was positioned *first* in one stimulus pair and *second* in another stimulus pair, as shown in Table 3.1.

Table 3.1: The ordering of stimuli in both first and second position in the stimulus pairs.

Stimuli in the stimulus pair	Order A	Order B
O / D	O_D	D_O
O / I	O_I	I_O
O / ID	O_ID	ID_O
D / ID	D_ID	ID_D
I / ID	I_ID	ID_I
D / I	D_I	I_D

O= original, D= duration manipulated, I= intonation manipulated, ID= intonation-duration manipulated.

In other words, for every combination of two stimuli, there were two stimulus pairs differing only in the order of the stimuli. Because every stimulus was positioned both first and second the number of stimulus pairs was doubled. It is important to note that within a stimulus pair the two stimuli were always from the same speaker. Table 3.1 shows all 12 stimulus pairs that were based on one speaker's original utterance. Table 3.2 shows all stimulus pairs used in the experiment. The left-hand column shows the 7 L1s and the right-hand column shows the 2 speakers from each L1. For each speaker there were 12 stimulus pairs, listed horizontally in the table.

Table 3.2: All 12 stimulus pairs from each of the 14 speakers.

L1	Speaker	Original/ Duration		Original/ Intonation		Original/ Int-dur		Duration / Int-dur		Intonation/ Int-dur		Duration/ Intonation	
French	Fr2	O_D	D_O	O_I	I_O	O_ID	ID_O	D_ID	ID_D	I_ID	ID_I	D_I	I_D
	Fr3	O_D	D_O	O_I	I_O	O_ID	ID_O	D_ID	ID_D	I_ID	ID_I	D_I	I_D
English	En2	O_D	D_O	O_I	I_O	O_ID	ID_O	D_ID	ID_D	I_ID	ID_I	D_I	I_D
	En3	O_D	D_O	O_I	I_O	O_ID	ID_O	D_ID	ID_D	I_ID	ID_I	D_I	I_D
German	Ge2	O_D	D_O	O_I	I_O	O_ID	ID_O	D_ID	ID_D	I_ID	ID_I	D_I	I_D
	Ge3	O_D	D_O	O_I	I_O	O_ID	ID_O	D_ID	ID_D	I_ID	ID_I	D_I	I_D
Russian	Ru1	O_D	D_O	O_I	I_O	O_ID	ID_O	D_ID	ID_D	I_ID	ID_I	D_I	I_D
	Ru4	O_D	D_O	O_I	I_O	O_ID	ID_O	D_ID	ID_D	I_ID	ID_I	D_I	I_D
Tamil	Ta1	O_D	D_O	O_I	I_O	O_ID	ID_O	D_ID	ID_D	I_ID	ID_I	D_I	I_D
	Ta2	O_D	D_O	O_I	I_O	O_ID	ID_O	D_ID	ID_D	I_ID	ID_I	D_I	I_D
Chinese	Chi6	O_D	D_O	O_I	I_O	O_ID	ID_O	D_ID	ID_D	I_ID	ID_I	D_I	I_D
	Chi7	O_D	D_O	O_I	I_O	O_ID	ID_O	D_ID	ID_D	I_ID	ID_I	D_I	I_D
Persian	Pe2	O_D	D_O	O_I	I_O	O_ID	ID_O	D_ID	ID_D	I_ID	ID_I	D_I	I_D
	Pe3	O_D	D_O	O_I	I_O	O_ID	ID_O	D_ID	ID_D	I_ID	ID_I	D_I	I_D

O= original, D= duration manipulated, I= intonation manipulated, ID= intonation-duration manipulated.

3.4 Perception experiment

The preceding section has described the stimuli and their organization into stimulus pairs. The following explains how stimulus pairs were presented to listeners in the perception experiments. A total of 14 listeners of both sexes and between 20 and 35 years of age participated in the perception experiment. The listeners were from all parts of Norway. Most of the listeners were university students. None reported hearing loss and none reported experience with foreign accented speech at a level out of the ordinary. The latter point is important because investigations have shown that experience with accented speech affects listeners' perceptions of it (Gass & Varonis, 1984). The listeners were paid for their participation.

As explained earlier, 12 stimulus pairs were generated on the basis of one original utterance. In the perception experiments, each stimulus pair was repeated 5 times for the sake of statistical reliability. As there were 12 stimulus pairs from each speaker, 2 speakers from each of the 7 L1s and also 5 repetitions of each stimulus pair for statistical purposes, this yielded a total of 840 stimulus pairs presented to each listener during the perception experiment. Because of the large test size, the experiment was split into four sessions for each listener. For each listener the four sessions were conducted over two days. On each day the listener sat for two listening sessions separated by a 30 minute break. Each listener was seated in a sound treated room in front of a computer screen. The stimulus pairs were presented through loudspeakers. Because some of the stimuli sounded somewhat unnatural due to the manipulations, the listeners were told that some of the utterances might sound strange due to technicalities in the sound recording process. They were told to try and ignore this technical artefact and focus uniquely on their task. They were not told that the utterances had been manipulated. The listeners' task was to judge which stimulus featured less of a foreign accent than the other in each of the stimulus pairs they were presented with. The computer screen in front of the listener was organized into five horizontal slots, as illustrated below.

1 FAR LESS ACCENTED THAN 2
1 LESS ACCENTED THAN 2
EQUAL DEGREE OF ACCENT
2 LESS ACCENTED THAN 1
2 FAR LESS ACCENTED THAN 1

All listeners seemed to find this test design comprehensible. When the listener clicked the screen to give his judgment, the next stimulus pair was presented automatically. In order for the experimenter to process the resulting data, the listeners' responses were later converted into positive and negative numbers ranging from –2 to 2, as illustrated below.

1 FAR LESS ACCENTED THAN 2 = **2**
1 LESS ACCENTED THAN 2 = **1**
EQUAL DEGREE OF ACCENT = **0**
2 LESS ACCENTED THAN 1 = **-1**
2 FAR LESS ACCENTED THAN 1 = **-2**

Informal inspection of each of the 14 listeners' responses later revealed that 13 listeners were very consistent in their responses, whereas one speaker displayed strikingly random responses. This listener's responses were therefore excluded on the suspicion that an unreported hearing loss or some other unknown factor had influenced her results. Thus the results from the perception experiments were based on 13 listeners' responses.

3.5 Results

The perception experiment tests the impact of the factor manipulation. However, two other factors could have influenced listeners' perceptions; these are the stimulus order in the stimulus pairs and the difference between the dialect of the N1 template speaker and the dialect of the individual listeners. The possible influences of these factors are briefly investigated in the following two sections.

3.5.1 Stimulus order

In Munro & Derwing 1995a, listeners judged the degree of foreign accent in extemporaneous versus read sentences. They found that stimuli that were presented second (whether extemporaneous or read) were perceived as having a stronger foreign accent. Inspection of the results from the present perception experiment suggested that there could in fact be a correlation between the stimulus order in the stimulus pair and the size of the effect on the perceived degree of accent. In all cases where a manipulated stimulus was paired with an original stimulus, the manipulated stimulus seemed to be judged as having even less of a foreign accent when it was positioned first. This is illustrated in Figure 3.1.

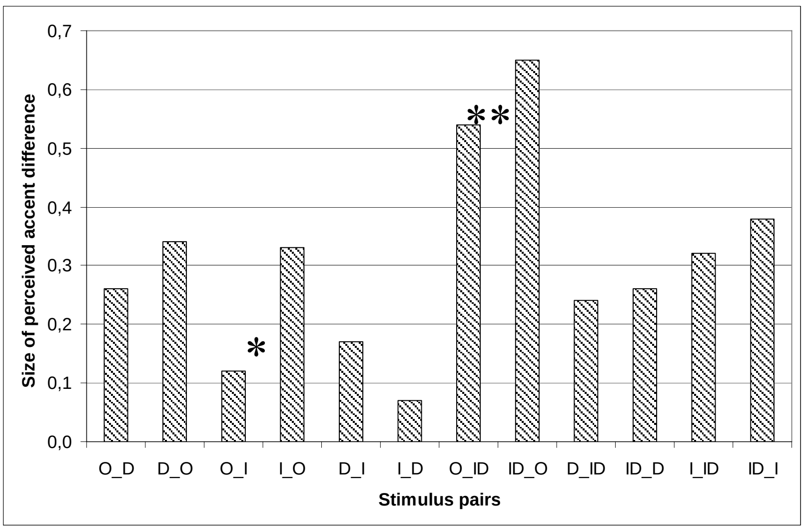

Figure 3.1: Size of the perceived accent difference between stimuli in each of the 12 stimulus pairs (n = 910 for each stimulus pair).

The vertical bars in Figure 3.1 show the size of the accent difference between the stimuli in each stimulus pair. *Size* is defined as the degree to which the manipulation affects the accent.

Remember from section 3.4 that the listeners rated the accent difference between the two stimuli in each stimulus pair. If the perceived accent difference is greatly affected by the manipulation, the size of the manipulation effect is large. If the perceived accent difference is moderately affected, then the manipulation effect is small.

Section 3.4 presented an explanation of how listeners' ratings were converted to both negative and positive numbers, yet all the numbers in Figure 3.1 are positive. The reason for this is explained here. As explained in 3.4, negative numbers showed that the second stimulus was less accented and positive numbers showed that the first stimulus was less accented. Because the manipulated stimuli were in fact always perceived as having less of a foreign accent (effects of manipulations will be investigated in later sections), all negative numbers have been converted to positive, such that only the size of the effect is shown in Figure 3.1.

For example, Figure 3.1 shows that the D stimulus is perceived as less accented in both the O_D (left) and the D_O (to its right) pair. The interesting point here is that D in the D_O pair is perceived as having even less of a foreign accent than D in the O_D pair. Are the stimuli in the first position consistently perceived as less accented than the stimuli in the second position? A Mann-Whitney test comparing the difference between the stimuli in one pair with the difference between the stimuli in another pair with one common stimulus (Table 1, Appendix A), such as O_D versus D_O, shows that only two comparisons (O_I / I_O* and O_ID / ID_O**, see asterisks in figure) are significantly different whereas the others are not. This is interpreted to mean that, in general, stimulus order does not affect the perceived degree of accent.

3.5.2 Listener factors

Research suggests that listeners are very similar in how they rate degree of foreign accent (Cunningham-Andersson & Engstrand, 1989; Thompson, 1991; Munro & Derwing, 1999; Piske, Flege & MacKay, 2001; Moyer, 1999; Abelin & Boyd, 2000). However, Almberg & Husby (2002) investigated foreign-accented Norwegian, and found that the foreign-accent ratings varied between those listeners who had a low tone dialect and those that had a high tone dialect (see Chapter 1, section 1.4.4 for an explanation of the Norwegian accents). In contrast, Bannert (1995) investigated foreign-accented *Swedish*, a language which is very closely related to Norwegian and has a similar accent system (Gårding, 1998), and he did *not* find that the listeners' own accent realizations had affected their perceptions.

The N1 speaker whose utterances provided the template for the manipulations was from the southeast area of Norway (see Chapter 2, section 2.1.4.1). However, the listeners were from all parts of Norway and therefore represented different dialects. The question here is whether the listeners' background as either low tone (abbreviated LH) speakers or high tone (abbreviated HL) speakers has affected their perceptions regarding the intonation manipulation.

Figure 3.2 shows the perceived accent difference between the original and intonation manipulated stimuli as perceived by HL listeners and LH listeners respectively.

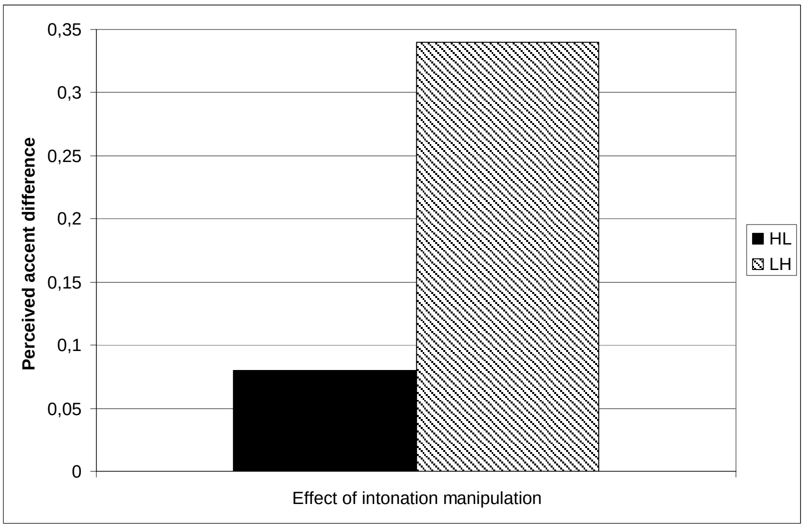

Figure 3.2: Perceived accent difference as measured in the O_I and I_O stimulus pairs (n= 1820) for HL (n= 5) and LH (n= 8) listeners.

The figure shows that the intonation manipulation reduced the foreign accent less for HL listeners (black bar) than for LH listeners (patterned bar). This could be because the manipulations were based on a LH dialect template. In order to investigate whether these differences were due to intonational differences between the listeners' dialects, or to other unidentified listener differences, Figure 3.3 compares the effect of duration manipulation between the HL listeners and the LH listeners.

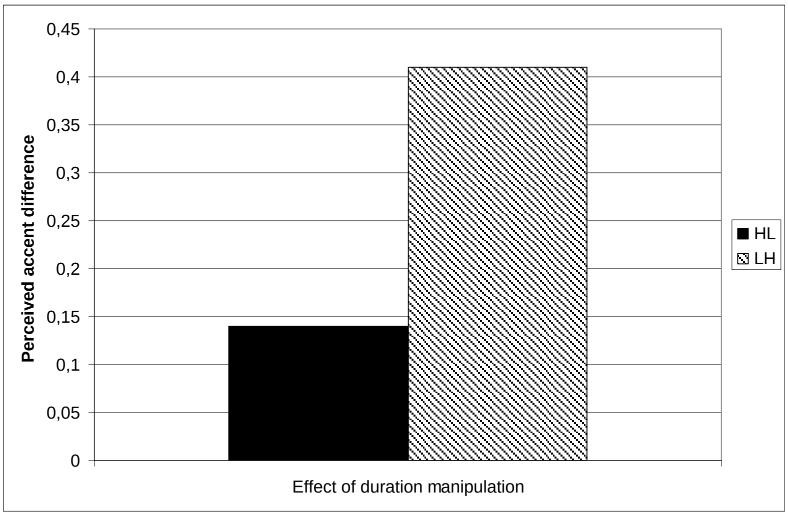

Figure 3.3: Perceived accent difference as measured in the O_D and D_O stimulus pairs (n= 1820) for HL (n= 5) and LH (n= 8) listeners.

The accent reduction effect of the manipulation was smaller for the HL group than for the LH group, not only for the comparison of original versus intonation manipulated stimuli (Figure 3.2), but also for the comparison of original versus duration manipulated stimuli (Figure 3.3). This suggests that the HL dialect listeners do not judge differently because they belong to a different accent dialect than the N1 template speaker. Different listeners perceive somewhat differently, for instance due to differing amount of N2 experience (e.g. Gass & Varonis, 1984). The difference between the HL and the LH listeners' judgements must therefore be due to factors other than the listeners' accent realizations.

3.5.3 Grouping of data

Because neither listener dialect nor stimulus order affected listeners' judgements, the results are presented across listeners and across stimulus orders using the stimulus pair abbreviations for stimulus order A as listed in Table 3.1 in section 3.3. In the data, stimulus order has been eliminated by pooling the results for the stimulus pairs that differ only in stimulus order. For instance, the results for O_D and D_O have been pooled. Also, negative numbers have been converted to positive. If, for a grouping of data, the effect for O_D was -0.26 (negative number= second stimulus is less accented) and the effect for D_O was 0.34 (positive number= first stimulus is less accented), the pooled effect was 0.30, and the positive number means that the duration manipulated stimulus had less of a foreign accent than the original.

3.5.4 Statistical tests

Remember from section 3.4 that the listeners responded by choosing between 5 values on a scale. Because the distance between the values on this scale cannot be said to be equal, the resulting data are ordinal scaled. In order to investigate the effects of the manipulations (described from section 3.5.6 onwards), statistical tests appropriate for ordinal scaled data have been used. Two statistical tests have been used, the Sign test and the Mann-Whitney test. The Sign test investigates the difference between two stimuli in a stimulus pair, for instance between O and D in the stimulus pair O_D. The Mann-Whitney test investigates the difference between the stimuli within one pair as compared to the difference between the stimuli within another pair across stimulus pairs with one common stimulus. For instance, the Mann-Whitney test allows the investigation of the accent reduction within the pair O_D as compared with the accent reduction in the pair O_I. The Sign test and the Mann-Whitney test therefore explore the effect of the manipulations from different angles. The Sign test was used to investigate all 6 stimulus pairs. Mann-Whitney tests were used as a supplementary test for selected stimulus pairs, in order to further investigate the relative impacts of the manipulations in support (or in refute) of the results from the Sign tests. The results from all statistical tests referred to in this chapter can be found in Appendix A.

3.5.5 Figures

The effects of the manipulations were investigated for each L1 in the subsequent sections. For each L1 the effects are presented in figures. The numbers on the x-axes in the figures show the listeners' ratings. If the term "degree of accent" were to have been used to specify the quantity in the x-axes, *increasing* x-values would indicate *decreasing* degree of accent. By using the term "native-like" in the figures, *increasing* x-values instead correspond to *increasing* native-likeness. The term native-like was therefore used as an equivalent to, and interchangeably with, the term degree of foreign accent. The figures were organized according to the results for the L1 in question in the following manner: For the tags on the y-axis, the stimulus that was rated as having less of a foreign accent (whether the difference was significant or not) was placed first. For instance, if, for a particular L1 group, the comparison D_I showed that I was perceived as having less of a foreign accent than D, the tag in the figure would read "I_D". Note that this system only applies to the stimulus pair tags in the figures. In all discussions the stimulus pairs still refer to the abbreviations for stimulus order A as shown in Table 3.1, section 3.3. For instance, if, for an L1, I was perceived as having less of a foreign accent than D, the tag in the figure would read "I_D". In the discussions, however, this stimulus pair will still be referred to as “D_I”. In this way the terminology for the discussion of the stimulus pairs remains constant across the sections for the different L1s. Information on the statistical significance of effects has been illustrated with asterisks inside each horizontal bar. The three levels of significance were as follows: * = p< 0.05, ** = p< 0.01, *** = p< 0.001.

3.5.6 English

The results are first investigated for the English L1 group. Figure 3.4 shows the results for this L1 group.

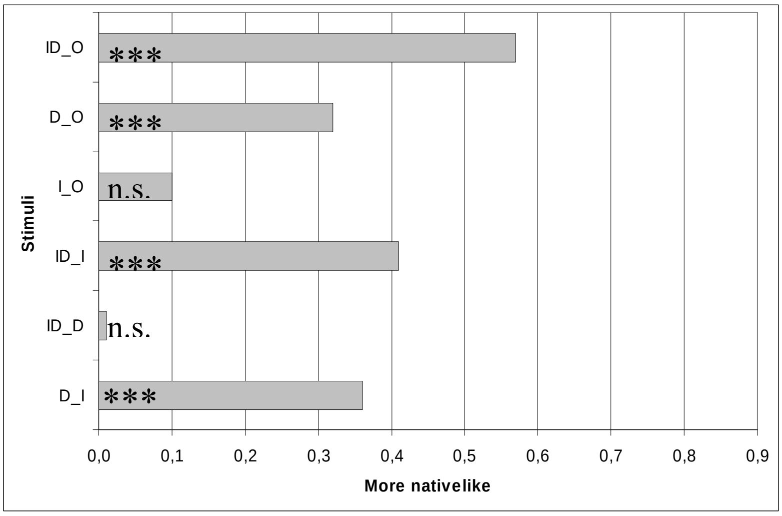

Figure 3.4: Results for the English L1 group (n= 1560). Stimuli on y-axis and effect on x-axis. For each comparison (i.e. each horizontal bar) the more native-like stimulus is written first. I = intonation manipulation. D = duration manipulation. ID = intonation-duration manipulation. O = original.

For the English L1 group, the comparison of the original stimulus to the intonation-duration manipulated stimulus showed a large difference. The rated difference was 0.57. A Sign test (Table 2, Appendix A) showed that the effect was highly significant (p< 0.001). In other words, the ID stimulus was perceived as having less of a foreign accent than the O stimulus.

The duration manipulated stimulus was perceived as having significantly less of a foreign accent than the original stimulus (p< 0.001). The rated difference between these stimuli was 0.32, which was smaller than the effect of the ID manipulation as shown in the previous section. However, when comparing the original stimulus with the intonation manipulated stimulus (rated difference 0.1) there was no significant difference in the perceived degree of foreign accent. This means that when compared with the original stimulus, only the duration manipulation, and not the intonation manipulation, affected the degree of foreign accent.

The intonation-duration manipulated stimulus was perceived as having less of a foreign accent when compared to the intonation manipulated stimulus ($p < 0.001$, rated difference 0.41), but was not perceived as significantly less accented than the duration manipulated stimulus (very small rated difference of 0.01). The difference between the stimuli in the pair I_ID was that duration manipulation has been added in the latter stimulus. This added manipulation reduced the degree of foreign accent, which suggests that duration manipulation was more important than intonation in English N2 foreign accent reduction. This was in keeping with the finding in the previous paragraph, where a significant accent reduction was found in the O_D comparison but not in the O_I comparison.

The comparison of the duration manipulated stimulus with the intonation manipulated stimulus showed a rated difference of 0.36, which was statistically significant ($p < 0.001$). The difference was such that the D stimulus was perceived as having less of a foreign accent than the I stimulus. This comparison thus lends further support to the interpretation that duration seemed to be the most important factor in contributing to the degree of accent for these two English N2 speakers, as shown in all the previous comparisons in this section.

3.5.7 French

Figure 3.5 shows the results for the French L2 group.

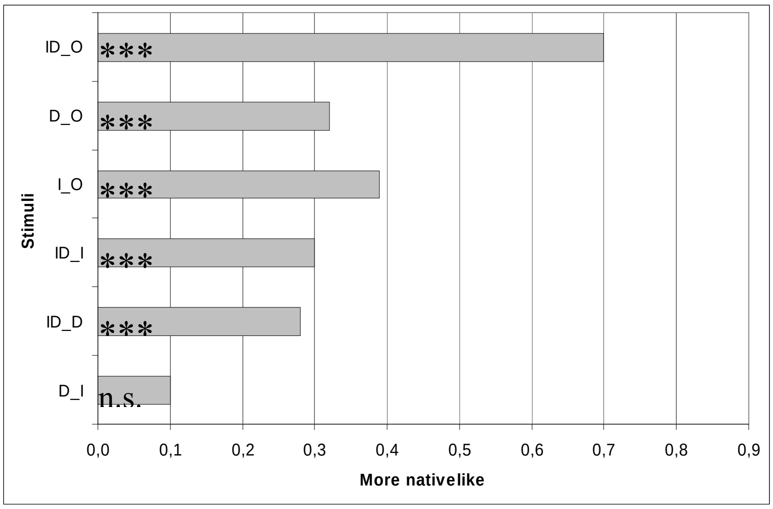

Figure 3.5: Results for the French L1 group (n= 1560). Stimuli on y-axis and effect on x-axis. For each comparison (i.e. each horizontal bar) the more native-like stimulus is written first. I = intonation manipulation. D = duration manipulation. ID = intonation-duration manipulation. O = original.

The French speakers' intonation-duration manipulated stimulus was perceived as having much less of a foreign accent than the original stimulus (rated difference 0.7). A Sign test (Table 2, Appendix A) showed that this difference was statistically highly significant ($p< 0.001$).

The comparison of the original stimulus with the duration manipulated stimulus showed that the manipulation reduced the degree of foreign accent significantly ($p< 0.001$). The intonation manipulated stimulus was also perceived as having less of a foreign accent as compared with the original ($p< 0.001$). As can be seen in Figure 3.5, the difference was greater between the original and the intonation manipulated stimuli (rated difference 0.39) than between the original and the duration manipulated stimuli (rated difference 0.32). For the English L1 group (previous section), only the duration manipulation affected the foreign accent when compared to the original stimulus. For the French L1 group however, both manipulations affected the accent when compared to the original.

The intonation-duration manipulated stimulus was perceived as having less of a foreign accent both relative to the intonation manipulated stimulus and relative to the duration manipulated stimulus. Both differences were highly significant ($p< 0.001$). As can be seen from Figure 3.5, the rated difference was somewhat larger between the ID and I stimuli (0.3) than between the ID and D stimuli (0.28), but a Mann-Whitney test comparing the accent difference within the stimulus pair I_ID with the accent difference within the stimulus pair D_ID showed that the difference between these pairs was not significant (Table 4, Appendix A).

The results from the comparisons of the 6 stimulus pairs described in the previous paragraphs do not give any clear indication as to which manipulation most efficiently reduces the degree of foreign accent in French N2 speech. A Mann-Whitney test was carried out for further investigation. This test compared the accent difference in the stimulus pair O_D with the accent difference in the stimulus pair O_I and showed that the difference was significantly greater for the latter pair ($p < 0.05$).

The results from the comparisons for the French L1 group are difficult to interpret. The results from the (Sign) tests on the accent reduction within each stimulus pair indicated that the two manipulations reduced the foreign accent to the same degree. However, one (Mann-Whitney) test indicated the superior role of intonation. The results for this L1 group are interpreted to mean that intonation affected the degree of foreign accent for these two French N2 speakers more than duration.

3.5.8 Tamil

Figure 3.6 shows the results for the Tamil L1 group.

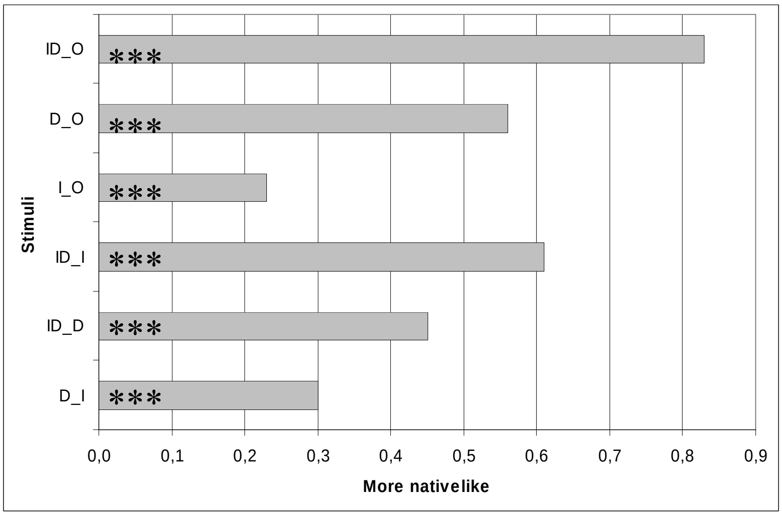

Figure 3.6: Results for the Tamil L1 group (n= 1560). Stimuli on y-axis and effect on x-axis. For each comparison (i.e. each horizontal bar) the more native-like stimulus is written first. I = intonation manipulation. D = duration manipulation. ID = intonation-duration manipulation. O = original.

The Tamil speakers' intonation-duration manipulated stimuli were perceived as having much less of a foreign accent than their original stimuli (rated difference 0.83). A Sign test (Table 2, Appendix A) showed that this difference was significant ($p < 0.001$).

The duration manipulated stimulus was perceived as having less of a foreign accent than the original stimulus ($p < 0.001$). The effect was smaller than for the O_ID comparison (previous paragraph), but still fairly large (rated difference 0.56). The intonation manipulated stimulus was also perceived as having less of a foreign accent than the original ($p < 0.001$), but the effect was smaller (rated difference 0.23).

The intonation-duration manipulation reduced the degree of foreign accent in the Tamil speakers' stimuli significantly ($p < 0.001$) and to a great extent (rated difference 0.61) when compared with the intonation manipulated stimulus. This double manipulation was also judged as having significantly less of a foreign accent ($p < 0.001$) when compared with the duration manipulated stimulus, although not to the same extent (rated difference 0.45). A Mann-Whitney test (Table 4, Appendix A) showed that there was no significant difference in accent reduction between the pairs I_ID and D_ID.

In the stimulus pair comparing the duration manipulated and the intonation manipulated stimuli, the former were perceived as having less of a foreign accent than the latter ($p < 0.001$). The rated difference between these stimuli was 0.3.

A Mann-Whitney test (Table 4, Appendix A) comparing the accent difference in the pair O_D with the accent difference in the pair O_I, lends further support to the interpretation that duration contributes more to the degree of foreign accent than intonation ($p < 0.001$).

All the comparisons of the Tamil group's stimuli thus indicated that duration was more important than intonation in the reduction of foreign accent in their N2 speech.

3.5.9 Chinese

Figure 3.7 shows the results for the Chinese L1 group.

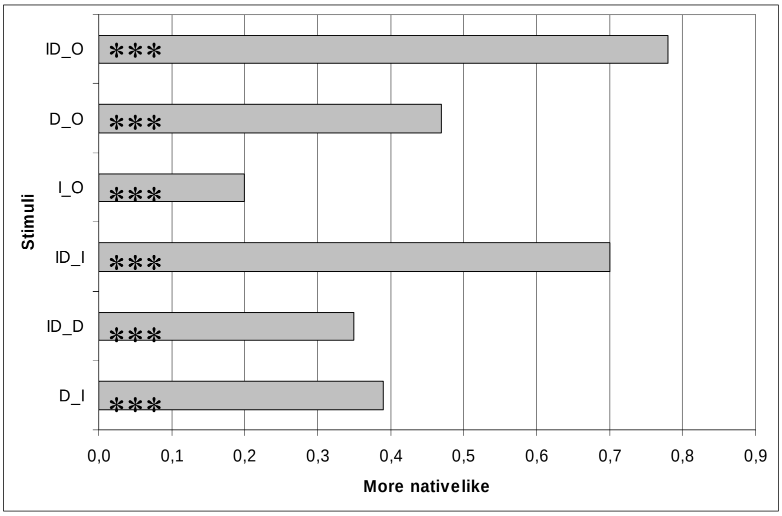

Figure 3.7: Results for the Chinese L1 group (n= 1560). Stimuli on y-axis and effect on x-axis. For each comparison (i.e. each horizontal bar) the more native-like stimulus is written first. I = intonation manipulation. D = duration manipulation. ID = intonation-duration manipulation. O = original.

The Chinese speakers' intonation-duration manipulated stimulus was rated as having much less of a foreign accent than their original stimulus (rated difference 0.78). This difference was significant ($p < 0.001$, Table 2, Appendix A).

Comparing the original stimulus with the duration manipulated stimulus showed that the latter was perceived as having less of a foreign accent ($p < 0.001$, rated difference 0.47). The intonation manipulated stimulus was also judged as less accented than the original ($p < 0.001$), but the difference was rated as smaller (0.2). A Mann-Whitney test (Table 4, Appendix A) showed that the accent reduction in the pair O_D was greater than the accent reduction in the pair O_I ($p < 0.001$).

The intonation-duration manipulated stimulus was perceived as having less of a foreign accent both relative to the intonation manipulated stimulus ($p < 0.001$) and to the duration manipulated stimulus ($p < 0.001$). However, the difference between the stimuli in the I_ID pair was rated as greater (0.7) than the difference between the stimuli in the D_ID pair (0.35). With the help of a Mann-Whitney test (Table 4, Appendix A), the accent reduction within the stimulus pair I_ID was found to be significantly greater than the accent reduction within the D_ID pair ($p < 0.001$). Because the difference between the stimuli in the I_ID pair was that of added duration manipulation, this result indicates that duration may be more effective than intonation. The results from this and the previous paragraphs therefore indicate that duration may be more important than intonation for this L1 group.

The comparison between the duration manipulated stimulus and the intonation manipulated stimulus showed that the former stimulus was perceived as having less of a foreign accent than the latter (p< 0.001). This difference was rated 0.39.

All the stimulus pair comparisons for the Chinese L1 group consistently indicated that duration was more important than intonation to allowing speech to be perceived as having less of a foreign accent.

3.5.10 Russian

Figure 3.8 shows the results for the Russian L1 group.

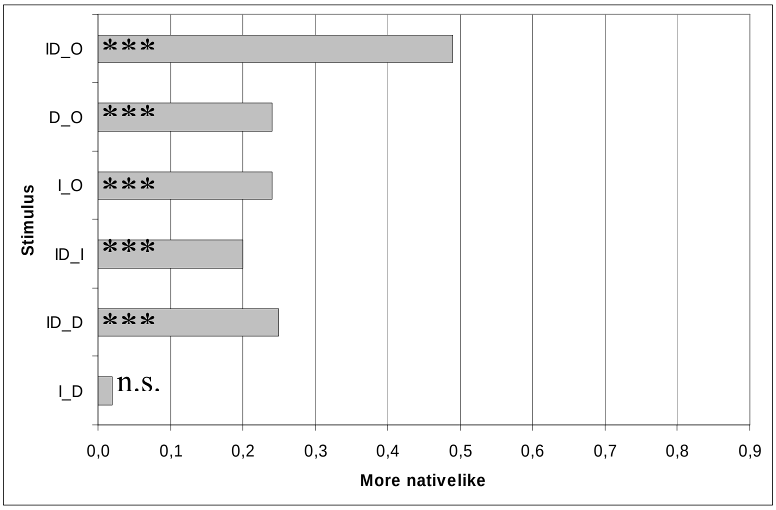

Figure 3.8: Results for the Russian L1 group (n= 1560). Stimuli on y-axis and effect on x-axis. For each comparison (i.e. each horizontal bar) the more native-like stimulus is written first. I = intonation manipulation. D = duration manipulation. ID = intonation-duration manipulation. O = original.

The Russian N2 speakers' intonation-duration manipulated stimulus received a much lower foreign accent score than the original stimulus (rated difference 0.49). This difference was a significant effect (p< 0.001, Table 2, Appendix A).

The duration manipulated stimulus was perceived as having less of a foreign accent than the original stimulus (p< 0.001). The intonation manipulated stimulus was also perceived as less accented than the original (p< 0.001). Both manipulations reduced the foreign accent to the same extent (rated difference within each stimulus pair 0.24).

The intonation-duration manipulation reduced the perceived foreign accent compared with both the intonation manipulated stimulus (p< 0.001) as well as with the duration manipulated stimulus (p< 0.001). However, the rated difference was larger for the D_ID comparison (rated difference 0.25) than for the I_ID comparison (rated difference 0.20), indicating that intonation manipulation may reduce the foreign accent to a greater extent than the duration manipulation.

The comparison of the intonation manipulated stimulus with the duration manipulation stimulus showed a very small difference (approximately 0.02) such that I was rated as having less of a foreign accent compared to D. This small difference was not significant, however. The D_I comparison thus gave no indication as to which manipulation reduced the degree of foreign accent most in Russian N2 speech.

A Mann-Whitney test (Table 4, Appendix A) was carried out in order to further explore the relation between the accent reducing effects of duration manipulation as compared to that of intonation manipulation. The accent difference in the stimulus pair O_D was compared with the accent difference in the stimulus pair O_I. The result showed that the difference between the effects of the two manipulations was not significant.

For the Russian L1 group, as judged from the various stimulus comparisons discussed above, the conclusion must be that durational and intonational aspects influenced the degree of foreign accent to the same degree.

3.5.11 German

The results for the German L1 group are shown in Figure 3.9.

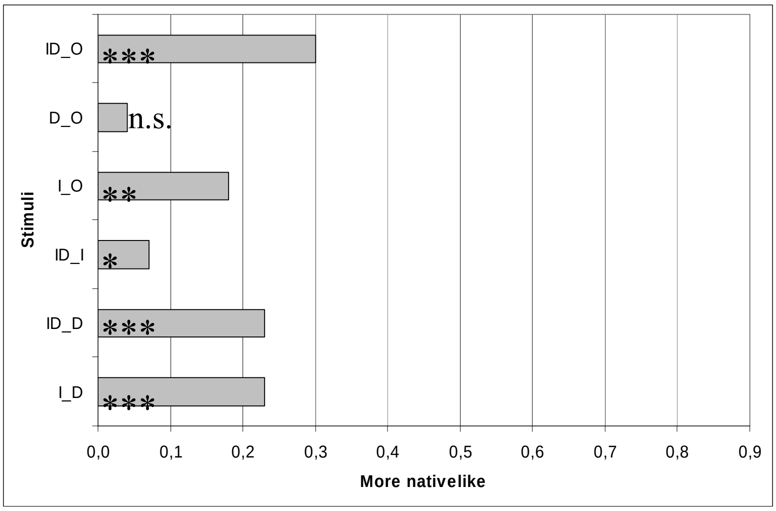

Figure 3.9: Results for the German L1 group (n= 1560). Stimuli on y-axis and effect on x-axis. For each comparison (i.e. each horizontal bar) the more native-like stimulus is written first. I = intonation manipulation. D = duration manipulation. ID = intonation-duration manipulation. O = original.

The German N2 speakers' intonation-duration manipulated stimulus was rated as having less of a foreign accent than the original stimulus. The rated difference between these stimuli was 0.3. Moreover, this difference was significant ($p < 0.001$, Table 2, Appendix A).

Contrary to the findings for all the previously investigated L1 groups, the Germans' duration manipulated stimulus was perceived as equally foreign accented as the original stimulus (the rated difference is only 0.04). However, the intonation manipulated stimulus reduced the amount of foreign accent significantly as compared with the original stimulus ($p < 0.01$). The difference between the O and I stimuli was 0.18. Intonation could thus be more important than duration for this L1 group.

The intonation-duration manipulated stimulus was perceived as having less of a foreign accent than both the duration manipulated stimulus ($p < 0.001$) and the intonation manipulated stimulus ($p < 0.05$). The difference between the D and ID stimuli was, however, larger (rated difference 0.23) than between the I and ID stimuli (rated difference 0.07). A Mann-Whitney test (Table 4, Appendix A) showed that the accent reduction in the D_ID pair was significantly larger than the accent reduction in the I_ID pair ($p < 0.001$). Because the difference between the stimuli in the D_ID pair consisted of added intonation manipulation in the latter stimulus, this result lends further support to the interpretation that intonation was more important than duration for accent reduction in German N2.

The intonation manipulated stimulus was perceived as having less of a foreign accent than the duration manipulated stimulus. The difference was rated as 0.23 and was significant (p< 0.001). This result thus points in the direction of intonation as more effective in accent reduction than duration, in accordance with the results from the previous paragraphs.

The results for the German L1 group indicated that intonational aspects were more important than durational aspects for foreign accent reduction. In fact, duration had no significant effect on the Germans' N2 accent. A Mann-Whitney test (Table 4, Appendix A) comparing the difference in the stimulus pair O_D with the difference in the stimulus pair O_I further supports this interpretation (p< 0.01).

3.5.12 Persian

Figure 3.10 shows the results for the Persian L1 group.

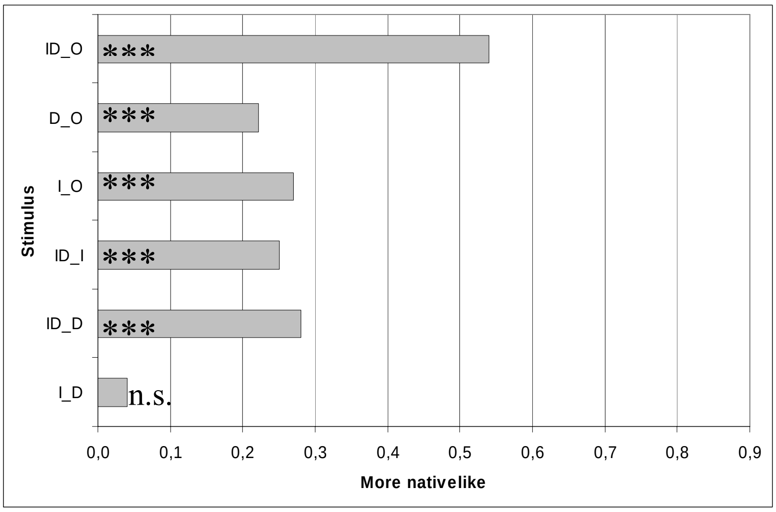

Figure 3.10: Results for the Persian L1 group (n= 1560). Stimuli on y-axis and effect on x-axis. For each comparison (i.e. each horizontal bar) the more native-like stimulus is written first. I = intonation manipulation. D = duration manipulation. ID = intonation-duration manipulation. O = original.

