Factors in on-line loanword adaptation

Christine Haunz, MA

A thesis submitted in fulfilment of requirements for the degree of Doctor of Philosophy

to

Linguistics and English Language
School of Philosophy, Psychology and Language Sciences
University of Edinburgh

30th March, 2007

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Christine Haunz

Abstract

This thesis investigates the factors influencing the adaptation of foreign words to English, beyond traditional phonological parameters such as sonority distance. The data examined were produced in an on-line adaptation task to study purely linguistic rather than orthographic or historical influences. The adapted words contain only lesser-studied phonotactic problems rather than segmental ill-formedness. The choice of Russian as a donor and English as a borrowing language allow the study of adaptations in a setting which allows a further strategy of alteration of ill-formed consonant clusters beyond vowel epenthesis and consonant deletion, namely the substitution of segments to change one cluster into another. In contrast to previous research, English production of Russian stimuli with initial consonant clusters showed that segment change is applied frequently, comparable to the amount of vowel epenthesis. Extensive variation was observed, both in ratio of successful production, and in the choice and distribution of adaptation strategy.

The factors in adaptation investigated were the sonority distance of the foreign clusters, as well as concepts which have received much recent attention within phonology, namely gradient grammaticality, similarity and frequency: English native speaker judgments were collected about the perceived grammaticality of foreign clusters and the similarity between targets and adaptations, while the frequency of possible adaptations in English was calculated from a corpus of spoken English. Results show that sonority cannot explain the variation in adaptation. Furthermore, frequency has no influence on the choice of adaptation; however, higher perceived badness results in a higher percentage of adaptations, and perceived similarity is decisive for the choice of adaptations. A comparison of similarity judgments of English and Russian listeners suggested that, in keeping with Steriade (2001), there are some cross-linguistically corresponding rankings of similarity; however, differences between languages due to phonotactics and phonetic detail were also found. In summary, the experiment results suggest that the adaptation of loanwords occurs in both in perception and production; furthermore, it is determined both by L1 specifics and cross-linguistic tendencies, an thus neither a straightforward application of L1 phonology nor completely independent of language background.

Acknowledgements

First of all I would like to thank the Carnegie Trust for the Universities of Scotland for supporting my research with a three-year PhD scholarship, without which it would have been impossible for me to embark on this project at the University of Edinburgh.

My gratitude goes to my supervisors, Mits Ota and Bob Ladd. Thank you to Mits for careful attention to detail and always making time, to Bob for knowing everything and suggesting connections to other lines of research, and to both for support and great patience.

I am also indebted to Andreas Mahler in Munich, my first linguistics teacher. Without his inspiring courses and appreciation of my linguistic thoughts I would now be a secondary school teacher of English and Latin and not a linguist.

A big thank you goes to the technical support staff at Edinburgh, both in the recording studio and in computing: Mike Bennett, Morag Brown, Eddie Dubourg, Cedric Macmartin, and Alan Whyte. I cannot count the times you helped me out.

Ian Underwood ensured smooth sailing on my journey into the unknown seas of LaTeX many a time - thank you so much! Thank you also to Cassie Mayo for introducing me to LaTeX and SenSyn, to her and Bert Remijsen for helping me make sense of statistics, and to Alexei Kochetov for input about literature on Russian and stimulus design.

I am much obliged to the Russian speakers who recorded my experiment stimuli: Galina Andreeva, Alexander Gutkin and Ari Krupnik in Edinburgh, and Nina Volskaya of the Phonetics department at St. Petersburg University; and I would like to express my appreciation to all the English and Russian speakers who agreed to sit through my experiments - there are pleasanter pastimes, I am sure!

Finally, I am deeply grateful to my family and friends for their unfailing support. Thank you to Marianne and Ian for giving me good and affordable places to stay during the write-up, to Ruth, Susana, Ivan, Hannele and my office mates for chats, giggles and cuppas, and for sharing the mental state of being a PhD student, to Neil for knowing when to push me to work, and to Chris for boundless optimism, warmth and computer helpfulness.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The aim of this thesis is to examine which factors influence the adaptation of loanwords, i.e. of foreign words entering a speaker's native language. Loanword research has been at the heart of the debate on the resolution of conflicts between two different phonologies; it lends itself to investigating the mechanisms which underlie such a resolution, since in a loanword situation there is a definite previous form of the word available - as it is produced in the language where it originates from - which can be compared to the adapted form in the borrowing language to study the mechanisms of change.

In the literature, there is disagreement on several issues in adaptation, such as whether the adaptation process is a phonological or a phonetic one, and whether the potential influence that a speaker's native language may exert on his or her perceptual system is a decisive or even relevant part in this process. The experimental strategy in this thesis is therefore initially to produce a corpus of adapted loanwords unaffected by non-linguistic influences such as historical developments or orthography; subsequently, it is tested whether factors which have been suggested in the literature on existing loanwords, such as phonological parameters, the native lexicon, L1-specific perception or phonetics, are responsible for the adaptation strategies.

The results of the experiments which are necessary to measure and evaluate factors in adaptation serve to contribute to debates in loanword research as well as to current topics in phonology such as the role of frequency and the relationship between perception and phonology.

1.1 Conflicting phonologies

A number of trends in phonology in recent years have tended away from treating phonology as its own self-contained discipline, and towards investigating the role of its relationship with aspects of language research that are outside what used to be thought of as phonological theory proper. First in line of these is of course phonetics, and the debate about the phonetic grounding (Pierrehumbert 2000) and, within the framework of Optimality Theory, the putative universality of phonological constraints (as put forward in Prince & Smolensky 1993), which may or may not be drawn from phonetic principles, has attracted contradiction (Stemberger & Bernhardt 1999). The interface between phonetics and phonology has been subject of discussion, as in Keating (1989) and Pierrehumbert (1990), or indeed the special issue on the phonetics-phonology interface of the *Journal of Phonetics* which contains Pierrehumbert's paper. Other, closely related areas of investigation have been the relationship between phonology and perception, as in Hume (1999) and Hume & Johnson (2001*b*), and between phonology and the lexicon (Johnson 2004, Broe 1993) as well as frequency of usage (Bybee 2001, Pierrehumbert 2003), and variability and gradience in phonology (Hayes 2000, Boersma & Hayes 2001).

A field of research which is well suited to adding new insights into the relationship of phonology to perception, as well as to frequency, is that of conflict between two phonologies - be it in a setting of historical change, multilingualism or multidialectism, in second language acquisition (or indeed, any kind of language acquisition) or, as in this thesis, loanwords. In the case of new loanwords, listeners are confronted with single words from a foreign language, which are subsequently incorporated into their own. This means that the speakers have to resolve a clash between a new system that provides a certain output (i.e. the word as it is pronounced in the foreign language), and a familiar one for which this output is not well-formed. These clashes often result in modifications to the words of the non-native language to be partly or completely consistent with the native language (L1), i.e. in what is termed *loanword adaptations*. These adaptations are far from an obvious straightforward application of L1 phonology, and so the study of the adaptation of foreign words provides material also for the new directions in phonology listed above. For example, the review of loanword literature will show that the relevance of perceptual similarity and phonetic detail and their influence on loanword adaptation are the most hotly discussed topics in loanword research today; this means that perception and phonetics are discussed with regard to a process which was long regarded as a purely phonological one.

This thesis will therefore examine on-line experimental productions of loanwords, the variation in the output, and the factors bearing on the adaptations. The factors under investigation are well-known phonological principles, as well as concepts which have only recently moved into focus. For example, the gradient grammaticality of foreign structures is explored, as well as the frequency of potential adaptations in the native lexicon, and the perceived similarity (the nature of which is compared cross-linguistically) between a foreign form and potential native adaptations. The implications of the results for loanword adaptation specifically, and the broader issues in current phonological theory, are discussed in chapter 6.

1.2 Loanwords

1.2.1 Types of adaptations

Loanwords are words of one language, termed the source (or donor) language, that enter, often through the mediation of bilingual speakers, a borrowing (or recipient) language. In the process of entering the borrowing language, phonetic, phonemic, phonotactic or prosodic characteristics of these words change in the vast majority of cases. As an example of a phonetic change, the first vowel of the French pronunciation of *souvenir* and its adapted form in English are both transcribed as /u/, yet are phonetically very different.