The Persian N2 speakers' intonation-duration manipulated stimulus was perceived as having significantly less of a foreign accent than their original stimulus as shown by a Sign test (p< 0.001, Table 2, Appendix A). The effect was fairly strong (rated difference 0.54).

The duration manipulated stimulus was judged as having less of a foreign accent than the original stimulus (p< 0.001). The intonation manipulated stimulus was also perceived as having less of a foreign accent than the original stimulus (p< 0.001). However, the difference between the stimuli in the O_I pair was larger (rated difference 0.27) than the difference between the stimuli in the O_D pair (rated difference 0.22). This indicates that intonation manipulation may reduce the foreign accent more than duration manipulation.

Intonation-duration manipulation rendered the N2 speech as having significantly less of a foreign accent than either intonation manipulation (p< 0.001) or duration manipulation (p< 0.001). The ratings between the I_ID and the D_ID stimulus pairs were fairly equal, but the difference was somewhat larger in the D_ID pair (rated difference 0.28) than in the I_ID pair (rated difference 0.25). However, this difference between the two pairs was not significant as shown by a Mann-Whitney test (Table 4, Appendix A).

A comparison of the intonation manipulated stimulus with the duration manipulated stimulus showed no significant difference in degree of foreign accent between the two (very small rated difference of 0.04).

Both duration manipulation and intonation manipulation thus reduced the foreign accent in the Persians' N2 speech. It remains unclear which manipulation was the more important. A Mann-Whitney test (Table 4, Appendix A) was conducted to compare the accent reducing effect between the stimuli in the O_D pair with that of the stimuli in the O_I pair. The result shows that the difference between the stimulus pairs was not statistically significant, however.

The results for the Persian L1 group seem to indicate that there was no difference in the accent reducing effects of duration manipulation and intonation manipulation. For the Persian speakers, their N2 speech was equally accent reduced by the two manipulations under investigation.

3.5.13 Summary

The results from the experiment on the degree of foreign accent have been described for each L1 group separately in sections 3.5.6 through 3.5.12 above. The results showed that the combined manipulation of duration and intonation (the ID manipulation) significantly reduced the degree of foreign accent. This was true for all the seven L1s investigated. The degree of accent reduction differed between the L1s, meaning that some L1 groups benefited more from this manipulation than other L1 groups. Figure 3.11 shows the amount of accent reduction caused by the ID manipulation as measured in the O_ID stimulus pair for each L1.

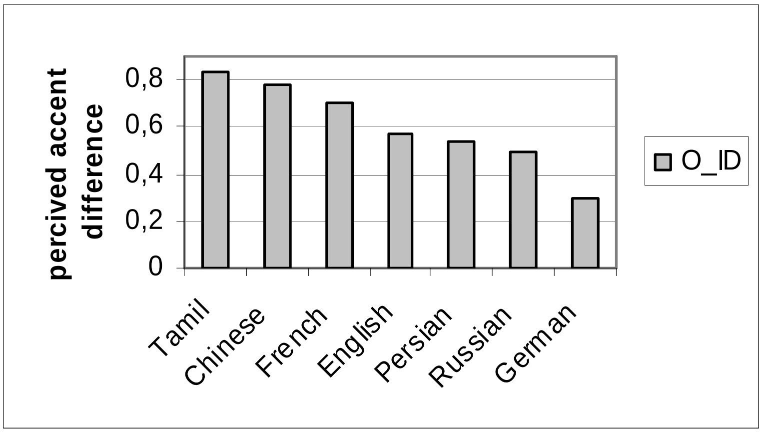

Figure 3.11: The accent reducing effect of the intonation-duration manipulation as measured in the O_ID stimulus pair for each L1.

Figure 3.11 shows that the ID manipulation reduced the degree of foreign accent the most for the Tamil group (rated difference 0.83) and the least for German group (rated difference 0.30). The fact that the German N2 was least affected by the ID manipulation can be explained in light of the finding that all L1 groups except German benefited from duration manipulation (see section 3.5.11).

Moreover, the separate manipulations of duration and intonation each contributed to the reduction of foreign accent. This was the case for all L1 groups except for English, where the degree of foreign accent remained unaltered despite intonation manipulation, and except for German, where duration manipulation had no accent reducing effect. When compared to the original stimulus, the ID stimulus always caused a larger accent reduction than either the D or the I stimulus alone. Both the D and the I manipulations respectively reduced the foreign accent to varying degrees across the different L1 groups. Figure 3.12 shows the amount of foreign accent reduction caused by the duration manipulation as measured in the O_D stimulus pair for each L1. (German is not included in the figure because duration did not significantly affect German N2).

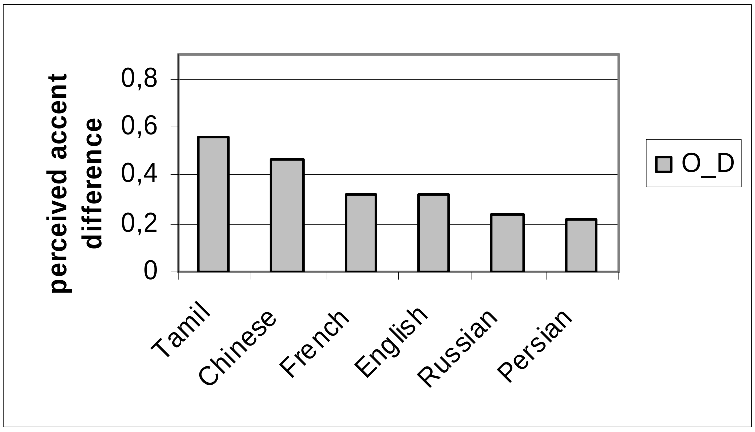

Figure 3.12: The accent reducing effect of the duration manipulation as measured in the O_D stimulus pair for each L1.

Figure 3.12 shows that Tamil (rated difference 0.56) was the L1 group that benefited most from duration manipulation, whereas the Persian L1 group gained the least from this manipulation (rated difference 0.22).

Figure 3.13 shows the accent reduction caused by intonation manipulation as measured in the O_I stimulus pair for each L1. (Because there was no significant effect of intonation manipulation for the English L1 group, this group was not included).

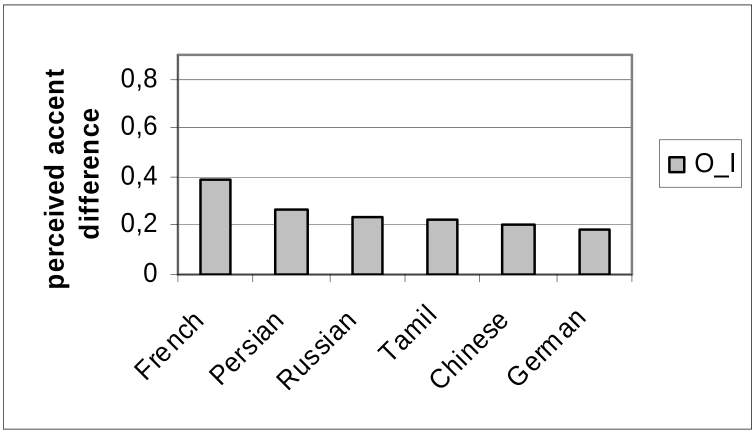

Figure 3.13: The accent reducing effect of the intonation manipulation as measured in the O_I stimulus pair for each L1.

In Figure 3.13 we see that the French N2 was affected the most by intonation manipulation, and that the German N2 benefited the least from this manipulation.

The aim of the investigation was to establish the relative impacts of durational and intonational aspects on the degree of foreign accent. These results are listed for each L1 group in Table 3.3.

Table 3.3: The most important manipulation for foreign accent reduction purposes for each L1 group.

L1	**Manipulation**
English	Duration
Tamil	Duration
Chinese	Duration
French	Intonation
German	Intonation
Russian	Equally important
Persian	Equally important

Table 3.3 shows that for the English, Tamil and Chinese L1 groups, duration was the most important remedy for reducing a perceived accent. For the French and German L1 groups on the other hand, intonation manipulation was the most important aspect. For the Russian and Persian L1 groups, however, no one manipulation can be singled out as more effective in reducing the perception of a foreign accent.

3.6 Production analyses

The previous sections have shown that durational and intonational aspects affected the degree of accent in the N2 speech of several L1 groups. It has also been shown that some L1 groups benefited more from durational adjustments whereas others benefited more from intonational adjustments. However, the manipulations were global in the sense that they were carried out over whole sentences. As for the duration manipulation, the manipulation consisted of adjusting all phoneme durations across utterances. The perceptual effect in terms of foreign accent reduction could be due to adjustments within particular types of segments. For instance, perhaps the perceptual effects of the duration manipulations were due mainly to vowel durations. As for the intonation manipulation, the manipulation changed the global utterance intonation contour. The perceptual effect of the intonation manipulation gave no information as to which parts of the utterance were most sensitive to intonational changes. For instance, were the changes in perceived foreign accent chiefly due to intonational adjustments in the stressed syllables? The remainder of this chapter is devoted to determining which manipulation details caused the perceptual effects in terms of foreign accent reduction described in sections 3.5.6 through 3.5.12 above. Information on statistical tests can be found in Appendix C.

3.6.1 Duration

This section attempts to relate the effect of the duration manipulation for each utterance to specific details in the manipulation of the utterances. The effect of the duration manipulation is called the *manipulation effect* and is here defined as the rated accent difference between the stimuli in the O_D stimulus pair (mean effect across all listeners). The specific details of the duration manipulation that were investigated are simply referred to as *factors*. These factors were in the form of segment type (vowels and consonants), V/C ratios, phonologically long vowels and articulation rate, and will be further explained in section 3.6.1.1.

The extent to which a factor (e.g. vowel durations) was adjusted as a result of the duration manipulation is called the *manipulation size*. Manipulation size was measured as the percent adjustment made to the particular factor. Adjustments to the articulation rate (also called *manipulation size*) were measured as the difference in number of phonemes per second between the O and D stimuli. Manipulation size was investigated for a correlation with the manipulation effect for each utterance. For instance, consider an utterance with a large duration manipulation effect (i.e. perceived as having a considerably reduced accent in the D stimulus as compared to the O stimulus). The manipulation size for the articulation rate is large (i.e. a large rate difference between the D and O stimulus). Because it is reasonable to assume that large effects should be due to large changes in the signal, it is plausible that the duration manipulation effect was affected by the large adjustment in articulation rate for this utterance.

For each utterance, and for each factor, the manipulation size and manipulation effect were investigated for correlations. In section 3.6.1.1 the various factors are defined before the results from the correlation analyses are presented in section 3.6.1.2.

3.6.1.1 Factors

The first factor that is defined is the *overall durational adjustment* of the utterances. This factor was measured as the mean percentage durational adjustment across all the segments in an utterance.

Research on the effect of L2 segment production accuracy (native listeners judged how accurate the segments were produced) on speech intelligibility (Bent, Bradlow & Smith, 2007) has found that vowel errors were more important than consonant errors. Perhaps the duration of vowels is similarly highly important for L2 degree of accent. Durational adjustment of the *vowels* and *consonants* were therefore defined as two factors.

In Chapter 1, section 1.4.3, it was explained that Norwegian has a two-way vowel quantity distinction such that there is phonological opposition between long and short vowels. Moreover, all stimuli in the present experiment were based on the same read sentence (see section 3.3), which contained three stressed, phonologically long vowels. The *duration of the three phonologically long vowels* was defined as a factor in the present investigation.

Speech rhythm was also defined as a factor. In section 1.4, Chapter 1, it was explained that there are as yet no universally standardized method of measuring speech rhythm, but that phonetic approaches typically measure ratios or intervals between successive units in the time domain. It was therefore decided to investigate speech rhythm through the measurement of V/C ratios. If this measurement reveals significant effects, then more elaborate methods of speech rhythm measurement can be applied. In the present investigation, V/C ratios were measured for the three stressed (and phonologically long) vowels and their following consonants.

The duration manipulation of the N2 utterances (Chapter 2, section 2.2.1) affected not only the internal durational organization of the utterances, but also the total utterance durations. This is because the sum of each phoneme's duration equals the total utterance duration. *Articulation rate*[4] was measured as the number of phonemes in an utterance divided by the utterance's total duration[5].

The factors are summarized below:

- All segments

[4] Rate will be referred to as articulation rate as opposed to speaking rate because the manipulations have not involved the remediation of pauses or similar disfluencies.
[5] Note that *articulation rate* is a different phenomenon from *duration across all segments*, the factor defined in the first paragraph of the present section. The former measures adjustments in terms of number of phonemes per time unit whereas the latter measures the extent to which the phonemes have been adjusted regarding duration.

- All consonants
- All vowels
- All phonologically long vowels
- V/C ratio
- Articulation rate

To sum up, for each utterance and for each factor the manipulation size was correlated with the manipulation effect. The analyses were performed in order to reveal which details (here called factors) in the duration manipulations that had caused the perceptual effects described in sections 3.5.6 through 3.5.12. The correlations were performed as regression analyses. Note that these regression analyses were multiple (with more than one predictor variable) only when the categories in the factors did not overlap. Only two factors did not overlap, i.e. the factor *vowels* and the factor *consonants*. For this reason, vowels and consonants were investigated in one multiple regression analysis whereas the remaining factors were investigated in separate regression analyses.

3.6.1.2 Results

The analyses was carried out in the form of multiple regression analyses to test whether there were correlations between the manipulation size for each of the defined factors (the extent to which the factors have been adjusted) and the manipulation effect (the accent reducing effect of the duration manipulation). Figure 3.14 shows the manipulation size across all segments (vertical bars) related to the manipulation effect (graph) for each utterance. The data were sorted in ascending order according to manipulation size. A trend line was drawn for the manipulation effect.

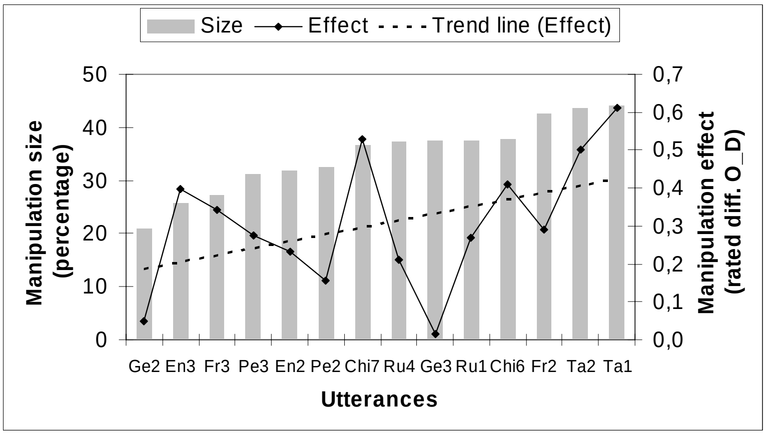

Figure 3.14: Manipulation size and manipulation effect across all segments for each utterance (n= 14). Data in ascending order according to manipulation size. A trend line was drawn for the manipulation effect.

In the display of the individual utterances' manipulation size related to the resulting manipulation effect, no clear relationship can be discerned. For instance, the German speaker Ge3's utterance was duration manipulated to a large extent, but the accent reducing effect of the manipulation was very small. Conversely, the English speaker En3's utterance was moderately duration manipulated, but the effect on the degree of accent was large. However, the trend line for the manipulation effect indicates a possible correlation. A regression analysis with duration across all segments as the predictor variable (Table 1, Appendix C) shows that there was no correlation with manipulation effect in terms of accent reduction. This could be interpreted to mean that the degree of foreign accent did not diminish linearly as a function of overall adjustment to the segment durations.

A multiple regression analysis was carried out with the consonants and the vowels as the two predictor variables in order to investigate the impacts of durational adjustments within each of these two segment groups. A significant correlation was found between manipulation size for the consonants and the manipulation effect in the form of accent reduction (Beta= 0.655; $p < 0.05$), but no effect was found for vowels. In other words, only the consonant durations and not the vowel durations affected the degree of foreign accent. Moreover, the correlation was such that when an N2 utterance's consonant durations were extensively adjusted, the effect on the degree of foreign accent was larger than when the utterance's consonant durations were less adjusted. The reason why consonant durations affected the degree of foreign accent significantly whereas vowel duration did not may be because consonants in general are less compressible than vowels. Listeners may therefore be more sensitive to deviations in consonant durations than in vowel durations.

Three multiple regression analyses were performed with phonologically long vowels, V/C ratio and articulation rate factors as predictor variables (Table 1, Appendix C), but none of these factors were found to correlate with the manipulation effect.

We turn now to the impact of adjustments to the articulation rate. As explained above, the duration manipulation automatically changes the utterance duration. If, for instance, most of the phoneme durations in an utterance are shortened, then the utterance duration is automatically shortened. The same number of phonemes is then uttered in a shorter period of time, which in turn means that the articulation rate becomes faster. Figure 3.15 shows the relation between changes in articulation rate (called "manipulation" in the figure although the adjustment of the articulation rate was merely an automatic side effect from the duration manipulation) shown in the vertical bars and the manipulation effect shown in the graph. A trend line was drawn for the manipulation effect.

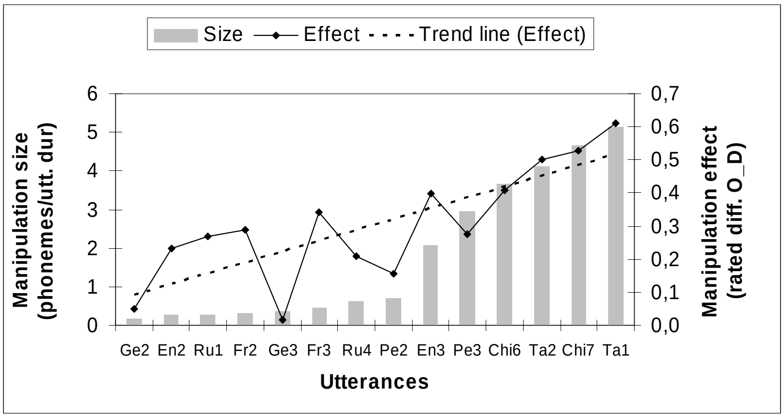

Figure 3.15: Manipulation size of articulation rate and manipulation effect for each utterance (n= 14). Data in ascending order according to manipulation size. A trend line was drawn for the manipulation effect.

The visual impression of Figure 3.15 strongly indicates a relation between the articulation rate and the manipulation effect in terms of accent reduction. The articulation rate did in fact significantly correlate with the manipulation effect (Beta= 0.842; $p < 0.001$). Almost all of the original N2 utterances were produced at a slower rate than the corresponding N1 utterances. The effect was therefore such that when the N2 articulation rate accelerated, the degree of foreign accent diminished.

3.6.2 Intonation

In section 3.6.1 above, various durational factors were correlated with the effect of the duration manipulation for each utterance. It was found that consonant durations and articulation rate were the durational factors that affected the degree of foreign accent in the N2 utterances. In this section, similar analyses will be carried out to identify the intonational changes responsible for the perceptual effects observed for the intonation manipulated N2 utterances.

Intonation analyses can often be carried out using phonological models. The Trondheim Model (TM, e.g. Nilsen 1992) is an intonation model specifically developed to describe Norwegian intonation. Therefore, if a phonological model should be used for these analyses, the TM would be the natural choice. However, the TM was found to be inadequate for the present analyses, as will be explained in the following. The TM describes how intonation and syntax interact in Norwegian speakers' encoding and Norwegian listeners' interpretation of *the meaning* of utterances. The model does not predict how the intonation affects *intelligibility* (if many or few words are identified by a listener) or the perceived *degree of foreign accent* of an utterance. The object of study in this investigation is not how intonation contributes to the meaning of utterances. There are additional reasons for why this model is unsuited for the analyses, and these reasons are related to the lack of detail in the model as explained in the following. The TM presupposes that the speaker is a native Norwegian, and that he therefore has the native speaker competence to modulate his intonation in specific ways according to what he wants to convey. In contrast, non-native speakers do not have this native speaker competence. That is why their speech can be identified as non-native. For example, the model presupposes the correct realizations of the word accents, as they can only be described as either accent 1 or accent 2 in the model. A non-native intonation may well fall within the categories of native Norwegian when described in the TM. For example, an accent phrase in a non-native utterance may be transcribed as accent 2, but the non-native pronunciation may still deviate from the native pronunciation in a more subtle way. This is not only a postulation based on the experimenter's own perceptions of the present N1 and N2 utterances, but relies on findings from previous studies: Mennen (2004) studied Dutch-accented Greek. Dutch and Greek have the same phonological structure in pre-nuclear rises, but the phonetic realizations are different. In the Dutch speakers' L2 Greek, their rises were phonetically deviant from the Greek L1 rises. Moreover, Atterer & Ladd (2004) studied

German-accented English, and found that the Germans carried over German patterns of F0 alignment into the English L2. The intonation analyses in this section therefore need to be fine-grained in the intonational dimension and in the durational dimension in order to be able to capture the *perceptually relevant* production deviations responsible for the range of accent reductions (this chapter) and for the range of increased intelligibility scores (next chapter) across the different N2 utterances. For the reasons explained in the above paragraphs, a detailed phonetic approach is deemed more appropriate than a coarse phonological approach.

Remember that the aim of these analyses is not to model N2 intonation, but to find and analyse those particular intonational differences that the listeners most likely have paid perceptual attention to. F0 range endpoints outline the shape of intonation curves, but this does not automatically entail that the F0 range endpoints themselves represent the perceptually relevant aspects of the curve. Previous research on tonal perception in Asian tone languages has shown that listeners from such languages pay perceptual attention primarily to *F0 slope*. For example, tonal perception research by Gandour & Harshman (1978) and Gandour (1983) showed that listeners from tone languages paid more perceptual attention to *F0 slope* and *F0 direction* than listeners from non-tone languages, and that listeners from non-tone languages instead relied on the *average F0* and the *endpoints of the F0 range*. Guion & Pederson (2007) also investigated tonal perception and found that listeners from tone languages relied on *F0 slope* and *average F0*, and that listeners from non-tone languages relied most heavily on *F0 mean* and secondarily on *F0 range endpoints*. Guion & Pederson (2007) thus supports Gandour & Harshman (1978) and Gandour (1983) in showing that listeners from a tone language rely on F0 slope in tonal perception. These investigations are however not in accordance regarding whether tone language listeners *in addition* use average F0 or F0 direction. For the present purposes, it was chosen to include F0 direction in the analyses because two of the three investigations referred to above showed that F0 direction was the additional perceptual aspect used by tone language listeners. Although Gandour & Harshman (1978), Gandour (1983) and Guion & Pederson (2007) studied *Asian tones* which are admittedly not identical to *Norwegian accents* (see Chapter 1, section 1.4.4 for an explanation of the Norwegian accents), the phenomena are related, and these results may therefore be relevant for the perception of Norwegian by Norwegian listeners. It is hypothesized that the Norwegian listeners in this investigation may have paid perceptual attention to the changes regarding F0 slopes and F0 directions in the intonation manipulated stimuli. The intonational analyses in this section therefore measure N1-N2 deviations in terms of F0 slope and F0 direction differences.

Three terms, namely *manipulation effect*, *manipulation size* and *factors,* were central in the discussions in section 3.6.1 above. The same terms will be used in these intonation analyses. The definition of these terms is repeated here. The manipulation effect refers to the accent reducing effect of the I stimulus compared to the O stimulus as observed in the O_I stimulus pairs (measured across all listeners). The details of the intonational adjustments that were investigated for correlation with the manipulation effect are called factors. Manipulation size was thus defined as the extent to which a factor was adjusted as a result of the intonation manipulation.

3.6.2.1 Factors

For each utterance, the same three 2-syllable content words were selected for the analyses of intonation manipulation adjustments: "**Bilen kjørte** forbi **huset** vårt". It was chosen to focus on discrete words because the researcher wanted to establish a method for analyses that could be used also for the intelligibility analyses in the next chapter. In that chapter, intelligibility will be measured as the success in identifying the discrete words in an utterance. It seems likely that a high word-identification score will be affected more by the content words than by the non-content-words because it is probably easier to guess the identity of the non-content words from having identified the content-words than vice versa (this is merely the researcher's postulation, and this matter will not be investigated in this thesis). In all of the three selected words, the first syllable initiated one of the two Norwegian accent contours (the Norwegian accents were explained in Chapter 1, section 1.4.4) in the pronunciation of the N1 template speaker, and these words therefore displayed considerable F0 movement in the N1 template. It should here be pointed out that the present is *not* an analysis of the Norwegian accents. Accent contours are not confined to discrete words, instead each accent contour is initiated by a stressed (primary stress) syllable and persists until the next stressed syllable, defining an accent phrase.

Figure 3.16 shows an N2 original and an N2 intonation manipulated intonation contour in a schematic form.

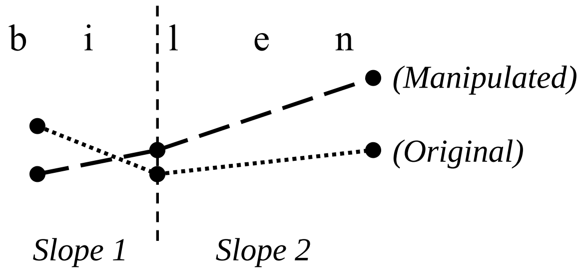

Figure 3.16: Schematic representations of an N2 original contour (dashed line) and the corresponding N2 intonation manipulated contour (dotted line) for one of the three words (bilen= "the car") selected for the intonation investigation. F0 was measured at three points (black dots) in the word, defining two slopes.

Figure 3.16 shows that F0 was measured at three coordinates in each word (black dots). The first and last coordinates define the beginning and end of the word whereas the middle coordinate corresponds to the intonational turning point of each particular curve. In the example, one contour (dashed line) has a rising slope 1 and a continuing rise through slope 2. This contour represents the manipulated N2 intonation. The other contour (dotted line) has a falling slope 1 and a rising slope 2. This represents the intonation original N2 intonation. The schematic representation in Figure 3.16 is a simple illustration of the two ways in which intonation contours can differ: a) in terms of the steepness of the slope (as in slope 2) and b) in terms of direction of the slope (as in slope 1).

It was decided to perform the analyses in two steps, Step A and Step B: In Step A, only the impacts of slope steepness adjustments were investigated. In Step B, attempts were made to add information about slope direction to the measurements in order to investigate the impacts of *slope steepness+slope direction* as one compound parameter. Step A and Step B of the investigation are further explained in the following.

Step A: In these analyses, only the slope steepness adjustments were measured. For each utterance, slope steepness adjustment was measured in each of the 6 syllables (2 syllables per word). The measurements were carried out in semitones per second. This measurement showed the extent to which the steepness of each syllable slope was adjusted as a result of the intonation manipulation. It was investigated whether any of these six slope steepness adjustments correlated with the intonation manipulation effect. These analyses were conducted to determine for which syllables the slope steepness adjustments affected the degree of foreign accent. It was expected that at least the F0 movements in the three stressed syllables would have affected the listeners' perceptions because stressed syllables are more perceptually salient and have tonal accent.

The above paragraph explains how intonational adjustment was measured in 6 syllables divided over 3 words. In this section, one more measure is defined that extends over a larger part of the utterance. The three selected words each began with a stressed syllable. In the N1 original, and consequently in the N2 intonation manipulated utterances, the second and third of these stressed syllables represented the F0 maximum and the F0 minimum of the utterance. In this respect, the three syllables together define the "intonational frame" of the intonation manipulated utterances. It was therefore decided to also measure the slopes between the first and second stressed syllables, and the slopes between the second and third stressed syllables. Measurements were made at the beginning of each stressed syllable (these coordinates are therefore identical to the first coordinate defining each of the three content words exemplified in Figure 3.16 earlier). Figure 3.17 shows the measurement of the slope steepness between words 1 and 2 ("bilen-kjørte") and between words 2 and 3 ("kjørte-huset").

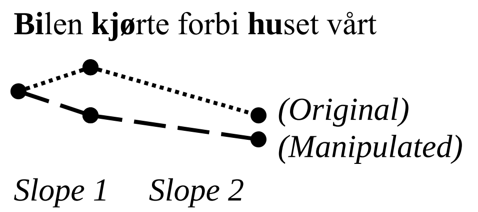

Figure 3.17: Schematic representations of an N2 original contour (dotted line) and the corresponding N2 intonation manipulated contour (dashed line) for the utterance "Bilen kjørte forbi huset vårt" (*The car drove past our house)*. F0 was measured at three points (black dots) over the utterance corresponding to the beginning of each of the three selected words.

Figure 3.17 is a schematic representation of how the F0 at the beginning of each of the selected words differs between an N2 original (dotted line) and an N2 intonation manipulated contour (dashed line).

The units in which the N1-N2 difference regarding F0 slopes and F0 directions were investigated are listed below.

- Across the 6 syllables
- word 1, syllable 1
- word 1, syllable 2
- word 2, syllable 1
- word 2, syllable 2
- word 3, syllable 1
- word 3, syllable 2
- Between word 1 and word 2
- Between word 2 and word 3

Step B: In Step A (above) the slope steepness adjustment was measured in 6 syllables divided over 3 words. An additional measure was included in which slope steepness adjustment was measured between the beginnings of the three words. Step B comprises not only slope steepness adjustment, but also slope direction adjustment. Slope direction was illustrated in Figure 3.16 earlier, which shows that the intonation manipulation involved changing the direction of Slope 1 from rising to falling. The analyses in Step B used the *same* slope steepness measures as in Step A, but the slope steepness measures from those slopes in which the direction was altered were given an arbitrary weighting by multiplying them by a factor of 2. This method was a means of acknowledging the importance of the direction of the intonation curve by weighting slopes that have been adjusted in terms of direction. The resulting compound parameter thus represents *both* slope steepness *and* direction adjustment. Figure 3.18 visualizes Step B.

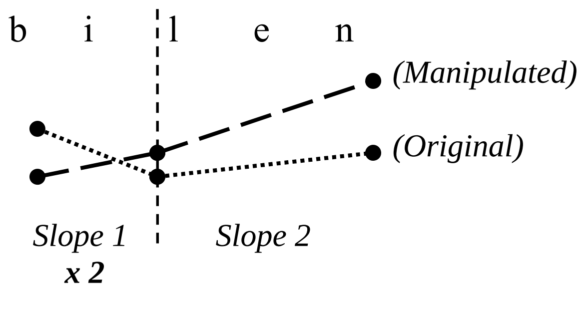

Figure 3.18: The two slopes (dotted and dashed) have different directions in Slope 1 and the same direction in Slope 2. For the analyses in Step B, the measurement of slope steepness in Slope 1 was therefore multiplied by 2.

Figure 3.18 shows that slopes that have been altered in direction as a result of the intonation manipulation have been multiplied by a factor of 2 for the analyses conducted in Step B.

3.6.2.2 Results

As explained above, each of the defined factors was investigated for correlation with the intonation manipulation effect. The correlations were performed as two multiple regression analyses in Step A and Step B respectively, one investigating the 6 slopes in the 3 words and the other investigating the slopes between the three words.

First, we look at the extent to which each utterance has been intonation manipulated, or in other words, the manipulation size for each utterance. This is defined as the mean intonational adjustment across each of the 6 syllables. The assumption is that if the manipulation size is large for an utterance, then the manipulation effect should also be large for this utterance. In Figure 3.19, the manipulation size across the 6 syllables for each utterance (vertical bars) and the resulting manipulation effect (graph) are shown. The figure does not comprise information about slope direction, only slope steepness.

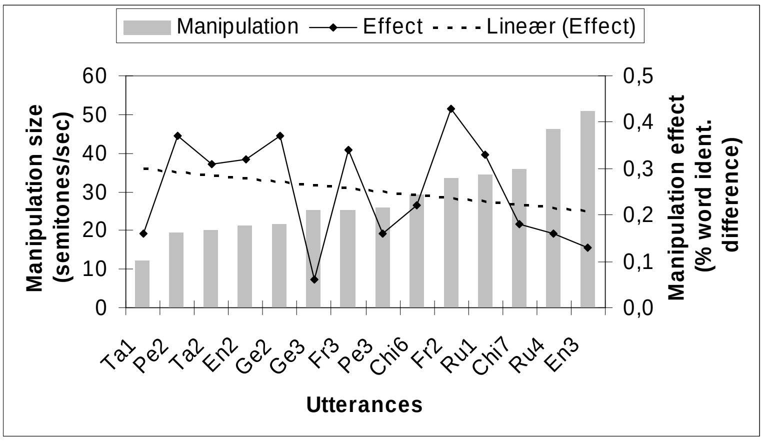

Figure 3.19: Manipulation size (across 6 syllables) and manipulation effect for each utterance (n= 14). Data in ascending order according to manipulation size. A trend line was drawn for the manipulation effect.

First, analyses corresponding to Part A are carried out, in which only the F0 slopes are investigated. In Figure 3.19, the trend line for the manipulation effect indicates that as the manipulation size (the utterance's overall intonational adjustment as measured across 6 syllables) increases, the perceptual effect decreases. If there is a significant correlation, it must therefore be negative. A regression analysis with the intonational adjustment across the 6 syllables as the predictor variable and the manipulation effect as the dependent variable

(Table 2, Appendix C) showed no effect. A multiple regression analysis with the 6 slope steepness adjustments as the predictor variables (Table 2, Appendix C) also showed no correlation. Another multiple regression analysis (Table 2, Appendix C) showed that there were no effects when investigating slope steepness adjustments between the three words. There were thus no effects to be found when investigating only the slope steepness adjustments. The following section discusses Part B of the investigation, in which slope direction was included.

As explained earlier, Part B of these intonation analyses was based on the same slope steepness adjustments as in Part A, but the slopes in which the direction was altered have now been multiplied by a factor. A regression analysis was carried out with adjustments across the 6 syllables as the predictor variables (Table 3, Appendix C) but there was no effect. A multiple regression analysis (Table 3, Appendix C) with each of the 6 syllables as the predictor variables also showed no effect. Lastly, no effect was found in a multiple regression analysis with the slopes between the three words as predictor variables (Table 3, Appendix C).

In this section, attempts have been made to relate the extent to which various parts of the utterances were adjusted (in terms of slope steepness and slope direction) with the resulting intonation manipulation effect for each utterance. The hope was to identify the intonational changes responsible for the accent reducing effects described in sections 3.5.6 through 3.5.12 earlier. Disappointingly, no correlations were found. This may be due to difficulties in determining the perceptually relevant aspects of the intonation contour and in finding an appropriate method of measurement to represent these aspects. On the other hand, the reason may be that the perception of intonational aspects, at least for the specific task of assessing the degree of foreign accent, is more holistic than the perception of durational aspects, such that local adjustments are perceptually relevant only cumulatively.

3.7 Similarity between speakers

The degree of accent-experiment in this chapter was based on two speakers from each of the 7 L1s. In this section, two speakers from the same L1 will be referred to as a *speaker pair*. For the analyses of *manipulation effect* (section 3.5), the perceptual data were pooled across each speaker pair because it was assumed that utterances spoken by speakers from the same L1 would be similarly affected by the manipulations. The *production analyses* (section 3.6.)

investigated manipulation size and manipulation effect for each individual speaker. This section brings together information from the manipulation effect analyses across the two speakers *and* from the production analyses for each individual speaker with the purpose of discussing the degree of similarity between the two speakers in each speaker pair.

Similarity within speaker pairs will be discussed in terms of manipulation size (a measure for the degree of N1-N2 deviance), magnitude of manipulation effect (the degree of accent reduction) and the relative impacts of intonation manipulation versus duration manipulation (the manipulation that *most* affects the degree of accent). The latter type of similarity is of prime importance because the analyses in this chapter were based on the assumption that utterances read by speakers from the same L1 will be *most* affected by the same manipulation, and the results were therefore pooled across the two speakers. However, if the two speakers from a particular L1 were in fact different in this respect, the accent reduction effects observed earlier for this L1 group can not be generalised to apply to the whole L1 population.

This section continues to use terms defined in the past two sections. These terms are briefly repeated here. *Durational manipulation size* refers to the overall durational adjustment across all segments in each utterance as defined in section 3.6.1. *Intonational manipulation size* refers to the overall intonational adjustment across 6 syllables in each utterance as defined in section 3.6.2. *Duration manipulation effect* is the accent reduction observed in the O_D stimulus pair across all listeners. *Intonation manipulation effect* is the accent reduction in the O_I stimulus pair across all listeners. In section 3.6.2, two types of manipulation size were used in the analyses, one with a weighting for deviating F0 slope-direction and the other without such weighting. In this section, intonational manipulation size was calculated without such weighting.

The following figures show the manipulation sizes (Figure 3.20) and the manipulation effects (Figure 3.21) for each speaker.

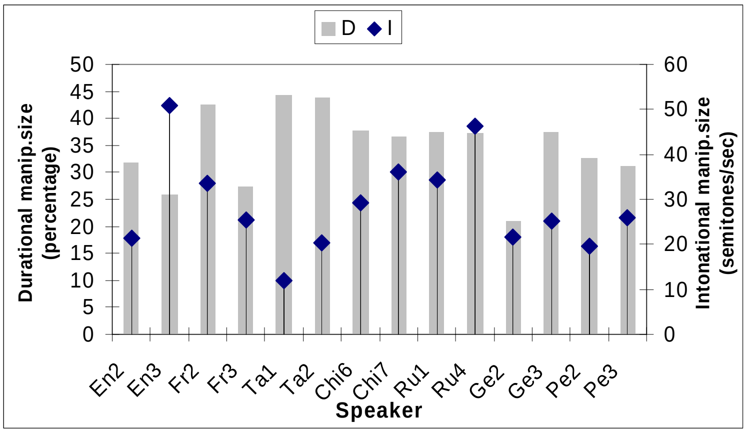

Figure 3.20: Durational (grey bars) and intonational (black lines with squares) manipulation size for each speaker.

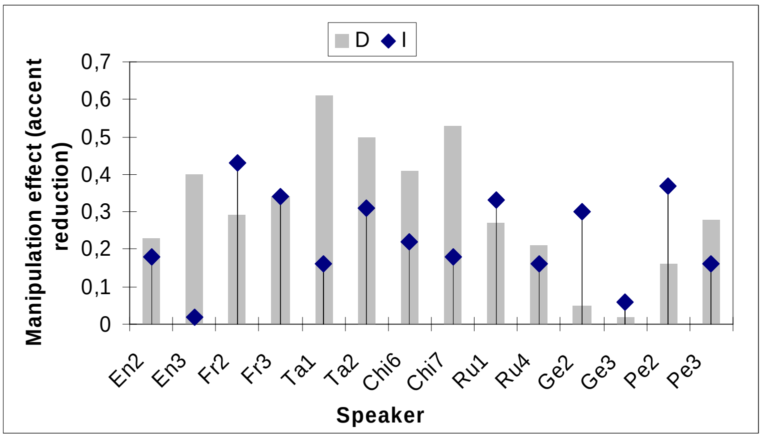

Figure 3.21: Durational (grey bars) and intonational (black lines with squares) manipulation effect for each speaker.

Regarding durational manipulation size, Figure 3.20 shows that the two speakers within the French, and the two speakers within the German speaker pair were relatively different from each other, that the two English speakers were more similar, and that the rest of the speakers showed a very high degree of within-pair consistency. There therefore seems to be good within-pair consistency for most L1 groups regarding N1-N2 durational deviation. As for intonational manipulation size, there was a large difference between the English speakers. The other speaker pairs showed small inter-speaker differences. However, the degree of discrepancy was very similar within *each* of these speaker pairs. This was interpreted to indicate that these small differences were within the range of "normal" variation between speakers from the same L1. The speakers in each speaker pair were thus *in general* fairly similar regarding their N1-N2 degree of production deviance. Figure 3.21 shows the manipulation effects. Regarding the duration manipulation effect, it can be seen that the English speaker pair (again) shows inter-speaker difference while the rest of the speaker pairs show inter-speaker similarity. The speakers have been more differently affected by the intonation manipulation. The largest intonation manipulation effect differences were between the two German speakers and between the Persian speakers. The most similar intonation manipulation effects were within the French pair and within the Chinese pair.

In general, the individuals in each speaker pair were fairly similar in terms of production, while they were more different in terms of manipulation effect. One would assume that a large manipulation size would cause a large perceptual effect, and similarly, that a small manipulation size would cause a small effect. An observation supporting this assumption is that both Tamil speakers had very large durational manipulation sizes (Figure 3.20) and also very large duration manipulation effects (Figure 3.21). An observation going against this assumption is that the two German speakers had similar intonation manipulation sizes (Figure 3.20), yet Ge2 had a much larger intonation manipulation effect than Ge3 (Figure 3.21). Also, the duration manipulation size of the Russian speaker pair was virtually identical to that of the Chinese speaker pair (Figure 3.20), yet the Russian pair was more moderately accent reduced by the duration manipulation than the Chinese pair (Figure 3.21). Remember that the two previous sections (3.6.1 and 3.6.2) investigated correlations between manipulation size and manipulation effect, and that no such correlation was found[6]. In other words, the degree of N1-N2 production deviance does not predict the degree of the manipulation effect.

The aim of the degree of accent-experiment described in this chapter was to investigate the relative importance of intonational versus durational aspects on degree of foreign accent. The analyses pooled the data across the two speakers from the same L1 because it was assumed that the foreign accent in utterances spoken by speakers from the same L1 would be reduced most by the same manipulation. It is therefore interesting to compare the individuals within each speaker pair regarding the relative impacts of the two manipulations. Table 3.4 shows the mean accent reduction from the duration manipulation (as measured in the O_D stimulus pair) and the mean accent reduction from the intonation manipulation (as measured in the O_I stimulus pair) for each speaker. Asterisks show the two L1 groups (Russian and Persian) in which the individual speakers did *not* gain most from the same manipulation.

[6] At least, no correlations were found with the particular measures of manipulation size which shows *overall* durational and *overall* intonational adjustment. Effects were however found for the durational aspects of consonant duration and articulation rate.

Table 3.4: Mean accent reduction ratings for each speaker in the O_D and O_I stimulus pairs. Asterisks show the speaker pairs that were *not* similar in terms of relative impact of the manipulations.

Speaker	D effect	I effect	Manipulation
En2	0.23	0.18	Duration
En3	0.40	0.02	Duration
Fr2	0.29	0.43	Intonation
Fr3	0.34	0.34	Equal
Ta1	0.61	0.16	Duration
Ta2	0.50	0.31	Duration
Chi6	0.41	0.22	Duration
Chi7	0.53	0.18	Duration
Ru1 *	0.27	0.33	Intonation
Ru4 *	0.21	0.16	Duration
Ge2	0.05	0.30	Intonation
Ge3	0.02	0.06	Intonation
Pe2 *	0.16	0.37	Intonation
Pe3 *	0.28	0.16	Duration

The table shows that the two speakers in the English, Tamil, Chinese and German speaker pairs were consistent regarding which manipulation that most affected their accent, whereas the Russian and Persian speaker pairs showed opposite effects (speaker pairs marked with asterisks in the table). In the French L1 group however, Fr3 was equally affected by both manipulations. The two French speakers can therefore not be regarded as either consistent or inconsistent. The cases in which there was within-pair inconsistency will be further discussed in the following.

Ru1 gained most from the intonation manipulation, while Ru4 gained most from the duration manipulation. The table shows that the difference between the O_D and O_I accent reductions were very similar for both speakers: for Ru1, this difference was (0.33-0.27=0.06) 0.06, and for Ru4 the difference was (0.21-0.16=0.05) 0.05. The manipulation analyses of the Russian group in section 3.5.10 earlier concluded that the two manipulations had affected the foreign accent of the Russian speakers' N2 to the same degree, but the information in Table 3.4 shows that it was impossible to establish the relative importance of the two manipulations because of opposite effects between the two speakers.

The Persian group also shows inter-speaker difference. Pe2 gained most from the intonation manipulation, whereas Pe3 gained most from the duration manipulation. For Pe2, the difference between the O_D and O_I accent reductions was (0.37-0.16=0.21) 0.21, and for

Pe3 this difference was 0.12. This means that the superior effect of intonation for Pe2 was somewhat larger (0.21) than the superior effect of duration for Pe3 (0.12). The relative effect of the manipulations for the Persian L1 group was investigated in section 3.5.12 earlier. Because it was not possible to establish the relative importance of the manipulations in that section, it was concluded that both manipulations probably affected the N2 from the two Persian speakers to the same degree. However, Table 3.4 shows that the two Persian speakers gained most from different manipulations.

Because there were inter-speaker conflicts regarding the relative impact of the manipulations for the Russian and Persian groups, the results from the manipulation analyses earlier cannot be assumed to reflect typical effects for these L1 populations. The results for the Russian and Persian groups in sections 3.5.10 and 3.5.12, showing equal effect of the manipulations, have been affected by opposite effects between the two speakers in each speaker pair. One could further interpret the *similar* impact of the manipulations between the speakers in each of the remaining L1 groups (English, French, Tamil, Chinese and German) such that the manipulation effects found for these L1 groups do show the typical effect for each particular L1 population, but because within-pair conflict has been shown for two of the L1 groups, and because there were only two speakers per L1, there is a possibility that the inter-speaker consistency for the rest of the L1 groups could be coincidental.

The N2 utterances from speakers of the same L1 were assumed to be similarly affected by the manipulations, but some of them have in fact been differently affected. How can one ensure that the speakers selected to represent an L1 will be similar in terms of accent reduction effects? In order to find speakers from the same L1 for whom manipulations may have the same effect, one tries to select listeners with similar non-native productions in terms of N1-N2 deviations. Let us look at specific examples. The two French speakers were on the same Norwegian course level (revisit Table 2.1, Chapter 2). This would predict similarity in N1-N2 production deviation. They were in fact similar regarding *intonational* N1-N2 deviation, but they were different regarding *durational* N1-N2 deviation (Figure 3.20). The same course level does therefore not assure the same N1-N2 deviance. Moreover, the same degree of N1-N2 deviance does not assure that the manipulation effect will be similar. For example, the German speakers' intonations were adjusted to the same degree (Figure 3.20), yet the perceptual result was much greater for Ge2 than for Ge3 (Figure 3.21). Also, for the English speakers' N2 productions, En2 had more durational problems than En3, whereas En3 had more intonational problems than En2. This would predict that duration manipulation would reduce the accent more for En2 than for En3, and that intonation manipulation would reduce the accent more for En3 than for En2, but the opposite is in fact the case as shown in Figure 3.21. More importantly, there are also problems in predicting the relative perceptual importance of the two manipulations. The two Persian speakers had fairly similar N1-N2 durational and intonation deviations. One would therefore predict that these speakers would gain most from the same manipulation, but Table 3.4 has shown that the Persian speakers differ in this respect.

This section has compared the speakers within the speaker pairs in order to assess their degree of similarity. It was shown that even in the cases where the N2 speakers were fairly similar regarding N1-N2 production deviations, their N2 speech could be differently affected by the manipulations, and they could even gain most from different manipulations. The question is whether these speaker differences were due to the selection of “atypical” speakers that are not representative of their L1 populations, or if these speaker differences reflect a general variability regarding the effects of manipulations. The degree of accent-experiment reported in this chapter, used one utterance for each of the speakers. The next chapter describes the intelligibility-experiment in which three utterances per speaker have been used. At the end of that chapter, analyses equivalent to the present analyses will assess inter-speaker similarity within the speaker pairs. The larger number of utterances will make it possible to look more closely at the variability, and will also make it possible to discuss intra-speaker consistency.

4. Intelligibility of N2 speech

The previous chapter investigated and found effects of durational and intonational aspects of speech on the perceived degree of foreign accent. This chapter presents an investigation of the same speech aspects for effects on intelligibility of N2 speech. Before the results from this experiment are described, the concept of intelligibility will be defined, some methodological issues will be discussed and the experimental design will be explained.

4.1 Method of measurement

There is no universally established definition of intelligibility, as there is no agreed upon method of measuring it (Field, 2005; Jenkins, 2000; Munro & Derwing, 2005; Pickering, 2006). In the following, a brief overview of methods used to measure intelligibility is given, a distinction between intelligibility and the related term comprehensibility is drawn, and the usage of terms for the present investigation is defined.

In some investigations, listeners have judged intelligibility impressionistically by rating how intelligible they felt utterances to be. Examples of this impressionistic approach can be found in Fayer & Krasinski (1987), Palmer (1976) and Bannert (1995). For example, Bannert investigated the impact of various phonetic corrections of foreign accented Swedish on its intelligibility. In order to assess the intelligibility, he asked the native Swedish listeners to rate how intelligible they felt the different corrected utterance versions were. However, the terms intelligibility and comprehensibility have often been used interchangeably in the literature (Smith & Nelson, 1985), and the impressionistic rating of intelligibility corresponds to how M. J. Munro and T. Derwing, two influential researchers in the field, have measured perceived comprehensibility (e.g. Munro & Derwing, 1995a; Derwing & Munro, 1997). Munro and Derwing have defined perceived comprehensibility as the effort expended by the listener when processing speech (non-native speech often requires more effort to process than native speech), and they measured perceived comprehensibility through listener ratings. Instead of *rating* intelligibility, a large number of studies have used sentence transcriptions. This means that the listeners write down the words of the utterances that they hear. Examples of studies that have used this approach are Benoît (1990), Bradlow & Bent (2002), Bradlow, Kraus & Hayes (2003), Hazan & Simpson (1998), Hazan & Markham (2004), Maassen & Povel (1984), Maassen & Povel (1985), Tajima, Port & Dalby (1997) and Osberger & Levitt (1979). The previously mentioned researchers Munro and Derwing are among those who have measured intelligibility through listener transcriptions, as in Derwing & Munro (1997). That particular investigation is methodologically interesting because they used listener ratings for perceived comprehensibility measures, and used listener transcriptions for intelligibility measures, and found different results for perceived comprehensibility and intelligibility using the same speech material. This shows that ratings and transcriptions can yield different results, and that perceived comprehensibility and intelligibility therefore should be kept apart. Smith & Nelson (1985) addressed the confusion in the field regarding the definition and usage of terms. They suggested defining intelligibility as the identification of word forms and sentences, comprehensibility as word and utterance meaning or the propositional content of messages (Smith & Nelson's term comprehensibility is therefore not identical to perceived comprehensibility as defined by Munro & Derwing, as explained above), and interpretability as the perceptions of speakers' intentions. For the present investigation, intelligibility is defined as word identification in line with Smith & Nelson (1985) and Munro & Derwing (e.g. 1995a).

It is important to bear in mind that intelligibility as measured through formal word identification scores is merely one aspect of bottom-up information that may contribute to ultimately *understanding* utterances. Identifying all the word forms in an utterance does not automatically mean that the listener understands the meaning of the utterance. On the other hand, failing to recognize a word may impede understanding. Understanding speech is thus a multifaceted process drawing on information on many levels, interacting with listeners' expectations and experiences. Intelligibility must therefore not be confounded with meaning and understanding in a wider sense.

After listeners' responses have been collected, the experimenter must further process the responses. Different procedures also exist for this step in the investigation. Some researchers have counted the number of sentences that the listener has perceived perfectly (Benoît, 1990; Benoît, Grice & Hazan, 1996), some have counted only the correctly reproduced key words (the content words) in the utterances (Bradlow & Bent, 2002; Bradlow, Kraus & Hayes, 2003; Bradlow, Toretta & Pisoni, 1996), and some have counted the number of correctly perceived words per sentence (Matsuura, Chiba & Fujieda, 1999; Maassen & Povel, 1984; Maassen & Povel, 1985; Osberger & Levitt, 1979). The latter method has been chosen for the experiment presented in this chapter, because word counts provide a more fine-grained assessment of the listeners' responses, and hence possibly yield more information than key word counts or sentence counts.

In summary, the present investigation has investigated intelligibility defined as the identification of word forms measured through listener transcriptions of utterances.

4.2 Stimuli

4.2.1 Sentences

In investigations of intelligibility, both meaningful and non-meaningful sentences have been presented to the listeners. Examples of studies where meaningful sentences have been used are Bradlow & Bent (2002) and Bradlow, Kraus & Hayes (2003), whereas non-meaningful sentences have been used in Hazan & Simpson (1998) and Benoît, Grice & Hazan (1996). Because non-meaningful sentences are very common in experiments on intelligibility, the reasons for using non-meaningful sentences are explained in the following. However, non-meaningful sentences are in fact unnecessary in this experiment because of the particular experimental design.

When a sentence is meaningful, parts of it may be guessed. If a listener hears for instance the sentence "The car drove past our house", and he does not directly perceive the second word of the sentence, his knowledge of the world may provide him with the correct word because it is likely that the word refers to a motor vehicle. Therefore, the listener may identify all the words in the sentence despite the fact that he did not actually perceive the second word. This guesswork is naturally a part of all normal, real-life communication.

In tests of intelligibility, the aim is often to measure the impact of a factor upon intelligibility. In the present experiment, for instance, one of the factors investigated was global utterance intonation. In order to investigate only this factor, all other factors had to be eliminated or controlled. In the literature, this problem has frequently been solved by the use of non-meaningful sentences, often Semantically Unpredictable Sentences, called SUS (Benoît, 1990). Benoît proposed a set of five basic syntactic structures. These structures are subsequently paradigmatically filled with words randomly selected from special word lists. An example of a SUS sentence would be "The bird wrote a red table". SUS sentences are therefore grammatically correct, but not meaningful, in order to prevent the listener from guessing parts of the sentence on the basis of semantic information.

In this experiment, the sentences were arranged in separate stimulus sets. Within one stimulus set, all the sentences were different. The same sentences were included across the different stimulus sets, but in different manipulated versions. There was a different listener group for each stimulus set (this is an oversimplification which will be refined later). Intelligibility was measured by comparing the manipulated versions of the same sentence across the different stimulus sets. In other words: Different sentences were *not* compared with each other; instead, different manipulations of the *same sentence* were compared. This rather complicated experimental design is only described briefly here, but will be thoroughly explained later. The intelligibility of one particular sentence was assumed to be fairly similarly influenced by guessing across the acoustically different sentence manipulations in the different stimulus sets. For this reason, SUS sentences were not necessary in the present experiment. Instead, meaningful sentences were used. An advantage with meaningful sentences is that they occur in real-life communication situations whereas non-meaningful sentences do not.

4.2.2 Noise

It was assumed that the different sentences in this experiment would be inherently difficult to perceive, especially when produced by different speakers and, what is more, by speakers from different L1s. The inherent intelligibility of the different sentences has several reasons. Different foreign accents presumably represent different levels of difficulty for listeners. For instance, a Chinese foreign accent presumably hinders intelligibility more than a German foreign accent to Norwegian listeners. Another reason is that in their L1, different speakers have different levels of inherent intelligibility (e.g. Hazan & Markham, 2004). Perhaps this inherent speaker intelligibility could be transferred to a speaker's L2. Because the present speech material consists of N2 speech, a factor would be the individual N2 level of attainment. Lastly, the different sentences were probably differently difficult to perceive because of their inherent "guessability" as discussed in section 4.2.1 above[7]. Different inherent intelligibility levels across the different sentences could result in very low intelligibility scores for one sentence and very high intelligibility scores for another sentence. When the intelligibility score of a particular sentence was compared across the different

[7] As explained earlier, the use of different sentences will not influence intelligibility measures because intelligibility was measured across different versions of the *same sentence*.

manipulations, comparisons will be hard if the intelligibility score was 100% for this sentence across all manipulations. Noise was therefore added to the sentences. For each sentence, the intelligibility was adjusted by calibrating the sound level of the noise as explained in the next paragraph. In the literature, different types of noise have been added to stimuli, for instance multi-talker babble (Hazan & Simpson, 2004), white noise (Bradlow, Kraus & Hayes, 2003; Bradlow & Bent, 2002) and speech shaped noise (Hazan & Simpson, 1998). For this experiment it was decided to use pink noise, partly because it is easy and quick to generate, and partly because the spectrum of pink noise greatly resembles the spectrum of speech. This type of noise therefore masks the speech well.

Stereo sound files were generated with speech in one channel and noise in the other channel, and were later played as mono files in the perception experiment. All sentences were adjusted to the same mean sound level (20 dB). In order to calibrate the noise level for each individual sentence, pilot tests were carried out in which 12 subjects listened to all the original sentences that were to be used in the experiment. These pilot tests were informal and unstructured, but the method will be explained briefly. The S/N ratio for each particular sentence was first calibrated based on the experimenter's own intuitions. The listeners were presented with each sentence only once (no repetition). The word identification scores indicated the appropriateness of the S/N ratio. The ratio was often readjusted between each listener's sessions. For instance, if an S/N ratio had yielded a very high word identification rate for the first two listeners, then the S/N ratio was adjusted before the third listener. If the third listener's results showed that the noise had been decreased too much, then the noise was increased somewhat before the next listener. In this way, the S/N ratio was frequently adjusted in the course of the pilot test. At the end of the pilot test, the final S/N ratio for each sentence was calibrated based on the results for the various S/N ratios for the various listeners. For each sentence in its original version, an S/N ratio was chosen that yielded approximately 30-40 % word intelligibility. This relatively low intelligibility level was sought for the original stimuli because it was judged to be most likely that the intelligibility level would increase for the various manipulated stimuli, not decrease.

4.2.3 Close-original stimuli

In the experiment on the degree of foreign accent (previous chapter), there were four types of stimuli, namely original, duration manipulated, intonation manipulated and intonation-duration manipulated. Initially, the stimuli in the present experiment comprised only these same four types of stimuli. However, after having launched the perception tests and informally begun viewing the intelligibility scores, it soon became clear that the intelligibility scores were much higher for the original stimuli than for any of the manipulated stimuli. In section 4.4.1, it will be shown that the intelligibility scores for the original stimuli were in fact significantly higher than the scores for the manipulated stimuli. This unexpected and unwanted effect is explained in the following, and the generation of new stimuli is described, which replaced the original stimuli in the continuation of the perception experiment.

The manipulated stimuli (see Chapter 2, section 2.2 for manipulation methodology) were in fact altered not only regarding their phonetic structures (durational and intonational patterns). The utterances were also subjected to PSOLA synthesis. The PSOLA synthesis itself could have distorted the signal to the extent that the intelligibility had been affected. It is however probably impossible to manipulate speech such that the speech remains unaffected by the synthesis method. The possible effect of the PSOLA synthesis seems inevitable and impossible to counteract, and will therefore not be further discussed. Instead, we will concern ourselves with two other manipulation-induced side effects, for which counteractions are possible by minor adjustments to the original stimuli. These two factors are explained in the following. The intonation contours were stylized in the intonation manipulated stimuli (Chapter 2, section 2.2.2.1), and the utterance durations were altered (as a side effect) in the duration manipulated stimuli (Chapter 2, section 2.2.1.3). The low intelligibility of the manipulated stimuli could therefore be due to the stylization (for the intonation manipulated stimuli) and the altered utterance durations (for the duration manipulated stimuli). Because the manipulated stimuli were less intelligible than the original stimuli, the original stimuli could not serve as baselines with which to compare the manipulated stimuli. New baseline stimuli therefore had to be generated. One set of stimuli called *close-original duration* stimuli was generated for comparison to the duration manipulated stimuli. Another set of stimuli called *close-original intonation* stimuli was created for comparison with the intonation manipulated stimuli. These two additional sets of stimuli are described in the following.

4.2.3.1 Close-original duration stimuli

As explained earlier (Chapter 2, section 2.2.1.3), the duration manipulations had in fact altered not only the phoneme durations, but as a side effect, also altered the utterance durations. This is because the sum of the phoneme durations equals the utterance duration.

For example, if most of the N2 phoneme durations were shortened in an utterance, then the duration of that entire utterance would be shortened as well.

Close-original duration stimuli were generated to serve as a baseline with which to compare the duration manipulated stimuli. The close-original duration stimuli were original N2 utterances in which only the utterance durations were changed. Each original N2 utterance was lengthened or shortened linearly so that its duration matched the utterance duration of the corresponding duration manipulated utterance. The close-original duration stimuli were thus N2 utterances in which the utterance durations had been linearly adjusted, whereas the duration manipulated stimuli were N2 utterances in which *both* the phoneme durations *and* the utterance durations (as a side effect) had been altered. Because the utterance durations were equal between the close-original duration stimuli and the duration manipulated stimuli, the measurement of intelligibility would only be affected by the manipulation of phoneme durations, and not by differences in utterance durations.

4.2.3.2 Close-original intonation stimuli

The intonation manipulated stimuli were generated by stylizing the N1 contours and copying them onto the corresponding N2 utterances (Chapter 2, section 2.2.2.2). This means that the intonation manipulated utterances were changed not only in the phonetic pattern of the contour, but also in that the manipulated contour had been stylized. Stylization means that an intonation contour has been represented by only a limited number of F0 coordinates, glossing over minor variations while retaining the larger variations. The stylization did not affect the perceptual impression of the utterances' intonation in the ears of the experimenter, a trained phonetician. However, because stylization was the only difference between the original and the intonation manipulated stimuli, and because the intonation manipulated stimuli had shown lower intelligibility scores than the original stimuli, it was assumed that the stylization must have had a perceptual effect (albeit not consciously detected). The close-original intonation stimuli were generated by stylizing the original N2 utterances. Close-original intonation stimuli were thus N2 utterances where the intonation contour was stylized, whereas intonation manipulated stimuli were N2 utterances where the intonation contour was both stylized and changed regarding its phonetic pattern.

4.2.4 Stimulus sets

There were a total of six different types of stimuli in the present experiment:

- Original stimuli
- Close-original duration stimuli
- Close-original intonation stimuli
- Duration manipulated stimuli
- Intonation manipulated stimuli
- Intonation-duration manipulated stimuli

The stimuli were organized into separate stimulus sets as illustrated in Table 4.1.

Table 4.1: Stimuli in the perception experiment that investigated intelligibility of N2 speech. There were two speakers from each L1 reading three sentences each.

L1s	**O** (n= 42)	**COD** (n= 20)	**COI** (n= 20)	**D** (n= 38)	**I** (n= 39)	**ID** (n= 20)
English 6 sentences	1-6	1-6	1-6	1-6	1-6	1-6
German 6 sentences	7-12	7-12	7-12	7-12	7-12	7-12
French 6 sentences	13-18	13-18	13-18	13-18	13-18	13-18
Tamil 6 sentences	19-24	19-24	19-24	19-24	19-24	19-24
Chinese 6 sentences	25-30	25-30	25-30	25-30	25-30	25-30
Persian 6 sentences	31-36	31-36	31-36	31-36	31-36	31-36
Russian 6 sentences	37-42	37-42	37-42	37-42	37-42	37-42
Total	= 42	= 42	= 42	= 42	= 42	= 42

O= original stimuli, COD= close-original duration stimuli, COI= close-original intonation stimuli, D= duration manipulated stimuli, I= intonation manipulated stimuli, ID= intonation-duration manipulated stimuli.

The top horizontal row shows each of the six stimulus sets. There were a total of 42 different sentences within one stimulus set. The same 42 sentences occurred across the different stimulus sets, but in six acoustically different versions. The left column shows the speakers. There were speakers from 7 different L1s, and there were 2 speakers from each of these L1s.

For example, the top left of the table shows the case for the English L1 group. For English, there were two speakers who read 3 different sentences each. One of the speakers read sentences 1-3 and the other read sentences 4-6. Consequently, different manipulations of the same sentence occur across all six stimulus sets. For instance, the original sentences 1-6 as read by the English speakers also occurred as close-original duration stimuli, as close-original intonation stimuli and so on across the six manipulations.

The original stimuli are included in the table because section 4.4.1 investigates the intelligibility differences between the original and the close-original duration and the close-original intonation stimuli, and also investigates the intelligibility differences between the original stimuli and the duration manipulated, intonation manipulated and intonation-duration manipulated stimuli. These analyses will be carried out in order to assess the appropriateness of replacing the original stimuli by the close-original stimuli (as explained in section 4.2.3). The original stimuli will subsequently be excluded from all further analyses.

The sentences differed in the number of words (see Appendix E). The word identification scores will be somewhat affected by the number of words in each sentence. For instance, if one word is missed in a 5-word sentence, the intelligibility drops to 80 %, but if one word is missed in a 9-word sentence, the intelligibility drops to 89 %. Intelligibility will however only be measured across two conditions, for instance across the COD and D conditions. The impact of sentence length is presumably equal in both conditions, and sentence length should therefore not affect the intelligibility differences across conditions.

The manipulations are abbreviated as follows: O= original stimuli, D= duration manipulated stimuli, I= intonation manipulated stimuli, ID= intonation-duration manipulated stimuli, COD= close-original duration stimuli, COI= close-original intonation stimuli. The abbreviations are used in cases where the fairly long stimuli labels would make a complicated discussion more difficult to follow.

4.3 Listeners and their intelligibility data

So far, the generation of stimuli and the organization of stimuli into stimulus sets have been discussed. In the intelligibility experiment, a total of 103 native Norwegians listened to these sentences and reproduced them in writing so that the impacts of the manipulations could be measured by the experimenter. All listeners reported normal hearing. None had experience with N2 speech on a level judged as extraordinary. The listeners were of both sexes and from all parts of Norway. This section describes the setup for how the listeners listened to the various stimulus sets, and for how the resulting data was further processed by the experimenter. This setup was fairly complicated, and the reader is encouraged to read the present section thoroughly as information provided here is crucial in order to understand the data analyses throughout this chapter.

Different listeners perceive foreign accented utterances with differing ease for many reasons (e.g. Gass & Varonis, 1984). For instance, different listeners will have different familiarity with listening to foreign accented speech, and they may also have different familiarity with foreign accented speech originating from different L1s. An extreme way of removing the factor *listener* would be to use the same listener group across all stimulus sets. The same listeners would then have to listen to 6 different manipulations of the same sentence. However, at the same time, listeners would become more and more influenced by learning the sentences as the number of times they heard the sentences increased. Naturally, this stepwise increased effect of learning would interfere with the measurement of the manipulation effect.

In the literature, this problem has been solved by using different listener groups for different stimulus sets. When there are different sentences within a stimulus set and different sentences across the different stimulus sets, this means that each listener is presented with the same sentence only once. Maassen & Povel (1985) investigated the intelligibility of deaf peoples' speech and used different listener groups for different stimulus sets. However, they also pointed out that differences between the listener groups could have influenced the results, because one group might consist of listeners for whom deaf speech was more intelligible than for listeners in another group. Even if one takes care to assemble in the same group listeners who are as similar as possible in terms of for instance experience with N2 speech, different listeners may still perceive somewhat differently. In other words, the approach with different listener groups for different stimulus sets removes sentence learning as a factor, but retains the listener as a factor.

Both approaches thus have advantages and disadvantages. For this reason, the present perception experiment was carried out using a combination of these two approaches as explained in the following.

In the present experiment, some listeners listened through only one stimulus set (*single-session listeners*), whereas others listened to first one stimulus set and then, immediately after, another stimulus set (*double-session listeners*). The data from the double-session listeners' first listening session was therefore free from learning effects. The data from the double-session listeners' second listening session may be influenced by learning effects. However, because the double-session listeners have listened to no more than two stimulus sets (not 6, which was the total number of stimulus sets), the learning factor was binary (present or non-present in the data) and could therefore be investigated as a factor in multifactor statistical tests.

The number of listeners was not equal among the stimulus sets. Also, the ratio between the number of intelligibility data influenced by learning effects (from second listening sessions) and the number of intelligibility data free from learning effects (from first listening sessions) was not equal between the stimulus sets. The discussion from now on will focus on intelligibility data rather than on the listeners. It is necessary to define three different groups of intelligibility data according to their function when intelligibility is compared between two stimulus sets. Intelligibility has been measured using each of these three data groups. The data groups are explained in the following.

The All Data group:

This group comprises *all* the intelligibility data, regardless of the unequal number of data between the stimulus sets, and regardless of the unequal ratio between data affected/not affected by learning across the stimulus sets.

Let us take the measurement of the intelligibility score difference between the original and the duration manipulated stimulus sets as an example. Across L1 groups, the comparison with All Data comprised 1765 intelligibility scores for the original stimuli and 1596 intelligibility scores for the duration manipulated stimuli. For the intelligibility data from the original stimuli, 678 were affected by learning and 1087 were free from learning effects. For the intelligibility data from the duration manipulated stimuli, 756 were affected by learning and 840 were free of learning effects. As explained earlier in this section, the skewed influence of learning effects across the stimulus sets was unproblematic because the learning effect was defined as a factor in multifactor statistical tests.

The Paired Data group:

All subjects who listened to the close-original stimuli listened to two stimulus sets. Moreover, these subjects were organized in the following way. Every subject who listened to the close-original duration (COD) stimuli listened to two stimulus sets, with the other being the duration manipulated (D) stimuli. Similarly, every subject who listened to the close-original intonation (COI) stimuli listened to two stimulus sets, with the other being the intonation manipulated (I) stimuli. Note that the Paired Data comprised all the COD and COI data, but only a subset of the D and I data. (In other words, there are more D and I data than those paired with the COD and COI data).

For the subjects who provided the Paired Data, half listened to the close-original stimuli in their first session and the manipulated stimuli in their second session, and the other half listened to the manipulated stimuli in their first session and the close-original stimuli in their second session. For the Paired Data, there is thus an equal number of data with and without learning effects, both for the comparisons between the COD and D stimuli and for the comparisons between the COI and I stimuli.

There are Paired Data only for the COD/D comparison and the COI/I comparison. The setup for the Paired Data is further clarified by the illustration in Table 4.2.

Table 4.2: Comparison of intelligibility data between COD and D stimuli and between COI and I stimuli with the Paired Data.

COD	COI	D	I
Listener 1-20	Listener 21-40	Listener 1-20	Listener 21-40

The illustration shows that the same subjects (listeners 1-20) have listened to both COD and D stimuli. Another group of subjects (listeners 21-40) have listened to COI and I stimuli. The Paired Data enables within-listener comparisons of data across the stimulus sets, thus eliminating the *listener* factor.

The Paired Data is a subgroup of All Data. In other words, the All Data group is comprised of data that were also analyzed as Paired Data.

The Rest Data group:

As explained above, Paired Data is a subgroup of All Data. This means that when an effect is found when investigating All Data and when investigating Paired Data, the effect for All Data could in fact be due to effects present only within the Paired Data. In order to find out whether effects found for All Data are due to effects only within the Paired Data, intelligibility will also be investigated for a third data group called the Rest Data. The Rest Data equals All Data minus Paired Data.

The relation between the All Data, Paired Data and Rest Data groups is illustrated in Figure 4.1.

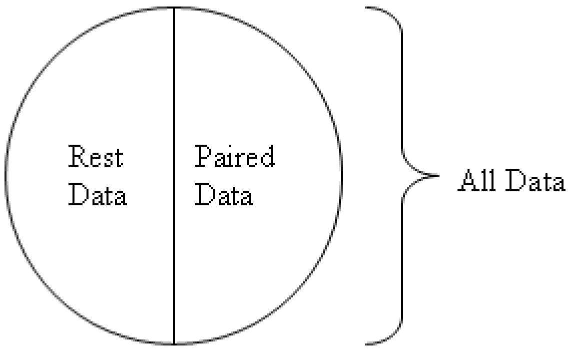

Figure 4.1: The three groups of intelligibility data.

To sum up, the data groups are as follows: The All Data group comprises *all* the data from each of the compared stimulus sets, regardless of the skewed number of data between the sets and the skewed influence of learning effects between the sets. The Paired Data group comprises only those data that originated from the same listeners across the COD and D conditions and across the COI and I conditions. The Rest Data are the All Data minus the Paired Data, and were used to investigate the reliability of effects found for All Data (i.e. whether effects found for All Data were actually due to effects present just in the subgroup called Paired Data). The advantage of the Paired Data is that all listener effects are *eliminated*, as opposed to merely *reduced* in approaches using homogeneous listener groups.

Note that there are Paired Data, and hence also Rest Data, only for the COD/D and COI/I comparisons.

4.4 Effects of manipulations

The aim of the investigation was to measure the impacts of two manipulations: intonation manipulation and duration manipulation. The impacts of the manipulations were investigated through a comparison of intelligibility scores across pairs of stimulus sets. The impact of intonation manipulation was measured through comparisons between the COI and the I data, and between the ID and the D data. The latter stimulus set pair differs only in the added intonation manipulation in the ID stimuli as compared with the I stimuli, and so any intelligibility differences between these data should be due to intonation manipulation. The impact of duration manipulation was investigated by comparing the scores between the COD and the D data, and between the ID and the I data. Because the only difference between the ID and the I stimuli is the added duration manipulation in the ID stimuli, the intelligibility score differences between them must be due to duration manipulation. The comparisons mentioned above have been used to investigate the manipulations that enhanced intelligibility in the N2 speech for the different L1 groups. Three types of results were thus possible for an L1 group: None of the manipulations enhanced intelligibility, one of the manipulations enhanced intelligibility or both manipulations enhanced intelligibility. In order to further investigate which manipulation enhanced intelligibility most, direct comparisons of the intelligibility scores between the I and D stimuli were undertaken.

When the effects of the manipulations are measured both across listeners as well as in a pair wise fashion, the results are considered particularly reliable. For this reason, the main focus was on the COI/I and the COD/D comparisons.

Before presenting the results from the perception experiment, the organization of the data in the figures will be explained. Effects on intelligibility were measured as intelligibility score differences between two stimulus sets. For the COD/D comparison and the COI/I comparison, intelligibility score differences were investigated within each of the three data groups defined in the above section. As already explained, there was an uneven number of data with and without learning effects across stimulus sets for the All Data and the Rest Data. While this imbalance was unproblematic in terms of measuring intelligibility differences due to multifactor statistical tests, it could distort the visual impression of the relative intelligibility between the stimulus sets. For instance, if a figure showed intelligibility scores across the COD and D stimuli, and the D stimuli had more data affected by learning effects than the COD stimuli, then the D intelligibility scores might look deceptively high. Therefore, figures showing data from the All Data group and the Paired Data group were based only on data free from learning effects (from first listening sessions). When the text refers to intelligibility score differences between stimulus sets, the differences were based on the same data upon which the figures are based, namely the data free from learning effects (except in Figure 4.2).

In order to perform statistical tests, the percentage of correctly perceived words per sentence was further converted into rationalized arcsine transform units (*rau units*, Studebaker, 1985). This is because rau units are more appropriate than percentage numbers for statistical tests, and because they are often used in intelligibility experiments (Bradlow & Bent, 2002; Hazan & Markham, 2004; Osberger & Levitt, 1979; Maassen & Povel, 1984 and Maassen & Povel, 1985).

The design of the present experiment has been explained in several of the previous sections. The goal of the experiment was to explore the roles of durational and intonational aspects of speech in N2 intelligibility. In the following sections, these effects are described and discussed. Detailed information about the outcomes of statistical tests can be found in Appendix B.

4.4.1 Original versus close-original stimuli

Section 4.2.3 explained that intelligibility scores were higher for the original stimuli than for any of the manipulated stimuli, which was the reason for generating the close-original duration stimuli and the close-original intonation stimuli. We therefore first investigated the intelligibility of the original versus each of the manipulated stimuli. Next, the intelligibility of the original versus the close original stimuli was investigated. If the close-original stimuli were in fact significantly less intelligible than the original stimuli, they were assumed to be appropriate for use in comparisons with manipulated stimuli.

Figure 4.2 shows the intelligibility scores for the original, the intonation manipulated, the duration manipulated and the intonation-duration manipulated stimuli.

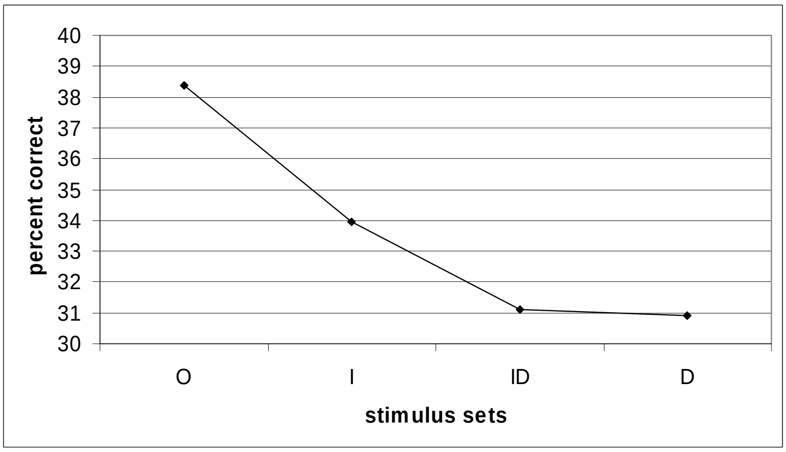

Figure 4.2: Manipulation effect for original data (n= 1765), duration manipulation data (n= 1596), intonation manipulation data (n= 1638) and intonation-duration manipulation data (n= 840). (All Data).

The graph in Figure 4.2 shows a steep decline from the score for the original stimuli to the left to the scores for the various manipulated stimuli to the right. In other words, the original stimuli seem to be more intelligible than the manipulated stimuli. The difference in intelligibility scores between the original and the intonation manipulated stimuli was 4.4 % (unit = percent correctly identified words per sentence), with the difference between the scores for the original and the intonation-duration manipulated stimuli at 7.3 %, and the difference between the scores for the original and the duration manipulated stimuli at 7.5 %.

Three separate analyses of variance investigated manipulation and learning effects across L1s, comparing the original stimuli with each of the other three manipulations (Tables 1-3, Appendix B). The results showed that the original stimuli were in fact significantly more intelligible than any of the three manipulations ($p< 0.001$ for all three comparisons). This result indicates that there had been effects from the manipulation method (stylization of intonation contour in the case of the I stimuli and utterance durations in the case of the D stimuli) on intelligibility, which therefore justified the generation of the close-original stimuli.

The question remained whether the close-original stimuli were more appropriate as baseline stimuli instead of the original stimuli. It was therefore relevant to investigate the relative intelligibility of the original stimuli versus each of the close-original stimuli. If the intelligibility of the close-original stimuli was higher than the original stimuli, then they cannot be considered more appropriate. If the close-original stimuli were less intelligible than the original stimuli, then this would be interpreted to indicate that the intelligibility of the close-original stimuli had been lowered as a result of the manipulation methods, and the close-original stimuli would then be judged as more appropriate to use as baselines with which to compare the manipulated stimuli. Figure 4.3 shows the intelligibility scores for the original, close-original intonation and close-original duration stimuli.

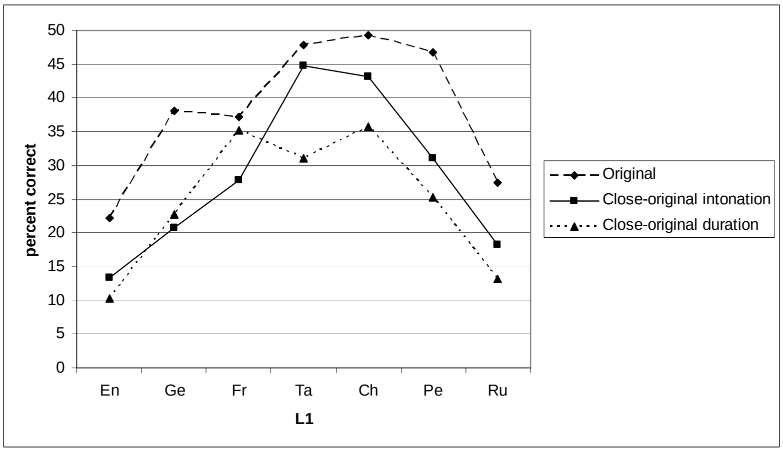

Figure 4.3: Manipulation effect for original data (n= 1765), close-original intonation data (n= 840) and close-original duration data (n= 840). (All Data).

Figure 4.3 shows that both the close-original intonation stimuli and the close-original duration stimuli had lower intelligibility scores than the original stimuli. This indicates that the unexpected results showing lower intelligibility for the manipulated stimuli than for the original stimuli were probably due to the manipulation methods themselves. The mean intelligibility score difference between the original and the close-original intonation stimuli was 10.0 %. This difference was tested by means of an analysis of variance for independent samples with L1, manipulation and learning effects as factors (Table 4, Appendix B), and was found to be highly significant ($F(1, 2619) = 79.998$; $p < 0.001$). For the separate L1s (Table 5, Appendix B), the differences between the manipulations ranged from 3.2 % for Tamil to 17.3 % for German. All these differences were significant, except for Tamil. The close-original intonation stimuli were thus generally less intelligible than the original stimuli.

We turn now to the comparison of the original stimuli with the close-original duration stimuli. The difference between these two stimulus sets was 13.6 %. An analysis of variance for independent samples with L1, manipulation and learning effects as factors (Table 6, Appendix B), showed that this difference was statistically significant ($F(1, 2577) = 161.681$; $p < 0.001$).

When each L1 was investigated separately, the differences between the manipulations ranged from 2.0 % for French to 21.5 % for Persian. An analysis of variance for independent samples for each L1 factors manipulation and learning effects as factors (Table 7, Appendix B), showed that all differences were significant except for French. The close-original duration stimuli were thus in general less intelligible than the duration manipulated stimuli.