Phonemic changes occur for example when the borrowing language's inventory lacks a phoneme in the loanword, such as in *gemütlich* /gemytlıç/, where English speakers substitute /u/ and /k/ for the German phonemes /y/ and /ç/, respectively. Further examples are the replacement of English interdentals by voiceless /s, t, f/ and voiced /z, d, v/ by learners with different language backgrounds, and the merger of the English lax high vowels /ɪ, ʊ/ with their tense counterparts /i, u/ by native speakers of languages with five-vowel systems, for example Spanish and Greek.

A further kind of loanword adaptation occurs on the level of syllable structure: when the phonotactics of the borrowing language does not allow sound combinations or sounds in certain contexts as they occur in the borrowed word, phonemes are altered, inserted or deleted to satisfy the requirements of the speakers' L1. For example, in Russian and German word-final obstruents are devoiced, which is often applied also in loanwords from English. In terms of syllable structure, there is an abundance of loanwords of English and French origin to be found in Japanese, or Bantu and Khoisan languages, where the borrowing languages mostly have CV syllable structure and therefore do not allow consonant clusters. In many of these cases vowels are epenthesised between the consonants, however there is also deletion of input consonants. For example, Zulu /ifulegi/ is a reflex of English *flag* (Khumalo 1984); the adaptation of French *voyou* (/vwaju/) in Fula (a Niger-Congo language) is /waju/ (Paradis 1995). There are also interactions between phonemic and phonotactic conflicts, as mentioned by Rochet (1995) for French learners of English in perceptual terms: the most frequent misperception of English /θ/ is /s/, whereas if the same sound is preceding /r/ word-initially, it is instead perceived as /f/, resulting in the - in both English and French - well-formed /fr/ rather than the ill-formed initial /sr/.

Finally, prosodic adaptations take place for example where stress is marked differently or unrecognisably for listeners of the borrowing language (e.g. Shinohara 2004), or where an output form of the source language, for example from English, has to be mapped onto tone languages such as Cantonese (e.g. Silverman 1992) or Yoruba (Kenstowicz 2004).

In some loanword situations, especially where a large amount of foreign vocabulary enters a language, the assimilation to the L1 system may not be complete. As early as 1948, Firth made this observation in his "polysystemic" descriptions of language. He notes that the lexicon of a given language may have different strata, each with a different phonology (Firth 1948). Pike & Fries (1949) discussed the issue, naming it "coexistent phonemic systems", using the example of a Central American language, Mazateco, which had been heavily influenced by Spanish, and as a result contained a subset of lexical items that did not conform to Mazateco, but were only incompletely assimilated. In the native vocabulary of Mazateco, stops in postnasal position are always voiced, but in the set of loans from Spanish this is not the case, for example in 'siento' (hundred). Thus native speakers of Mazateco work with two conflicting phonologies. The same phenomenon occurs in Japanese, as discussed by Ito & Mester (1999) (described below): the whole Japanese lexicon can be divided into four extensive sets of lexical items that each comply with a different grammar.

1.2.2 Adaptation scenarios

There are degrees of a loanword being a part of the borrowing language's lexicon, ranging from an elicited form (in an experiment situation), where the speaker has no knowledge of the source language, to a fully established loanword that has been an integral part of the borrowing language for centuries. A further variable is the degree of community bilingualism: Haugen (1950)[1] divides the continuum into three stages:

1. the pre-bilingual period, where only a small number of adults are bilingual and there is a high degree of variability in adapted forms
2. the adult bilingualism period, where most adults are bilingual and adapted forms show increased uniformity,
3. childhood bilingualism, where children grow up bilingually and thus do not have to learn the second language as adults

Furthermore, a non-bilingual scenario can be added, such as, in the very early stages of language contact, monolingual speakers who only adopt single words from

[1] quoted in Paradis & LaCharité (1997)

a language without having acquired or being in the process of acquiring the source language.

To determine whether the status of the majority of speakers in terms of bilingualism is relevant to loanword research, it is necessary to establish whether knowledge of the source language has any influence on adaptations (cf. Silverman 1992 in section 1.3.2). Either way, it could be argued that in most communities there is a sizable proportion of monolingual speakers, in which case the loanword form that will eventually be established will be shaped by the monolingual, not the bilingual speakers. In this study the issue has been excluded by investigating only adaptations by speakers without knowledge of the source language.

1.3 Debates

Many of the following issues are interrelated. Some of them are specific to loanwords; some have been discussed mainly regarding phonology proper, some mainly within second language acquisition (SLA). However, all pertain to the question of loanword phonology, or rather, the relationship of loanword adaptations to phonological systems.

1.3.1 The relationship of loanwords to second language acquisition

Silverman (1992, described in 1.3.2) claims that even bilinguals will not employ their knowledge of the source language's phonology when they perceive and adapt loanwords; this means that, for example, in a Cantonese setting, a Cantonese/English bilingual would perceive English words like a monolingual Cantonese speaker does. Silverman cites evidence of an experiment in which English/Spanish bilinguals used different VOT boundaries depending on which language the experiment was set in, i.e. switched languages (Elman et al. 1977).[2] Thus, loanwords and second language acquisition may be different in whether knowledge of the source or second language is influencing the non-native speaker productions or not.

On the other hand, it is reasonable to assume that the patterns will be similar: what is often produced non-targetlike by learners in language acquisition is likely to be changed in loanword adaptation, too. For example, native speakers of German often substitute /ɛ/ for the /a/ or /æ/ of English *band*, both when producing English utterances as learners, and when using *band* as a loanword in a German context. Furthermore, the changes observed between source form and adapted form may afford insights into the phonology of the borrowing language which cannot be gained from native data (in optimality-theoretic terms, hidden constraint rankings

[2]However, this is not generalisable for all loanword situations, since the speakers were merely shifting the boundary on a continuum which was the basis of classification for both languages, but it is not a case of phonemes being present in one language, but not the other.

that account for variable degrees of assimilation, as in Davidson 2000, Ito & Mester 1999; examples are described in the next section). This additional information may account for acquisition difficulties specific to learners with a specific first language.

Nevertheless, it is important to remember that the aims of loanword adaptation and second language acquisition are different: whereas in adaptation, the goal is communication with fellow speakers of the borrowing language, the goal of second language acquisition is to converse with speakers of the foreign language. This means that there is little incentive to switch out of one's native system with loanwords; speakers rather incorporate foreign elements to fit. In acquiring a language, learners are attempting to modify their grammar permanently by utilising the information contained in the foreign words they encounter, and to produce foreign words target-like in order to be understood by native speakers of the relevant language. Thus, while it would be a mistake to ignore findings from second language acquisition in loanword research, we need to keep in mind that there can be no one-to-one translation of results.

1.3.2 L1 phonology and loanword phonology

Loanword processes are caused by the borrowing language's phonology in that the result is greater conformity of the loanword to this phonology. Nevertheless, there is disagreement as to the relationship between the native phonology and a putative loanword phonology (in fact, there is disagreement even on whether loanword phonology is phonology at all). The question whether these adaptations are purely grammar-driven, or if other factors such as perception also play a role, will be discussed below.

Traditionally, loanword research has viewed the adaptation process as a purely phonological one, with the input of the loanword production process being the phonological surface form of the word in the donor language, the output the surface form in the borrowing language, and the differences between the two grammars resulting in the mapping of one onto the other, through generative rules. In contrast to using specific phonological rules producing the adapted form of the loanword, it has been debated whether there is a cross-linguistically applied strategy (Paradis & LaCharité 1997), i.e. one that is active in all loanword adaptation processes, or whether the process of adaptation is mostly identical to applying the phonology of the borrowing language.

The claims regarding the relationship between loanword phonology and the native phonology, and between L1 specificity of adaptation and universal strategies, are to some extent connected to the framework one works in, and range from a view of loanword phonology as the native phonology "in action", i.e. consisting of exactly the same grammar (as following naturally from an optimality-theoretic background), to the idea that loanword phonologies are separate from L1, and cross-linguistically more similar to each other than the native phonology of a language is to its loanword counterpart. The latter view implies that there are universal, not language-specific, processes at work. An intermediate view is that, as Singh (1995) put it, loanword phonology is not exactly the same, but a derivative extension of native phonology.