In this section, the intelligibility differences between the original and the close-original intonation stimuli and between the original and the close-original duration stimuli are investigated. Each of the close-original stimuli sets was shown to be less intelligible than the original stimuli. This indicates that the close-original stimuli were more appropriate than the original stimuli for use as baselines with which to compare the manipulated stimuli. For this reason, only the close-original stimuli, and not the original stimuli, are compared with the manipulated stimuli in the remainder of the chapter.

4.4.2 Intonation manipulation

In this section, the effect of the intonation manipulation on N2 intelligibility is investigated. The investigation was carried out by comparing intelligibility scores across the close-original intonation and the intonation manipulated stimuli, and also across the intonation-duration manipulated stimuli and the duration manipulated stimuli.

First, the close-original intonation and the intonation manipulated stimuli were compared. As explained earlier (section 4.3), intelligibility was investigated for the COI and I comparison with three different groups of data called All Data, Paired Data and Rest Data. The All Data comprised all the data for the two stimulus sets regardless of whether the listeners listened to one or two manipulations. The All Data was therefore analyzed across listeners. Figure 4.4 shows the intelligibility scores for the close-original intonation stimuli and the intonation manipulated stimuli for All Data.

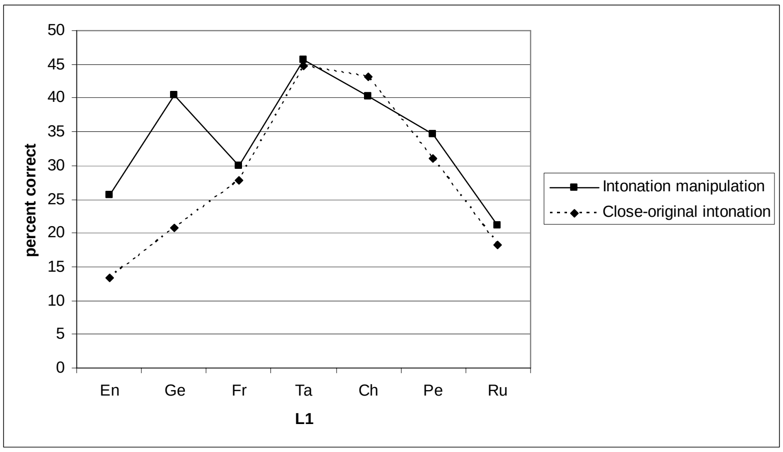

Figure 4.4: Manipulation effect for close-original intonation data (n= 840) and intonation manipulation data (n= 1638). (All Data).

Figure 4.4 shows that, in general, the overall intelligibility score was higher for the intonation manipulated stimuli than for the close-original intonation stimuli. The mean difference between the sets of stimuli was 5.6 %. An analysis of variance for independent samples with L1, manipulation and learning effects as factors (Table 8, Appendix B), showed that this difference was statistically significant ($F (1, 2492) = 28.325$; $p< 0.001$). The test also showed that there was significant interaction between L1 and manipulation ($F (6, 2492) = 5.482$; $p< 0.001$), meaning that the differences between the manipulations varied according to L1.

The difference was greatest for English (12.4 %) and German (20.1 %). When each L1 was analyzed separately by means of an analysis of variance for independent samples with manipulation and learning effects as factors (Table 9, Appendix B), only English ($F (1, 356) = 20.233$; $p< 0.001$) and German ($F (1, 356) = 39.840$; $p< 0.001$) showed significant differences between the manipulations. For the Chinese L1 group, however, the intelligibility score was somewhat higher (3.0 %) for the close-original intonation stimuli than for the intonation manipulated stimuli, but this difference was not significant. The results for the All Data thus showed that the intonation manipulation significantly enhanced the N2 intelligibility for the English and German L1 groups.

We turn now to the subgroup of data called the Paired Data. The Paired Data originated from the same group of subjects listening to both the close-original intonation stimuli and the intonation manipulated stimuli. Therefore, the Paired Data were compared in a pair wise manner within listeners, eliminating the effect of different listeners. Figure 4.5 shows the intelligibility scores for the close-original intonation stimuli and the intonation manipulated stimuli for the Paired Data.

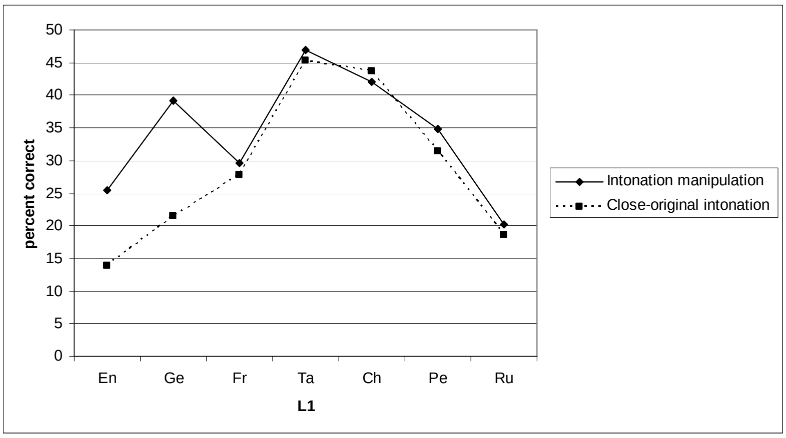

Figure 4.5: Manipulation effect for close-original intonation data (n= 840) and intonation manipulation data (n= 840). (Paired Data).

Figure 4.5 strongly resembles Figure 4.4. The Paired Data showed that the intonation manipulated stimuli had an overall higher intelligibility score than the close-original intonation stimuli. The difference between the means for the two manipulations was 5.1 %. An analysis of variance for repeated measures with L1, manipulation and learning effects as factors (bottom of Table 10, Appendix B), showed that the effect of intonation manipulation was significant ($F (1, 838) = 24.891$; $p< 0.001$). The figure shows that the differences varied between the L1s.

The separate L1s showed differences between the manipulations ranging from 3.0 % for Persian to 1.6 % for Tamil. The L1s with the largest differences were English (11.5 %) and German (17.7 %). An analysis of variance for repeated measures for each L1 separately, with manipulation and learning effects as factors (Table 10, Appendix B), showed that the effect was significant only for English ($F (1, 118) = 16.100$; $p< 0.001$) and German ($F (1, 118) = 34.543$; $p< 0.001$) L1 groups.

At this point, the role of intonation in N2 intelligibility has been investigated, first across listeners for All Data and then within listeners for the subgroup Paired Data. Both types of comparison showed the same results: When the N2 utterances' intonation contours were replaced with the N1 intonation contour, the N2 utterances became more intelligible. However, when the different L1s were investigated separately, it was shown that only the

English and German speakers' N2 benefited from the intonation correction. The other L1s did not show a significant effect.

In the above sections, the effect of intonation manipulation was measured by comparing intelligibility scores across the close-original intonation and the intonation manipulated stimuli. These data were categorized into three different data groups, and in the above sections the groups called All Data and Paired Data were investigated. In the following, the ID/D comparison is investigated. For this comparison, only All Data were used. (Remember from section 4.3 that a subgroup of D data had been paired with the COD data for the COD/D Paired Data comparisons. This subgroup of D data was however irrelevant for all other comparisons than the COD/D comparison). The difference between the ID and the D stimuli was that intonation manipulation was present in the ID stimuli while absent in the D stimuli. Any differences between them should therefore be due to intonation manipulation.

Figure 4.6 shows the intelligibility scores for the duration manipulated stimuli and the intonation-duration manipulated stimuli.

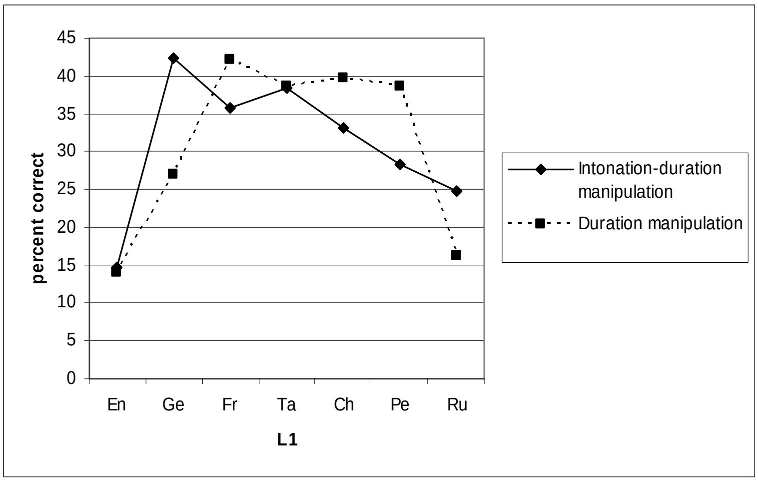

Figure 4.6: Manipulation effect for duration manipulation data (n= 1596) and intonation-duration manipulation data (n= 840). (All Data).

Figure 4.6 indicates that the data for the two manipulations were fairly similar. Across all L1s, the mean difference between the stimulus sets was only 0.3 %. An analysis of variance for independent samples across L1s with L1, manipulation and learning effects as factors (Table 20, Appendix B), showed that this small difference was not significant.

For the English, German and Russian L1 groups, the intonation-duration manipulated stimuli had higher intelligibility scores than the duration manipulated stimuli. For French, Tamil, Chinese and Persian, however, the duration manipulated stimuli had higher intelligibility scores than the intonation-duration manipulated utterances. An analysis of variance with manipulation and learning effects as factors for each L1 separately (Table 21, Appendix B), showed that two of the L1s that had higher intelligibility score for the intonation-duration manipulated stimuli, German and Russian, reached significance alone. For German, the difference between the stimulus sets was 19.0 % ($F (1, 344) = 22.155$; $p < 0.001$). For Russian, the difference was 11.0 % ($F (1, 344) = 6.884$; $p < 0.01$). Also, two of the L1s with higher intelligibility score for the duration manipulated stimuli, Chinese and Persian, showed significant differences between the manipulations. For Chinese, the difference was 9.8 % ($F (1, 344) = 4.096$; $p < 0.05$). For Persian, the difference was 13.2 % ($F (1, 344) = 6.577$; $p < 0.05$).

These unexpected results could be interpreted to mean that the addition of intonation manipulation (in the ID stimuli) significantly enhanced intelligibility for two groups, German and Russian, while it reduced intelligibility for another two groups, Chinese and Persian. Moreover, these results were in partial conflict with the results found in the COI/I comparisons in the previous paragraphs of the present section, where it was shown that intonation manipulation enhanced N2 intelligibility not only for the German group, but also for the English group. However, the confusing outcome of the ID/D comparison could be due to unwanted factors as suggested in the following.

The first factor for discussion is stylization. Remember from section 4.2.3 that the I stimuli had lower intelligibility scores than the O stimuli, and that the O stimuli therefore were deemed unsuitable for comparisons with the I stimuli. The reason for the low scores for the I stimuli was presumably due to the manipulation method (PSOLA synthesis and/or intonation curve stylization) used for the I stimuli. That was the reason for generating the COI stimuli (section 4.2.3.2), which have stylized contours, and that were used in comparison with the I stimuli. In the ID/D comparison, the ID stimuli have stylized intonation contours whereas the D stimuli did not. The stylization in the ID stimuli could have lowered the intelligibility in these stimuli. Another possible confounding factor complicates the direct comparison between the findings in the present and previous paragraphs, namely the fact that both of the stimuli being compared are *manipulated* stimuli. In the comparisons between the close-original and the manipulated stimuli, the change from non-manipulated speech to manipulated speech was investigated. The results from that investigation showed whether the change made the N2 speech significantly more intelligible *as compared to the non-manipulated N2 speech*. In the ID/D comparison, there was no stimulus set that represented non-manipulated speech. The ID/D comparisons can strictly speaking only show whether one manipulated pronunciation was more intelligible than another manipulated pronunciation. Thus, for a given L1 group, the ID stimuli could be significantly more intelligible than the D stimuli, but none of the manipulations may necessarily enhance N2 intelligibility significantly for this group.

The comparisons of the ID stimuli with the D stimuli were possibly corrupted by unwanted factors as discussed above. These unwanted factors make it hard to interpret the results from this section. Because the COI/I comparisons a) involve stimuli that represent non-manipulated N2 speech (the COI stimuli) and b) were investigated both across listeners (with All Data) and within listeners (with Paired Data), the results from the COI/I comparisons are regarded as more reliable than the results from the ID/D comparisons. Because the results from the ID/D comparison are inconclusive, I have chosen to focus exclusively on the results from the COI/I comparisons in the following.

4.4.2.1 Reliability

In section 4.4.2 above, it has been shown that intonation manipulation enhances the intelligibility of N2 speech. Only the L1 groups English and German however reach significance alone (as explained at the end of the previous section, the results from the ID/D comparison will not be further discussed). This effect was found for All Data and for Paired Data. Remember from section 4.3 that the Paired Data is a subgroup of All Data, and that the Rest Data is defined as the All Data minus the Paired Data. The effects found for All Data could be due to effects present only within the Paired Data. In order to investigate the reliability of the results for the All Data, the Rest Data were examined for effects of intonation manipulation. The intelligibility scores for the close-original intonation and the intonation manipulated stimuli for the Rest Data are shown in Figure 4.7.

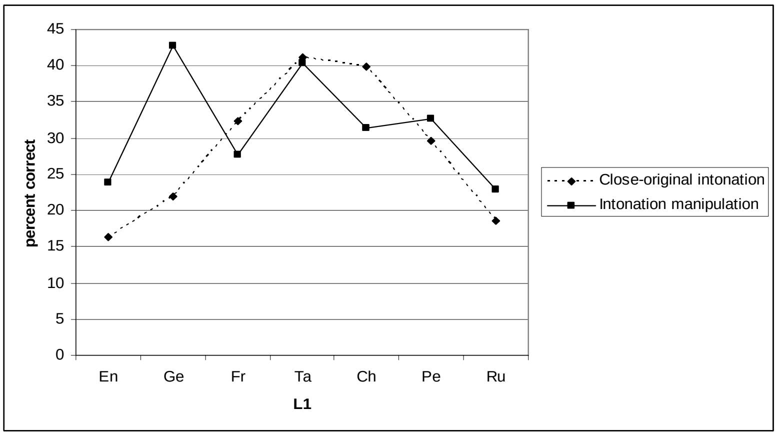

Figure 4.7: Manipulation effect for close-original intonation data (n= 840) and intonation manipulation data (n= 798). (Rest Data).

Figure 4.7 shows that some L1s had a higher intelligibility score for the close-original intonation stimuli, whereas other L1s had a higher score for the intonation manipulated stimuli. When comparing the means for the two manipulations, there was a difference of 5.4 %, such that the intonation manipulated stimuli yielded a higher score than the close-original intonation stimuli. An analysis of variance for independent samples with L1, manipulation and learning effects as factors (Table 11, Appendix B), showed that the difference between the manipulations was significant (F (1, 1652) = 20.529; $p< 0.001$). The intonation manipulated stimuli were thus more intelligible than the close-original intonation stimuli. There was also a significant interaction between manipulation and L1 (F (6, 1652) = 4.154; $p< 0.001$), meaning that the effect of the manipulation varied according to the L1.

The graphs in Figure 4.7 show that the L1s from the English and German groups had the largest differences among the manipulations. For English, the difference was 7.4 %, and for German, the difference was 20.7 %. An analysis of variance for independent samples for each L1 separately, with manipulation and learning effects as factors (Table 12, Appendix B), showed that the effect of intonation manipulation was in fact present only for English (F (1, 236) = 15.941; $p< 0.001$) and German (F (1, 236) = 33.375; $p< 0.001$).

In summary, the previous and present sections showed the same results for All Data, Paired Data and Rest Data. Intonation manipulation enhanced the intelligibility of N2 speech significantly, but when the effect was investigated within the different L1s, significant effects were found only for the English and German L1 groups.

4.4.2.2 Learning effects

As explained earlier (section 4.3), many of the listeners listened to two stimulus sets. One would assume that hearing the same sentences in a second listening session would improve the word identification scores. In this section, the data for the COI and I stimuli are investigated for learning effects. (Learning effects are not be investigated within the ID and the D data used in the ID/D comparison in section 4.4.2, because no reliable effects of manipulation were found with this comparison). Remember from section 4.3 that for the Paired Data, an equal number of subjects listened to the stimulus sets in the order COI – I as in the order I – COI. This design balanced the impact of learning effects across the two stimulus sets, and thereby eliminated learning effects as a factor when the intelligibility scores were compared between stimulus sets. However, learning effects were investigated not only for All Data and Rest Data, but also for Paired Data.

Before the analysis can be evaluated, however, it is first necessary to explain how information about learning effects was extracted from the Paired Data. The following explanation is based on COI and I data, but the same method of measuring learning effects is of course also valid for the Paired Data COD/D comparisons later in this chapter. The Paired Data originated from the same listeners listening to two stimulus sets, one set in the first listening session and another set in the second listening session. Half of the listeners listened first to the I stimuli and second to the COI stimuli. The other half listened first to the COI stimuli and second to the I stimuli. Figure 4.8 shows how the intelligibility scores for the Paired Data presumably varied as a result of both manipulation (COI or I) and the order in which the manipulation was presented (COI - I or I - COI). The figure provides fictitious intelligibility scores for the sake of discussion.

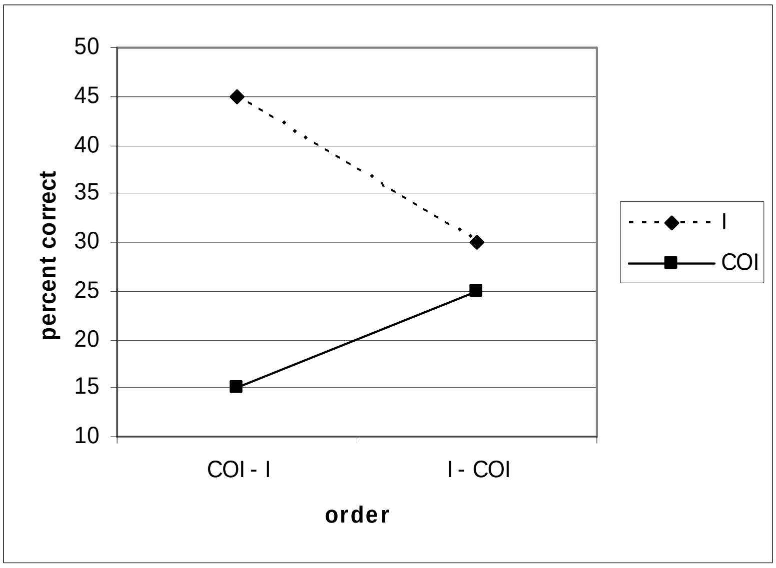

Figure 4.8: Example with fictitious data showing the intelligibility scores for the COI and I manipulations according to their order of presentation.

In Figure 4.8, the (fictitious) scores for the two manipulations are shown both with and without learning effect. When the manipulations are presented to the listeners in the order COI - I, the resulting COI data are free from learning while the I data may be affected by learning. Conversely, when the order is I-COI, the I data are free from learning and the COI data may be affected by learning. In this example, the mean intelligibility score for the COI stimuli from the COI - I order was 15 %. The score for the COI stimuli from the I - COI order was 25 %. This second score is slightly higher because of learning effects. As for the I stimuli, the score was 30 % for the I – COI order, and rises to 45 % for the I – COI order. The increase can be attributed to learning effects. There is thus an increase in intelligibility score due to learning effects for both manipulations. We see from Figure 4.8 that the learning effects in these data are revealed by an interaction between manipulation and order. For the Paired Data, significant *interaction between manipulation and order* was therefore interpreted as the presence of *learning effects*.

Note that this particular way of measuring learning effects is relevant only for the Paired Data, not for the All Data or the Rest Data where learning effects were investigated simply by comparing the data across the two listening sessions. Because learning effects were investigated in a special way for the Paired Data, in the following, one figure represents the All Data and the Rest Data whereas a separate figure represents the Paired Data. Note also that because there are no different groups of close-original data, the same close-original data is presented in both of these figures.

In section 4.4.2, the effects of intonation manipulation were found only for the English and German L1 groups separately. The effects of manipulations are the main focus of the present thesis work, not learning effects. Therefore, only the individual L1s English and German were investigated for learning effects. The aim was to evaluate the reliability of the manipulation effects that were found for these two L1s. First, learning effects were investigated for the English L1 group.

For All Data and Rest Data, Figure 4.9 shows the intelligibility scores for the close-original intonation and the intonation manipulated stimuli from the first and second listening sessions.

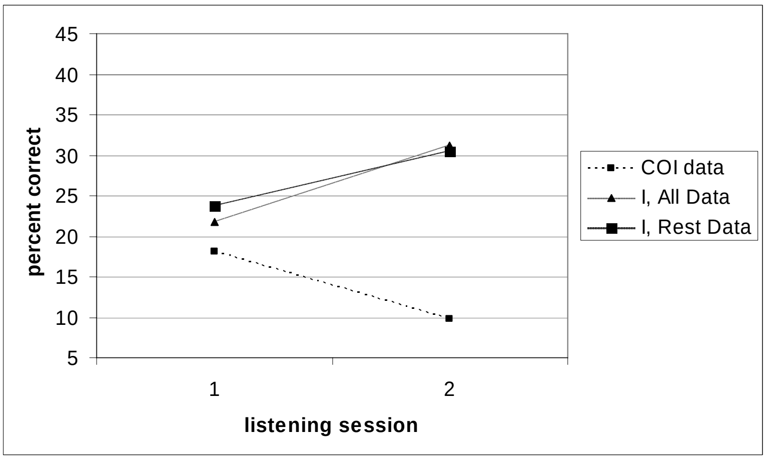

Figure 4.9: Learning effects for close-original intonation data and intonation manipulation data. COI 1st session (n= 60), COI 2nd session (n= 60), I (All Data) 1st session (n= 138), I (All Data) 2nd session (n= 96), I (Rest Data) 1st session (n= 78), I (Rest Data) 2nd session (n= 36). English L1 group.

Figure 4.9 shows that the intelligibility score for the COI data decreased (6.6 %) from the first to the second listening sessions. The intonation manipulated All Data and Rest Data, on the other hand, had higher intelligibility scores in the second listening session. The increase for the All Data was 9.4 %, and the increase for the Rest Data was 6.7 %. We start by examining the All Data. An analysis of variance for independent samples with manipulation and learning effects as factors (Table 9, Appendix B), showed that there were no learning effects for the English L1 group. Next, learning effects are investigated in the subgroup of data called Rest Data. An analysis of variance for independent samples with manipulation and learning effects as factors (Table 12, Appendix B), showed that there were no learning effects within these data either. The fact that no learning effects were found is probably because across sessions 1 and 2, the intelligibility score increased for the intonation manipulated data whereas it decreased for the close-original intonation data. We turn now to learning effects within the Paired Data. Figure 4.10 shows the COI and I scores for the Paired Data.

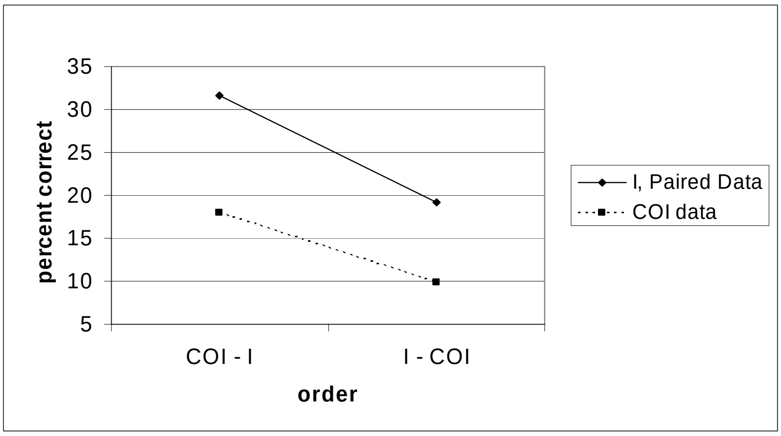

Figure 4.10: Learning effects for close-original intonation data and intonation manipulation data. COI order COI-I (n= 60), COI order I-COI (n= 60), I (Paired Data) order COI-I (n= 60), I (Paired Data) order I-COI (n= 60). English L1 group.

Figure 4.10 shows that the I data increased as a result of order of presentation. As explained earlier, there were no different groups for COI data, so the same COI data that were presented in Figure 4.9 are presented in Figure 4.10. The COI data show a decrease in the intelligibility score. An analysis of variance for repeated measures with manipulation and learning effects as factors (Table 10, Appendix B), showed that there was no significant interaction between manipulation and learning effects. This means that there were no learning effects. As was the case in the previous paragraph, the reason why no learning effects have been found is probably because the decrease in the COI data neutralized the increase in the intonation manipulated data.

In summary, no learning effects have been found in the data for the COI/I comparison for the English L1 group. The result from the present section shows that the intelligibility enhancing effect of intonation found in section 4.4.2 earlier was not influenced by learning effects, which therefore adds to the robustness of the intonation manipulation effect for the English L1 group.

The above paragraphs showed that there were no learning effects for the data in the COI/I comparison for the English L1 group. In this section, the same data is investigated for the

German L1 group. Figure 4.11 shows the COI data and the I data both as All Data and Rest Data.

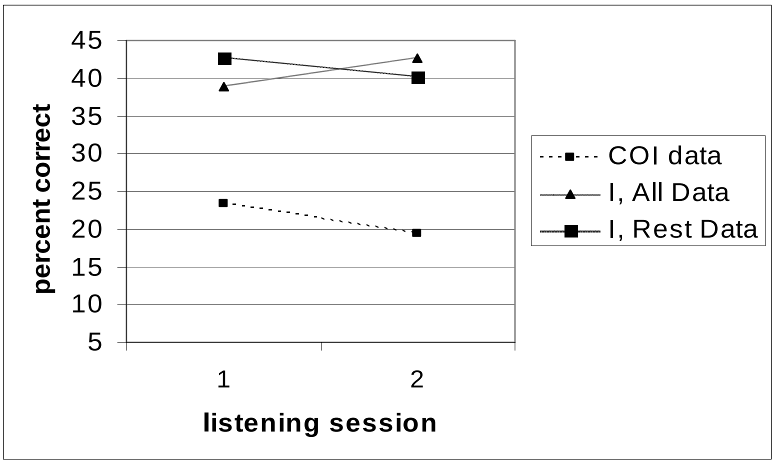

Figure 4.11: Learning effects for close-original intonation data and intonation data. COI 1st session (n= 60), COI 2nd session (n= 60), I (All Data) 1st session (n= 138), I (All Data) 2nd session (n= 96), I (Rest Data) 1st session (n= 78), I (Rest Data) 2nd session (n= 36). German L1 group.

As was the case for the English L1 group, the German L1 group also showed a drop in the intelligibility score (4.0 %) from the first to the second listening session for the COI data. The figure further shows that between the first and second sessions, the score for the All Data increased (3.8 %) whereas the score for the Rest Data decreased (2.6 %). An analysis of variance for independent samples with manipulation and learning effects as factors (Table 9, Appendix B), comparing the COI data with the I data grouped as All Data, showed that there were no significant learning effects for the German L1 group. The results from the COI/I comparison with All Data were therefore not affected by learning effects. Now we turn to learning effects within the subgroup of data called Rest Data. An analysis of variance for manipulation and learning effects as factors (Table 12, Appendix B), comparing the COI data with the I data grouped as Rest Data, showed that there were no learning effects for the German group. No learning effects were thus found for the COI/I data when the I data were grouped as either All Data or Rest Data. Figure 4.12 shows the intelligibility scores for the COI and I data for the Paired Data.

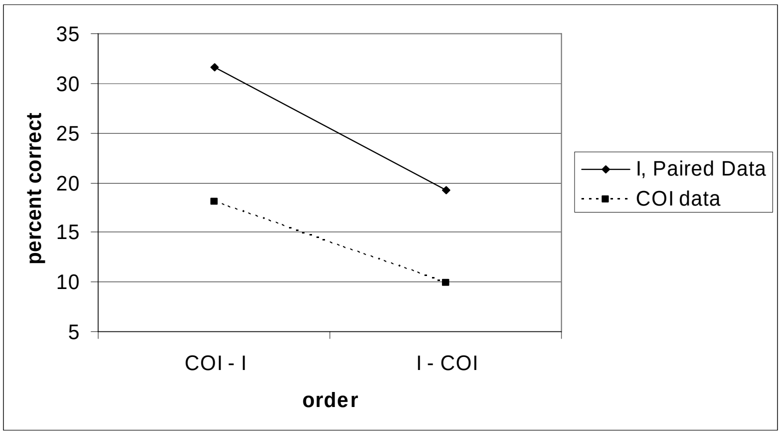

Figure 4.12: Learning effects for close-original intonation data and intonation manipulation data. COI order COI-I (n= 60), COI order I-COI (n= 60), I (Paired Data) order COI-I (n= 60), I (Paired Data) order I-COI. German L1 group.

Remember that there were no different groups for COI data. The same COI data that were presented in Figure 4.11 are therefore also presented in Figure 4.12. The figure shows that the I data grouped as Paired Data had higher intelligibility scores when they were heard second. An analysis of variance for repeated measures with manipulation and learning effects as factors, comparing the COI data with the I data grouped as Paired Data (Table 10, Appendix B), showed that there was no interaction between learning effects and manipulation for the German group; no interaction means that there were no learning effects in these data.

The findings from the present section show that no learning effects have influenced the data in the COI/I comparisons with regards to the German L1 group. Earlier in section 4.4.2, it was found that intonation manipulation significantly enhanced N2 intelligibility for the German L1 group. The findings in the present section show that this manipulation effect was not influenced by learning effects.

Table 4.3 summarizes the learning effects found in the data for the COI/I comparisons using All Data, Rest Data and Paired Data.

Table 4.3: Learning effects in the data for the COI/I comparisons.

L1s	COI/I (All Data)	COI/I (Rest Data)	COI/I (Paired Data)
English	n.s.	n.s.	n.s.
German	n.s.	n.s.	n.s.

Table 4.3 shows that there were no learning effects for the English and German L1 groups. Earlier in section 4.4.2, COI/I comparisons were carried out which showed that intonation manipulation significantly enhanced N2 intelligibility for English and German accented speech respectively. The results from the present section show that these manipulation effects were not influenced by learning effects. This in turn adds to the reliability of the manipulation effects.

4.4.3 Duration manipulation

In section 4.4.2 above, it was shown that intonation manipulation significantly influenced N2 intelligibility for the English and German L1 groups. In this section, the impact of duration manipulation on N2 intelligibility is investigated. This manipulation was investigated by comparing the intelligibility scores across the close-original duration stimuli and the duration manipulated stimuli, and also across the intonation-duration stimuli and the intonation manipulated stimuli (the only difference between the ID and the I stimuli was the added duration manipulation in the ID stimulus, and any difference between them should therefore be due to duration manipulation).

First, the difference between the close-original duration and the duration manipulated data is investigated. Figure 4.13 shows the intelligibility scores for the COD and D manipulated stimuli for All Data.

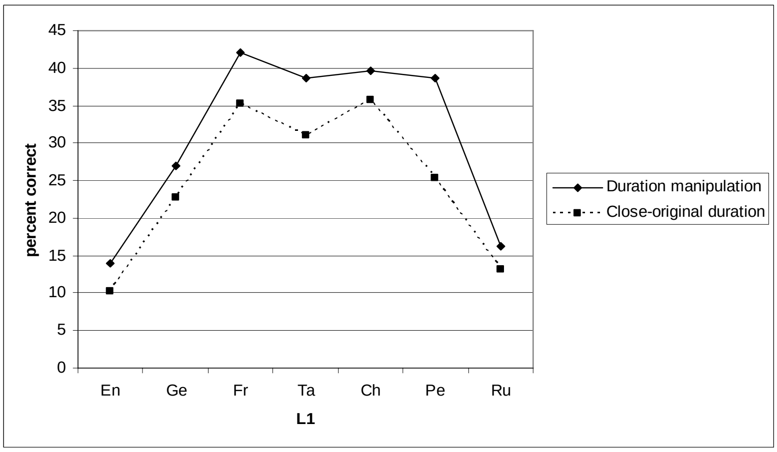

Figure 4.13: Manipulation effect for close-original duration (n= 840) and duration manipulation data (n= 1596). (All Data).

Figure 4.13 shows that the duration manipulated stimuli yielded higher intelligibility scores than the close-original duration stimuli. This was true for all L1s. The mean difference between the two stimulus sets was 3.8 %. An analysis of variance for independent samples for L1, manipulation and learning effects as factors (Table 13, Appendix B), showed that the intelligibility enhancing effect of duration manipulation was statistically highly significant ($F_{(1, 2408)} = 27.832$; $p < 0.001$).

The figure shows that the French, Tamil and Persian L1s had the largest differences between the stimulus sets. An analysis of variance for each L1 separately, with manipulation and learning effects as factors (Table 14, Appendix B), showed that duration manipulation significantly enhanced intelligibility only for these three L1s. For French the difference was 6.9 % ($F_{(1, 344)} = 5.326$; $p < 0.05$), for Tamil the difference was 7.6 % ($F_{(1, 344)} = 4.250$; $p < 0.05$) and for Persian the difference was 13.3 % ($F_{(1, 344)} = 11.295$; $p < 0.01$).

When the effect of duration manipulation was investigated with All Data, the effects of duration manipulation were found for the French, Tamil and Persian L1 groups.

We turn now to investigate the effect of duration manipulation for the Paired Data (the different data groups have been explained in section 4.3). Figure 4.14 shows the intelligibility scores for the close-original duration and the duration manipulated stimuli for the Paired Data.

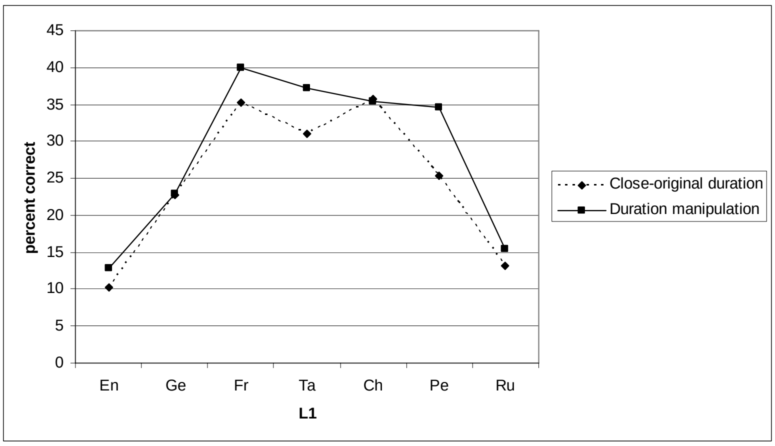

Figure 4.14: Manipulation effect for close-original duration data (n= 840) and duration manipulation data (n= 840). (Paired Data).

The Paired Data showed the same pattern as the All Data earlier: In general, the intelligibility scores were higher for the duration manipulated stimuli than for the close-original duration stimuli. The mean difference between the stimulus sets was 3.6 %. An analysis of variance for repeated measures with L1, manipulation and learning effects as factors (bottom of Table 15,

Appendix B), showed that the duration manipulation effect was statistically significant ($F(1, 838) = 14.332$; $p < 0.001$).

Figure 4.14 also shows that the differences were greatest for the French (4.7 %), Tamil (6.2 %) and Persian (9.4 %) L1s. An analysis of variance for repeated measures for each L1 separately, with manipulation and learning effects as factors (Table 15, Appendix B), showed that only three L1s reached significance on their own. These were French ($F(1, 118) = 4.364$; $p < 0.05$), Tamil ($F(1, 118) = 3.989$; $p < 0.05$) and Persian ($F(1, 118) = 7.613$; $p < 0.01$).

In summary, the COD/D comparisons showed the same results regardless of whether the data were investigated across listeners (with the All Data) or within listeners (with the Paired Data). The results showed that duration manipulation enhanced N2 intelligibility for the French, Tamil and Persian L1 groups. There was no effect of duration manipulation for any of the remaining L1 groups.

In previous paragraphs in this section, the effect of duration manipulation was measured by comparing the intelligibility scores across the close-original duration and the duration manipulated stimuli. These comparisons were conducted for All Data, Paired Data and Rest Data. In the following, the impact of duration manipulation is investigated through the ID/I comparison. Investigations are carried out only with All Data. (Remember from section 4.3 that a subgroup of I data was paired with the COI data, but that this subgroup was used only in the COI/I comparison). Figure 4.15 shows the intelligibility scores for the intonation manipulated stimuli and the intonation-duration manipulated stimuli.

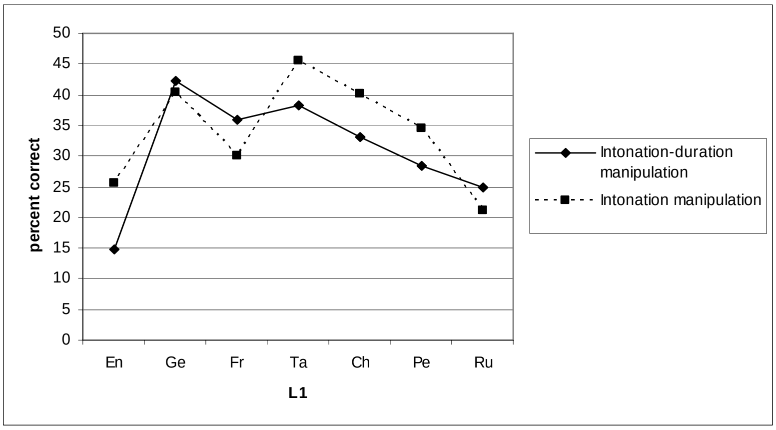

Figure 4.15: Manipulation effect for intonation manipulation data (n= 1638) and intonation-duration manipulation data (n= 840). (All Data).

Figure 4.15 shows that for the German, French and Russian L1 groups, the intonation-duration manipulated stimuli had higher intelligibility scores than the intonation manipulated stimuli, whereas the English, Tamil, Chinese and Persian L1 groups showed the opposite tendencies. Across all L1s, the difference between the stimulus sets was 5.1 %. An analysis of variance for independent samples with manipulation and learning effects as factors (Table 18, Appendix B), showed that there was a significant difference between the stimulus sets such that the stimuli that were only intonation manipulated were more intelligible than the stimuli that were intonation-duration manipulated ($F (1, 2450) = 14.549$; $p< 0.001$). The difference was highly significant. The expectation was that the added duration manipulation in the ID stimulus would enhance intelligibility as compared with the I stimulus. Before discussing the reasons behind this unexpected result, we examine the results for the separate L1s.

An analysis of variance for each of the L1s separately (Table 19, Appendix B), shows that there were significant differences between the stimulus sets for the four L1s which showed higher intelligibility for the intonation manipulated stimuli than for the intonation-duration manipulated stimuli. For the English L1 group, the difference between the stimulus sets was 15.3 % ($F (1, 350) = 16.784$; $p< 0.001$). For Tamil, the difference was 13.4 % ($F (1, 350) = 11.205$; $p< 0.01$). For Chinese, the difference was 10.6 % ($F (1, 350) = 7.737$; $p< 0.01$). For Persian, the difference was 8.4 % ($F (1, 350) = 4.186$; $p< 0.05$).

These results show that the intonation manipulated stimuli were more intelligible than the intonation-duration manipulated stimuli. This is surprising because one would expect that the combined manipulation of both intonation and duration would make the utterances even more intelligible than the intonation manipulation alone. One would at least not expect the intonation-duration manipulated utterances to be significantly *less* intelligible than the intonation manipulated utterances. At the end of section 4.4.2, a similar investigation was carried out, examining the impact of intonation manipulation through the comparisons of the ID and D stimuli. This investigation also yielded results that were unexpected and counterintuitive. However, it was hypothesized that certain unwanted factors could have influenced the results from the ID/D comparison, and similar reasons could account for the unexpected results from the ID/I comparison in the present section as explained in the following.

In section 4.2.3, it was explained that the O stimuli were more intelligible than the D stimuli. The reason was probably that the utterance durations in the D stimuli had been altered as a side effect from the duration manipulation. The utterance durations therefore differed between the O and D stimuli. For this reason, the COD stimuli were generated (section 4.2.3.1) in which the utterance durations were adjusted to match the utterance durations of the corresponding D stimuli. When the ID and I stimuli were compared in the present section, the effect of the D manipulation could have been obscured by the difference regarding utterance durations between the ID and I stimuli. Another factor that could account for the unexpected results in the ID/I comparison has already been suggested as a confounding factor in the ID/D comparison at the end of section 4.4.2. This factor concerns small intonational changes automatically induced by the D manipulation. When an utterance was D manipulated, lengthened portions automatically obtained less steep intonation slopes, and conversely, shortened portions obtained steeper intonation slopes (Chapter 2, section 2.2.1.3). Although these small intonational changes were judged as negligible and unimportant by the experimenter, it is of course possible that they have affected the results in the ID/I comparison, because such intonational changes were present in the ID stimuli but not in the I stimuli.