Singh's view of loanword phonology, as indeed of child phonology and native phonology, is as a part of a larger phonological competence, a pool of knowledge, which any speaker of any language is endowed with. Loan phonology cannot necessarily be predicted from native phonology. Evidence for his position is an adaptation process in Yoruba, where a foreign final voiced obstruent is adapted as its voiceless counterpart, although the native vocabulary does not contain any final obstruents.[3] Singh concludes that infants are born endowed with all possible processes, of which they have to discard those which are not part of their native phonology. This is closely related to Stampe's Natural Phonology (e.g. Stampe 1979), and is also reminiscent of the OT idea of a universal set of innate constraints, except that it speaks of processes rather than constraints. Singh's definition of phonology and its acquisition is the following:

> "What is learnt in learning the phonology of a language is the set of conditions that trigger repair, or, to put it differently, the phonology of a language is, primarily, the set of conditions that inhibit these repairs - what is normally called the phonology of a language as a constrained, derivative, delimitation of the phonological competence of its speakers."

Kager (1999) describes problems for rule-based theories as lying mainly in the adaptation of foreign syllable structure and in the duplication of rules in native and loan phonology with the same effect, for example devoicing final consonants in both native and loanwords. Additionally, there is the question how loanword rules deal with sounds that are not contained in the native language. Unless it is stipulated that, underlyingly, a language may contain all possible segments and features (and the processes changing them, as in Singh), or that rules are heavily feature-based rather than specifically about phonemes, sounds such as /θ/ or /x/ pose a problem, as there are no rules dealing specifically with these sounds in many

[3] Within optimality theory, however, a solution for this case is rather straightforward: in the native system faithfulness would be ranked below a constraint against final obstruents, which in turn would be harmonically below a constraint against final *voiced* obstruents. For loanwords, faithfulness is promoted to an intermediate rank, making the effect of the constraint against final voiced obstruents visible.

languages. Thus, for illegal segments, or at least for segments with inactive features, specific loan rules would have to be proposed.

In contrast to the traditional rule-based theories, Paradis & LaCharité (1997) propose a Theory of Constraints and Repair Strategies (TCRS) to deal with loanword adaptation. TCRS is a loan phonology theory suggesting that adaptations are aimed at fulfilling language-specific constraints; the (serial) derivations are, however, bound by inviolable, universal meta-constraints about adaptation behaviour, such as the Preservation Principle (favouring epenthesis over deletion), the Minimality Principle (demanding as little change as possible on the lowest phonological level, for example on the segmental level rather than the syllabic one), and the Threshold Principle (if a universal threshold of two repairs within a domain would be overstepped, deletion instead of epenthesis occurs). These work according to a Phonological Hierarchy, which defines this minimality in a hierarchy of phonological constituents; syllables are more highly ranked than segments, which are in turn higher than features with dependent feature, which is higher than a terminal feature. Repairs should apply at the lowest possible level and in minimal numbers. If the universal threshold of two repairs is violated, deletion occurs rather than change. This results in the favouring of epenthesis over deletion, which is exemplified in the overwhelming part of their data. The theory can be exemplified through the different adaptation strategies for French *classe* /klas/ and *voyou* /vwaju/ to Fula /kalas/ and /waju/, respectively. In one case, an illegal cluster is resolved through vowel epenthesis, in the other one of the consonants is deleted. Since both /k/ and /l/ are legal Fula segments, the two steps of repair for /klas/ consist in nucleus insertion and the spreading of the following vowel. /v/, however, is not a phoneme of Fula; hence the third step of adapting /v/ to a legal segment would be necessary. However, this procedure exceeds the threshold of two repair steps and is therefore abandoned in favour of /v/ deletion.

TCRS holds to some extent an intermediate place between rule- and constraint-based theories. The question of defining what exactly constitutes a repair step, as well as the rigid universal threshold of two repairs not allowing for different settings, could be perceived as weaknesses of this model.

When evaluating their claims, one should also bear in mind that much of the idea of universality in loanword adaptation may reflect the history of political, economic and cultural imperialism: a substantial amount of loanwords originate in English or French, adapted into the various native languages of the peoples in the empires, a great many of which are CV languages or avoid clusters, whereas in English and French much more marked syllable structures are possible. Thus many adaptation processes are similar in that they break up clusters and resyllabify codas as onsets, and putative universal strategies have much the same effect as language-specific ones. Indeed, one would expect a certain similarity among loanword adaptation, since, if there is an adaptation, the direction is from marked to more unmarked structures, from as from complex to single onsets, but not the reverse.

With the rise of Optimality Theory a view opposite to the one exemplified by Paradis' TCRS has gained ground, i.e. that loanword phonology is not a cross-linguistically applied strategy which exists in addition to L1 phonology, but that loanword phonology is merely the native phonology "in action", where the input is evaluated by the same hierarchy of constraints, possibly with a promotion of faithfulness in the hierarchy applied to foreign words (Ito & Mester 1999, Davidson 2001).

For example, Lombardi (2000) deals with the issue of differential substitution (mostly in second language acquisition), where a foreign sound is altered differently by speakers of different languages or in diverse contexts. In the case of the English interdentals, Weinberger (1990) had used underspecification theory to claim that the respective substitutions are the most underspecified segments of the borrowing languages. Conversely, Lombardi argues that in some languages it is required that a constraint which militates against a changing of the value of [continuant] is split in two for continuants and stops, which results in /θ/ retaining frication in one case, but being altered to a stop in the other. Lombardi's explanation is in accord with the data. Alternatively, differential substitution may be stemming from differential perception, as discussed in 1.3.4.

Ito & Mester (1999) also approach the issue from an optimality theoretic background. They show that in Japanese, beginning with the native core vocabulary (Yamato), via the set words of Chinese origin, to recent loans for instance from English each sublexicon respects a smaller set of the restrictions of Japanese phonology, i.e. remains closer (more faithful) to the word in the donor language than Japanese phonology would suggest. In terms of optimality theory, Ito & Mester's adopted framework, this means that faithfulness (F) constraints are ranked differently with respect to markedness (M) constraints for different parts of the lexicon. For all of them, within the sets of M and F constraints the hierarchy is the same. However, F constraints are ranked differently with respect to markedness, so the strata are schematically defined by the following hierarchies[4]:

[4]The numbers express the ranking within F and M constraints, respectively; for example, F1 >> F2, i.e. F1 is ranked higher than F2.

1. M1 >> M2 >> M3 >> M4 >> all F
2. M1 >> M2 >> M3 >> F1 >> M4 >> F2
3. M1 >> M2 >> F1 >> M3 >> F2 >> M4
4. M1 >> F1 >> M2 >> M3 >> F2 >> M4

The first stratum will apply to the core lexicon, the other strata are more and more peripheral, i.e. they observe continually fewer of the markedness constraints. This way, new additions to the vocabulary will be in the most peripheral stratum, and their form will be preserved rather than assimilated.

Following on, Ito & Mester (1999) apply this to the issue of impossible nativisation, and determine L1 constraint rankings which cannot be determined on the basis of L1 data. These are rankings which only have an influence on native speakers' output when they are confronted with a conflicting system, for example incoming loanwords. One example is that of the German adaptation of English *story*, /stɔɹi/, where two possible repairs are /st/ > /ʃt/ and /ɹ/ > /ʀ/, yielding the fully nativised /ʃtɔʀi/. However, not all combinations of these repairs are possible. The concept of impossible nativisations is introduced, which are theoretically possible, partially adapted forms (here: /ʃtɔɹi/), which do not occur, while other partially adapted forms exist (here: /stɔʀi/). From these hidden rankings can be established, in this case *ɹ> *[sC (the constraint against the approximant /r/ is ranked above the constraint against initial /s/ followed by a consonant).

In a similar vein, Davidson (2000) attempts to establish these hidden rankings by asking English speakers to produce Polish CC clusters which are prohibited in English. Her method is to attempt to keep speakers in their monolingual English mode whilst eliciting renditions of Polish nonsense words, thus ensuring the employment of the English grammar rather than of what she terms a "foreign register". Her results show convincingly that, of a selection of CC clusters illegal in English, not all are equally unacceptable and prone to adaptation, but some are more easily imitated by English speakers than others. Again, a hierarchy of hidden rankings is established, however with a large number of constraints composed of local conjunctions of up to four separate lower-ranked constraints. The output forms differ only in the rank that is assigned to faithfulness in any particular instance, which is highly dependent on the speaker and the situation.