In order to eliminate the unwanted factor of different utterance durations between the ID and the I stimuli, a possible solution could have been to generate a separate stimulus set of I stimuli with adjusted utterance durations for the comparison with the ID stimuli. Unfortunately, problematic and inconclusive results from the ID/I comparisons mean that these results must be excluded from further discussion. In the following discussion, the effects of duration manipulation are based solely on the COD/D comparison.

In the present section, duration manipulation was found to enhance intelligibility across L1s, but it has also been shown that this general effect was due to effects only in the individual French, Tamil and Persian L1 groups.

4.4.3.1 Reliability

In the previous section, it was found that duration manipulation enhanced N2 intelligibility across all L1s, but when the individual L1 groups were investigated, significant effects were shown only for the French, Tamil and Persian groups. The significant effect of duration manipulation was present in the All Data and in the Paired Data. The third group of data, called Rest Data, equals All Data minus Paired Data. The relationship between the three groups of data was explained in section 4.3. Because the Paired Data is part of the All Data, the results for the All Data could be due to effects present only in the Paired Data. In this section, the reliability of the results for the All Data is investigated by examining whether the Rest Data shows an effect of duration manipulation. Figure 4.16 shows the intelligibility scores for the close-original duration and the duration manipulated stimuli for the Rest Data.

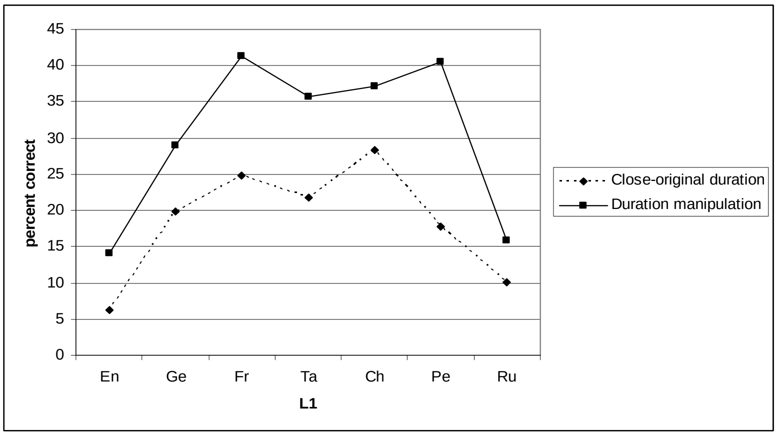

Figure 4.16: Manipulation effect for close-original duration data (n= 840) and duration manipulation data (n= 756). (Rest Data).

The figure shows that the duration manipulated stimuli had higher intelligibility scores than the close-original duration stimuli. This was true for all L1 groups. The difference between the means for the two stimulus sets was 5.7 %. An analysis of variance for independent samples with L1, manipulation and learning effects as factors (Table 16, Appendix B), showed that the effect of duration manipulation was highly significant ($F (1, 1568) = 43.888$; $p< 0.001$).

The figure shows that Russian had the smallest mean difference between the stimulus sets, and so it is not surprising that an analysis for independent samples for each L1 separately (Table 17, Appendix B) showed that there were effects for all L1 groups except Russian. For English, the difference was 7.8 % ($F (1, 224) = 5.122$; $p< 0.05$); for German, the difference was 9.0 % ($F (1, 224) = 6.361$; $p< 0.05$); for French, the difference was 16.6 % ($F (1, 224) = 6.446$; $p< 0.05$); for Tamil, the difference was 14.0 % ($F (1, 224) = 5.175$; $p< 0.05$); for Chinese, the difference was 8.9 % ($F (1, 224) = 5.396$; $p< 0.05$); and for Persian, the difference was 22.8 % ($F (1, 224) = 17.306$; $p< 0.01$).

In the earlier parts of this section, duration manipulation effects were found for the French, Tamil and Persian groups using All Data and Paired Data. The Rest Data contained even more L1 groups that showed significant effects from duration manipulation. In the Rest Data, all the L1 groups except Russian showed effects from duration manipulation. This means that in the subgroup called Paired Data, there were effects for French, Tamil and Persian, and in the subgroup called Rest Data there were effects for English, German, French, Tamil, Chinese and Persian. The fact that the results were not identical across the three data groups makes it difficult to give one simple answer as to which L1 groups significantly benefited from duration manipulation. Perhaps the subgroup of listeners for the Rest Data comprised individuals that were especially sensitive to duration manipulation. Still, because the Paired Data and the All Data were investigated both within listeners and across listeners, the identical results from these two experiments are deemed more reliable than the deviant results from the Rest Data, which were investigated only across listeners. Therefore, the results from the All Data and Paired Data are judged as more reliable than the results from the Rest Data. The conclusions regarding the effects of duration manipulation are therefore based exclusively on the results from the All Data and Paired Data.

On the basis of the results from the present section, it is possible to conclude that the intelligibility of French, Tamil and Persian accented N2 speech is enhanced by duration manipulation.

4.4.3.2 Learning effects

Because the sentences were identical across all stimulus sets, the intelligibility in the second listening session could have been influenced by learning effects. In the present section, the data in the COD/D comparisons is investigated for learning effects in order to examine the reliability of the observed manipulation effects. Remember from section 4.4.2.2 that when the statistical tests investigating learning effects in the Paired Data showed significant interaction between manipulation and the order in which the manipulation was presented to the listeners, this result must be interpreted as showing significant learning effects. Because learning effects were investigated in a special way for the Paired Data, the Paired Data are presented in one figure while the All Data and the Paired Data are presented in a separate figure.

In section 4.4.3, COD/D comparisons showed that duration manipulation enhanced N2 intelligibility for the French, Tamil and Persian L1 groups. This section investigates whether the intelligibility data for these three L1 groups were affected by learning effects. The same COD data were used for comparisons with the three different groups of D data (All Data, Paired Data and Rest Data). The same COD data therefore occurred in both the figure showing the All Data and Rest Data, and in the figure presenting the Paired Data. For the three L1 groups under investigation, the intelligibility score increased from the first to the second listening session, both for the COD data and for the D data. Therefore, one common figure is presented, in which data were pooled across the three L1 groups. Figure 4.17 shows the intelligibility scores across the two listening sessions for the COD data and for the D data grouped as All Data and as Rest Data.

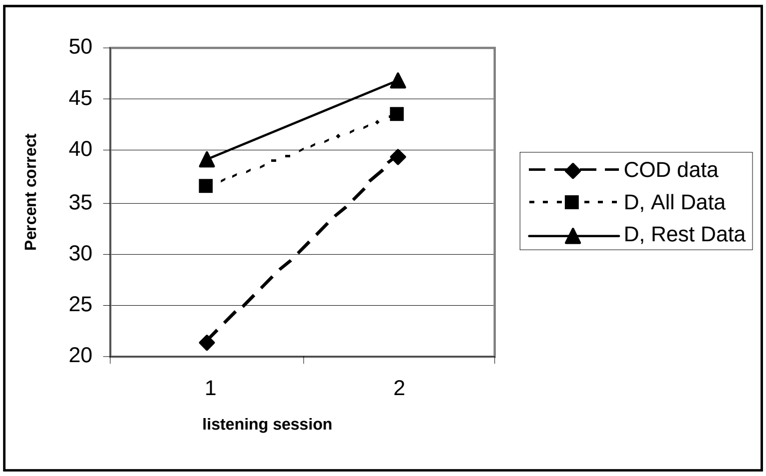

Figure 4.17: Learning effects for close-original duration data and duration manipulation data. COD 1st session (n= 360), COD second session (n= 360), D (All Data) 1st session (n= 720), D (All Data) 2nd session (n= 648), D (Rest Data) 1st session (n= 360), D (Rest Data) 2nd session (n= 288). Across the French, Tamil and Persian L1s.

The figure shows that all intelligibility scores increased in the second listening session. The increase for the COD data was 18.0 %, the increase for the D data grouped as All Data was 7.0 % and the increase for the D data grouped as Rest Data was 7.6 %. The numbers in the figure are mean scores pooled over the French, Tamil and Persian L1s. Three analyses of variance for independent samples for manipulation and learning effect as factors (Table 14, Appendix B), show that the All Data in the COD/D comparison were affected by learning effects for each of these L1 groups: French ($F (1, 344) = 17.003$; $p < 0.001$), Tamil ($F (1,344) = 15.789$; $p < 0.001$) and Persian ($F (1, 344 = 9.719$; $p < 0.01$). When the Rest Data were investigated with equivalent analyses of variance (Table 17, Appendix B), learning effects were also found for these data: French ($F (1, 224) = 14.651$; $p < 0.001$), Tamil ($F (1,224) = 14.044$; $p < 0.001$) and Persian ($F (1, 224 = 6.367$; $p < 0.05$).

When the All Data and Rest Data in the COD/D comparison were investigated for learning effects, such effects were thus found for all the three L1 groups under investigation, namely French, Tamil and Persian. We turn now to learning effects in the COD/D data used in the paired comparisons. Figure 4.18 shows the COD data and the D data grouped as Paired Data.

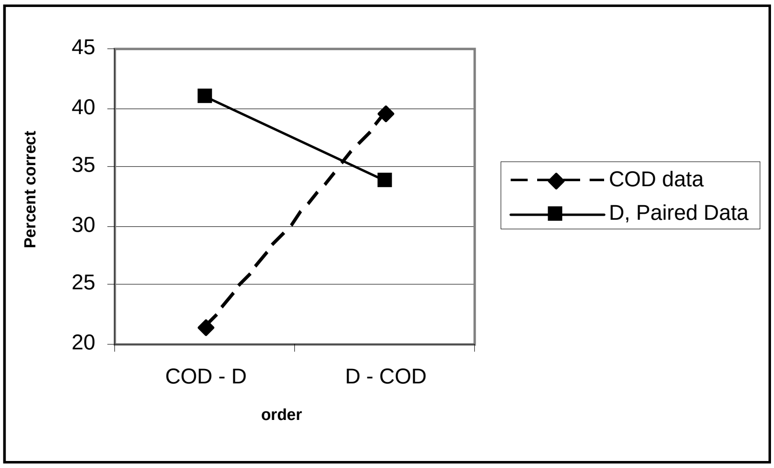

Figure 4.18: Learning effects for close-original duration data and duration manipulation data. COD order COD-D (n= 360), COD order D-COD (n= 360), D (Paired Data) order COD-D (n= 360), D (Paired Data) order D-COD (n= 360). Across the French, Tamil and Persian L1s.

The figure shows that the intelligibility score increased for both the COD data and the D data when they were heard second. The increase for the COD data was 18.0 %, and the increase for the D data grouped as Paired Data was 7.1 %. The figure shows data pooled across the three L1s. Three different analyses of variance were carried out, however, in which the factors were manipulation and order of presentation (Table 15, Appendix B). The results show that the paired data in the COD/D comparison were affected by learning effects only for the French L1 group ($F(1, 118) = 4.054$; $p < 0.05$). Remember, however, that any learning effects present in the Paired Data have not interfered with the measurement of the manipulation effect. This is because there were a perfectly balanced number of observations from the first and second listening sessions for these data, which resulted in an equal degree of learning for both manipulations.

Table 4.4 summarizes the results from the present section.

Table 4.4: Learning effects in the data for the COD/D comparisons.

L1s	COD/D (All Data)	COD/D (Rest Data)	COD/D (Paired Data)
French	Sign.	Sign	Sign.
Tamil	Sign.	Sign.	n.s.
Persian	Sign.	Sign	n.s.

Table 4.4 shows that, in general, the data in the COD/D comparisons were affected by learning effects. For the Paired Data comparison, however, there were learning effects only for the French group, not for the Tamil and Persian groups. In the case of the All Data and the Rest Data COD/D comparisons, it is possible that the learning effect could have affected the measurement of the manipulation effects. Because learning effects have not corrupted the COD/D Paired Data comparisons, the manipulation effects found with these data are reliable. The conclusion at the end of section 4.4.3, claiming duration manipulation effects for the French, Tamil and Persian groups, thus remains valid.

4.4.4 Relative effects of intonation and duration

Effects of intonation manipulation have been investigated in section 4.4.2 and found for the English and German L1 groups. Effects of duration manipulation have been investigated in section 4.4.3 and found for the French, Tamil and Persian L1 groups. These results should mean that for the English and German groups, intonation is more important to address than duration, and for the French, Tamil and Persian groups, duration should be more important than intonation. In this section, the relative importance of intonation and duration are further investigated by directly comparing the intonation manipulated stimuli with the duration manipulated stimuli. Figure 4.19 shows the intelligibility scores for the intonation manipulated stimuli and the duration manipulated stimuli.

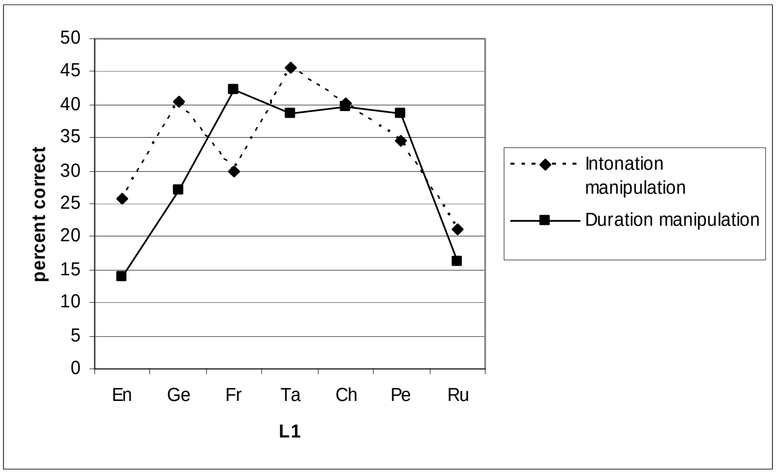

Figure 4.19: Manipulation effect for intonation manipulation data (n= 1638) and duration manipulation data (n= 1596). (All Data).

Figure 4.19 shows that the data from the two groups overlapped. The mean difference between the stimulus sets was 4.8 %. An analysis of variance for independent samples (Table 22, Appendix B), showed that this difference was statistically significant such that intonation manipulated stimuli were more intelligible than duration manipulated stimuli ($F(1, 3206) = 18.041$; $p < 0.001$).

The figure shows that for English, German, Tamil, Chinese and Russian, the scores for the intonation manipulated stimuli were higher than for the duration manipulated stimuli. For French and Persian, on the other hand, the duration manipulated stimuli had higher intelligibility scores than the intonation manipulated stimuli. An analysis of variance for independent samples for each separate L1 (Table 23, Appendix B), showed significant differences between the manipulations for English ($F(1, 458) = 28.876$; $p < 0.001$), German ($F(1, 458) = 28.078$; $p < 0.001$), Tamil ($F(1, 458) = 10.711$; $p < 0.01$) and Russian ($F(1, 458) = 6.121$; $p < 0.05$) such that the intonation manipulated stimuli were more intelligible than the duration manipulated stimuli. There was also a significant difference for French ($F(1, 458) = 16.980$; $p < 0.001$) such that the duration manipulated stimuli were more intelligible than the intonation manipulated stimuli.

The results thus showed that intonation was the most important aspect to address for the English, German, Tamil and Russian groups, while duration was the most important aspect for the French group. For the Chinese and Persian groups, there was no difference among the intelligibility of the manipulations. For some of the L1 groups, these results were not in concordance with the results from the investigations in sections 4.4.2 and 4.4.3. However, let us first discuss the results that were in agreement. It had earlier been found that intonation manipulation, but not duration manipulation, enhanced English and German N2 intelligibility. This finding was supported by the finding in the present section showing that the intonation manipulated stimuli were more intelligible than the duration manipulated stimuli for these two L1 groups. French N2 was earlier found to become more intelligible with duration manipulation, but not with intonation manipulation. This result was also in accordance with the result in the present section showing that the duration manipulated stimuli were more intelligible than the intonation manipulated stimuli for this L1 group. For these three L1 groups therefore, the results were in agreement across the findings in sections 4.4.2 plus 4.4.3 and the findings in the present section. As for the remaining L1 groups, the relations between the present and earlier results were confusing, and in some cases contradictory. For the Tamil L1 group, the findings were definitely in conflict. It had earlier been found that there were no effects of intonation manipulation for this group, but that there was an effect of duration manipulation. In this section, however, the intonation manipulated stimuli yielded higher intelligibility than the duration manipulated stimuli for this L1 group. We have earlier seen that the Persian L1 group benefited significantly from duration manipulation and not from intonation manipulation, but in this section we see that the two manipulations were equally intelligible. As for the Chinese group, no effects have been found earlier. Yet, in this section it has been shown that the duration manipulated stimuli were significantly more intelligible than the intonation manipulated stimuli for the Chinese N2.

It was expected that the I/D comparisons would show which manipulation was more important for the individual L1 groups, while in accordance with the previous results from sections 4.4.2 and 4.4.3. There are several possible reasons for the conflicting results between the results in the present and previous sections.

Towards the ends of sections 4.4.2 and 4.4.3, unexpected results arose from the ID/D and ID/I comparisons. One possible reason for those unexpected results was that none of the stimuli in the pairs represented natural N2 speech. Consequently the results from those comparisons could not show which manipulation actually enhanced N2 intelligibility as compared with natural N2 speech. This possible explanation could also apply to the present unexpected results: Because none of the stimuli in the I/D comparison represented a baseline, the results from this comparison cannot show which manipulation enhanced intelligibility relative to natural N2 speech. In the COI/I comparison and the COD/D comparison, the close-original stimuli represented natural N2 speech, and the results from those comparisons therefore showed which manipulation enhanced intelligibility. This is one plausible reason for the conflicting results between the results in the present and earlier sections. In conjunction with the ID/D and ID/I investigations at the ends of sections 4.4.2 and 4.4.3, yet another possible confounding factor was suggested, namely small intonational changes automatically induced by the duration manipulation. In Chapter 2, section 2.2.1.3, it was explained that the duration manipulation had a side effect on intonation. This is because when a portion of an utterance was lengthened, the intonation slope for that portion became less steep, and vice versa; when a portion of an utterance was shortened, the intonation slope for that portion became steeper. This small intonational change in the duration manipulated stimuli was judged as imperceptible and hence unimportant on the basis of the experimenter's own informal listening. This small intonational discrepancy was nonetheless present between stimuli that were duration manipulated and stimuli that have not been duration manipulated. These small intonational differences between the D and I stimuli could thus have affected the results from the I/D comparison in the present section. Yet more factors can tentatively be suggested for confounding the results from the I/D comparison. Stylization has been suggested as a possible confounding factor in influencing results from the ID/D comparison, and utterance duration could have influenced the equally unexpected results from the ID/I comparison. In the case of the I/D comparison, both these factors could have influenced the results. The I stimuli are different from the D stimuli in that the former have stylized intonation curves (remember from section 4.2.3 and 4.4.1 that stylization in itself lowers intelligibility), and the latter have utterance durations that were affected by the duration manipulation process (remember from section 4.2.3 and 4.4.1 that differences regarding utterance durations affect intelligibility).

The comparisons of the duration manipulated data with the intonation manipulated data were possibly contaminated by unwanted factors as discussed above. The results for the ID/D, the ID/I and the I/D comparisons (sections 4.4.2, 4.4.3 and the present section) were judged to be too problematic to make reasonably reliable inferences regarding the relative impacts of duration manipulation versus intonation manipulation, so for this reason the results from these three comparisons must be disregarded.

4.4.5 Summary

The previous sections have investigated the impacts of intonation manipulation and duration manipulation on N2 intelligibility. The results showed that both manipulations significantly enhanced N2 intelligibility when measured across all L1s, but when each L1 group was measured separately, only the English and German groups benefited from intonation manipulation, and only the French, Tamil and Persian groups benefited from duration manipulation. However, the degree to which a manipulation enhanced N2 intelligibility differed among the L1 groups. This section will address the degree to which intonation manipulation enhanced N2 intelligibility for the English and German L1 groups (measured as the intelligibility score difference between the COI and I stimuli across all listeners) and the degree to which duration manipulation enhanced N2 intelligibility for the French, Tamil and Persian groups (measured as the intelligibility score difference between the COD and D stimuli across all listeners). Figure 4.20 shows the degree to which N2 intelligibility was enhanced for the English and German L1 groups.

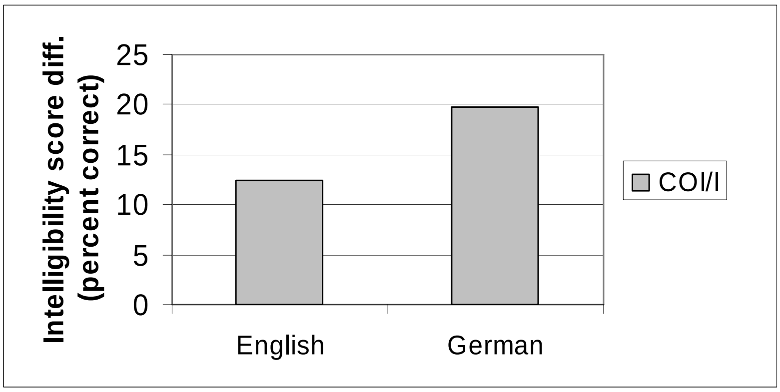

Figure 4.20: The intelligibility enhancing effect of the intonation manipulation as measured in the COI/I comparison for the English and German L1 groups. COI (n= 120) and I (n= 234) for each L1 group.

It can be seen from the figure that the difference between the intelligibility scores of the COI and I stimuli was greater (by 7.3 %) for the German than for the English L1 group. In Figure 4.21, we look at the degree to which duration manipulation enhanced N2 intelligibility for the French, Tamil and Persian L1 groups.

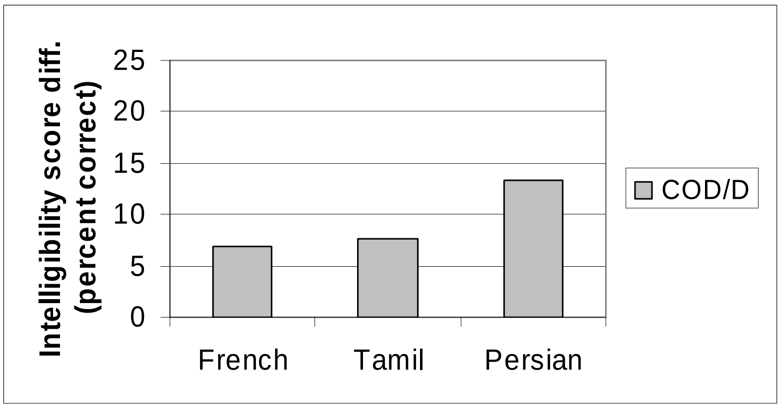

Figure 4.21: The intelligibility enhancing effect of the duration manipulation as measured in the COD/D comparison for the French, Tamil and Persian L1 groups. COD (n= 120) and D (n= 234) for each L1 group.

The figure shows that the Persian L1 group's N2 speech benefited more from the duration manipulation than the Tamil and French groups. The intelligibility score difference between the Persian and the Tamil groups was 5.7 %. The small difference between the Tamil and French groups amounted to 0.8 %.

The aim of the investigation has been to establish the relative importance of intonation manipulation compared to duration manipulation. Table 4.5 shows which manipulation most effectively enhances N2 intelligibility for the different L1 groups.

Table 4.5: The most important manipulation for intelligibility enhancement for each L1 group.

L1	**Manipulation**
English	Intonation
German	Intonation
French	Duration
Tamil	Duration
Persian	Duration
Chinese	*No manipulation effects*
Russian	*No manipulation effects*

The table shows that English and German N2 speech benefit more from intonation manipulation, whereas French, Tamil and Persian N2 benefit more from duration manipulation. None of the manipulations significantly affected the N2 intelligibility for the Chinese and Russian L1 groups.

The experimental design of the intelligibility experiment involved some listeners listening to the same sentences twice, with an initial listening session immediately followed by a second listening session. It was investigated whether learning effects could have influenced the intonation manipulation effects (section 4.4.2) and the duration manipulation effects (section 4.4.3). The general trend was for learning effects to affect intelligibility data from the second listening sessions only for the data in the COD/D comparisons, but not for the in the COI/I comparisons. The author can think of no plausible reason for this finding, so this skewed effect of learning should perhaps be regarded as due to chance.

4.5 Production analyses

The previous sections of this chapter have shown that the intelligibility of N2 speech was influenced by durational and intonational patterns in the N2 speech. Because the manipulations were carried out globally over whole utterances, the results give no information as to which *details* in the manipulations that caused the perceived effects on intelligibility. In the case of duration manipulation, some segment types might be more sensitive to durational changes than other segment types. For instance, consonant duration might be more important for N2 intelligibility than vowel duration (as was found in equivalent analyses in section 3.6 in the previous chapter). As for intonation manipulation, certain parts of the utterance could be more sensitive to intonational adjustments than other parts. For example, the intonation in stressed syllables could be more important than the intonation in unstressed syllables. This section examines how adjustments to specific details in the utterances have contributed to the perceptual effects in terms of intelligibility as observed in sections 4.4.2 and 4.4.3.

Towards the end of the previous chapter (Chapter 3, section 3.6), a similar investigation was carried out in order to relate manipulation details to manipulation effects in terms of the *degree of foreign accent*. The methodology and terminology from that investigation will also be used in the present investigation. The reader is referred to Chapter 3, section 3.6 for detailed accounts of the methodology and terminology, as they are reviewed only briefly in the following. First, details of duration manipulation are investigated in section 4.5.1 before we move on to investigate details of the intonation manipulation in section 4.5.2. The statistical details from all analyses can be viewed in Appendix D.

4.5.1 Duration

In this section, specific details of the duration manipulation are investigated for correlation with the duration manipulation effect in terms of intelligibility. These details are referred to as *factors*. The factors are in the form of vowels and consonants and will be defined in section 4.5.1.1. The term *Manipulation size* refers to the extent to which a factor (such as vowels) has been altered as a result of the duration manipulation. For each utterance, the manipulation size was calculated as the percent adjustment or alteration of the factor. The *manipulation effect* is the intelligibility score difference (as measured in the All Data group, see section 4.3) between the COD and D stimuli.

One important difference between the investigation regarding the degree of foreign accent in Chapter 3, section 3.6 and this investigation of intelligibility is that the present investigation shows learning effects for certain intelligibility data. As discussed in section 4.3 in the present chapter, the number of data affected by learning effects (i.e. data from the second listening session) varied across the stimulus sets. Here, the impact of each factor is analyzed in two regression analyses, one analysis with the manipulation effect as calculated with data *affected* by learning effects (from both listening sessions), and one analysis with the manipulation effect as calculated with data *unaffected* by learning effects (from the first listening sessions).

To sum up, for each utterance, the manipulation size of each factor is investigated for correlation with the utterance's duration manipulation effect. The expectation is that the manipulation size for a factor will correlate positively with the manipulation effect. In other words, when an utterance has been considerably manipulated, it is likely that the effect on intelligibility will also be considerable.

4.5.1.1 Factors

The factors in this investigation are identical to the factors defined in Chapter 3, section 3.6.1.1, and the reader is referred to that section for explanations of each factor, as they will only be listed in this section. However, one factor was added for the present analyses, namely pauses. As explained earlier (Chapter 2, section 2.2.1.3), pauses were intended to be left unmodified in the duration manipulation process, but in a few cases it was nonetheless necessary to shorten pauses in order to maintain the naturalness of the utterances. As there were fewer utterances in the experiment on the degree of foreign accent (14 utterances) than in the experiment on intelligibility (42 utterances), the chance of having these cases in the latter experiment was greater, and in fact, no pauses were adjusted in the utterances for the experiment on the degree of foreign accent. That is why pauses were not defined as a factor in Chapter 3, section 3.6.1.1. In the present chapter, one pause in each of 9 utterances has been shortened in order to retain naturalness. Pauses are therefore here defined as a factor.

The following should also be noted. In Chapter 3, section 3.6.1.1, the articulation rate was defined as a factor, and articulation rate is also investigated here. Note, however, that the articulation rate was adjusted only to a very limited extent for the stimuli in the present chapter as compared with the stimuli in the previous chapter. The reason is that in the investigation of the degree of foreign accent in the previous chapter, the impact of duration was measured through comparisons between the O and D stimuli. The utterance durations differed between the O and D stimuli, and consequently also the articulation rate. The articulation rate was affected in each utterance, and was thus clearly a relevant factor to investigate in section 3.6.1. In the present chapter, which investigates N2 intelligibility, the impact of duration was measured through comparisons between the COD and the D stimuli. The utterance durations were in fact similar between the COD and D stimuli and had therefore not affected the articulation rate (see section 4.2.3.1). However, pauses were adjusted in some of the D stimuli whereas the pauses in COD stimuli were left intact. When a pause was shortened in a D stimulus, this caused the same number of phonemes to occur in a shorter period of time, accelerating the articulation rate. Intelligibility score differences between those COD and D stimuli for which pauses were shortened in the D stimuli could thus have been affected by the articulation rate. Although the removal of pauses affected the articulation rate in only a few utterances (and moreover, the articulation rate was only slightly affected), the articulation rate is also defined as a factor here.

All the factors are listed below. As explained earlier, the reader should consult Chapter 3, section 3.6.1.1 for explanations of each factor.

- All segments
- All consonants
- All vowels
- All phonologically long vowels
- V/C ratio

- Articulation rate
- Pauses

The factors have been investigated for correlation with the manipulation effect by the use of regression analyses. *Multiple* regression analyses have been used only in those cases where the categories in the factors did not overlap. Only the vowels and consonants did not overlap, and these two factors were therefore investigated in a multiple regression analysis, whereas the remaining factors were investigated in separate regression analyses.

4.5.1.2 Results

The previous section explained how details in the duration manipulation, called factors, were investigated for their effect on intelligibility, called the manipulation effect. The investigation was conducted by correlating the manipulation size for a factor with the manipulation effect for the particular utterance. The first factor that was investigated was the overall durational adjustment across all segments. The expectation is that the size of overall durational adjustment of an utterance should correlate with the size of the perceptual effect for that utterance. Figure 4.22 shows the manipulation size across all segments for each utterance (vertical bars) related to each utterance's manipulation effect (graph). A trend line has been drawn for the manipulation size. The manipulation effect in the figure is based on data without learning effects.

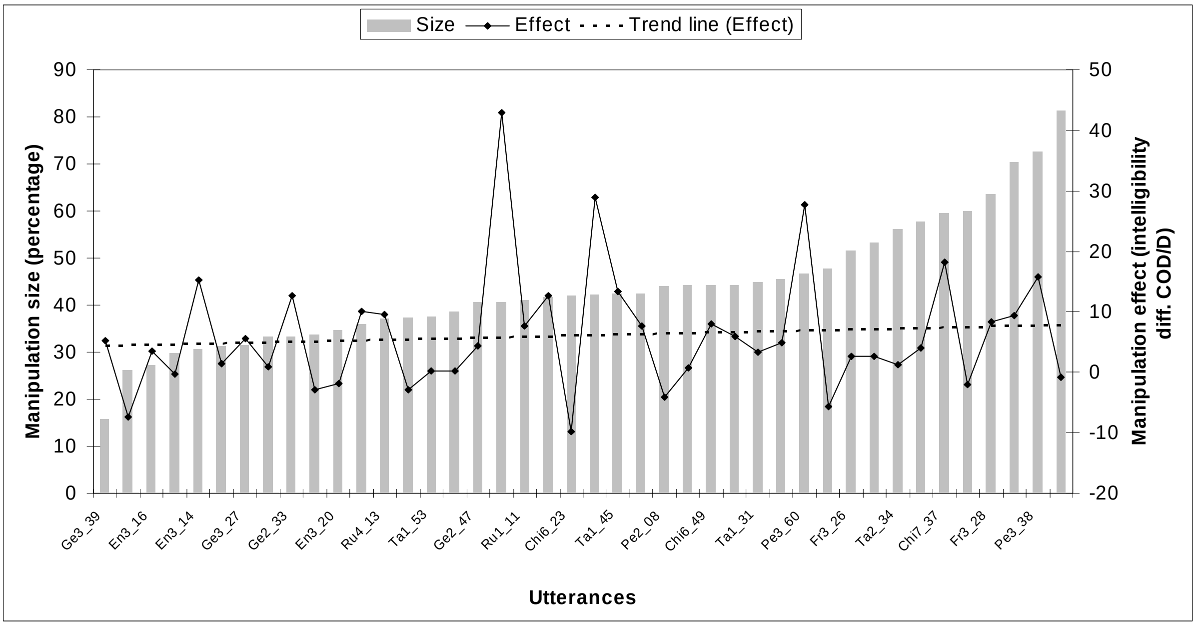

Figure 4.22: Manipulation size and manipulation effect across all segments for each utterance (n= 42). Data in ascending order according to manipulation size. A trend line was drawn for the manipulation effect. Positive numbers mean that the D stimulus is more intelligible than the COD stimulus.

In Figure 4.22, the y-axis showing the manipulation effect (on the right) is expressed with both negative and positive numbers. The positive numbers show that the D stimulus had a higher intelligibility score than the COD stimulus, and vice versa, the negative numbers show that the COD stimulus had a higher intelligibility score than the D stimulus. The figure gives the impression that the relation between the manipulation size and the resulting manipulation effect was random. The flat trend line indicates that the manipulation effect was unaffected by manipulation size. A regression analysis with a duration adjustment across all segments as the predictor variable, and the manipulation effect as measured with learning effects as the dependent variable (Table 1, Appendix D), showed that there was in fact no correlation. An equivalent analysis with the manipulation effect based on data free of learning effects (Table 2, Appendix D), also showed no correlation. This way of measuring durational adjustments across segments thus yielded no effect. A multiple regression analysis with consonants and vowels as predictor variables showed that there were no correlations between any of these factors and the manipulation effect, regardless of whether the analyses were carried out with the manipulation effect based on data *with* (Table A) or *without* learning effects (Table B). Further regression analyses were conducted to investigate each of the remaining factors (phonologically long vowels, V/C ratio and articulation rate) as predictor variables, but none of the analyses showed any correlation between the factor and the effect, whether the dependent variable (manipulation effect) comprised data *with* or *without* learning effects.

The investigation above thus showed that none of the defined factors correlated with the manipulation effect. However, the investigation so far has been carried out for all L1s. Remember from the summary in section 4.4.5 that the duration manipulation significantly enhanced N2 intelligibility only for three L1 groups, namely the French, Tamil and Persian groups, suggesting that correlation effects might be found if only these three L1 groups are analyzed. New regression analyses were therefore conducted with data for these L1 groups only, but again, no correlations were found, regardless of the factor investigated and regardless of whether the manipulation effect was based on data *with* (Table 3, Appendix D) or *without* (Table 4, Appendix D) learning effects [8].

No correlations have thus been found between the manipulation sizes of the various factors and manipulation effects. In the equivalent investigation in Chapter 3, section 3.6.1, effects had in fact been found for the durations of consonants and for the articulation rate. The reason why no effects have been found for the articulation rate in the present investigation could be because the articulation rate was adjusted in only a few utterances, and only to a limited extent within these utterances, as explained in the previous section. Another reason could be that as the articulation rate increases in *natural speech*, coarticulation effects also increase. The increased coarticulation provides the listener with perceptual cues to the identity of a particular segment well beyond that segment's boundaries. Kühnert & Nolan (1999) suggested that the increased coarticulation observed in fast speech may actually be necessary in order to perceive it. A third possible reason why no intelligibility effects were found for the accelerated articulation rate in the duration manipulated stimuli could be that native listeners require more time to process non-native speech (Munro & Derwing, 1995b). Lastly, one could imagine a combination of the two latter reasons such that a native listener needs more coarticulatory aid and more time to successfully perceive foreign accented speech (which is in general lacking in redundancy) when it is also produced at a fast rate.

4.5.2 Intonation

In the previous sections, durational details, called factors, were investigated for their impacts on N2 intelligibility. However, no effects were found. In this section, the roles of intonational details are investigated.

[8] In the previous chapter, additional analyses have *not* been carried out within the L1 groups that were significantly accent reduced due to the duration manipulations. This was because there were only 2 utterances per L1 in that chapter, and so there were too few data for meaningful statistical testing.

Section 3.6.2 in the previous chapter described intonation analyses equivalent to those that will be carried out in the present section. In that section, it was explained why a phonetic approach was chosen over a phonological approach. The reader is referred to that section in order to review those explanations as they will not be repeated here. The phonetic analysis method described in that section will also be used here. This section continues the use of the terms *factors*, *manipulation size* and *manipulation effect*: The intonational details are called factors. The extent to which a factor has been adjusted in the intonation manipulation process is called the manipulation size. The manipulation effect is the intelligibility score difference between the COI and the I stimuli (measured with All Data, see section 4.3), as observed in section 4.4.2. For each utterance, the manipulation size of each factor is investigated for correlation with the utterance's intonation manipulation effect. The factors have been defined and measured in the same way as in Chapter 3, section 3.5.2. The reader is referred to that section for thorough explanations of the factors, as they will only briefly be repeated here. The intonational N1-N2 deviations were examined in the form of F0 slopes and F0 direction in units that will be described in the next section.

4.5.2.1 Factors

In line with the procedures in the previous chapter section 3.6.2, three 2-syllable content words from each utterance were analysed in terms of N1-N2 F0 slope and F0 direction deviations. Each word was segmented into two syllables. Figure 4.23 (repeated from Chapter 3, section 3.6.2.1) illustrates the segmentation of a word into two syllables.

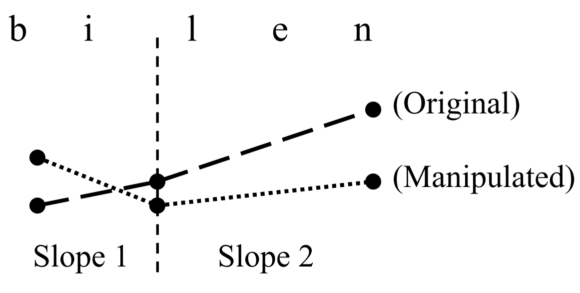

Figure 4.23: Schematic representations of an N2 original contour (dashed line) and the corresponding N2 intonation manipulated contour (dotted line) in one of the three words (bilen= "the car") selected for the intonation investigation. F0 was measured at three points (black dots) in the word, defining two slopes.

Three coordinates in each word were measured. The first and last coordinates define the beginning and end of the word, whereas the middle coordinate corresponds to the turning point of the particular intonation curve. The F0 slope of each syllable was then measured (semitones per second). For each utterance, 6 slopes (across the 3 words) were measured. The measurements were undertaken across the 6 syllables, and within each of the 6 syllables. Measurements were also carried out between the onsets of words 1 and 2, and between the onsets of words 2 and 3 as illustrated in Figure 4.24 (repeated from Chapter 3, section 3.6.2.1).

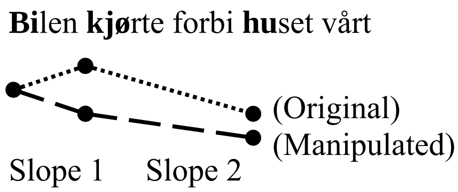

Figure 4.24: Schematic representations of an N2 original contour (dashed line) and the corresponding N2 intonation manipulated contour (dotted line) for the sentence "Bilen kjørte forbi huset vårt" (*The car drove past our house)*. F0 was measured at three points (black dots) over the utterance corresponding to the beginning of each of the three selected words.

Figure 4.24 illustrates the measurement between the first syllables of word 1 and word 2, and between the first syllables of word 2 and word 3.

The list below summarizes the units in which the F0 slope and the F0 direction were measured:

- Across the six syllables
- word 1, syllable 1
- word 1, syllable 2
- word 2, syllable 1
- word 2, syllable 2
- word 3, syllable 1
- word 3, syllable 2
- Between word 1 and word 2

- Between word 2 and word 3

At this point, the following difference between the stimuli in Chapter 3 and the stimuli in the present chapter should be pointed out. In Chapter 3, the utterances were the *same sentence*, whereas in the present chapter, the utterances were *different sentences*. For the intonation analyses in Chapter 3, section 3.6.2, the same three words were investigated across all utterances. Because each utterance in the present investigation was a different sentence, three different words were investigated for every utterance. In a few utterances, it was not possible to find as many as three 2-syllable content words. In such cases, measurements were carried out over two 1-syllable words. The following sentence serves as an example: “De jaget sauene langt vekk” (*They chased the sheep far away*). In this sentence, intonation changes were measured in the three units “jaget” (*chased*), “sauene” *(the sheep)* and “langt vekk” (*far away*).

The analyses were carried out in two steps: Step A and Step B. In Step A, the adjustment of each factor in terms of slope steepness was investigated for correlation with the manipulation effect for each utterance. In the intonation manipulation process, some slopes had been adjusted not only in terms of slope steepness, but also in direction (from falling to rising or from rising to falling). In Step B, the slopes for which the direction was altered were weighted by multiplication with an arbitrary factor (a factor of 2). The factors analyzed in Step B were therefore the same slope measures analyzed in Step A, but the slopes that were altered in direction received a weighting. Step A and Step B therefore investigated impacts of adjustments to the steepness of the slopes and the directions of the slopes.

4.5.2.2 Results

In the following, analyses corresponding to Step A of the investigation (only slope steepness adjustments) will be discussed first, before we move on to Step B of the investigation (in which adjustments in terms of slope direction have been included). Figure 4.25 shows the manipulation size across the 6 intonation slopes (columns) related to the manipulation size (graph) for each sentence. The data are in ascending order according to manipulation size, and a trend line has been drawn for the manipulation effect.

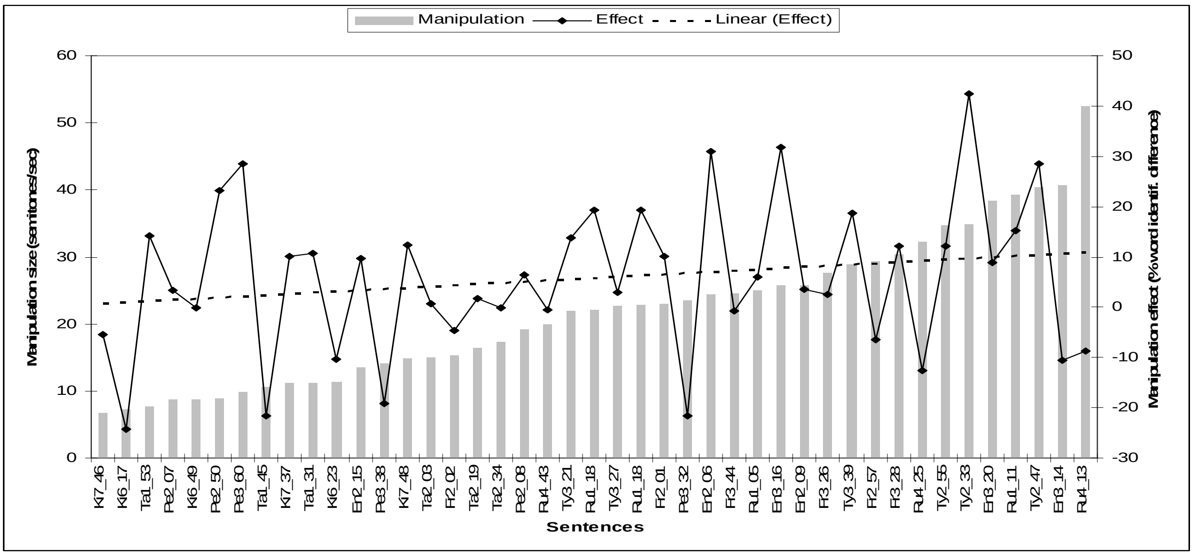

Figure 4.25: Manipulation size across 6 syllables and manipulation effect for each utterance (n= 42). Data in ascending order according to manipulation size. A trend line was drawn for the manipulation effect. Positive numbers mean that the I stimulus was more intelligible than the COI stimulus.

Figure 4.25 does not give the impression of any linear relation between the size of the manipulation and the size of the manipulation effect. The trend line for the effect (dotted line) is fairly horizontal and therefore suggests that the effect of the manipulation stays unaffected by the size of manipulation. Two analyses were carried out in order to see whether the size of overall intonational adjustment correlated with the size of the manipulation effect. Both regression analyses had the manipulation size across the 6 slopes as the predictor variable, but the manipulation effect was *with* learning effects in the first analysis (Table 5) and *without* learning effects in the second analysis (Table 6). No correlation effects were found in these tests. Next, one multiple regression analysis was performed with each of the 6 slopes as predictor variables and the manipulation effect *with* learning effects as the dependent variable (Table 5). This showed no effect. Another multiple regression analysis with each of the 6 slopes as predictor variables and the manipulation effect *without* learning effects was carried out (Table 6). This test revealed a correlation between the degree of slope adjustment in the second syllable of the third word (word 3, slope 2) and the size of the intonation manipulation effect. However, no effects were found when slope adjustments between words 1 and 2 and between words 2 and word 3 were investigated (Tables 5 and 6).

In section 4.4.2, it was found that only the English and German L1 groups benefited significantly from intonation manipulation. Analyses equivalent to those in the previous paragraph were therefore carried out only within the data pooled across these two L1 groups. These analyses showed correlation effects for two slopes: For word 3, slope 2 (Beta= 0.658,

$p < 0.05$), when the manipulation effect was calculated *without* learning effects (Table 6), and for word 1, slope 2 (Beta= -0.593, $p < 0.05$) when the manipulation effect was *with* learning effects (Table 7). The former correlation had also been found in one of the analyses across L1s (previous paragraph), but in that analysis, the manipulation effect was calculated without learning effects.

The analyses performed so far correspond to Step A of the investigation, comprising only adjustments to slope steepness. Step B analyses were also conducted, in which data had been weighted for slope direction adjustment (see previous section). An investigation equivalent to that in Step A was performed, investigating correlation effects for all factors, both across L1s (Tables 9 and 10) as well as across the English and German L1 groups (Tables 11 and 12), and both with dependent variable *with* (Tables 9 and 11) and *without* (Tables 10 and 12) learning effects, but no correlation effects were found with any of these analyses.

This section has shown that the intonation manipulation effects in terms of enhanced intelligibility was due to F0 slope changes in the second syllable of the first word and the second syllable of the third word. These are unstressed syllables. It may seem counterintuitive that there should be effects for the *unstressed* syllables and not for the *stressed* syllables, seeing as the stressed syllables are more perceptually salient and initiate the accent phrase's accent contour in the N1 template, but remember that the manipulated utterances are foreign-accented, and the phonetic realizations of both stress and accent are likely to have been produced in deviant manners in these utterances.

4.6 Similarity between speakers

The intelligibility experiment was based on utterances from two speakers from each of 7 L1 groups. In this section, the two speakers from the same L1 will be called a *speaker pair*. The *manipulation analyses* (section 4.4) were carried out with perceptual data pooled across the two speakers from each speaker pair, whereas the *production analyses* (section 4.5) investigated correlations between manipulation size and manipulation effect for each individual utterance. This section brings together information from the manipulation analyses and the production analyses in order to discuss the degree of similarity between the two speakers from the same L1.

Similarity between the speakers in each speaker pair will be assessed by comparing manipulation size (N1-N2 production deviance), manipulation effect (intelligibility enhancement from the manipulations) as well as relative impact of the manipulations for each of the two speakers (the manipulation that most affected the intelligibility). Similarity in terms of relative impact of the manipulations is of particular interest because the two speakers in a speaker pair must be similar in this respect if the results from the manipulation analyses can be considered to reflect effects typical for each particular L1 group.

In the degree of accent-experiment in the previous chapter, one utterance of the same sentence was used from each of the 14 speakers. In the intelligibility experiment in this chapter, each speaker uttered three different sentences. In this experiment there were thus a total of 42 utterances. Figure 4.26 shows the durational and intonational manipulation sizes expressed as means across the three utterances for each speaker.

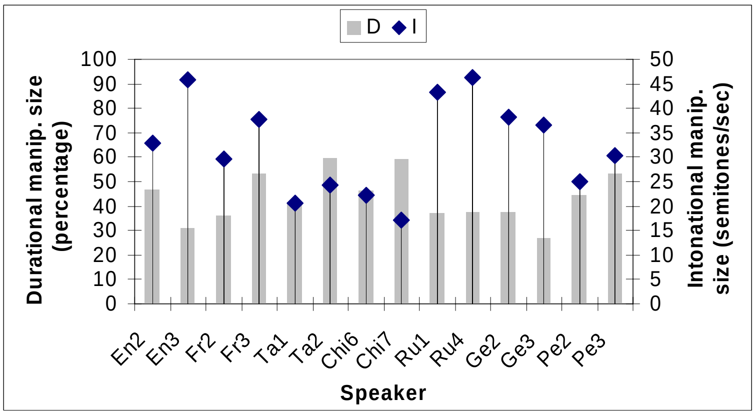

Figure 4.26: Mean duration (grey bars) and intonation (black lines with squares) manipulation size for each speaker.

The figure shows that there was some duration manipulation size difference between the speakers in each speaker pair. The largest differences were within the English and French speaker pairs, whereas the greatest consistency was within the Russian speaker pair. There were also intonation manipulation size differences. Again, the largest differences were within the English and the French speaker pairs. The Russian and German speaker pairs were very similar regarding intonation manipulation size. As explained above, Figure 4.26 presents means across three utterances for each speaker. With as few as three utterances per speaker to choose between, it is in fact possible to find pairs of utterances that show great similarity between the two speakers, and likewise it is possible to find pairs of utterances that show great difference within each speaker pair. Figure 4.27 shows utterances selected to show *similarity* in terms of manipulation size for the two speakers in each speaker pair. Figure 4.28 selects utterances that show inter-speaker *differences* for each speaker pair. Because it is not the same utterances that show similarity for both duration and intonation, and it is not the same utterances that show difference for both duration and intonation, two x-axes are used in each of the figures: The bottom x-axes show the durational manipulation size with values on the left-hand y-axes; the top x-axes show the intonational manipulation size with values on the right-hand y-axes.

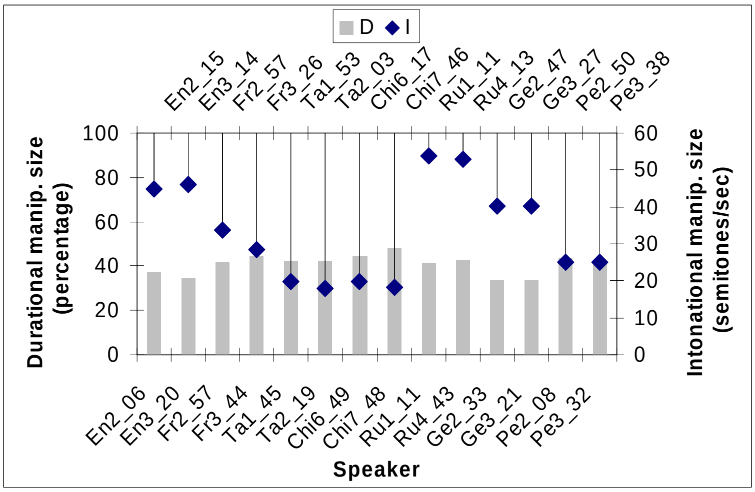

Figure 4.27: Selected utterances that show duration (grey bars) and intonation (black lines with squares) manipulation size *similarity* within each speaker pair. The utterances that show durational similarity are on the primary x-axis (bottom) with values along the left-hand y-axis. The utterances that show intonational similarity are on the secondary x-axis (top) with values along the right-hand y-axis.

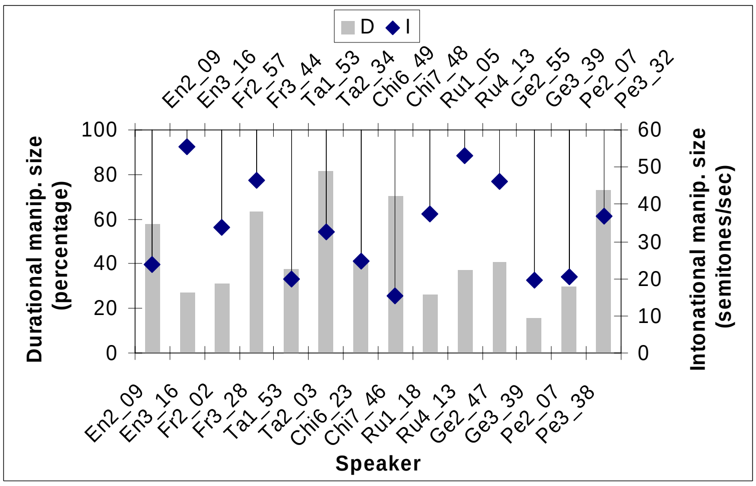

Figure 4.28: Selected utterances that show duration (grey bars) and intonation (black lines with squares) manipulation size *difference* within each speaker pair. The utterances that show durational similarity are on the primary x-axis (bottom) with values along the left-hand y-axis. The utterances that show intonational similarity are on the secondary x-axis (top) with values along the right-hand y-axis.

Based on Figure 4.27, one would have to say that the speakers within each speaker pair were very similar in terms of both durational and intonational manipulation size. In contrast, the selected utterances in Figure 4.28 give the impression that the speakers in each speaker pair were very different regarding durational and intonational manipulation size. The fact that such great similarity as well as such great difference can be shown by choosing between merely three utterances for each speaker shows the variation in production across the speakers' different utterances. This variation could be due to the fact that each utterance is a different sentence, thus representing different information structures, different segmental compositions, and different intonation contours, but there could also be speaker-internal factors. It seems likely that there should be speaker-internal effects because L2 learners have more unstable and less robust perceptual representations than do native speakers. This could be reflected in unstable productions with a constantly varying approximation to the target pronunciation.

We turn now to assess speaker similarity in terms of manipulation *effects*. Figure 4.29 shows manipulation effects as means across the three utterances for each speaker. The values for the duration manipulation effect run along the left-hand y-axis, while the values for the intonation manipulation effect run on the right-hand y-axis.

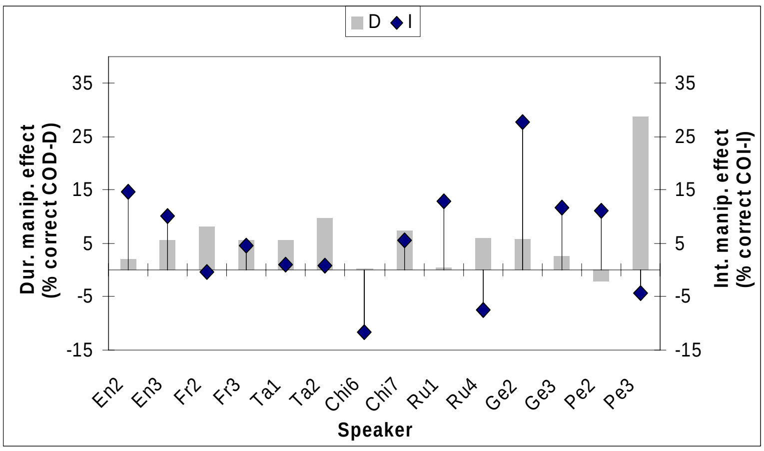

Figure 4.29: Mean durational (grey bars) and intonational (black lines with squares) manipulation effect for each speaker. Duration manipulation effect values are on the left-hand y-axis; intonation manipulation effect values are on the right-hand y-axis.

In terms of manipulation effects, it is difficult to make assessments about general similarity or general difference between the speakers in each speaker pair based on the information in Figure 4.29. There seems to be much variation. There was however great similarity between the Tamil speakers regarding intonational manipulation effect. In the previous paragraph it was shown that it was possible to select a pair of utterances that showed similarity and a pair of utterances that showed difference between the speakers in each speaker pair. It was not quite as easy to find such matching utterances regarding manipulation effect for absolutely all speaker pairs. This indicates that the speakers were generally more similar in terms of production than in terms of manipulation effect. However, *fairly* similar utterances for most of the speaker pairs are shown in Figure 4.30, and different utterances are shown in Figure 4.31. Regarding the figures, note that some of the effects were very small, and for that reason they do not show well in the figures. For instance, in Figure 4.30, the durational manipulation effect for Pe2_07 was -0.30, and in Figure 4.31, the duration manipulation effect was 0.22 for Ta1_53, 0.74 for Ru1_05, and 0.20 for Pe2_50. Negative numbers show that the manipulated utterance had a lower intelligibility score than the close-original utterance.

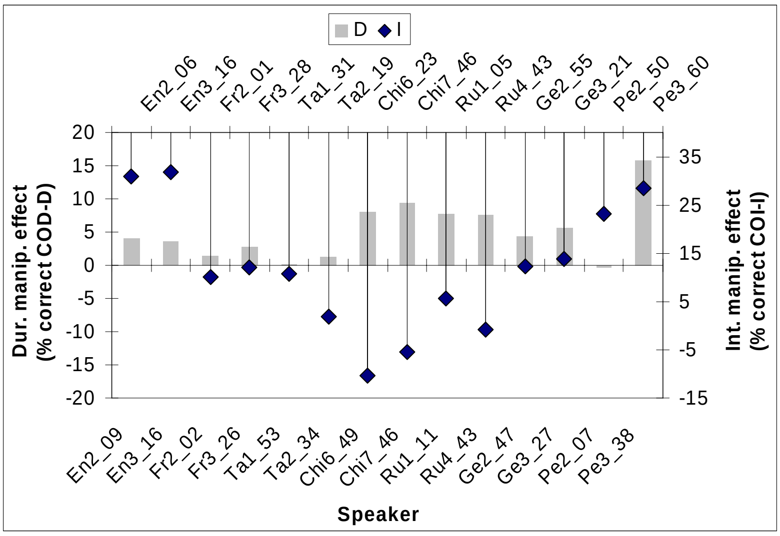

Figure 4.30: Selected utterances that show duration (grey bars) and intonation (black lines with squares) manipulation effect *similarity* within each speaker pair. The utterances that show durational similarity are on the primary x-axis (bottom) with values on the left-hand y-axis. The utterances that show intonational similarity are on the secondary x-axis (top) with values on the right-hand y-axis.

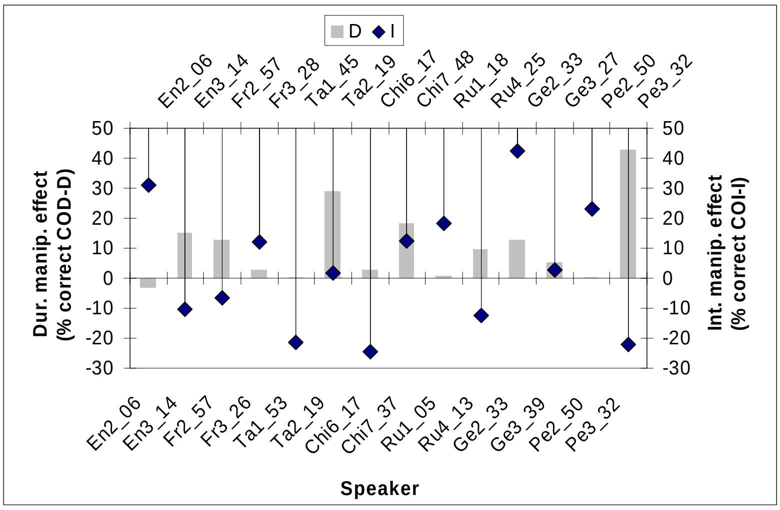

Figure 4.31: Selected utterances that show duration (grey bars) and intonation (black lines with squares) manipulation effect *difference* within each speaker pair. The utterances that show durational similarity are on the primary x-axis (bottom) with values on the left-hand y-axis. The utterances that show intonational similarity are on the secondary x-axis (top) with values on the right-hand y-axis.

Figure 4.30 gives the impression that there were fairly similar manipulation effects across the speakers in each speaker pair. Regarding durational manipulation size, there was good consistency within all speaker pairs except within the Persian speaker pair, for which a large inter-speaker difference can be seen. In fact, the effect for Pe2_07 was slightly negative (-0.30), reflecting that the mean COD intelligibility score was somewhat higher than the mean D intelligibility score for this sentence. (All three utterances for Pe2 had small negative effects, and all utterances for Pe3 had large positive effects. The difference between Pe2 and Pe3 shown in Figure 4.30 was actually the smallest difference between these speakers). As for intonation manipulation effect, the English, French and German pairs have the most similar effects. In Figure 4.31, the utterances were selected to show differences between the speakers in each speaker pair. When comparing Figures 4.30 and 4.31, we see for instance that the two English N2 utterances in Figure 4.30 have virtually identical intonation manipulation effects, but that the English N2 utterances in Figure 4.31 have widely different intonation manipulation effects, even to the extent that for one sentence the COI stimulus has a higher mean intelligibility than the I stimulus. The German speakers' utterances in Figure 4.30 show similar intonation manipulation effects, but their utterances in Figure 4.31 show a very large difference in the degree of the intonation manipulation effect.

So far in this section, we have assessed the degree of similarity between the speakers in each speaker pair in terms of manipulation size, or in other words the N1-N2 degree of production deviation, and in terms of manipulation effect. We have seen that it is possible to select utterances that show a very high degree of inter-speaker consistency regarding manipulation size, but that it is more difficult to select utterances that show great inter-speaker consistency regarding manipulation effect for all speaker pairs. This indicates that the speakers were more similar in terms of N1-N2 production deviation than in terms of manipulation effect. When comparing the manipulation size with the manipulation effect, one would expect a large manipulation size to cause a large manipulation effect and a small manipulation size to cause a small manipulation effect. When we compare the mean manipulation size (Figure 4.26) with the mean manipulation effect (Figure 4.29), we see for example that the German Ge2 had a large intonation manipulation size, and that the effect of this manipulation was also large. However, Ge3 had an equally large intonation manipulation size, yet for him the effect was much smaller. Further, both Russian speakers' utterances were extensively intonation manipulated. Yet, the effect of the manipulation was small for Ru1, and even negative for Ru4 (*negative* means that COI had a higher intelligibility score than I). Remember from the past two sections (4.5.1 and 4.5.2) that correlation analyses were carried out in which no correlation was found between manipulation size and manipulation effect. It is therefore not the case that the degree of N1-N2 deviation predicts the degree of the manipulation effect.

This chapter has investigated the relative effects of intonation and duration on intelligibility. Because it was assumed that utterances from speakers sharing the same L1 would gain most from the same manipulation, the data was pooled across the two speakers from the same L1. It is therefore important to investigate whether the two speakers within each speaker pair were actually similar in this respect. If not, there are certain implications for the interpretation of the manipulation effect analyses earlier in the chapter. Table 4.6 shows the mean intelligibility enhancement in the COD-D and the COD-I comparisons for each utterance. Asterisks mark the speaker pairs in which the speakers showed opposite manipulation effects.

Table 4.6: Mean intelligibility score for each utterance in the COD_D and COI_I comparisons. Asterisks show the speaker pairs that were *not* similar in terms of relative impact of the manipulations.

Speaker	Utterance	D effect	I effect	Mean D effect	Mean I effect	Manipulation
En2	En2_06	-2.94	30.95			
	En2_09	3.96	3.52			
	En2_15	4.89	9.63	1.97	14.70	Intonation
En3	En3_14	15.26	-10.51			
	En3_16	3.49	31.75			
	En3_20	-1.79	8.94	5.65	10.06	Intonation
Fr2	Fr2_01	10.07	10.13			
	Fr2_02	1.45	-4.72			
	Fr2_57	12.72	-6.41	8.08	-0.33	Duration
Fr3	Fr3_26	2.70	2.47			
	Fr3_28	8.36	12.09			
	Fr3_44	5.91	-0.77	5.66	4.60	Duration
Ta1	Ta1_31	3.34	10.66			
	Ta1_45	13.39	-21.54			
	Ta1_53	0.22	14.20	5.65	1.11	Duration
Ta2	Ta2_03	-0.84	0.73			
	Ta2_19	28.89	1.79			
	Ta2_34	1.28	-0.18	9.78	0.78	Duration
Chi6	Chi6_17	2.66	-24.32			
	Chi6_23	-9.83	-10.33			
	Chi6_49	8.08	-0.10	0.30	-11.59	Duration
Chi7	Chi7_37	18.20	10.05			
	Chi7_46	9.30	-5.40			
	Chi7_48	-5.72	12.30	7.26	5.65	Duration
Ru1 *	Ru1_05	0.74	5.60			
	Ru1_11	7.68	14.84			
	Ru1_18	-7.42	18.21	0.33	12.88	Intonation
Ru4 *	Ru4_13	9.50	-9.08			
	Ru4_25	0.92	-12.45			
	Ru4_43	7.58	-0.70	6.00	-7.41	Duration
Ge2	Ge2_33	12.71	42.35			
	Ge2_47	4.36	28.52			
	Ge2_55	0.20	12.22	5.76	27.70	Intonation
Ge3	Ge3_21	-2.97	13.78			
	Ge3_27	5.64	2.91			
	Ge3_39	5.16	18.64	2.61	11.78	Intonation
Pe2 *	Pe2_07	-0.30	3.40			
	Pe2_08	-4.07	6.52			
	Pe2_50	-2.02	23.18	-2.13	11.03	Intonation
Pe3 *	Pe3_32	42.84	-22.12			
	Pe3_38	15.74	-19.19			
	Pe3_60	27.71	28.53	28.76	-4.26	Duration

The fifth and sixth columns from the left show each speaker's mean effect from the duration manipulation and intonation manipulation respectively, and the rightmost column shows the manipulation that most affected each speaker's intelligibility. The table shows that when the data are pooled across each speaker's three utterances, all the speaker pairs showed inter-speaker consistency regarding which manipulation that most affected the intelligibility, except for the Russian speaker pair in which Ru1 was most affected by intonation and Ru4 was most affected by duration, and the Persian speaker pair in which Pe2 gained most from intonation and Pe3 gained most from duration. These speaker pairs are marked with asterisks in the table. The case of the Russian and Persian speaker pairs will now be discussed in more detail.

Ru1 was most affected by intonation, whereas Ru4 was most affected by duration. All three utterances from Ru4 received *lower* scores when they were intonation manipulated. In contrast, all three utterances from Ru1 received *higher* scores from the intonation manipulation. Remember from Figure 4.26 that the two Russian speakers were very similar in terms of intonational manipulation size, and that this manipulation affected Ru1 positively and Ru4 negatively as shown in Figure 4.29. The information in the table shows that the tendency was the same for all three utterances from each speaker. The situation is parallel for the Persian speaker pair. Pe2 gained most from intonation manipulation, while Pe3 gained most from duration manipulation. All Pe2's utterances received lower scores from the duration manipulation, whereas all Pe3's utterances received higher scores from this manipulation. As for the intonation manipulation, it lowered the scores for two of Pe3's utterances, while it yielded higher scores for all three utterances from Pe2. Remember from Figure 4.26 that there weren't any large differences between the two Persian speakers regarding either durational or intonational manipulation size, but that the duration manipulation effect was negative for Pe2 and positive for Pe3, and that the intonation manipulation effect was positive for Pe2 and negative for Pe3. The information in the table shows that these opposite manipulation effects between the two speakers were present in all utterances (except Pe3_60 which was positively affected by the intonation manipulation).

Note that similar assessments with different utterances[9] for each speaker were carried out in section 3.6.3 in the previous chapter, and in that section inter-speaker conflict was also shown only within the Russian speaker pair and within the Persian speaker pair.

[9] The exception is Ru1_05 which was used in both the degree of accent- experiment in the previous chapter and in the intelligibility experiment in the present chapter.

There were thus two speaker pairs which showed inter-speaker conflict regarding the manipulation that most affected intelligibility. For both the Russian speaker pair and the Persian speaker pair, the two speakers were in almost perfect opposition in the sense that all three utterances from one speaker showed the opposite effect from all three utterances from the other speaker (the exception was Pe3_60). This could indicate that there may be intra-speaker consistency regarding the relative effect of the manipulations, such that e.g. certain Russian N2 speakers consistently gain most from duration manipulation whereas certain other Russian N2 speakers gain most from intonation manipulation. However, the results for the individual utterances from the speakers from the *other* L1 groups go against this hypothesis because many of these speakers indicate intra-speaker *in*consistency. For example, in the French speaker pair, each speaker has one sentence for which intonation is most important, and they each have two utterances for which duration is most important. A similar example is provided by the Tamil speaker pair in which Ta1 has two utterances that gain most from intonation, and one utterance that gains most from duration, whereas Ta2 has one utterance that gains most from intonation and two utterances that gain most from duration.

With the limited number of utterances per speaker, it is impossible to make very firm assessments of whether there are consistent differences between speakers (which is indicated by the intra-speaker consistency within the Russian speaker pair and within the Persian speaker pair), or whether there is inconsistency within each speaker (as indicated by the intra-speaker inconsistency observed for example in the French and Tamil speaker pairs). A larger number of speakers per L1 together with a larger number of utterances per speaker would be necessary in order to examine these matters more thoroughly.

5. Summary, discussion and conclusions

This thesis has presented two experiments investigating the relative contributions of durational and intonational aspects to perceived degree of foreign accent (Chapter 3) and intelligibility (Chapter 4). This final chapter summarizes and discusses the results from the experiments. Section 5.1 provides a summary of the speech corpus and stimulus generation, section 5.2 summarizes and discusses the degree of foreign accent-experiment, and section 5.3 summarizes and discusses the intelligibility experiment. Note that the summaries will only comprise aspects judged essential to obtain an overview of the experiments and their results. For instance, certain aspects regarding the experimental design of the intelligibility experiment will be omitted. Section 5.4 extracts general conclusions, section 5.5 discusses problems regarding manipulation encountered in the course of the investigation, and finally, section 5.6 provides suggestions for future research.

5.1 Summary of stimulus generation

The aim of the present investigation was to establish the relative contributions of durational and intonational aspects of N2 speech to native listeners' perceptions in terms of degree of foreign accent (Chapter 3) and in terms of intelligibility (Chapter 4). The speech corpus consisted of Norwegian sentences read by 14 N2 speakers. There were 2 speakers from each of the 7 L1 groups English, French, Tamil, Chinese (Mandarin), Russian, German and Persian. One Norwegian speaker was used as a native Norwegian template. This native speaker had a Southeast Norwegian dialect, the dialect which traditionally has represented the unmarked version of spoken Norwegian. The Norwegian speaker had read exactly the same sentences as the N2 speakers, which made it possible to transfer the durational and intonational patterns from each N1 utterance to each corresponding N2 utterance. As for the duration manipulation, each N2 segment was lengthened or shortened to match the corresponding segment in the N1 utterance. As for the intonation manipulation, the N1 intonational contour was stylized and superimposed onto the corresponding N2 utterance. The N2 utterances were also manipulated regarding *both* duration and intonation. Three manipulated stimuli were thus generated. The original N2 utterances were also used as stimuli in the experiments. The stimuli were called original (O), duration manipulated (D), intonation manipulated (I), and intonation-duration manipulated (ID). These abbreviations will be used in the following.

5.2 The degree of accent-experiment

5.2.1 Summary

An experiment was conducted in order to investigate the contributions of intonation and duration to native listeners' judgments of degree of foreign accent. For this experiment, only one read sentence (Bilen kjørte forbi huset vårt= *The car drove past our house*) was used in its original and various manipulated forms. Pairs of stimuli were put together in sound files with a short pause between each single stimulus. This stimulus pairing enabled the listeners to judge the difference in degree of accent between the two stimuli. Stimulus order in the stimulus pairs was balanced. 13 native Norwegian listeners participated in the experiment. The listeners' task was to judge which stimulus in the stimulus pair featured less of a foreign accent than the other.

The results from this experiment showed firstly that the combined manipulation of duration and intonation in the ID stimulus significantly reduced foreign accent for all seven L1 groups. The results also showed that almost all L1 groups significantly benefited from both the duration manipulation and the intonation manipulation. This was true for all L1 groups except the English L1 group which was not affected by the intonation manipulation, and the German L1 group which was not affected by the duration manipulation. There were differences between the different L1 groups regarding the magnitude of the effects.

Subsequent analyses investigated which aspects of the duration manipulation had caused the accent reductions. The aspects that were analysed regarding durational adjustment were a) all segments, b) all consonants, c) all vowels, d) all phonologically long vowels, e) V/C ratio and f) articulation rate. The results from these analyses showed that the adjustment of consonant durations and the adjustment of articulation rate were responsible for the accent reduction effects. The effect for the articulation rate was such that a faster rate was associated with less foreign accent.

Analyses were also carried out to investigate which aspects of the intonation manipulation had caused the accent reduction effects. Manipulation-induced changes regarding F0 slope and F0 direction were analysed in the three stressed content words in the utterance: **Bilen kjørte** forbi **huset** vårt = ***The car drove*** *past our* ***house***. The intonational changes were analysed a) across the 6 syllables (two syllables in each of the three words), b) in each of the 6 syllables and c)

between the onsets of each stressed syllable in the utterance (BI-len KJØ-rte forbi HU-set vårt). None of these measures were found to correlate with the intonation manipulation effects. It was suggested that listeners may judge the degree of foreign accent holistically across an utterance such that local intonational changes are perceptually relevant only cumulatively.

The degree of accent-experiment was carried out with data pooled across the two speakers representing the same L1, called *speaker pairs*, because it was assumed that speakers from the same L1 would be similar regarding the manipulation that most affected their N2 degree of accent. At the end of the chapter on degree of foreign accent, investigations were carried out in order to find out whether the two speakers in each speaker pair were in fact similar to each other. Similarity was assessed in three ways: as the degree of N1-N2 production deviation, the magnitude of the manipulation effect, and the manipulation that most affected the degree of accent. In general, the speakers were more similar in terms of N1-N2 production deviation than in the magnitude of the manipulation effects. The most interesting type of similarity was the relative importance of the manipulations because if the two speakers were not similar in this respect, this would have certain implications for the manipulation effects measured with data pooled across the two speakers. Table 5.1 shows the manipulation that most affected the N2 degree of accent for each of the L1 groups when the data was pooled across the two speakers, and it also shows the two L1 groups in which the N2 speech had gained most from *different* manipulations across the two individual speakers (marked with *).

Table 5.1: The manipulation that most affected the degree of accent for each L1 group as measured with data pooled across the two speakers from each speaker pair. The L1 groups in which the speakers' N2 speech gained most from different manipulations are marked with asterisks.

L1	**Most important manipulation**
French	Intonation
German	Intonation
English	Duration
Tamil	Duration
Chinese	Duration
Persian *	Equally important
Russian *	Equally important

The table shows that for the L1 groups Persian and Russian, the results pooled across the two speakers had indicated that the two manipulations were equally important for the degree of foreign accent. The table also shows that in each of these two speaker pairs, the individual speakers had gained most from different manipulations. These opposite effects between the two speakers in each speaker pair was the reason why it had not been possible to identify one manipulation as more important than the other when the data had been pooled across the two listeners. For one of the French speakers, the two manipulations had affected his accent to the same degree. For the French L1 group, the superior role of intonation for accent reduction was therefore due to an effect present for only one speaker. Moreover, because there were only two speakers per L1, one can not dismiss the possibility that the inter-speaker consistency regarding the relative impact of the two manipulations in the remaining speaker pairs could be due to chance. The fact that there were opposite effects for the two Persian speakers and for the two Russian speakers, makes the interpretation of the results from the degree of accent-experiment difficult.

5.2.2 Discussion

A summary of the degree of accent-experiment was presented above. In this section, the findings from this experiment will be related to the findings in the literature as presented in the introductory chapter (section 1.2.2, Chapter 1). It is difficult to relate the findings from this investigation to the findings in the literature, mainly because of differences regarding the languages involved, in other words the target language and the L1. This problem was pointed out in the introductory chapter (section 1.3.2, Chapter 1), and now seems even more relevant in the light of this investigation's finding that speakers from different L1 groups gain most from different manipulations. For instance, the English N2 gained most from the duration manipulation while the German N2 gained most from the intonation manipulation.

The results from this investigation showed that some L1 groups gained most from duration manipulation while other L1 groups gained most from intonation manipulation. Specifically, the Russian and Persian groups were found to gain equally from the two manipulations (when the data were pooled across the two listeners), the English, Tamil and Chinese groups' N2 gained most from the duration manipulation, whereas the French and German groups' N2 gained most from the intonation manipulation. This discussion focuses primarily on the previous investigations which are most comparable in terms of target language and L1s. Boyd, Abelin & Dorriots (1999) investigated Swedish L2 speech produced by speakers from the L1s Hungarian, Spanish, Persian and Russian. They used only one speaker per L1. The present investigation had Norwegian as the target language, and Norwegian is very closely related to Swedish. Moreover, two of the L1 groups were the same, namely the Persian and Russian groups. Boyd, Abelin & Dorriots (1999) did not find any evidence of different weightings for different aspects of the L2 speech when they compared the impacts of segmental, prosodic and phonotactic deviances on the degree of foreign accent as rated by many (54) listeners. Bannert (1995) and Almberg & Husby (2000) also used only one speaker to represent each of their L1s. Almberg & Husby (2000) investigated Russian accented Norwegian and found that duration was most important, while Bannert (1995) investigated Russian and Punjabi accented Swedish and found that intonation was most important. In light of the finding from this thesis that there is variability across different speakers from the same L1, even to the extent that they sometimes gain most from different manipulations (as was the case for the Russian and Persian speakers), the results from Boyd, Abelin & Dorriots (1999), Almberg & Husby (2000), and Bannert (1995) all seem unreliable because they only used one speaker to represent each L1. However, the methodology in this thesis most resembles that of Almberg & Husby (2000), for instance in the choice to directly compare the impacts of durational and intonational aspects, and in the choice to investigate complete utterances as opposed to sentence fragments as in Bannert (1995). Also, the target language is identical across the present investigation and Almberg & Husby's investigation. It would therefore seem likely that the results would be similar across the two investigations. The results are in fact not similar. Almberg & Husby (2000) found that durations were of superior importance for their speaker, while the present investigation found that the manipulations had equal effect across the two speakers. However, remember that there were opposite effects for the two Russian speakers in this investigation such that one gained most from duration manipulation while the other gained most from intonation manipulation. There were also opposite effects for the two Persian speakers. Because of these opposite effects, it is impossible to compare the results for the Russian and Persian L1 groups with the results from previous investigations which make general statements for these L1 groups.

Of the remaining investigations presented in the introductory chapter that studied the relative contributions of durational and intonational aspects to the degree of foreign accent, none studied an L1 that was also studied in the present investigation. In general, the literature suggests that durational aspects, particularly in the form of speaking rate, affect degree of foreign accent more than intonational aspects. This can be confirmed by the present results in the sense that 3 L1 groups (English, Tamil and Chinese) were primarily affected by duration while 2 L1 groups (French and German) were primarily affected by intonation. Also, analyses showed that the important durational aspects were *articulation rate*[10] and *consonant durations*. The investigation by Gut (2007) seems particularly reliable in the sense that she investigated a total of 101 speakers from a total of 41 L1 groups (she did not specify the L1s). Her main finding was that speaking rate was the most important aspect that affected the L2 degree of accent. Other investigations support the finding that speaking rate is of paramount importance for degree of accent. Trofimovich & Baker (2006) also used a fairly large number of subjects in their study of 10 native listeners' perceptions of 30 Korean speakers' foreign-accented English, and they too found that speaking rate was of particular importance. The rest of the investigations that are relevant to discuss here used fewer subjects and are in this respect less reliable. We will nonetheless have a brief look at Wayland (1997) and Kamiyama (2004). Wayland found that intonational aspects were more important than durational aspects for the degree of foreign accent, but the type of durational aspects Wayland investigated were VOT and vowel durations. VOT was not explicitly investigated in the present investigation, but vowel durations were investigated and found *not* to correlate with the degree of foreign accent-ratings (section 3.6.1, Chapter 3). Kamiyama (2004) found that speaking rate did not affect the degree of accent, but this finding is not reliable because the investigation was based on only one single utterance, and because the utterance was not a complete sentence but a short fragment of a sentence lacking a verb. Perhaps a listener needs to hear a complete utterance in order to get a clear impression of the speaking rate.

The clearest way in which the findings from the present investigation relate to previous findings in the literature, is in demonstrating the great importance of speaking rate (in this experiment *articulation rate*, i.e. pauses excluded from measurements) on the degree of foreign accent.

5.3 The intelligibility experiment

5.3.1 Summary

This experiment was based on speech material produced by the same 14 speakers as in the previous experiment. Three utterances from each speaker were included. Each of the seven L1s was represented by 6 utterances. Thus, a total of 42 different sentences were used in the intelligibility experiment. Each N2 utterance was mixed with noise in order to avoid ceiling

[10] The measure of rate in this thesis is called articulation rate rather than speaking rate because it did not include pauses.

effects. The listeners' task was to write down the words they perceived of each utterance, and intelligibility was measured as percentage word-identification per sentence. The stimuli in this experiment were grouped in different stimulus sets, one for each type of stimulus. The effect of intelligibility was measured as the difference in word-identification score across two stimulus sets, for instance across the duration manipulated and the intonation manipulated stimuli.