According to Ito & Mester and Davidson, in the optimality theory view, loanwords can provide evidence for the constraint ranking within the native phonology, and predict which nativisation processes are possible and which are not. A solely rule-based grammar cannot predict which of its rules may be abandoned in loanword adaptation and which cannot. Similarly, Paradis' Theory of Constraints and

Repair Strategies, as an account with inviolable universal claims about the nature of adaptation, cannot predict differences such as in the Japanese strata (for example, between the older, Chinese borrowings, and the recent, mostly English ones). Rule-based theories usually concentrate on one level of borrowings rather than explaining differences between levels. optimality theory, on the other hand, may not predict which adaptations will take place, but can predict what is possible and what is not, as in the above example of English *story* being adapted in German.

1.3.3 The input - perception and production of loanwords

At first glance it seems clear what the input to a borrower's native phonology is: the output of a word in a different language. It is not quite as easy as it seems, however: the output that borrowers are confronted with is an acoustic signal and not a phonological representation - i.e. not the same kind of representation as the input to the native language production phonology.

The question which arises is whether or not an additional transformation is necessary - preceding the application of loanword phonology - to arrive at the phonological representation of the foreign word, which is then modified by loanword phonology. Does the process of mapping the incoming, acoustically detailed speech signal onto a linguistic representation - i.e. what will be referred to as *perception* in this thesis - result in a change from the source language surface form? If perception takes place without difficulty like the perception of native words, then the input representation is already fixed. If, on the other hand, the perception process works differently for native and foreign input, for example if it filters out certain structures which do not comply with the borrower's phonology, further questions must be answered: what is the result of this filtering process? How is it influenced by the first language? How are its results converted into a phonological representation? Here loanword research can be informed by results from other areas, such as second language acquisition, and then existing loanword data can be re-evaluated to establish the relationship of loanword adaptations to phonology.

Until recently for many researchers, apart from Silverman (1992), this had not been an issue at all, in that they tacitly assume that borrowers can utilise at least features, segments and stress patterns of the original output, if not syllable structure. On the other hand, in research on second language acquisition, perception plays an important role: the perception of foreign phonemes is influenced considerably by the learners' native phonology and thus a major source of mistakes and difficulties with the phonological system of the learned language. Learners tend to project new phones onto phonemes of their first language; for example, French learners often perceive English [θ] as /s/ Rochet (1995). These had been mostly ignored in loanword research, although it is all the more likely to be of interest for loanwords, which are adapted not only by bilinguals and language learners, but also by monolinguals.

However, in the wake of Silverman's article splitting the adaptation process into Perceptual and Operational Levels, the question of the input has developed into the key issue of loanword research, and is hotly discussed in current work. Rather than explicitly referring to the input to loan (or L1) phonology, the issue is now called the perception/production debate: depending on the viewpoint taken, the input may not be a relevant concept in loanword adaptation anymore.

There are three main points of view: one, loanwords are altered in production only (as was previously assumed); two, adaptations occur both in perception and production; and finally, adaptations occur in perception only.

Production only

The traditional point of view is stated expressly in Paradis and LaCharité's loanword theory, TCRS, as well as by Jacobs & Gussenhoven (2000) in a reanalysis of Cantonese loanword data discussed in Silverman (1992) and Yip (1993). They claim that humans possess a "universal phonological vocabulary", with the aid of which borrowers analyse the foreign word as spoken by native speakers, and translate it into their own representation, which matches the output of the source language exactly. Whether the native language has the sounds and features in question in its inventory or grammar is irrelevant, as the speakers' competence includes all possible sounds of all languages equally. This extreme position is hardly tenable: firstly, although infants are very good at discriminating all speech sounds until about six months of age, this ability decreases the more children are exposed to input of the native language (Werker & Tees 1984). Secondly, it contradicts abundant second-language evidence that language learners do have difficulty perceiving some non-native contrasts. Moreover, the realisations of the same phoneme in different languages differ greatly in their phonetic detail, such as spectral specifications, or voice onset times (VOT) for consonants.

Like Jacobs and Gussenhoven's, Steriade's (e.g. Steriade 2005) view of loanword adaptation does not include a direct contribution of perception to the adaptation: in her model the input to the production phonology is not altered from the surface form in the source language. However, the production phonology which does effect the adaptations is, in her optimality-theoretic view, heavily influenced by perceptual considerations: the ranking of the output-output faithfulness constraints which militate against adaptations is dependent on how perceptible the relevant change is. The more perceptible the alteration, the higher the corresponding faithfulness constraint is ranked. This results in less perceptible changes such as devoicing being cross-linguistically predominant, for example in loanwords. Miao (2005) uses the same approach and adds experimental support for a correspondence between adaptations and what are perceived to be the most similar forms, for loanwords adapted into Mandarin Chinese.

Kang (2003) and Yip (2002) water down the strictness of Steriade's approach to some extent by, whilst still viewing production as the locus of adaptation, making limited reference to the phonetic surface form of the foreign word. Yip (2002) adds a family of freely rankable, loanword-specific MIMIC constraints to her production grammar, which "relates the output to a specific sub-type of input, a demonstrably foreign form" (p.5). For example, MIMIC-QUALITY favours the retention of similar formant values of a foreign vowel and its adapted form. Kang (2003), in her analysis of loanwords in Korean, proposes a constraint family promoting perceptual similarity, called BESIMILAR, in favour of input and output strings corresponding in certain characteristics, such as consonant releases. Again, this necessitates that the input of loanword adaptation is unlike a standard phonological underlying representation in generative phonology, but that some phonetic information is retained in the input to the production grammar. These models form a kind of intermediate step between the "production only" and the "perception and production" approaches, by making reference to perceptual phonetic similarity between foreign and adapted form, albeit without adaptation taking place during the perception process. This solution has a theoretical drawback: since it must contain a certain amount of phonetic information, e.g. about consonant releases, the characteristics of the loanword input differs from the L1 underlying forms, whilst nevertheless applying the same production phonology (plus added faithfulness constraints to them).

Perception and production

The view that perception contributes to adaptation directly was first put forward by Silverman, and Yip (1993), analysing the same data set of Cantonese loanwords, the former in terms of rules, the latter within optimality theory. They both subscribe to a Perceptual (acoustic) Level (Scansion 1) taking place before any phonological adaptations, and selecting in terms of the segment inventory of the native language: segments that cannot be mapped onto "native feature matrices" are represented as a native segment that they closely resemble. This level acts as a filter, which only well-formed segments pass through and provides syllable nodes for salient consonants such as /s/. For example, since Cantonese does not have contrastive voicing, the result of the perceptual scan of a voiced stop, such as in English *game*, is voiceless, as in [kɛm]. Only then the Operational Level (Scansion 2) scans this sequence of well-formed native segments, which may, however, violate structural constraints, for example by being in a disallowed syllable position, and effects necessary structural changes. Silverman and Yip agree that phonotactic differences between two languages do not influence the perception of foreign words, and are only resolved in production, for example by vowel epenthesis breaking up clusters or consonant deletion (for some non-salient problematic consonants which have not been assigned a syllable node in Scansion 1), and the occlusivisation of fricatives to avoid disallowed fricative codas.

Silverman's work has unquestionably made an important point: before applying a phonological grammar to loanwords, the role of perception must be made clear. Intuitively, this distinction between a perceptual and operational level equals roughly a distinction between segmental (as well as tonal) and syllable-structural adaptations. On the other hand, this kind of clear-cut division may not always be possible, and the application of phonological feature matrices during the perception of phonetic material may be conceptually problematic. Also, this would appear to pose a challenge for optimality theory, as used by Yip: the segment inventory and phonology are not understood as two independent entities, but crucially depend on each other. If there were two levels, this would mean that the same grammar is at work twice, once with the "inventory part" of the grammar (i.e. filtering out illegal segments during the Perceptual Level), and subsequently with the phonotactic part (repairing syllable structure in the Operational Level), which is problematic in this framework.

Salience is also a difficult issue, most obviously for Yip, as Jacobs & Gussenhoven (2000) remark: she employs a constraint PARSE (salient), which is treated as universal - as constraints are in optimality theory, and also because it is employed in loanword adaptation cross-linguistically -, although research in second language suggests an influence of L1 on speech perception (see section 1.3.4/2). Also, if the perceptual level is indeed pre-grammatical, as Silverman claims, it is not immediately obvious how optimality-theoretic constraints fit in to decide on the salience of elements of the input.