The original stimuli were found to be unsuitable for this experiment as the original stimuli were more intelligible than any of the manipulated stimuli. This was due to side effects from the manipulations which will be discussed in section 5.5. For that reason, two close-original stimulus sets were generated to replace the original stimuli: The close-original duration stimuli (COD) were generated for comparison with the duration manipulated stimuli and the close-original intonation stimuli (COI) were generated for comparison with the intonation manipulated stimuli.

When intelligibility was measured across the COI and I stimulus sets, the results showed that the intonation manipulation significantly enhanced N2 intelligibility for the English and German L1 groups. The comparison of the COD and D stimulus sets showed that the duration manipulation significantly enhanced the N2 intelligibility for the French, Tamil and Persian L1 groups. Unexpectedly, the I/D, ID/D and ID/I comparisons showed results that were not entirely compatible with the results from the COI/I and COD/D comparisons. It was suggested that this incompatibility was due partly to various side effects from the manipulations. These manipulation problems will be discussed in section 5.5. Because of these problems, it was decided to discard the unreliable results from the I/D, ID/D and ID/I comparisons, and only use the results from the more reliable COD/D and COI/I comparisons (see e.g. 4.4.3, Chapter 4). The magnitude of the intelligibility enhancing effect of the manipulations differed according to the particular L1 group.

Subsequent analyses were carried out to identify those aspects of the duration manipulation that had caused the intelligibility enhancement. Durational adjustment was analysed in a) all segments, b) all consonants, c) all vowels, d) all phonologically long vowels, e) V/C ratio, f) articulation rate and g) pauses. Similar analyses were carried out at the end of Chapter 3 investigating degree of foreign accent for the same durational aspects except *pauses*. Pauses were added as a factor for the intelligibility analyses because pauses had been shortened in some of the utterances used in the intelligibility experiment in order to maintain naturalness in the duration manipulated stimuli. Surprisingly, no effects were found for any of these durational aspects. A possible reason why articulation rate was found to significantly affect degree of foreign accent, but not intelligibility, is that a listener may need increased coarticulation effects in order to perceived faster rates, especially when the speech is foreign accented and thus in general less redundant.

Analyses were also carried out to identify those intonational aspects that caused the intelligibility enhancement. F0 slope and F0 direction in three content words per sentence were analysed. The analyses were undertaken a) across the 6 syllables (2 syllables per word), b) in each of the 6 syllables, and c) between the onsets of each stressed syllable. The results of these analyses showed that the intelligibility enhancement had been due to F0 slope changes in the second syllable of the first word and in the second syllable of the third word. These syllables were unstressed. This result may therefore indicate that N2 speakers' deviant F0 slopes in unstressed syllables can interfere with native listeners' identifications of N2 words.

The manipulation analyses in the intelligibility experiment had been carried out with data pooled across the two speakers representing the same L1, called a *speaker pair*, because it was assumed that the two speakers would be similar in the sense that their N2 speech would gain most from the same manipulation. At the end of Chapter 4, investigations were carried out in order to assess the similarity across the speakers in each speaker pair. Similarity was investigated in three ways: as the degree of N1-N2 production deviation, as the magnitude of the manipulation effect, and as the manipulation that most affected the intelligibility. The results from these similarity assessments showed that there was some variability across the two speakers in a speaker pair regarding the N1-N2 production deviation and the magnitude of the manipulation effects. In general, there was more inter-speaker consistency in the speaker pairs regarding N1-N2 production deviation than regarding magnitude of effects. There did not seem to be any systematic relationship between N1-N2 production deviation and magnitude of manipulation effects. The similarity analyses also showed that not all speaker pairs consisted of speakers who had gained most from the same manipulation. Table 5.2 shows the manipulation that most affected intelligibility for each L1 group, and the two speaker pairs in which the individual speakers gained most from different manipulations (marked with *).

Table 5.2: The manipulation that most affected the intelligibility for each L1 group as measured with data pooled across the two speakers from each speaker pair. The L1 groups in which the speakers' N2 speech gained most from different manipulations are marked with asterisks.

L1	Most important manipulation
French	Duration
German	Intonation
English	Intonation
Tamil	Duration
Chinese	Equally important
Persian *	Duration
Russian *	Equally important

The table shows that when the data was pooled across the two Russian speakers, the manipulations affected intelligibility equally for the Russian L1 group. An asterisk shows that the individual speakers in the Russian L1 group had gained most from different manipulations. This means that the equal effect of the manipulations for the Russian L1 group was caused by opposite effects between the two individual speakers. The table further shows that when the data was pooled across the two Persian speakers, duration was found to affect intelligibility more than intonation. An asterisk shows that the two Persian speakers actually gained from different manipulations. This means that the superior effect of duration for the Persian L1 group was due to a superior D effect for only one speaker. The other Persian speaker instead had a superior I effect. The fact that there were opposite effects between the two speakers in the Russian speaker pair and between the two speakers in the Persian speaker pair, makes the results for the Russian and Persian L1 groups inconclusive.

5.3.2 Discussion

In this section, the results from the intelligibility experiment will be related to the literature presented in the introductory chapter (section 1.2.3, Chapter 1) that is judged as relevant. Unfortunately, only two of those investigations actually compared the effects of intonational and durational aspects on intelligibility. These two investigations were Almberg & Husby (2000), who investigated Russian-accented Norwegian, and Bannert (1995), who investigated Punjabi- and Persian-accented Swedish. Both of these investigations used listener ratings of perceived comprehensibility. It is difficult to directly compare these studies with the present investigation because the literature has shown that such ratings can yield different results than intelligibility as measured through transcriptions (e.g. Matsuura, Chiba & Fujieda, 1999; Derwing & Munro, 1997). Almberg & Husby (2000) investigated Russian-accented N2, and thus shares the target language and one L1 with the present investigation. Their result showed that durational aspects affected the perceived comprehensibility more than intonational aspects. They used only one speaker, as opposed to the two speakers in the present investigation. In this investigation, it was found that when the data were pooled across the two Russian speakers the two manipulations affected intelligibility equally, but further analyses showed that the two Russian speakers had in fact gained most from different manipulations. For one of the Russian speakers, duration was thus more important than intonation in line with the result from Almberg & Husby (2000). In his study, Bannert (1995) found that the Punjabi- and Persian-accented Swedish was more affected by intonational aspects than by durational aspects. In the present investigation, one of the Persian speakers had in fact gained most from the intonation manipulation while the other Persian speaker had gained most from the duration manipulation. Because of the opposite effects for the individual speakers in the Persian speaker pair and in the Russian speaker pair, it is difficult to make meaningful comparisons with the literature regarding these particular L1 groups. The remaining investigations that were presented in the literature review of the introductory chapter did not compare durational and intonation aspects. Instead, these studies mainly investigated durational aspects, and showed significant effects of such aspects on intelligibility. Across these studies, the perceptual effect of speaking rate was frequently studied (Anderson-Hsieh & Koehler, 1988; Derwing & Munro, 1997; Munro & Derwing, 1998; Munro & Derwing, 2001). One of these investigations, Anderson-Hsieh & Koehler (1988), showed that faster speaking rates had adverse effects on comprehensibility as measured through questions on text content. They used naturally varied speaking rate as opposed to the digitally manipulated speaking rate in the present study. Because it seems likely that faster rates can increase the amount of error and pronunciation inaccuracy in L2 speech, it is unclear whether that investigation actually measured the effects of speaking rate. The other investigations of speaking rate showed that a moderately accelerated speaking rate was beneficial in terms of perceived comprehensibility (Derwing & Munro, 1997; Derwing & Munro, 1998: Derwing & Munro, 2001) but that it had no effect on intelligibility as measured through word-identification scores (Derwing & Munro, 1997). The present investigation also found that there was no effect of speaking rate on intelligibility as measured through word-identification scores, and this result is therefore in accordance with the results from Derwing & Munro (1997).

5.4 General conclusions

When the data were pooled across the two speakers, the results from the two experiments were as follows. The degree of foreign accent-experiment showed that the French and German L1 groups gained most from the intonation manipulation, whereas the English, Tamil and Chinese groups gained most from the duration manipulation. The Russian and Persian groups were equally affected by the two manipulations. As for intelligibility, the German and English L1 groups gained most from the intonation manipulation, while the French, Tamil and Persian L1 groups gained most from the duration manipulation. However, subsequent analyses showed that there were opposite effects between the two speakers in the Persian speaker pair and between the two speakers in the Russian speaker pair for both experiments. The results are therefore inconclusive for these two L1 groups. Moreover, because there were only two speakers per L1 group, the consistency between the speakers within each of the remaining speaker pairs regarding could be due to chance. This problem will be further discussed in section 5.6.

The results from the experiments have also shown that N2 speech which is significantly accent reduced does not necessarily become significantly more intelligible. For instance, the degree of accent for French N2 was significantly reduced as a result of the intonation manipulation (Table 5.1), but the intelligibility remained unaffected (Table 5.2). In fact, only German and Tamil N2 speech were simultaneously accent reduced and intelligibility enhanced: German N2 was both accent reduced and intelligibility enhanced from the intonation manipulation, and Tamil N2 was both accent reduced and intelligibility enhanced from the duration manipulation (Tables 5.1 and 5.2). A robust finding from previous investigations was that there is no clear relationship between the degree of foreign accent and intelligibility of foreign accented speech (e.g. Munro & Derwing, 2005). For instance, listeners often judge an utterance as heavily foreign-accented, yet they transcribe it perfectly and do not rate it as difficult to understand (Munro, 2008). The findings from this thesis may be taken to support this view of a partial independence between degree of foreign accent and intelligibility.

In Chapter 1, section 1.3.2, hypotheses about the outcomes of the investigation were put forth. The hypotheses are repeated here:

A. Both intonational and durational aspects will affect the degree of foreign accent.
B. Durational aspects will affect the degree of foreign accent more than intonational aspects.
C. Both intonational and durational aspects will affect intelligibility.
D. Durational aspects will affect intelligibility more than intonational aspects.

These hypotheses were not L1-specific. Rather, they were statements about the relative effects of the manipulations across L1 groups. As the results from the present investigation showed that speakers from different L1 groups gained most from different manipulations, the hypotheses can be neither confirmed nor refuted. Moreover, for two of the L1 groups, Russian and Persian, the results are inconclusive because of opposite results between the two individual speakers representing each L1.

5.5 Manipulation problems

Several methodological challenges have emerged in the course of this investigation. One problem has been particularly difficult. This problem regards the side effects caused by the manipulations as will be discussed here. The controlled manipulation of specific aspects of speech in order to observe the manipulation's effect on listeners' perceptions is at the core of modern experimental phonetic methodology. Ideally, a manipulation affects only the isolated factor that the experimenter wants to study. However, there are side effects of manipulation. It has earlier been described how the duration manipulation had side effects both on utterance duration, affecting speaking rate, and on intonation, affecting F0 slopes (section 2.2.1.3, Chapter 2). Also, the stylization in the intonation manipulation process reduced intelligibility (section 4.4.2, Chapter 4). It has also been explained how these manipulation side effects caused difficulties in the intelligibility investigation (sections 4.4.2 and 4.4.3, Chapter 4). It has been shown that it was possible to counteract two of these manipulation side effects through adjusting the stimuli with which the manipulated stimuli were compared. The close-original duration stimuli were designed to counteract the duration manipulation's side effect on utterance duration, and the close-original intonation stimuli were designed to counteract the intonation stylization effect. In contrast, the duration manipulation's side effect on intonation seems very difficult to counteract, and this has not been attempted. The synthesis method can also be regarded as a manipulation side effect. The researcher has earlier suggested that the PSOLA synthesis itself may have been partly responsible for the speech signal degradation (section 4.2.3, Chapter 4). A manipulated change in a signal therefore seems inextricably linked to its manipulation method. It is important to be aware of such manipulation side effects, because there is a possibility that such side effects could affect measurements.

5.6 Future directions

The results from this investigation unexpectedly showed that N2 utterances from speakers sharing the same L1 do not always gain most from the same manipulation (section 3.7, Chapter 3; section 4.6, Chapter 4). This was the case for the two Persian speakers and the two Russian speakers. There were some indications that there may be intra-speaker consistency in the sense that a speaker from a certain L1 consistently gain most from one manipulation while another speaker from the same L1 consistently gain most from the other manipulation, but as there were a limited number of utterances per speaker in this investigation- 1 utterance per speaker in the degree of foreign accent-experiment, and 3 utterances per speaker in the intelligibility experiment- it is impossible to make very firm assessments of such intra-speaker consistency based on the present material. It is interesting to investigate *why* utterances spoken by speakers from the same L1 do not always gain most from the same manipulation, regardless of whether these differences reflect intra-speaker consistency or not. A possible explanation could be a perceptual interaction between N1-N2 deviations. Ideas about such perceptual interaction will be briefly outlined in the following. Although two L2 speakers share the same L1, their L2 speech does not deviate from the L1 speech in completely identical ways. As an example, consider two speakers of N2 that we can call A and B. Let us assume that measurements show that these two speakers' realizations of the Norwegian phonologically long vowels deviate from an N1 template to the same degree. One would expect that when their phonologically long vowels were manipulated (i.e corrected), the perceptual effect on for instance the degree of foreign accent would be the same for both speakers. Let us further assume that speaker A has very deviant vowel spectra, and that speaker B instead has very deviant intonation, and that intonation affects degree of foreign accent more than vowel spectra. The manipulation of the phonologically long vowels could make A's deviant vowel spectra and B's deviant intonation perceptually more salient. If intonation is more important for degree of foreign accent than spectrum, the manipulation could cause B's N2 speech to be *less* accent reduced than A's N2 speech from the same amount of duration manipulation. In short, when one deviation is removed, another deviation could become perceptually more salient. This hypothesis implies that the impact of a manipulation can not be predicted unless all of the utterance's N1-N2 deviations *and* the relative perceptual importance of these deviations are known. In order to investigate the existence of perceptual interaction between deviations, and in order to investigate the existence of intra-speaker consistency regarding relative effects of manipulations, it is necessary to use a large number of speakers from each L1, and a large number of utterances from each speaker.

6. References

Aaltonen, O., Eerola, O., Hellström, A., Uusipaikka, E., & Lang, A. (1997). Perceptual magnet effect in the light of behavioural and psychophysiological data. *Journal of the Acoustical Society of America, 101,* 1090-1105.

Abelin, Å., & Boyd, S. (2000). Voice quality, foreign accent and attitudes to speakers. *Proceedings of Fonetik 2000,* Skøvde, 21-24.

Abercrombie, D. (1949). Teaching pronunciation. *English language teaching, 3,* 113-122.

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine publishing company.

Akahane-Yamada, R., Tohkura, Y., Bradlow, A.R., & Pisoni, D.B. (1996). Does training in speech perception modify speech production? *Proceedings of ICSLP 1996,* Philadelphia, 606-609.

Almberg, J., & Husby, O. (2000). The relevance of some acoustic parameters for the perception of a foreign accent. In A. James, & J. Leather (Eds.), *New Sounds 2000*. (Pp. 1-10). University of Klagenfurt.

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning*, *38*(4), 561-613.

Anderson-Hsieh, J., Johnson, J., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning*, *42*, 529-555.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/. *Journal of Phonetics, 32,* 233-250.

Aoyama, K., & Guion, S. G. (2007). Prosody in second language acquisition. Acoustic analyses of duration and F0 range. In O-S. Bohn & M. J. Munro (Eds.), *Language experience in second language learning: In honor of James Emil Flege*. (Pp. 281-297). Amsterdam: John Benjamins.

Archibald, J. A. (1994). A formal model of learning L2 prosodic phonology. *Second Language Research, 10,* 215-240.

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German. *Journal of Phonetics, 32,* 177-197.

Baker, W., Trofimovich, P., Mack, M., & Flege, J. E. (2002). The effect of perceived phonetic similarity on non-native sound learning by children and adults. In B. Skarabela, S. Fish, & A. Do (Eds.), *Proceedings of the 26th annual Boston University Conference on Language Development.* (Pp. 36-47). Somerville, MA: Cascadilla Press.

Bannert, R. (1995). Intelligibility and acceptability in foreign accented Swedish: the effects of rhythmical and tonal features. *PHONUM, 3*, 7-29.

Beckman, M. E. (1992). Evidence for speech rhythms across languages. In Y. Tohura, E. Vatikiotis-Bateson and Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure*. (Pp. 457-463). Tokyo: Omsha and Amsterdam: IOS Press.

Beebe, L. M. (Ed.). (1988). *Issues in second language acquisition: Multiple perspectives.* Boston, MA: Heinle & Heinle.

Benoît, C. (1990). An intelligibility test using semantically unpredictable sentences: towards the quantification of linguistic complexity. *Speech Communication 9*(4), 293-304.

Benoît, C., Grice, M., & Hazan, V. (1996). The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using Semantically Unpredictable Sentences. *Speech Communication, 18*, 381-392.

Bent, T., Bradlow, A., & Smith, B. (2007). Phonemic errors in different word positions and their effects on intelligibility of non native speech. All's well that begins well. In O-S. Bohn & M. J. Munro (Eds.), *Language experience in second language learning: In honor of James Emil Flege*. (Pp. 331-347). Amsterdam: John Benjamins.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in Cross-linguistic research.* (Pp. 233-277). Timonium MD: York Press.

Best, C. T., McRoberts, G. W, & Sithole, N. N. (1988). The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 245-260.

Best, C. T., & Avery, R. A. (1999). Left hemisphere advantage for click consonants is determined by linguistic significance. *Psychological Science, 10,* 65-69.

Best, C. T. & Tyler, M. D. (2005). Nonnative and second-language speech perception: Commonalities and complementarities. In O-S. Bohn & M. J. Munro (Eds.), *Language experience in second language learning: In honor of James Emil Flege*. (Pp. 13-34). Amsterdam: John Benjamins.

Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In O-S. Bohn & M. J. Munro (Eds.), *Language experience in second language learning: In honor of James Emil Flege*. (Pp. 99-116). Amsterdam: John Benjamins.

Blaauw, E. (1995). On the perceptual classification of spontaneous and read speech. Doctoral dissertation, University of Utrecht. Utrecht: OTS Dissertation series.

Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. *Language, 26*, 86-125.

Boersma, P., & Weenink, D. (2004). Praat - A system for doing phonetics by computer. http://www.fon.hum.uva.nl/praat/.

Bohn, O. -S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition, 14,* 131-158.

Bond, Z.S., Stockmal, V., & Markus, D. (2003). Sentence durations and accentedness judgments. *Journal of the Acoustical Society of America, 113*, 2330-2334.

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis*. (Pp. 133-159). Mahwah, NJ: Erlbaum.

Bonneau, A., Camus, M., Laprie, Y., & Colotte, V. (2004). A computer-assisted learning of English prosody for French students. *Proceedings of ICALL 2004*, Venice, paper 028.

Boula de Mareüil, P., & Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica, 63,* 247-267.

Boyd, S., Abelin, Å., & Dorriots, B. (1999). Attitudes to foreign accent. *Proceedings of FONETIK 1999*, Department of Linguistics, University of Gothenburg, 31-35.

Bradlow, A., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, *112*(1), 272-284.

Bradlow, A., Toretta, G., & Pisoni, D. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication 20*, 255-272.

Bradlow, A., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities. *Journal of Speech, Language and Hearing Research*, *46*, 80-97.

Brennan, E., & Brennan, J. (1981). Accent scaling and language attitudes: Reactions to Mexican American English speech. *Language and Speech, 24*, 207-221.

Caldognetto, E. M., Zmarich, C., & Ferrero, F. (1997). A comparative acoustic study of spontaneous and read Italian speech. *Proceedings of Eurospeech 1997*, vol. 2, Rhodos, 779-782.

Cunningham-Andersson, U., & Engstrand, O. (1989). Perceived strength and identity of foreign accent in Swedish. *Phonetica, 46,* 138-154.

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

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favour of a broad framework for pronunciation instruction. *Language Learning, 48*, 393-410.

Derwing, T. M., & Rossiter, M. J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System, 30*, 155-166.

Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency and complexity of L2 accented speech. *Applied Language Learning, 13*(1), 1-17.

Dommelen, W. A. van (1998). Production and perception of preaspiration in Norwegian. *Proceedings of FONETIK 1998*, Department of Linguistics, University of Stockholm, 20-23.

Dommelen, W. A. van (2000): Preaspiration in Norwegian: some results on speaker age, gender, and linguistic factors. *Proceedings of FONETIK 2000*, University of Skövde, 141-144.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303-306.

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning, 37*, 313-326.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*(3), 399-424.

Fintoft, K., Mjåvatn, T., Møllegård, E., & Ulseth, B. (1978): Toneme patterns in Norwegian dialects. In E. Gårding, G. Bruce & R. Bannert (Eds.), *Nordic Prosody. Papers from a symposium*. (Pp. 197-206). Lund: Lunds universitet.

Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America, 76,* 692-707.

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.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of Americ*a, 84, 70-79.

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in Cross-linguistic research.* (Pp. 233-277). Timonium MD: York Press.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America, 9*(1), 370-389.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995a). Effects of age of second language learning on the production of English consonants. *Speech Communicatio*n*, 16*, 1-26.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995b). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97,* 3125–3134.

Flege, J. E., Bohn, O. –S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25,* 437-470.

Flege, J. E., & MacKay, I. R. A (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26,* 1-34.

Fox, R. A., & McGory, J. T. (2005). Second language acquisition of a regional dialect of American English by native Japanese speakers. In O-S. Bohn & M. J. Munro (Eds.), *Language experience in second language learning: In honor of James Emil Flege*. (Pp. 117-134). Amsterdam: John Benjamins.

Gaies, S., Gradman, H., & Spolsky, B. (1977). Toward a measurement of functional proficiency: Contextualisation of the noise test. *TESOL Quarterly 11,* 51-57.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of phonetics, 11*, 49-175.

Gandour, J., Harshman, R. (1978). Cross-language difference in tone perception: A multidimensional scaling investigation. *Language and Speech, 21*, 1-33.

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning, 34*, 65-89.

Gottfried, T. L. (2007). Music and language learning. Effect of musical training on learning L2 speech contrasts. In O-S. Bohn & M. J. Munro (Eds.), *Language experience in second language learning: In honor of James Emil Flege*. (Pp.221-237). Amsterdam: John Benjamins.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7*. (Pp. 515-546). Berlin: Mouton.

Gut, U. (2007). Foreign accent. In C. Müller, (Ed.), *Speaker classification*. (pp. 75-87). Heidelberg: Springer.

Guion, S. G., Flege, J. E., Liu, S. H., & Yeni-Komshian G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics, 21,* 205-228.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O-S. Bohn & M. J. Munro (Eds.), *Language experience in second language learning: In honor of James Emil Flege*. (Pp. 57-77). Amsterdam: John Benjamins.

Gynan, S. (1985). Comprehension, irritation and error hierarchies. *Hispania, 68*, 160-165.

Gårding, E. (1998). Intonation in Swedish. In D. Hirst & A. diCristo (Eds.), *Intonation systems: a survey of twenty languages*. (Pp. 112-130). Cambridge: Cambridge University Press.

Han, M. S. (1962). The feature of duration in Japanese. *Onsei no kenkyuu, 10*, 65-80.

Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*, 201-223.

Hart, J. 't, Collier, R., & Cohen, A. (1990). A perceptual study of intonation: An experimental-phonetic approach.

Hazan, V. (2002). Apprentissage des langues. *Proceedings of 14èmes Journées d'Étude sur la parole,* Nancy, 1-5.

Hazan, V., & Simpson, A. (1998). The effect of cue-enhancement on the intelligibility of nonsense word and sentence materials presented in noise. *Speech Communication, 24,* 211-226.

Hazan, V., and Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *Journal of the Acoustical Society of America*, *116*, 3108-3118.

Helgason, P., Stölten, K., & Engstrand, O. (2003). Dialectal and sociophonetic aspects of preaspiration. *Proceedings of ICPhS 2003*, Barcelona, 17-20.

Hognestad, J. K. (1997). *Tonemer i en høytonedialekt. En undersøkelse med utgangspunkt i Egersund bymål*. Oslo: Det norske samlaget.

Holm, S. (2003). Individual use of acoustic parameters in read and spontaneous speech. *Proceedings of Fonetik 2003*, Department of Philosophy and Linguistics, University of Umeå, 157-161.

Huckvale, M. (2006). The new accent technologies: recognition, measurement and manipulation of accented speech. In P. Zhang, T.-W. Xie, S. Lin, J.H. Xie, A.C. Fang, & J. Xu (Eds.), *Research and application of digitized Chinese teaching and learning.* (Pp. 28-37). Beijing: Language and Culture Press.

Husby, O., & Kløve, M. H. (1998). *Andrespråksfonologi: Teori og metodikk*. Oslo, Gyldendal.

James, C. (1998). *Errors in language learning and use*. London: Longman.

Jilka, M. (2000). Testing the contribution of prosody to the perception of foreign accent. *New Sounds, 4*, 199-207.

Jenkins, J. (2000). *The phonology of English as an international language*. England: Oxford University Press.

Jesney, K. (2004). *The use of global foreign accent rating in studies of L2 acquisition*. A report prepared for the Language Research Centre, University of Calgary.

Kalin, R., & Rayko, D. S. (1978). Discrimination in evaluative judgments against foreign accented job candidates. *Psychological Reports, 43*, 1203-1209.

Kamiyama, T. (2004). Perception of foreign accentedness in L2 prosody and segments: L2 Japanese speakers learning L2 French. *SP-2004*, 721-724.

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54,* 482-499.

Koopmans- van Beinum, F. J. (1980). Vowel contrast reduction: An acoustic and perceptual study of Dutch vowels in various speech conditions (Doctoral dissertation, University of Amsterdam).

Kristoffersen, G. (2000). *The phonology of Norwegian*. Oxford: Oxford University Press.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, USA, 97,* 11850-11857.

Kuhl, P. K, & Iverson, P. (1995). Linguistic experience and the ”perceptual magnet effect”. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in Cross-linguistic research.* (Pp. 121-154). Timonium MD: York Press.

Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. In W. J. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques*. (Pp. 7-30). Cambridge: Cambridge University Press.

Lado, R. (1957). *Linguistics across cultures.* Ann Arbor, MI: University of Michigan Press.

Lane, H. (1963). Foreign accent and speech distortion. *Journal of the Acoustical Society of America, 35*, 451-453.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Levis, J. M. (2005). Changing contexts and paradigms in pronunciation teaching. *TESOL Quarterly, 39*(3), 369-378.

Lippi-Green, R. (1997). *English with an accent*. New York: Routledge.

Magen, H. (1998). The perception of foreign accented speech. *Journal of Phonetics, 26*, 381-400.

Maassen, B., & Povel, D-J. (1984). The effect of correcting fundamental frequency on the intelligibility of deaf speech and its interaction with temporal aspects. *Journal of the Acoustical Society of America 76*(6), 1673-1681.

Maassen, B., & Povel, D-J. (1985). The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech. *Journal of the Acoustical Society of America 78*(3), 877-886.

MacKay, I. R. A, Meador, D., & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica, 58,* 103-125.

McAllister, R. (2007). Strategies for realization of L2-Categories. English /s/ - /z/. In O-S. Bohn & M. J. Munro (Eds.), *Language experience in second language learning: In honor of James Emil Flege*. (Pp. 153-166). Amsterdam: John Benjamins.

Major, R. C. (1986a). The ontogeny model: Evidence from L2 acquisition of Spanish *r*. *Language Learning, 36*, 453-503.

Major, R. C. (1986b). Paragoge and degree of foreign accent in Brazilian English. *Second Language Research, 2* (1), 53-71.

Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Erlbaum.

Markham, D. (1997). *Phonetic imitation, accent, and the learner.* Lund: Lund University Press.

Matsuura, H., Chiba, R., & Fujieda, M. (1999). Intelligibility and comprehensibility of American and Irish Englishes in Japan. *World Englishes, 18*(1), 49-62.

McAllister, R. (1997). Perceptual foreign accent: L2 users' comprehension ability. In J. Leather & A. James (Eds.), *Second-Language Speech.* (Pp. 119-132). Berlin: Mouton de Gruyter.

McAllister, R. (2001). Experience as a factor in L2 phonological acquisition. *Lund University, Dept. of Linguistics Working Papers 49,* 116-119.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of phonetics, 32,* 543-563.

Morrison, G. S. (2003). Perception and production of Spanish vowels by English speakers. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences: Barcelone 2003.* (Pp. 1533-1536). Adelaide, South Australia: Causal Productions.

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

Munro, M. J. (1995). Nonsegmental factors in foreign accent. *Studies in Second Language Acquisition, 17*(1), 17-34.

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign accented speech. *Studies in Second Language Acquisition, 20*, 139-154.

Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal, 20*(2), 38-51.

Munro, M. J. (2008). Foreign accent and speech intelligibility. In J. G. H. Edwards & M. L. Zampini (Eds.), Phonology and second language acquisition. (Pp. 193-218). Amsterdam: John Benjamins.

Munro, M. J., & Derwing, T. M. (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing, 11*(3), 253-266.

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*, 73-97.

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of foreign-accented speech. *Language and Speech, 38*, 289- 306.

Munro, M. J.,& Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign accented speech. *Language Learning, 48*(2), 159-182.

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning, 49* (suppl. 1), 285-310.

Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition, 23*, 451-468.

Munro, M. J., & Derwing, T. M., & Burgess, C. S. (2001). The detection of foreign accent in backwards speech. In M. -J. Solé, D. Recasens & J. Romero (Eds.*), Proceedings of the 15th International Congress of Phonetic Sciences*. (Pp. 535-538). Barcelona: Universitat Autonoma de Barcelona.

Munro, M. J., & Derwing, T. M. (2005). Second language accent and pronunciation teaching: A research based approach. *TESOL Quarterly, 39*(3), 379-398.

Nilsen, R. A. (1992). Intonasjon i interaksjon: Sentrale spørsmål i norsk intonologi. Doctoral dissertation, University of Trondheim.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Livonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., Sinkkonen, J., and Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385,* 432-434.

Osberger, M. J., & Levitt, H. (1979). The effect of timing on the intelligibility of deaf children's speech. *Journal of the Acoustical Society of America, 66*(5), 1316-1324.

Palmer, J. (1976). Linguistic accuracy and intelligibility. *Proceedings of the International Congress of Applied Linguistics 1976*, Stuttgart, Hochschul Verlag, 505-513.

Pater, J. V (1997). Metrical parameter missetting in second language acquisition. In S. J. Hannahs & M. Young-Scholten (Eds.), *Focus on phonological acquisition*. (Pp. 235-261). Philadelphia: John Benjamins.

Patkowski, M. (1990). *Age and accent in second language: A reply to James Emil Flege.* Applied Linguistics, 11, 73- 89.

Pennington, M. C. (1996). *Phonology in English language teaching*. London: Longman.

Pickering, L. (2006). Current research on intelligibility in English as a lingua franca. *Annual Review of Applied Linguistics, 26*, 219-233.

Pike, K. L. (1945). *The intonation of American English*. Michigan: Michigan University Press.

Piske, T., Flege, J. E., & Mackay, I. R. A. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*, 191-215.

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/ - /th/ perception: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America, 109,* 2190-2201.

Purcell, E. & Suter, R. (1980). Predictors of pronunciation accuracy: A re-examination. *Language Learning, 30,* 271-287.

Rajadurai, J. (2007). Intelligibility studies: a consideration of empirical and ideological issues. *World Englishes, 26*(1), 87-98.

Schmidt, A. M., & Flege, J. E. (1996). Speaking rate effects on stops produced by Spanish and English monolinguals and Spanish/ English bilinguals. *Phonetica, 53,* 162-179.

Scovel, T. (1988). *A time to speak: A psycholinguistic investigation into the critical period for human speech.* New York: Harper & Row.

Scovel, T. (1995). Differentiation, recognition and identification in the discrimination of foreign accents. In J. Archibald, (Ed.), *Phonological acquisition and phonological theory*. (Pp. 169- 181). Hillsdale NJ: Lawrence Erlbaum & Associates.

Smith, L., & Nelson, C. (1985). International intelligibility of English: Directions and resources. *World Englishes, 4,* 333-342.

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. (Pp. 3- 45). Timonium, MD: York Press.

Strange, W. (2007). Cross-language phonetic similarity of vowels. Theoretical and methodological issues. In O-S. Bohn & M. J. Munro (Eds.), *Language experience in second language learning: In honor of James Emil Flege*. (Pp. 35-55). Amsterdam: John Benjamins.

Studebaker, G. A. (1985). A rationalized arcsine transform. *Journal of Speech and Hearing Research, 28*, 455-462.

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign accented English. *Journal of Phonetics, 25*, 1-24.

Thompson, I. (1991). Foreign accent revisited: The English pronunciation of Russian immigrants. *Language Learning, 41,* 177-204.

Thorén, B. (2003). Can V/C-ratio alone be sufficient for discrimination of V:C/VC: in Swedish? A perception test with manipulated durations. *PHONUM, 9,* 49-52.

Trofimovich, P. & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28*(1), 1-30.

Trudgill, P. (2005). Speaker-listener equilibrium: Segmental phonological models in EFL. In K. Dziubalska-Kolaczyk & J. Przedlacka (Eds.), *English pronunciation models: a changing scene*. (Pp. 213-228). Bern: Peter Lang.

Varonis, E. M., & Gass, S. (1982). The comprehensibility of non-native speech. *Studies in Second Language Acquisition, 4,* 1114-136.

Wayland, R. (1997). Non-native production of Thai: Acoustic measurements and accentedness ratings. *Applied Linguistics, 18*(3), 345-373.

Werker, J. F. & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual review of psychology, 50*, 509-535.

Werner, S. (2000). Modelle deutscher Intonation: zu Vergleichbarkeit und empirischer Relevanz von Intonationsbeschreibungen (Doctoral dissertation, University of Joensuu). Joensuun yliopisto.

Wik, P. (2004). Designing a virtual language tutor. *Proceedings of FONETIK 2004*, University of Stockholm, 136-139.

Appendix A

The tables show results from statistical tests investigating the degree of foreign accent as discussed in Chapter 3. The boundary for statistical significance is $p < 0.05$. Statistical significance is marked with grey shading.

Table 1: The table shows the results from a Mann Whitney test investigating the perceived accent difference between the stimuli in one stimulus pair as compared to the perceived accent difference between the stimuli in another stimulus pair.

Stimulus pairs	Sig.
O_D / D_O	0.161
O_I / I_O	0.002
D_I / I_D	0.211
O_ID / ID_O	0.046
D_ID / ID_D	0.743
I_ID / ID_I	0.223

Table 2: The table shows the results from a Sign Test. Stimulus order across stimulus pairs is pooled.

L1	Stimulus pair	Most native-like stimulus	Sig.
English	O_D + D_O	D	0.000
English	O_I + I_O	I	0.133
English	D_I + I_D	D	0.000
English	O_ID + ID_O	ID	0.000
English	D_ID + ID_D	ID	0.939
English	I_ID + ID_I	ID	0.000
French	O_D + D_O	D	0.000
French	O_I + I_O	I	0.000
French	D_I + I_D	D	0.102
French	O_ID + ID_O	ID	0.000
French	D_ID + ID_D	ID	0.000
French	I_ID + ID_I	ID	0.000
German	O_D + D_O	D	0.280
German	O_I + I_O	I	0.002
German	D_I + I_D	I	0.000
German	O_ID + ID_O	ID	0.000
German	D_ID + ID_D	ID	0.000
German	I_ID + ID_I	ID	0.024
Russian	O_D + D_O	D	0.000
Russian	O_I + I_O	I	0.000
Russian	D_I + I_D	I	0.782
Russian	O_ID + ID_O	ID	0.000
Russian	D_ID + ID_D	ID	0.000
Russian	I_ID + ID_I	ID	0.000
Tamil	O_D + D_O	D	0.000
Tamil	O_I + I_O	I	0.000

Tamil	D_I + I_D	D	0.000
Tamil	O_ID + ID_O	ID	0.000
Tamil	D_ID + ID_D	ID	0.000
Tamil	I_ID + ID_I	ID	0.000
Chinese	O_D + D_O	D	0.000
Chinese	O_I + I_O	I	0.000
Chinese	D_I + I_D	D	0.000
Chinese	O_ID + ID_O	ID	0.000
Chinese	D_ID + ID_D	ID	0.000
Chinese	I_ID + ID_I	ID	0.000
Persian	O_D + D_O	D	0.000
Persian	O_I + I_O	I	0.000
Persian	D_I + I_D	I	0.505
Persian	O_ID + ID_O	ID	0.000
Persian	D_ID + ID_D	ID	0.000
Persian	I_ID + ID_I	ID	0.000

Table 3: The table shows the results from a Sign Test. Stimulus orders are separate. Negative difference= the stimulus in second position is less accented. Positive difference= the stimulus in first position is less accented.

L1	Stimulus pair	Most native-like stimulus	Sig.
English	O_D	D	0.000
English	D_O	D	0.000
English	O_I	I	0.598
English	I_O	I	0.005
English	D_I	D	0.000
English	I_D	D	0.000
English	O_ID	ID	0.000
English	ID_O	ID	0.000
English	D_ID	ID	0.236
English	ID_D	ID	0.326
English	I_ID	ID	0.000
English	ID_I	ID	0.000
French	O_D	D	0.000
French	D_O	D	0.000
French	O_I	I	0.003
French	I_O	I	0.000
French	D_I	D	0.032
French	I_D	D	0.921
French	O_ID	ID	0.000
French	ID_O	ID	0.000
French	D_ID	ID	0.000
French	ID_D	ID	0.002
French	I_ID	ID	0.000
French	ID_I	ID	0.000
German	O_D	D	0.115

German	D_O	D	1.000
German	O_I	I	0.073
German	I_O	I	0.017
German	D_I	I	0.044
German	I_D	I	0.001
German	O_ID	ID	0.003
German	ID_O	ID	0.000
German	D_ID	ID	0.083
German	ID_D	ID	0.000
German	I_ID	ID	0.180
German	ID_I	ID	0.100
Russian	O_D	D	0.000
Russian	D_O	D	0.000
Russian	O_I	I	0.294
Russian	I_O	I	0.000
Russian	D_I	I	0.845
Russian	I_D	I	0.492
Russian	O_ID	ID	0.000
Russian	ID_O	ID	0.000
Russian	D_ID	ID	0.025
Russian	ID_D	ID	0.000
Russian	I_ID	ID	0.006
Russian	ID_I	ID	0.000
Tamil	O_D	D	0.000
Tamil	D_O	D	0.000
Tamil	O_I	I	0.576
Tamil	I_O	I	0.000
Tamil	D_I	D	0.000
Tamil	I_D	D	0.053
Tamil	O_ID	ID	0.000
Tamil	ID_O	ID	0.000
Tamil	D_ID	ID	0.000
Tamil	ID_D	ID	0.000
Tamil	I_ID	ID	0.000
Tamil	ID_I	ID	0.000
Chinese	O_D	D	0.000
Chinese	D_O	D	0.000
Chinese	O_I	I	0.550
Chinese	I_O	I	0.000
Chinese	D_I	D	0.000
Chinese	I_D	D	0.008
Chinese	O_ID	ID	0.000
Chinese	ID_O	ID	0.000
Chinese	D_ID	ID	0.000
Chinese	ID_D	ID	0.000
Chinese	I_ID	ID	0.000
Chinese	ID_I	ID	0.000
Persian	O_D	D	0.000

Persian	D_O	D	0.000
Persian	O_I	I	0.000
Persian	I_O	I	0.001
Persian	D_I	I	0.062
Persian	I_D	I	0.396
Persian	O_ID	ID	0.000
Persian	ID_O	ID	0.000
Persian	D_ID	ID	0.000
Persian	ID_D	ID	0.011
Persian	I_ID	ID	0.000
Persian	ID_I	ID	0.002

Table 4: The table shows the results from a Mann Whitney test for independent samples comparing the perceived accent difference between the stimuli in one stimulus pair as compared with the perceived accent difference between the stimuli in another stimulus pair. This is done for stimulus pairs with one common stimulus. The stimulus orders are pooled.

L1	Stimulus pairs	Most native-like stimulus	Sig.
English	O_D / O_I	O_D	0.003
English	O_D / O_ID	O_ID	0.000
English	O_I / O_ID	O_ID	0.000
English	ID_D / ID_I	ID_I	0.000
English	D_O / D_I	D_I	0.000
English	I_D / I_O	I_D	0.000
French	O_D / O_I	O_I	0.020
French	O_D / O_ID	O_ID	0.000
French	O_I / O_ID	O_ID	0.000
French	ID_D / ID_I	ID_I	0.645
French	D_O / D_I	D_O	0.000
French	I_D / I_O	I_O	0.000
German	O_D / O_I	O_I	0.003
German	O_D / O_ID	O_ID	0.000
German	O_I / O_ID	O_ID	0.121
German	ID_D / ID_I	ID_D	0.000
German	D_O / D_I	D_I	0.000
German	I_D / I_O	I_D	0.569
Russian	O_D / O_I	*equal*	0.194
Russian	O_D / O_ID	O_ID	0.000
Russian	O_I / O_ID	O_ID	0.001
Russian	ID_D / ID_I	ID_D	0.112
Russian	D_O / D_I	D_O	0.010
Russian	I_D / I_O	I_O	0.005
Chinese	O_D / O_I	O_D	0.000
Chinese	O_D / O_ID	O_ID	0.000
Chinese	O_I / O_ID	O_ID	0.000
Chinese	ID_D / ID_I	ID_I	0.000
Chinese	D_O / D_I	D_O	0.000

Chinese	I_D / I_O	I_D	0.000
Tamil	O_D / O_I	O_D	0.000
Tamil	O_D / O_ID	O_ID	0.000
Tamil	O_I / O_ID	O_ID	0.000
Tamil	ID_D / ID_I	ID_I	0.066
Tamil	D_O / D_I	D_O	0.000
Tamil	I_D / I_O	I_D	0.000
Persian	O_D / O_I	O_I	0.075
Persian	O_D / O_ID	O_ID	0.000
Persian	O_I / O_ID	O_ID	0.000
Persian	ID_D / ID_I	ID_D	0.124
Persian	D_O / D_I	D_O	0.021
Persian	I_D / I_O	I_O	0.002

Appendix B

The following tables show the results from statistical tests on intelligibility as presented in Chapter 4. All statistics are based upon the percent correct word identification converted into rau units. The boundary for statistical significance is $p < 0.05$. Statistical significance is marked with grey shade.

Table 1: Original (O) and intonation manipulated (I) stimuli for All Data. The table shows a 3 factorial analysis of variance for independent samples with factors manipulation (O and I), learning effect and L1.

Factors	F	df	Error df	Sig.
L1	56.575	6	3375	0.000
Manipulation	18.325	1	3375	0.000
Learning effect	106.522	1	3375	0.000
L1 x Manipulation	5.522	6	3375	0.000
L1 x Learning effect	3.647	6	3375	0.001
Manipulation x Learning effect	0.040	1	3375	0.841
L1 x Manipulation x Learning effect	0.750	6	3375	0.609

x = interaction

Table 2: Original (O) and duration manipulated (D) stimuli for All Data. The table shows a 3 factorial analysis of variance for independent samples with factors manipulation (O and D), learning effect and L1.

Factors	F	df	Error df	Sig.
L1	34.723	6	3333	0.000
Manipulation	52.917	1	3333	0.000
Learning effect	40.710	1	3333	0.000
L1 x Manipulation	7.184	6	3333	0.000
L1 x Learning effect	0.760	6	3333	0.601
Manipulation x Learning effect	3.314	1	3333	0.069
L1 x Manipulation x Learning effect	0.670	6	3333	0.674

x = interaction

Table 3: Original (O) and intonation-duration manipulated (ID) stimuli for All Data. The table shows a 3 factorial analysis of variance for independent samples with factors manipulation (O and ID), learning effect and L1.

Factors	F	df	Error df	Sig.
L1	34.723	6	2577	0.000
Manipulation	52.917	1	2577	0.000
Learning effect	40.710	1	2577	0.000
L1 x Manipulation	7.184	6	2577	0.000
L1 x Learning effect	0.760	6	2577	0.186
Manipulation x Learning effect	3.314	1	2577	0.052

L1 x Manipulation x Learning effect	0.670	6	2577	0.994

x = interaction

Table 4: Close-original intonation (COI) and original (O) stimuli for All Data. The table shows a 3 factorial analysis of variance for independent samples with factors manipulation (O and COI), learning effect and L1.

Factors	F	df	Error df	Sig.
L1	58.244	6	2619	0.000
Manipulation	79.998	1	2619	0.000
Learning effect	15.118	1	2619	0.000
L1 x Manipulation	3.095	6	2619	0.005
L1 x Learning effect	2.001	6	2619	0.062
Manipulation x Learning effect	22.384	1	2619	0.000
L1 x Manipulation x Learning effect	0.605	6	2619	0.727

x = interaction

Table 5: Close-original intonation (COI) and original (O) stimuli for All Data. The table shows a 2 factorial analysis of variance for independent samples with factors manipulation (O and COI) and learning effect.

L1	Factors	F	df	Error df	Sig.
English	Manipulation	9.591	1	374	0.002
	Learning effect	0.057	1	374	0.811
	Manipulation x Learning effect	4.731	1	374	0.030
German	Manipulation	36.847	1	374	0.000
	Learning effect	0.632	1	374	0.427
	Manipulation x Learning effect	4.174	1	374	0.042
French	Manipulation	13.019	1	374	0.000
	Learning effect	0.001	1	374	0.974
	Manipulation x Learning effect	10.111	1	374	0.002
Tamil	Manipulation	0.659	1	374	0.417
	Learning effect	6.557	1	374	0.011
	Manipulation x Learning effect	0.212	1	374	0.646
Chinese	Manipulation	5.411	1	374	0.021
	Learning effect	12.201	1	374	0.001
	Manipulation x Learning effect	2.473	1	374	0.117
Persian	Manipulation	25.813	1	374	0.000
	Learning effect	4.544	1	374	0.334
	Manipulation x Learning effect	2.727	1	374	0.100
Russian	Manipulation	7.104	1	374	0.008
	Learning effect	0.491	1	374	0.484
	Manipulation x	2.273	1	374	0.132

	Learning effect				

x = interaction

Table 6: Close-original duration (COD) and original (O) stimuli for All Data. The table shows a 3 factorial analysis of variance for independent samples with factors manipulation (O and COD), learning effect and L1.

Factors	F	df	Error df	Sig.
L1	50.817	6	2577	0.000
Manipulation	161.681	1	2577	0.000
Learning effect	97.147	1	2577	0.000
L1 x Manipulation	3.652	6	2577	0.001
L1 x Learning effect	1.682	6	2577	0.121
Manipulation x Learning effect	1.516	1	2577	0.218
L1 x Manipulation x Learning effect	0.889	6	2577	0.502

x = interaction

Table 7: Close-original duration (COD) and original (O) stimuli for All Data. The table shows a 2 factorial analysis of variance for independent samples with factors manipulation (O and COD) and learning effect.

L1	Factors	F	df	Error df	Sig.
English	Manipulation	22.215	1	368	0.000
	Learning effect	7.970	1	368	0.005
	Manipulation x Learning effect	0.089	1	368	0.766
German	Manipulation	32.678	1	368	0.000
	Learning effect	5.477	1	368	0.020
	Manipulation x Learning effect	0.178	1	368	0.674
French	Manipulation	1.306	1	368	0.254
	Learning effect	24.387	1	368	0.000
	Manipulation x Learning effect	3.625	1	368	0.058
Tamil	Manipulation	34.150	1	368	0.000
	Learning effect	20.930	1	368	0.000
	Manipulation x Learning effect	2.529	1	368	0.113
Chinese	Manipulation	21.462	1	368	0.000
	Learning effect	20.919	1	368	0.000
	Manipulation x Learning effect	0.397	1	368	0.529
Persian	Manipulation	46.624	1	368	0.000
	Learning effect	18.992	1	368	0.000
	Manipulation x Learning effect	0.217	1	368	0.642
Russian	Manipulation	22.828	1	368	0.000
	Learning effect	5.220	1	368	0.023
	Manipulation x Learning effect	0.013	1	368	0.909

x = interaction

Table 8: Close-original intonation (COI) and intonation manipulated (I) stimuli for All Data. The table shows a 3 factorial analysis of variance for independent samples with factors manipulation (COI and I), learning effect and L1.

Factors	F	df	Error df	Sig.
L1	46.008	6	2492	0.000
Manipulation	28.325	1	2492	0.000
Learning effect	16.627	1	2492	0.000
L1 x Manipulation	5.482	6	2492	0.000
L1 x Learning effect	3.007	6	2492	0.006
Manipulation x Learning effect	24.250	1	2492	0.000
L1 x Manipulation x Learning effect	1.321	6	2492	0.244

x = interaction

Table 9: Close-original intonation (COI) and intonation manipulated (I) stimuli for All Data. The table shows a 2 factorial analysis of variance for independent samples with factors manipulation (I and COI) and learning effect.

L1	Factors	F	df	Error df	Sig.
English	Manipulation	20.233	1	356	0.000
	Learning effect	0.340	1	356	0.560
	Manipulation x Learning effect	6.353	1	356	0.012
German	Manipulation	39.840	1	356	0.000
	Learning effect	0.035	1	356	0.851
	Manipulation x Learning effect	1.863	1	356	0.173
French	Manipulation	1.767	1	356	0.185
	Learning effect	0.774	1	356	0.380
	Manipulation x Learning effect	16.646	1	356	0.000
Tamil	Manipulation	0.065	1	356	0.799
	Learning effect	8.915	1	356	0.003
	Manipulation x Learning effect	0.951	1	356	0.330
Chinese	Manipulation	0.101	1	356	0.751
	Learning effect	19.579	1	356	0.000
	Manipulation x Learning effect	6.021	1	356	0.015
Persian	Manipulation	1.615	1	356	0.205
	Learning effect	2.236	1	356	0.136
	Manipulation x Learning effect	1.016	1	356	0.314
Russian	Manipulation	1.125	1	356	0.290
	Learning effect	0.014	1	356	0.905
	Manipulation x	0.544	1	356	0.461

	Learning effect				

x= interaction

Table 10: Close-original intonation (COI) and intonation manipulated (I) stimuli for Paired Data. The table shows a 2 factorial analysis of variance for repeated measures with factors manipulation (I and COI) and learning effect.

L1	Factors	F	df	Error df	Sig.
English	Manipulation	16.1	1	118	0.000
	Learning effect	6.042	1	118	0.015
	Manipulation x Learning effect	0.613	1	118	0.435
German	Manipulation	34.543	1	118	0.000
	Learning effect	4.052	1	118	0.046
	Manipulation x Learning effect	1.593	1	118	0.209
French	Manipulation	1.006	1	118	0.318
	Learning effect	11.790	1	118	0.001
	Manipulation x Learning effect	5.854	1	118	0.017
Tamil	Manipulation	0.063	1	118	0.802
	Learning effect	1.632	1	118	0.204
	Manipulation x Learning effect	18.625	1	118	0.000
Chinese	Manipulation	0.216	1	118	0.643
	Learning effect	3.241	1	118	0.074
	Manipulation x Learning effect	38.345	1	118	0.000
Persian	Manipulation	1.400	1	118	0.239
	Learning effect	1.176	1	118	0.280
	Manipulation x Learning effect	5.329	1	118	0.023
Russian	Manipulation	0.780	1	118	0.379
	Learning effect	2.710	1	118	0.102
	Manipulation x Learning effect	1.211	1	118	0.273
All	Manipulation	24.891	1	838	0.000
	Learning effect	22.600	1	838	0.000
	Manipulation x Learning effect	41.898	1	838	0.000

x = interaction

Table 11: Close-original intonation (COI) and intonation manipulated (I) stimuli for Rest Data. The table shows a 3 factorial analysis of variance for independent samples with factors manipulation (COI and I), learning effect and L1.

Factors	F	df	Error df	Sig.
L1	32.026	6	1652	0.000
Manipulation	20.529	1	1652	0.000
Learning effect	3.895	1	1652	0.049

L1 x Manipulation	4.154	6	1652	0.000
L1 x Learning effect	3.171	6	1652	0.004
Manipulation x Learning effect	7.182	1	1652	0.007
L1 x Manipulation x Learning effect	1.272	6	1652	0.267

x = interaction

Table 12: Close-original intonation (COI) and intonation manipulated (I) stimuli for Rest Data. The table shows a 2 factorial analysis of variance for independent samples with factors manipulation (COI and I) and learning effect.

L1	Factors	F	df	Error df	Sig.
English	Manipulation	15.941	1	236	0.000
	Learning effect	0.016	1	236	0.898
	Manipulation x Learning effect	3.115	1	236	0.079
German	Manipulation	33.375	1	236	0.000
	Learning effect	0.879	1	236	0.350
	Manipulation x Learning effect	0.013	1	236	0.910
French	Manipulation	1.865	1	236	0.173
	Learning effect	0.004	1	236	0.950
	Manipulation x Learning effect	7.315	1	236	0.007
Tamil	Manipulation	0.081	1	236	0.776
	Learning effect	6.080	1	236	0.014
	Manipulation x Learning effect	0.537	1	236	0.464
Chinese	Manipulation	0.093	1	236	0.761
	Learning effect	12.180	1	236	0.001
	Manipulation x Learning effect	3.603	1	236	0.059
Persian	Manipulation	0.886	1	236	0.347
	Learning effect	0.535	1	236	0.465
	Manipulation x Learning effect	0.111	1	236	0.739
Russian	Manipulation	0.386	1	236	0.535
	Learning effect	1.412	1	236	0.236
	Manipulation x Learning effect	0.244	1	236	0.622

x = interaction

Table 13: Close-original duration (COD) and duration manipulated (D) stimuli for All Data. The table shows a 3 factorial analysis of variance for independent samples with factors manipulation (D and COD), learning effect and L1.

Factors	F	df	Error df	Sig.
L1	56.449	6	2408	0.000
Manipulation	27.832	1	2408	0.000

Learning effect	64.429	1	2408	0.000
L1 x Manipulation	1.048	6	2408	0.392
L1 x Learning effect	1.397	6	2408	0.212
Manipulation x Learning effect	8.185	1	2408	0.004
L1 x Manipulation x Learning effect	0.776	6	2408	0.589

x = interaction

Table 14: Close-original duration (COD) and duration manipulated (D) stimuli for All Data. The table shows a 2 factorial analysis of variance for independent samples with factors manipulation (D and COD) and learning effect.

L1	Factors	F	df	Error df	Sig.
English	Manipulation	3.468	1	344	0.063
	Learning effect	5.649	1	344	0.018
	Manipulation x Learning effect	1.293	1	344	0.256
German	Manipulation	2.411	1	344	0.121
	Learning effect	2.528	1	344	0.113
	Manipulation x Learning effect	0.069	1	344	0.793
French	Manipulation	5.326	1	344	0.022
	Learning effect	17.003	1	344	0.000
	Manipulation x Learning effect	6.948	1	344	0.009
Tamil	Manipulation	4.250	1	344	0.040
	Learning effect	15.789	1	344	0.000
	Manipulation x Learning effect	2.431	1	344	0.120
Chinese	Manipulation	1.396	1	344	0.238
	Learning effect	12.947	1	344	0.000
	Manipulation x Learning effect	0.053	1	344	0.819
Persian	Manipulation	11.295	1	344	0.001
	Learning effect	9.719	1	344	0.002
	Manipulation x Learning effect	1.424	1	344	0.234
Russian	Manipulation	1.843	1	344	0.175
	Learning effect	4.944	1	344	0.027
	Manipulation x Learning effect	0.271	1	344	0.603

x = interaction

Table 15: Close-original duration (COD) and duration manipulated (D) stimuli for Paired Data. The table shows a 2 factorial analysis of variance for repeated measures with factors manipulation (COD and D) and learning effect.

L1	Factors	F	df	Error df	Sig.
English	Manipulation	1.446	1	118	0.232

	Learning effect	0.984	1	118	0.323
	Manipulation x Learning effect	5.642	1	118	0.019
German	Manipulation	0.141	1	118	0.708
	Learning effect	0.251	1	118	0.617
	Manipulation x Learning effect	2.615	1	118	0.109
French	Manipulation	4.364	1	118	0.039
	Learning effect	4.054	1	118	0.046
	Manipulation x Learning effect	21.351	1	118	0.000
Tamil	Manipulation	3.989	1	118	0.048
	Learning effect	1.861	1	118	0.175
	Manipulation x Learning effect	21.899	1	118	0.000
Chinese	Manipulation	0.064	1	118	0.801
	Learning effect	0.089	1	118	0.766
	Manipulation x Learning effect	15.945	1	118	0.000
Persian	Manipulation	7.613	1	118	0.007
	Learning effect	0.667	1	118	0.416
	Manipulation x Learning effect	18.974	1	118	0.000
Russian	Manipulation	2.413	1	118	0.123
	Learning effect	0.039	1	118	0.845
	Manipulation x Learning effect	8.641	1	118	0.004
All	Manipulation	14.332	1	838	0.000
	Learning effect	4.698	1	838	0.030
	Manipulation x Learning effect	88.287	1	838	0.000

x = interaction

Table 16: Close-original duration (COD) and duration manipulated (D) stimuli for Rest Data. The table shows a 3 factorial analysis of variance for independent samples with factors manipulation (COD and D), learning effect and L1.

Factors	F	df	Error df	P
L1	44.776	6	1568	0.000
Manipulation	43.888	1	1568	0.000
Learning effect	54.439	1	1568	0.000
L1 x Manipulation	1.327	6	1568	0.242
L1 x Learning effect	1.250	6	1568	0.278
Manipulation x Learning effect	3.780	1	1568	0.052
L1 x Manipulation x Learning effect	0.809	6	1568	0.563

x = interaction

Table 17: Close-original duration (COD) and duration manipulated (D) stimuli for Rest Data. The table shows a 2 factorial analysis of variance for independent samples with factors manipulation (COD and D) and learning effect.

L1	Factors	F	df	Error df	Sig.
English	Manipulation	5.122	1	224	0.025
	Learning effect	4.682	1	224	0.032
	Manipulation x Learning effect	0.713	1	224	0.399
German	Manipulation	6.361	1	224	0.012
	Learning effect	3.386	1	224	0.067
	Manipulation x Learning effect	0.094	1	224	0.759
French	Manipulation	6.446	1	224	0.012
	Learning effect	14.651	1	224	0.000
	Manipulation x Learning effect	3.912	1	224	0.049
Tamil	Manipulation	5.175	1	224	0.024
	Learning effect	14.044	1	224	0.000
	Manipulation x Learning effect	0.984	1	224	0.322
Chinese	Manipulation	5.396	1	224	0.021
	Learning effect	13.138	1	224	0.000
	Manipulation x Learning effect	0.166	1	224	0.684
Persian	Manipulation	17.306	1	224	0.000
	Learning effect	6.367	1	224	0.012
	Manipulation x Learning effect	1.931	1	224	0.166
Russian	Manipulation	1.256	1	224	0.264
	Learning effect	2.481	1	224	0.117
	Manipulation x Learning effect	0.456	1	224	0.500

x = interaction

Table 18: Intonation manipulated (ID) and intonation-duration manipulated (I) stimuli for All Data. The table shows a 3 factorial analysis of variance for independent samples with factors manipulation (ID and I), learning effect and L1.

Factors	F	df	Error df	Sig.
L1	28.125	6	2450	0.000
Manipulation	14.549	1	2450	0.000
Learning effect	43.014	1	2450	0.000
L1 x Manipulation	4.558	6	2450	0.000
L1 x Learning effect	1.912	6	2450	0.075
Manipulation x Learning effect	3.952	1	2450	0.047
L1 x Manipulation x Learning effect	1.052	6	2450	0.390

x = interaction

Table 19: Intonation manipulated (ID) and intonation-duration manipulated (I) stimuli for All Data. The table shows a 2 factorial analysis of variance for independent samples with factors manipulation (ID and I) and learning effect.

L1	Factors	F	df	Error df	Sig.
English	Manipulation	16.784	1	350	0.000
	Learning effect	8.439	1	350	0.004
	Manipulation x Learning effect	0.064	1	350	0.800
German	Manipulation	0.043	1	350	0.837
	Learning effect	4.750	1	350	0.030
	Manipulation x Learning effect	0.468	1	350	0.494
French	Manipulation	2.193	1	350	0.140
	Learning effect	8.031	1	350	0.005
	Manipulation x Learning effect	3.228	1	350	0.073
Tamil	Manipulation	11.205	1	350	0.001
	Learning effect	11.790	1	350	0.001
	Manipulation x Learning effect	0.061	1	350	0.804
Chinese	Manipulation	7.737	1	350	0.006
	Learning effect	17.103	1	350	0.000
	Manipulation x Learning effect	4.831	1	350	0.029
Persian	Manipulation	4.186	1	350	0.041
	Learning effect	2.968	1	350	0.086
	Manipulation x Learning effect	0.638	1	350	0.425
Russian	Manipulation	0.408	1	350	0.523
	Learning effect	0.012	1	350	0.914
	Manipulation x Learning effect	0.188	1	350	0.665

x = interaction

Table 20: Duration manipulated (ID) and intonation-duration manipulated (D) stimuli for All Data. The table shows a 3 factorial analysis of variance for independent samples with factors manipulation (ID and D), learning effect and L1.

Factors	F	df	Error df	Sig.
L1	37.701	6	2408	0.000
Manipulation	0.082	1	2408	0.774
Learning effect	21.938	1	2408	0.000
L1 x Manipulation	7.011	6	2408	0.000
L1 x Learning effect	0.634	6	2408	0.703
Manipulation x Learning effect	0.047	1	2408	0.829
L1 x Manipulation x Learning effect	0.460	6	2408	0.838

x = interaction

Table 21: Duration manipulated (ID) and intonation-duration manipulated (D) stimuli for All Data The table shows a 2 factorial analysis of variance for independent samples with factors manipulation (ID and D) and learning effect.

L1	Factors	F	df	Error df	Sig.
English	Manipulation	0.115	1	344	0.735
	Learning effect	4.267	1	344	0.040
	Manipulation x Learning effect	0.768	1	344	0.381
German	Manipulation	22.155	1	344	0.000
	Learning effect	4.496	1	344	0.035
	Manipulation x Learning effect	0.651	1	344	0.420
French	Manipulation	3.446	1	344	0.064
	Learning effect	1.547	1	344	0.214
	Manipulation x Learning effect	0.054	1	344	0.816
Tamil	Manipulation	0.362	1	344	0.548
	Learning effect	6.998	1	344	0.009
	Manipulation x Learning effect	0.129	1	344	0.719
Chinese	Manipulation	4.096	1	344	0.044
	Learning effect	5.936	1	344	0.015
	Manipulation x Learning effect	0.355	1	344	0.551
Persian	Manipulation	6.577	1	344	0.011
	Learning effect	1.820	1	344	0.178
	Manipulation x Learning effect	0.287	1	344	0.592
Russian	Manipulation	6.884	1	344	0.009
	Learning effect	0.325	1	344	0.569
	Manipulation x Learning effect	0.837	1	344	0.361

x = interaction

Table 22: Intonation manipulated (I) and duration manipulated (D) stimuli for All Data. The table shows a 3 factorial analysis of variance for independent samples with factors manipulation (I and D), learning effect and L1.

Factors	F	df	Error df	Sig.
L1	55.737	6	3206	0.000
Manipulation	18.041	1	3206	0.000
Learning effect	66.890	1	3206	0.000
L1 x Manipulation	10.630	6	3206	0.000
L1 x Learning effect	2.914	6	3206	0.008
Manipulation x Learning effect	4.541	1	3206	0.033
L1 x Manipulation x Learning effect	0.884	6	3206	0.506

x = interaction

Table 23: Intonation manipulated (I) and duration manipulated (D) stimuli for All Data. The table shows a 2 factorial analysis of variance for independent samples with factors manipulation (I and D) and learning effect.

L1	Factors	F	df	Error df	Sig.
English	Manipulation	28.876	1	458	0.000
	Learning effect	6.667	1	458	0.010
	Manipulation x Learning effect	1.621	1	458	0.204
German	Manipulation	28.078	1	458	0.000
	Learning effect	2.804	1	458	0.095
	Manipulation x Learning effect	0.016	1	458	0.899
French	Manipulation	16.980	1	458	0.000
	Learning effect	13.989	1	458	0.000
	Manipulation x Learning effect	3.599	1	458	0.058
Tamil	Manipulation	10.711	1	458	0.001
	Learning effect	13.598	1	458	0.000
	Manipulation x Learning effect	0.678	1	458	0.448
Chinese	Manipulation	0.678	1	458	0.411
	Learning effect	36.364	1	458	0.000
	Manipulation x Learning effect	3.951	1	458	0.047
Persian	Manipulation	0.841	1	458	0.360
	Learning effect	6.080	1	458	0.014
	Manipulation x Learning effect	0.041	1	458	0.840
Russian	Manipulation	6.121	1	458	0.014
	Learning effect	1.704	1	458	0.192
	Manipulation x Learning effect	0.336	1	458	0.562

x = interaction

Appendix C

The tables show results from statistical tests investigating effects of manipulation details on degree of foreign accent as discussed in Chapter 3. The boundary for statistical significance is $p < 0.05$. Statistical significance is marked with grey shade.

Table 1: Regression analyses correlating size of duration manipulation with size of manipulation effect for various details of the manipulation.

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
All segments	0.012	0.006	0.460	1.790	0.098
All consonants	0.014	0.005	0.655	3.034	0.011
All vowels	0.003	0.003	0.202	0.938	0.368
Phon. long V.	0.001	0.003	0.071	0.246	0.810
V/C ratio	0.000	0.001	0.036	0.124	0.904
Articulation rate	0.078	0.014	0.842	5.398	0.000

Dependent variable: Effect of duration manipulation (rated difference between the stimuli in the O_D stimulus pair across all listeners).

Table 2: Regression analyses correlating size of intonation manipulation with size of manipulation effect within and between words.

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Across 6 slopes	-0.003	0.003	-0.276	-0.995	0.340
Word 1, slope 1	0.002	0.005	0.368	0.420	0.692
Word 1, slope 2	0.003	0.003	0.497	0.911	0.404
Word 2, slope 1	-0.003	0.004	-0.622	-0.774	0.474
Word 2, slope 2	-0.005	0.004	-0.609	-1.261	0.263
Word 3, slope 1	-0.003	0.004	-0.444	-0.646	0.547
Word 3, slope 2	0.004	0.004	0.766	1.161	0.298
Word 1 - word 2	-0.019	0.017	-0.742	-1.092	0.325
Word 2 - word 3	0.033	0.034	0.866	0.966	0.378

Dependent variable: Effect of intonation manipulation (rated difference between the stimuli in the O_I stimulus pair across all listeners).

Table 3: Regression analyses correlating size of intonation manipulation with size of manipulation effect within and between words. Manipulations that have included changing the direction of the slope (upwards/downwards) have been weighted with an arbitrary factor (multiplied with a factor 2).

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Across 6 slopes	-0.001	0.001	-0.244	-0.873	0.400
Word 1, slope 1	0.001	0.002	0.176	0.355	0.737
Word 1, slope 2	0.001	0.001	0.478	1.588	0.173

Word 2, slope 1	-0.001	0.001	-0.532	-1.163	0.297
Word 2, slope 2	-0.002	0.001	-0.453	-1.317	0.245
Word 3, slope 1	-0.002	0.001	-0.571	-1.344	0.237
Word 3, slope 2	0.002	0.001	0.638	1.899	0.116
Word 1 - word 2	-0.005	0.003	-0.506	-1.504	0.193
Word 2 - word 3	0.008	0.005	0.550	1.585	0.174

Dependent variable: Effect of intonation manipulation (rated difference between the stimuli in the O_I stimulus pair across all listeners).

Appendix D

The tables show results from statistical tests investigating effects of manipulation details on intelligibility as discussed in chapter 3. The boundary for statistical significance is $p < 0.05$. Statistical significance is marked with grey shade.

Table 1: Regression analyses correlating size of duration manipulation with size of manipulation effect for various details of the manipulation.

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
All segments	-0.016	0.045	-0.054	-0.342	0.734
All consonants	-0.032	0.047	-0.119	-0.695	0.491
All vowels	0.071	0.080	0.151	0.885	0.382
Phon. long V.	0.079	0.094	0.133	0.837	0.408
V/C ratio	-0.009	0.019	-0.070	-0.445	0.659
Pause	0.025	0.056	0.069	0.435	0.666
Articulation rate	0.215	6.265	0.005	0.034	0.973

Dependent variable: Intelligibility score difference COD/D (data with learning effects).

Table 2: Regression analyses correlating size of duration manipulation with size of manipulation effect for various details of the manipulation.

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
All segments	-0.015	0.063	-0.038	-0.240	0.812
All consonants	-0.036	0.065	-0.094	-0.549	0.586
All vowels	0.074	0.112	0.114	0.664	0.511
Phon. long V.	0.141	0.130	0.171	1.086	0.284
V/C ratio	-0.022	0.027	-0.130	-0.831	0.411
Pause	0.061	0.078	0.124	0.787	0.436
Articulation rate	-13.801	46.420	-0.047	-0.297	0.768

Dependent variable: Intelligibility score difference COD/D (data without learning effects)

Table 3: Regression analyses correlating size of duration manipulation with size of manipulation effect for various details of the manipulation. Across the French, Tamil and Persian groups.

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
All segments	-0.093	0.100	-0.315	-0.932	0.371
All consonants	-0.025	0.110	-0.081	-0.228	0.824
All vowels	0.013	0.206	0.024	0.061	0.953
Phon. long V.	0.080	0.327	0.106	0.245	0.811
V/C ratio	-0.071	0.062	-0.355	-1.154	0.273
Pause	-0.058	0.136	-0.164	-0.425	0.679
Articulation rate	1.443	8.390	0.086	0.172	0.872

Dependent variable: Intelligibility score difference COD/D (data with learning effects).

Table 4: Regression analyses correlating size of duration manipulation with size of manipulation effect for various details of the manipulation. Across the French, Tamil and Persian groups.

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
All segments	-0.114	0.152	-0.263	-0.751	0.468
All consonants	-0.025	0.167	-0.055	-0.150	0.883
All vowels	0.041	0.312	0.054	0.130	0.899
Phon. long V.	0.060	0.495	0.054	0.120	0.906
V/C ratio	-0.083	0.094	-0.283	-0.887	0.394
Pause	-0.021	0.205	-0.041	-0.103	0.920
Articulation rate	2.360	16.855	0.070	0.140	0.895

Dependent variable: Intelligibility score difference COD/D (data without learning effects).

Table 5: Regression analyses correlating size of intonation manipulation with size of manipulation effect within and between words.

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Across 6 slopes	0.231	0.213	0.167	1.083	0.285
Word 1, slope 1	-0.199	0.201	-0.307	-0.989	0.330
Word 1, slope 2	0.116	0.197	0.196	0.589	0.560
Word 2, slope 1	-0.040	0.228	-0.058	-0.175	0.862
Word 2, slope 2	0.078	0.284	0.094	0.273	0.787
Word 3, slope 1	0.063	0.153	0.086	0.411	0.684
Word 3, slope 2	-0.197	0.196	-0.347	-1.004	0.322
Word 1 - word 2	-0.524	0.548	-0.178	-0.957	0.345
Word 2 - word 3	-0.101	0.505	-0.034	-0.200	0.843

Dependent variable: Intelligibility score difference COI/I (data with learning effects).

Table 6: Regression analyses correlating size of intonation manipulation with size of manipulation effect within and between words.

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Across 6 slopes	0.053	0.197	0.042	0.270	0.789
Word 1, slope 1	0.136	0.162	0.231	0.838	0.408
Word 1, slope 2	-0.024	0.160	-0.044	-0.149	0.883
Word 2, slope 1	0.166	0.184	0.266	0.903	0.373
Word 2, slope 2	-0.126	0.230	-0.168	-0.546	0.589
Word 3, slope 1	-0.056	0.124	-0.084	-0.454	0.653
Word 3, slope 2	0.340	0.159	0.658	2.143	0.039
Word 1 - word 2	0.262	0.443	0.098	0.592	0.558
Word 2 - word 3	-0.210	0.409	-0.079	-0.515	0.610

Dependent variable: Intelligibility score difference COI/I (data without learning effects).

Table 7: Regression analyses correlating size of intonation manipulation with size of manipulation effect within and between words. For the English and German groups.

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Across 6 slopes	0.415	0.452	-0.224	-0.917	0.373
Word 1, slope 1	-0.228	0.237	-0.258	-0.962	0.357
Word 1, slope 2	-0.430	0.190	-0.593	-2.269	0.044
Word 2, slope 1	0.245	0.202	0.428	1.211	0.251
Word 2, slope 2	-0.018	0.338	-0.014	-0.054	0.958
Word 3, slope 1	-0.037	0.330	-0.035	-0.111	0.914
Word 3, slope 2	-0.037	0.116	-0.086	-0.321	0.754
Word 1 - word 2	-1.892	0.910	-0.926	-2.078	0.129
Word 2 - word 3	-1.856	1.006	-0.807	-1.844	0.162

Dependent variable: Intelligibility score difference COI/I (data with learning effects).

Table 8: Regression analyses correlating size of intonation manipulation with size of manipulation effect within and between words. For the English and German groups.

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Across 6 slopes	0.899	0.578	0.362	1.554	0.140
Word 1, slope 1	-0.012	0.322	-0.010	-0.036	0.972
Word 1, slope 2	-0.127	0.258	-0.131	-0.495	0.631
Word 2, slope 1	0.217	0.275	0.284	0.790	0.446
Word 2, slope 2	-0.302	0.460	-0.179	-0.658	0.524
Word 3, slope 1	-0.423	0.448	-0.300	-0.943	0.366
Word 3, slope 2	0.349	0.157	0.602	2.219	0.048
Word 1 - word 2	-1.892	0.910	-0.926	-2.078	0.129
Word 2 - word 3	-1.856	1.006	-0.807	-1.844	0.162

Dependent variable: Intelligibility score difference COI/I (data without learning effects).

Table 9: Regression analyses correlating size of intonation manipulation with size of manipulation effect within and between words. Manipulations that have included changing the direction of the slope (upwards/downwards) have been weighted with an arbitrary factor (multiplied with a factor 2).

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Across 6 slopes	0.068	0.080	0.131	0.848	0.401
Word 1, slope 1	-0.087	0.056	-0.277	-1.541	0.133
Word 1, slope 2	0.084	0.054	0.298	1.558	0.128
Word 2, slope 1	-0.012	0.059	-0.038	-0.209	0.836
Word 2, slope 2	0.074	0.077	0.191	0.965	0.341
Word 3, slope 1	0.064	0.075	0.168	0.852	0.400
Word 3, slope 2	-0.057	0.044	-0.209	-1.282	0.208
Word 1 - word 2	-0.211	0.294	-0.124	-0.718	0.478
Word 2 - word 3	0.206	0.265	0.131	0.777	0.443

Dependent variable: Intelligibility score difference COI/I (data with learning effects).

Table 10: Regression analyses correlating size of intonation manipulation with size of manipulation effect within and between words. Manipulations that have included changing the direction of the slope (upwards/downwards) have been weighted with an arbitrary factor (multiplied with a factor 2).

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Across 6 slopes	0.068	0.080	0.131	0.848	0.401
Word 1, slope 1	-0.129	0.073	-0.314	-1.763	0.087
Word 1, slope 2	0.070	0.069	0.191	1.012	0.319
Word 2, slope 1	-0.002	0.076	-0.005	-0.030	0.977
Word 2, slope 2	0.067	0.101	0.131	0.663	0.512
Word 3, slope 1	0.147	0.102	0.285	1.449	0.157
Word 3, slope 2	-0.112	0.059	-0.303	-1.909	0.065
Word 1 - word 2	-0.406	0.387	-0.180	-1.050	0.301
Word 2 - word 3	0.243	0.344	0.117	0.707	0.485

Dependent variable: Intelligibility score difference COI/I (data without learning effects).

Table 11: Regression analyses correlating size of intonation manipulation with size of manipulation effect within and between words. Manipulations that have included changing the direction of the slope (upwards/downwards) have been weighted with an arbitrary factor (multiplied with a factor 2). For the English and German groups.

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Across 6 syllables	-0.197	0.151	-0.309	-1.301	0.212
Word 1, slope 1	-0.295	0.216	-0.929	-1.367	0.265
Word 1, slope 2	0.354	0.184	1.450	1.921	0.150
Word 2, slope 1	0.133	0.120	0.449	1.107	0.349
Word 2, slope 2	0.315	0.267	0.776	1.178	0.324
Word 3, slope 1	0.118	0.119	0.394	0.990	0.395
Word 3, slope 2	-0.321	0.170	-1.189	-1.888	0.155
Word 1 - word 2	-1.000	0.472	-1.000	-2.118	0.124
Word 2 - word 3	-0.521	0.458	-.470	-1.137	0.338

Dependent variable: Intelligibility score difference COI/I (data with learning effects).

Table 12: Regression analyses correlating size of intonation manipulation with size of manipulation effect within and between words. Manipulations that have included changing the direction of the slope (upwards/downwards) have been weighted with an arbitrary factor (multiplied with a factor 2). For the English and German groups.

Predictor variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Across 6 slopes	-0.388	0.211	-0.418	-1.839	0.085
Word 1, slope 1	-0.592	0.545	-1.417	-1.086	0.391
Word 1, slope 2	0.573	0.552	1.825	1.038	0.408
Word 2, slope 1	0.186	0.184	0.498	1.011	0.418

Word 2, slope 2	0.806	1.087	1.533	0.742	0.536
Word 3, slope 1	0.104	0.243	0.264	0.430	0.709
Word 3, slope 2	-0.698	0.567	-1.904	-1.232	0.343
Word 1 - word 2	-1.549	0.965	-1.200	-1.605	0.250
Word 2 - word 3	-0.405	0.817	-0.289	-0.496	0.669

Dependent variable: Intelligibility score difference COI/I (data without learning effects).

Appendix E

This appendix shows the 60 sentences in the speech material.

1. To barn matet de tamme dyrene.
2. En rotte løp over matten.
3. Jeg serverte spagetti med tomater.
4. Alle barna ropte ”hei” til sauene.
5. Bilen kjørte forbi huset vårt.
6. Verdien sank på grunn av råten.
7. Han takket nei til dessert etter maten.
8. Kjelleren sank i verdi på grunn av rotter og råte.
9. Han surret strikken rundt fingeren så hardt at han hylte.
10. Den fornøyde bukken spiste kartet.
11. Den sure damen spiste sjokolade.
12. Han analyserte grammatikken i setningen.
13. Det sure barnet hylte høyt.
14. Barna hylte fordi ballongen sprakk.
15. Hun kjører gjerne pene biler.
16. Den fine pennen er et minne om møtet.
17. Råten i hylla ble verre.
18. De jaget sauene langt vekk.
19. Hun spiser piller og pastiller.
20. De skyter med dyre piler.
21. Hun kjøpte garn og perler.
22. Været ble verre etter møtet.
23. Maten i hylla ble sur.
24. Bukken og sauen fikk maten.
25. Hun møtte mange høye folk.
26. Noen surret en snor rundt boka.
27. Han møtte forfatteren av den farlige boka.
28. De to kundene hylte til hverandre under møtet.
29. Den tamme katten var kjærlig og noe lat.
30. Hun kjøpte pennen og en ny genser.
31. De pene jentene spiste pastillene.
32. Boka skal være i hylla.
33. Noen skjøt rottene med piler.
34. Han surret tauet rundt mattene.
35. De kjørte da skyene ble svarte.
36. Salaten serveres nøyaktig klokka sju.
37. De dyre pillene skulle tas med maten.
38. Boka har fått høy verdi.
39. Den høye mannen fanget sauen.
40. Bukken spiste sjokoladen i stedet for maten.
41. Pennen lå i hylla under møtet.
42. Hvordan blir været i morgen?
43. Været ble pent da skyene forsvant.
44. Han tok pennen og skrev ut en resept på pillene.

45. Problemet med rottene ble verre.
46. Den pene boka lå på matten.
47. Den sure mannen skjøt pila mot ballongen.
48. Han sa ”hei” til den pene damen da han møtte henne.
49. Råten i veggen var verre enn før.
50. Hun surret skjerfet rundt seg i det sure været.
51. Den gamle bukken kastet på hodet.
52. Den svarte katten la seg kjærlig i fanget hennes.
53. Noen av de dyre mattene var svarte og gule.
54. De nye pillene smakte verre enn den gamle medisinen.
55. Jenta mistet den lange, fine pennen i heisen.
56. Pila falt fort i bakken.
57. Bilen de kjørte var mye verdt.
58. Det kostet mange penger å fjerne råten.
59. Vi møtte en bonde med en bukk og en geit.
60. Den ærlige kjæresten fortalte alt.