A further problem arises for Silverman's analysis (and for the "perception only" claim described below) when one considers languages with very small phoneme inventories: speakers of Hawaiian would, however many languages they learn, only be able to perceive a minimal amount of contrasts when they are speaking Hawaiian. For example, [s] would always be perceived as /k/, even if the speaker is proficient in English. Similarly, according to Silverman, speakers of English would not be able to perceive clicks or rather, the difference between clicks and plosives, which they have been shown to do (Best et al. 1988). Also, speakers of languages that do not allow nasal vowels should not be able to perceive the nasal feature, as the nasal vowel would simply be represented as an oral vowel. The adaptation of a nasal vowel to VN, however, is abundant in loanword data (cf. Paradis & LaCharité 1997; for example: French "changer" [ʃãʒe] to Fula [sansude]). Furthermore, it is not clear how long the reach of ill-formedness is: the segment [ç], for example, is a part of the British English pronunciation of the word "tune", [tçun]. Many speakers are not aware of this, however, and deny having this sound in their language. It must be considered whether [ç] is filtered out in perception when an English speaker adapts a loanword.

In spite of these issues, the division of the adaptation labour into perception as well as production (and, on occasion, factors beyond both such as orthography, and visual articulatory information) has found further proponents; the resulting models draw for example on the phonetic realisation of the foreign word, and the goal of maximising its phonetic similarity to the adapted word, to explain adaptations. Kenstowicz & Suchato (2006) write, "the adaptation process can take into account a variety of factors to achieve the best match to the source word including phonetics as well as orthography. The adapter is not a passive recipient of the speech perception module but exercises active control over the native grammar in shaping the loan as well as possibly calling on implicit knowledge of phonetic similarity to fashion adaptations that lack a precedent in the native system."

The best way of modelling this division of labour, or rather of the perceptual part of the adaptation, is not obvious. For example, it can be perceived of as pre-grammatical (Silverman 1992), as grammatical (Kenstowicz 2001), or as unspecified in this respect, referred to as a *perceptual module*, which turns out an expressly non-native percept Yip (2006). The concept of this percept, which "includes some reflex of most of the non-native segments, but it may differ from the percept of a native speaker of the donor language", is responsible for the mention of Yip's work in both this and the previous section: as noted before, her work constitutes an intermediate, if theoretically challenging approach.

Kenstowicz (2001, also see Broselow 2005, for an account of stress adaptation), on the other hand, suggests a separate, *grammatical* perceptual mapping with a constraint ranking different from the production grammar; specifically, with differing rankings of faithfulness constraints, while the markedness constraints remain in the same order (much like the strata described in Ito & Mester 1999). This separation is used to account for the split strategy for final sC clusters in Fon: while there is epenthesis after /s/, the final C is deleted (e.g. English *post* is realised as /posu/). According to Kenstowicz, the poor available cues to the final C result in a ranking of Dep (against inserting material not present in the input) over Max (against deleting material from the input) in the perception grammar, which leads to the input of the production grammar with a deleted final C (/pos/). This grammar has a ranking of DEP below MAX, and therefore results in vowel epenthesis during production.

Perception only

Other researchers would argue that perception is indeed relevant for loanword adaptation, and that adaptations occur during the perceptual process. In contrast to the work previously discussed, however, no role at all is assigned to the production grammar. It is here that the input ceases to be a useful concept, since there is no production grammar that would further alter it. Instead, all adaptations are claimed to be of a perceptual nature. The advocates of the "perception only" view of loanword adaptation do not agree on the nature of this adaptation.

In Boersma's *functional phonology*, the perception process itself is phonological (e.g. Boersma 1998, 2000), and the output of the perception grammar is well-formed in the recipient language. This means that all adaptation takes place in perception (see section 1.3.4/4), whilst evaluating the similarity of the phonetic input and the optimal phonological representation thereof in L1 (Boersma 2004). A problematic issue with this proposal - as with Yip's and Kang's, although more explicitly - is the matter of the phonetics/phonology interface, and the use of continuous phonetic detail in phonology, which as a rule has been conceived of as categorical.

In effect, this model makes the same predictions for loanwords as the psycholinguistic model by Dupoux and Peperkamp and their colleagues (e.g. Dupoux et al. 1999, Peperkamp 2003): Dupoux and Peperkamp claim, like Boersma, that all adaptations occur in perception. However, while for Boersma the adaptation process is a purely phonological one, the interpretation of "perceptual adaptation" which Dupoux and Peperkamp favour is: perceptual phonetic approximation of the foreign output to native categories before a phonological representation is formed. Dupoux et al.'s study shows that Japanese listeners (as opposed to French controls) tend to perceive an epenthetic vowel in a categorical perception experiment for an /ebzo/-/ebuzo/ continuum even near the vowelless end point, due to obstruent codas being ill-formed in Japanese. Peperkamp compares Japanese adaptations of [n]-final words from French and English, and finds that epenthesis of a vowel after [n], which is not phonotactically necessary, since Japanese allows nasal codas, depends on duration and release of the nasal. Peperkamp concludes that all loan adaptations are perceptual, and of a phonetic nature, since they are assimilations of the signal to the most similar native phonetic categories. Therefore, Peperkamp's view of similarity is, in contrast to Boersma's phonological approach, an entirely phonetic one.

An issue for both Boersma and Dupoux/Peperkamp arises if adapters demonstrably perceive foreign words successfully, i.e. without the adaptations with which they alter those words when producing them. I will return to this issue when discussing the outcome of the experiments conducted for this thesis.

1.3.4 Perception and phonology

The views on the locus of adaptations are closely connected to views on the role of perception in phonology (or of phonology in perception). In some cases the above approaches to loanwords inform broader views on the interplay of perception and phonology. In other cases a previously held view on perception informs the standpoints regarding loanword adaptation. This section serves to make these views on perception explicit, and give an overview of further relevant research.

The differing viewpoints regarding the relationship between phonology and perception can be divided into four broadly defined positions:

1. perception and phonology are independent from each other
2. phonology influences perception
3. perception influences phonology
4. perception *is* phonology

Note that these positions are not necessarily mutually exclusive, but may highlight different aspects of the relationship between perception and phonology, e.g. universal as opposed to language-specific influences, or differing theoretical views of the same kind of data.

Perception and Phonology are independent

This view is, to some extent, the implicit traditional one. Standard generative grammar is only concerned with the production grammar, mapping the underlying representation to a phonological surface form, which is then phonetically implemented. The reverse direction was not subject to discussion. Instead, articulation was the basis for example for most phonological features, such as labial, plosive or approximant, and for processes and tendencies in phonology. There was only a comparatively smaller body of research on the perceptual side, for example on the scarcity of front rounded vowels (e.g. Liljencrants & Lindblom 1972), and work on vowel systems in general (e.g. Stevens 1989).

While the influence of perceptual facts on phonology in general has been subject to study, the influence of a specific phonological system on a listener's perception, specifically in the context of loanwords, was implicitly rejected, as described. An explicit formulation of the viewpoint can be found in Jacobs & Gussenhoven (2000) quoted above, denying a role of perception in loan phonology and thus a perceptual system which is influenced by one's native phonology, in favour of a "universal phonological vocabulary". In contrast to this, in recent years phonologist have increasingly devised ways of incorporating perception into their models of phonology. Three diverging approaches are described in the next sections.

Phonology influences Perception

This view is related to arguments in favour of the perceptual adaptation (to whichever extent) of loanwords. If one's native phonology influences perception, then incoming foreign words will not be perceived in the same way as they are by a native speaker of the source language.

On this issue, a look at research on perception in second language acquisition is helpful. Its findings are especially relevant in a study of loanwords; as described in section 1.3.1, in most cases the mistakes or assimilations that are recorded in second language research are true for loanwords also, as the latter includes speakers who are familiar with the source language as well as monolinguals. This correspondence is all the more true since for most loanwords it is not the case that the whole population that adopts them is bilingual in the relevant languages. Therefore, even if the loanword is initially introduced by bilinguals, the final established form is in most cases identical to an adaptation by a monolingual speaker.[5]

In general, second language acquisition research has found that speakers' native phoneme inventories have considerable effect on how foreign sounds are perceived, and are thus a major source of difficulty with the new system (Best 1995, Best et al. 2003, Flege 1995, Kuhl & Iverson 1995, Rochet 1995, Strange 1995).

There have been a large number of studies which showed divergence in perception depending on the first language of the subjects and also word position. Flege (1995) and Rochet (1995) report a number of perception findings to this effect: French /y/ is perceived as /i/ by native Portuguese speakers, but as /u/ by English speakers, the English dental /θ/ is perceived as /t/ by Italians and Russians, as /s/ by Japanese speakers and as /d/ or /s/ - depending on position - by Dutch speakers. Thus the reason for differential substitution might be sought in the phonetic closeness of the respective sounds in the two languages to English [θ], and the differences between Japanese, Italian and Russian [s] and [t] sounds (which all

[5]This is not to be confused with Silverman's assumption that loanwords are shaped by monolingual perception because bilinguals cannot use their knowledge. If the bilingual speakers of the community are able to incorporate sounds from one language to the other, then a highly proficient bilingual may adapt loanwords less drastically and thus provide a counterexample to Silverman.

the languages have), and no longer in phonological explanations such as Lombardi (2000).

Flege (e.g. Flege 1987*a*,*b*, 1993) proposes a speech learning model (SLM), according to which categories must be created for L2 sounds for successful acquisition. The difficulty of creating these categories depends crucially on the perceived similarity of L2 and L1 sounds. If an L2 phone is sufficiently different from any native categories, the learner will be able to distinguish own and foreign, and establish a new category. However, if the L2 sound is too close to a category in L1, it will be perceptually assimilated to the native sound (by a mechanism termed *Equivalence Classification*), and successful acquisition is more difficult. For example, French /y/ falls in the former group for English learners, since the English inventory does not contain this phoneme. /u/, on the other hand, is a phoneme in both languages, whilst being phonetically different, with a difference in F2, and the English realisation being more diphthongal in many varieties. Hence, English speakers are less successful at producing the French [u] sound correctly.

Kuhl and Iverson's (1995) model of second language perception is based on the notion of the "perceptual magnet". It also does not rely on phonological feature values, but on the articulation space and phonetic similarity. The first language warps the perceptual space such that spectral differences in the vicinity of native phonemes are perceived to be smaller than elsewhere, effectively non-existent in terms of identification. This leads to the identification of foreign sounds with native ones. The discrimination of foreign sounds is problematic, if they are assimilated to the same native category. The segmental adaptations are thus not seen as phonological, but phonetic, and dependent on the native grammar only in that the grammar determines the native inventory. This process seems more plausible for vowels, as in this case there is the objective similarity measure of distance in the vowel space, whereas for consonants more factors (often with binary values rather than along continua) come into play. It is difficult to produce a [t] -[n] continuum. Similarly, in terms of tongue articulation, [t] would be more similar to [i] than to [k], and, [t] and [d] would be exactly the same, having the same manner and place of articulation. On a vague basis of similarity, however, this model is difficult to prove or falsify compared to feature-based models: naturally, segments that are similar to each other share a large number of features. The question is, however, the treatment of "dissimilar" sounds: are they perceived in acoustic terms only? How are they translated into a representation, if they are not simply omitted?

A further model of the perception of non-native sounds is Best's (1995) Perceptual Assimilation Model (PAM). This model works on the idea that, wherever possible, sounds are assimilated to existing (native) articulatory categories (as good, acceptable, or bad exemplars of the relevant category), and where not, the speakers focus on auditory or phonetic properties, and the sound may or may not be recognised as a speech sound. Discrimination of the non-assimilated sounds on an acoustic basis is not, then, a problem. However, a way for these new sounds to become stored as mental representations is not described. The construction of representational categories provides another problem in loanword research with monolinguals. According to L2 research, it takes some time for speakers to build new categories; subjects in an on-line adaptation task are highly unlikely to be able to categorise sounds which cannot be assimilated or, in Flege's terms, perceptually linked to native sounds, which leaves open the question whether the subject may just discard these sounds.

One way of modelling the perception of foreign sounds is feature-based: as Matthews & Brown (1998) show (also endorsed by Rose (1999) in his loanword model; Flege 1995 disagrees), there is evidence that L2 learners find it easier to discriminate segments which are ill-formed in their L1 if the features distinguishing these sounds are active in their language, whereas inactive features pose a lot more problems. For example, Matthews and Brown claim that for Japanese learners of English the /l/-/r/ distinction with the distinguishing feature [coronal] is harder to discriminate than /f/-/v/ with the distinguishing feature [voice] which also discriminates for example native Japanese /p/ and /b/. This approach suggests that borrowers may filter out segments with unused (in L1) features in their perception before adapting a loanword. Again, the evidence as to whether Japanese speakers can perceive a difference between /l/ and /r/, is not conclusive (especially as it is founded in L2 research, i.e. after explicit instruction); furthermore, English speakers can distinguish between a stop and a click (cf. Best 1995) although the feature [click] or [velaric] does not serve a function in the English language.[6] This points in the direction of a universal feature perception (but not necessarily production), where the activity of a given feature exerts a limited influence in terms of ease of perception.

These models of perception in second language acquisition show a clear influence of the native phonology on perception, but deal with segmental problems only. However, there is evidence that even phonotactic constraints (or rules), which, according to Silverman, are employed after the Perceptual Level, influence perception. Context-dependent perception may be due to a change in acoustic characteristics by coarticulation, but also to phonotactic constraints of the borrowing language: Rochet (1995) cites anecdotal evidence that French speakers perceive /θ/ as /s/,

[6]This may, however, be due to clicks being processed as extralinguistic or paralinguistic by English speakers.

except before /r/, where it is perceived as /f/ instead, since the initial sequence /sr/ is illegal in French (as it is in English). These context-dependent differences in perception could be due to a change in acoustic characteristics due to coarticulation, but also due to phonotactic restrictions influencing the perception of a foreign sequence.

A first step in examining native language context-dependent influence on perception was taken by Massaro & Cohen (1983). They investigated the perception a /ri/-/li/ continuum, depending on context, i.e. preceding consonants, only some of which combinations resulted in well-formed English onsets (e.g. /tr/ vs. */tl/, but */sr/ vs. /sl/) and found English phonotactics to influence perception. Pitt (1998) and Moreton (2002) build on Massaro and Cohen's work and experimentally test frequency-based and structural models of the bias against non-native clusters, finding more support for phonotactic knowledge than frequency affecting listener's perception.

A further study by Hallé et al. (1998) specifically looked at initial /dl/ and /tl/ and found French speakers to assimilate them perceptually to the well-formed clusters /gl/ and /kl/. What is problematic with this study is that only native speakers of French were subjects, i.e. there was no control group with a native language allowing initial /dl/ and /tl/. This is all the more problematic since the acoustic characteristics of the sequence with a coronal and a velar stop are very similar to each other. However, the omission of a control group has since been remedied with a study comparing French and Israeli listeners, Hallé et al. (2003), finding less perceptual assimilation in the Israeli group.

In specific reference to loanwords, Dupoux et al. (1999) conducted categorical perception experiments with Japanese and French speakers to test whether the Japanese phonotactics, i.e. in this case the prohibition on consonant clusters, has an effect on Japanese listeners' perception of words containing clusters. Their stimuli were continua with end points such as /ebzo/ vs. /ebuzo/, recorded by native speakers of Japanese and French and then modified to exclude traces of vowel between the two consonants. They found that French participants reported the presence of a vowel in a gradually rising fashion, starting from around 10% at the /ebzo/ end, whereas Japanese had a very high rate of reported vowels, ranging from 70% to 90% across the continuum. Dupoux et al. conclude that this is a case of native phonotactics changing the perception of a sequence that is impossible in the listener's first language: Japanese is predominantly a CV language, and consequently consonant clusters are broken up by an intervening vowel, even in perception. The methodological issue with Dupoux et al.'s study is that Japanese has a wealth of loanwords, mainly from English, and a long-standing tradition in which the strategy to adapt consonant clusters to achieve a CV syllable structure is always the same, that is epenthesis of a vowel. This means that any Japanese speaker is aware of this convention of adaptation, and this knowledge might play a dominant role in this automatic assumption of a vowel between two consonants by Japanese listeners. The goal of this thesis, however, is to determine the origin of the strategies of adaptation, i.e. the phonological reasons why an adaptation strategy is chosen, rather than their effects.

Perception influences Phonology

The previous section gives an overview of the vast amount of evidence of perception being affected by one's native phonology - predominantly centred around the perception of L2 segments in second language acquisition, and a limited amount of work on native phonotactics being relevant for perception. Conversely, in recent years the body of research has been growing which examines the relevance of perception for phonology, i.e. ways in which the perceptibility of sounds and contrasts is reflected in their use in phonology. For example, the acoustic cues available for a voicing contrast between two consonants are better pre-vocalically than pre-consonantally or word-finally. As a consequence of this, languages with final devoicing, i.e. a neutralisation of the voicing contrast word-finally, but not medially or initially, are common (e.g. Russian or German), while the reverse case is not found. The general idea is therefore that contrasts tend to be employed where they are easily perceptible, but eroded where they are difficult to perceive, i.e. "licensed by cue" (Steriade 1994*a*,*b*). This is also endorsed by Côté (1997*a*,*b*, 2001) in her work on consonant clusters (within optimality theory) and epenthesis and deletion processes (e.g. in different varieties of French), drawing for example on the preference for consonants to be vowel-adjacent because of added transition cues.

Steriade (Steriade 2001, 2003, 2005, also see Hayes & Steriade 2003, Jun 1995) has developed the "Licensing by Cue" principle and modelled this perceptual influence in optimality-theoretic terms, specifically with reference to faithfulness, as the *P-map*. Aspects of the P-map are described in more detail in chapter 5. The general approach is that cues available for a certain contrast in - crucially - a certain context determine the perceptibility of this contrast, which in turn determines the ranking of the corresponding faithfulness constraint in the production phonology. This leads to less perceptible modifications being more frequently used to resolve conflicts. For example, if a final voiced obstruent like /b/ is disallowed, than the least perceptible modification (i.e. devoicing) is chosen to solve the problem. Other possibilities, such as deleting the obstruent, or adding a vowel to make it non-final, are not employed. According to Steriade (p.5), "for at least some phonological properties and perhaps for all, there appears to exist a cross-linguistically minimal modification". Thus, while a perceptual influence in phonology and one's perception being shaped by L1 phonology are not a priori mutually exclusive, Steriade assigns little relevance to the latter.

Hume and Johnson, on the other hand, who work on perceptual influences on phonology (Hume 1998, 1999, Hume & Johnson 2001*b*,*a*), investigate the reverse direction as well (Hume & Johnson 2003). Hume & Johnson (2001*a*) give an overview of perceptual motivation in phonology, including vowel epenthesis to strengthen the cues to a voicing contrast in Maltese, and consonant metathesis from word-final /sk+t/ to /kst/ in Faroese, improving the perceptibility for the first stop consonant. Conversely, weak consonants may be deleted if perceptual "gain" is less than the articulatory effort. Finally, it is desirable that alternations to a form should be minimally perceptually salient, thus echoing the idea of minimal adaptations in loanword phonology.

Further work exploring connections between perceptibility and phonology comes e.g. from Flemming (Flemming 1997, 2001, 2003; also Flemming 2005 for an overview of the field) on the dispersion theory of contrast, describing the interplay of different tendencies in phonological systems, namely maximising the number of contrasts, minimising effort in terms of articulation, and maximising the distinctiveness of contrasts (e.g. for vowel systems in Flemming 2003), thus combining articulatory and perceptual influences on phonological systems.

Perception is *Phonology*

This view is was first proposed in Boersma's *functional phonology* (e.g. Boersma 1998, 2000), and a variant of "perception influences phonology". Rather than attributing to the native system a non-specified influence (which, following from previous research, is not necessarily the same on different levels of the phonological hierarchy), Boersma instead claims that it is simply the native phonology through which we perceive incoming acoustic signals. This idea is echoed in Kenstowicz (2001) and Broselow (2005), who propose a perceptual grammar in loanword adaptation (see section 1.3.3).

According to Boersma, the incoming auditory event is subject to structural as well as correspondence constraints when mapped to a certain representation; the latter constraints are responsible for a kind of similarity between different kinds of representations. In an application of this model, loanword adaptations are claimed to be perceptual as well as phonological (Boersma 2004), i.e. that the hierarchy of the mapping and structural constraints in the borrowing language determines the changes made to the loanwords. However, the development of a perceptual optimality-theoretic grammar, for example to model cross-linguistic differences is vowel perception, requires a large amount of phonetically detailed constraints, such as spectral and durational constraints like "F1 of 349 Hz is not [ɪ]" or "74ms duration is not /i/" (Escudero & Boersma 2004). The problematic issues of this approach are discussed further in the light of the experimental results of this thesis.

1.3.5 Phonology and phonetics

A further issue on which loanword researchers disagree is the question whether adaptation is a matter of phonology or phonetics. Once again, traditionally, as the term "loanword phonology" suggests, this was not a subject of debate. Adaptation was perceived to be an application of a phonological process or rule in production. Similarly, Paradis' Theory of Constraints and Repair Strategies with its loanword phonology based on meta-constraints, is based on phonology only; for example, the phonological hierarchy and feature organisation plays a decisive part in the model, for example through the adaptation strategy favouring the change of a terminal feature over that of a feature with dependents, which is higher in the phonological hierarchy (Paradis & LaCharité 1997).

Within optimality theory, one viewpoint is that adaptation occurs through the simple application of the L1 constraint ranking, or a slight modification thereof, as shown for Japanese lexical strata by Ito & Mester (1999). As a consequence of the slight upwards shift of faithfulness, the adapted forms - as well as impossible adaptations (Ito & Mester 2001) - can serve to gain more information about constraint rankings, i.e. the phonology, of the borrowing language, which are not clear from native data. The same approach was taken by Davidson (2001). The constraints whose rankings can be shown through varying faithfulness in adaptation depend on the kind of ill-formedness which is adapted, and can be for example segmental or prosodic in nature. In terms of phonotactic problems, most loanword research has dealt with CV-only borrowing languages, showing - in optimality-theoretic terminology - the effects of NoCoda (against syllable codas) and *complex (against consonant clusters), which are also obvious from the native data (however, see the example of codas in Yoruba above). A more interesting case is a scenario like in Davidson's experiment, where both donor and borrowing language allow clusters, but differ with respect to the role of sonority difference (SD), which lends itself to testing the role of the parameter of Minimal Sonority Distance (MSD; for further description see chapter 2), and the question whether higher-ranked faithfulness in loanword adaptation will leave clusters with higher sonority distance unchanged, whilst adapting those with lower sonority distance. This question will be dealt with in chapter 3.

What has become more numerous in recent years (in contrast to the application of what is essentially L1 to foreign input), especially within optimality theory, are phonological accounts of loanword adaptation which include a way of explicitly making reference to phonetic considerations, specifically the phonetic salience of the foreign structures, and their phonetic similarity to adaptations. The implementation of these is closely connected with the specific views on the influence of perception on adaptation, and the relationship of perception and phonology.

For example, there is the claim of a perception grammar effecting all or part of the changes, as in Kenstowicz (2001) and Broselow (2005). Boersma (e.g. Boersma 2004) sees the locus of adaptation only in a perception grammar which, as described earlier, evaluates similarity and maps the incoming signal to an optimal corresponding native phonological form. This is accomplished with the aid of constraints which prohibit or enforce the identification of certain acoustic properties with a given phonological form. For example, a constraint prohibiting an F1 value of 300Hz to be identified with the vowel /i/ is ranked lower than a constraint against identification as /ɪ/. Conversely, Steriade's P-map, while representing the availability of phonetic cues, is conceived of as part of the production phonology. A further shift towards phonetics can be observed for example in Yip (2006), splitting the task of adaptation between a perceptual module and phonological production which takes into account phonetic salience, and features loanword-specific constraints.

All the above models regard loanword adaptation as a completely or partially phonological process, which is - for a majority of approaches - informed by phonetic and acoustic-perceptual considerations. In contrast to these, Peperkamp and Dupoux (e.g. Peperkamp et al. 2003, Peperkamp & Dupoux 2001 deny that loanword adaptation has any phonological component at all. Instead, aside from further influences such as orthography (Vendelin & Peperkamp 2006), they consider the adaptation a purely phonetic approximation to the foreign word.

1.3.6 Similarity

In the description of approaches to loanword adaptation, the aim of loanword adaptation is generally recognised to be an assimilation (partial or complete) of ill-formed structures to well-formed native reflexes which differ minimally from the original foreign word. This means, the adapted form is desired to be different from the foreign word, where necessary to improve the fit to the borrowing language's phonology, yet it should also be maximally similar to the foreign form.

Before any attempts to formalise this force to maximise similarity, a formalisation, or at least a precise account of the nature of similarity is necessary. However, while the concept of similarity is used widely, not only in loanword research, but also for example in the field of speech recognition, definitions often appeal to intuitions. Beyond this, various coarse- and fine-grained measures of similarity have been employed, based on entire segments, features, or the phonological organisation of features, and differing as to whether similarity is universally the same or language-dependent.

A more elaborate and empirical approach to similarity is Steriade's P-map, with the concept of cross-linguistically consistent knowledge of the perceptibility of contrasts, based on cue availability. Additionally, the evidence showing influence of L1 phonology on perception suggests that one's native language also influences how similar two structures are perceived to be. Chapter 5 reviews applications and models of similarity in more detail, and tests predictions following from them.

1.3.7 Grammaticality, gradience and variation

Traditionally, grammaticality has been viewed as an absolute concept, with structures either existing in a language and being well-formed, or not existing and therefore being ill-formed. However, more recently this view has been undermined by an extensive body of work.

The concept of gradient grammaticality, i.e. the idea that not all structures of a language are perceived to be equally acceptable for native speakers, as described for example in Sorace & Keller (2005), was first applied to variable acceptability of syntactic structures. In terms of variation, pioneering work was done in optimality-theoretic terms by Kiparsky (1993), Anttila (e.g. Anttila 1997), and Hayes and colleagues (e.g. Hayes & MacEachern 1998, Hayes 2000; for an application to learning, see also Boersma & Hayes 2001). The principle of modelling variation and gradience in Optimality Theory is fairly straightforward: rather than strictly ranking rules, the hierarchy itself may be variable. This can occur for example through unranked constraints, or partially ranked constraints. Rather than being perceived of as having no dimensions, a constraint can be thought of as a line, or a band. This can result in overlapping with neighbouring constraints. In any given instance, therefore, either of the overlapping constraints may serve as the higher-ranking one. The more overlap there is between two constraints, the nearer the probability of either constraint being ranked for an evaluation scenario is to 50%. This principle can account for both variation in the output, and for gradience in well-formedness judgments.

A specific case of this variation are loanwords, where variation may occur in different ways. Firstly, there can be variation in terms of the extent to which a word is assimilated to the recipient language, or conversely, how many of its ill-formed characteristics (and which ones) it retains, as demonstrated by Ito & Mester (1999) -

in optimality theory terms, to what extent faithfulness constraints are raised against markedness for loanwords. Furthermore, there can be variation regarding the choice of adaptation for a given markedness violation. As a simplistic example, a constraint against a certain segment, such as *ʃ, could either be remedied by altering a feature of the offending consonant (e.g. to /s/) or deleting it, depending on whether IDENT-PLACE (against changing the place of articulation of a consonant) or MAX-C (against the deletion of a consonant) are ranked higher in the given instance.

Another relevant factor in the case of loanwords may be a gradient *ill*-formedness. One way of conceiving of this is again in terms of optimality theory constraint rankings. Foreign words violating higher ranked constraints of the recipient language may be regarded as more ill-formed.

On the other hand, gradience may not be determined by ranked constraints, but by experience or lack of experience with exemplars of categories or structures, and statistical calculations of frequency. Exemplar-based approaches (e.g. Bybee 2001) aim to demonstrate that frequency or "language use" is a relevant force in the system; for example, phonetic reduction applies predominantly to high-frequency items. In the context of loanwords, frequency and structure of the native lexicon (Broe 1993, Johnson 2004) can be thought of as interesting in two different ways: one, since all ill-formed words or sequences have a frequency of zero, a frequency-based model would suggest that they all be regarded equally ungrammatical and unacceptable; two, if frequent items are considered to be strong exemplars and prominent in the cognitive system, the choice of adaptation for a foreign word may be determined by the frequency of the available repairs in the borrowing language, such that higher frequency native structures also occur more frequently in adaptations.

While there is some evidence that, in the case of illegal clusters, processing is influenced more by knowledge of phonotactic rules than by frequency (Pitt 1998, Moreton 2002), the question what determines grammaticality generally remains unresolved for the time being; the lexicon, the frequency of exemplars in the lexicon, and knowledge of abstract generalisations (phonological constraints) may be involved to different extents on different levels. For this reason, native speaker judgments were (as with similarity) chosen as the best method to obtain a reliable measure for the purposes of this thesis.

1.4 The goal - determining the relative influence of different factors in on-line loanword adaptation

The goal of this thesis is to investigate the factors which have been proposed to be involved in the adaptation of loanwords, and to add further knowledge to the points of disagreement surrounding this process. Specifically, the role of phonological parameters, the borrowing language lexicon, and a possible influence of a listener's native language on perception are examined.

Loanword adaptation is a case of conflicting phonologies, with a clear conflict. There is no instruction of how to resolve this conflict, and the process is oriented towards a resolution in favour of the borrowing language, rather than attainment of L2. Thus loanwords, and especially on-line loanwords (i.e. those produced in an experiment, see next section), are a suitable setting to inform us with respect to some of the debates raised above.

On-line adaptations are likely to display variability, and allow us to match this variability with the factors whose influence we are trying to determine, such as phonological knowledge or experience of L1, or the similarity of the foreign and the adapted form. The questions guiding the research for this thesis have been extracted from the debates currently active in loanword research, and are the following:

- Is loanword adaptation a phonological or a phonetic process?
- Does perceived similarity influence loanword adaptation, and if so, is it determined universally or L1-specifically?
- Are loanwords adapted in perception or production, or both?
- Is there substantial variability in on-line adaptations, and if so, is it connected to gradient grammaticality and/or frequency?

The first issue concerns the contributions of phonology and phonetics to adaptation, which will be explored in different ways. Firstly, an aspect of L1 phonological influence is studied through testing the role of sonority distance on the adaptation of the ill-formed consonant clusters. Secondly, the plausibility of phonetic explanations of the choice of adaptations (as in Peperkamp's work) is evaluated. Finally, the relevance of phonological parameters on the judgments of grammaticality and similarity (as factors on adaptation) is examined.

Two aspects of the concept of similarity will be surveyed: initially, its role in adaptation is tested, through examining the extent of the correlation between judgments of perceived similarity and preferences in adaptation. Furthermore, comparing similarity judgments cross-linguistically allows conclusions firstly with regard to similarity being determined cross-linguistically, or on a language-specific basis, and secondly with regard to the level or levels (i.e. phonetic detail, the segment inventory or phonotactics) on which a potential L1 influence operates.

The question of the input to adaptation, and a potential influence of L1-guided perception in determining this input, are a further central point of this thesis; the split of the on-line production in a production-oriented and a perception-oriented task shows where L1 perception may apply[7]. With similarity as a factor in perception, cross-linguistic differences in judgments may point to perceptual adaptations, i.e. an input which has been modified before production, and thus an influence of L1 on perception.

Finally, the role of gradience, variability and frequency in loanword adaptation are explored at several junctions, initially through the varying output of the production experiment. Furthermore, the relevance of frequency in the native lexicon for gradient grammaticality is examined, and both frequency and grammaticality are tested directly for their influence on adaptation.

To summarise, the following experiments are conducted in this thesis:

- on-line production of consonant clusters which are ill-formed in the participants' L1 (English); the experiment is split in a production-oriented and a perception-oriented task
- judgments of perceived grammaticality from English listeners
- judgments of perceived similarity from English and Russian (the source language) listeners

Taken together, these experiments will allow further insight into the nature of similarity, determinants of gradient grammaticality, and the relationship between perception and phonology. It will be outlined in the predictions section of this chapter in which way the results of each of these experiments will serve to provide evidence for and against specific viewpoints on the debated issues.

1.4.1 Methodological decisions

On-line vs. established loanwords

Silverman (1992) claims that borrowers' knowledge and familiarity with the source language plays no role in their perception if they are in *L1 mode*. There is, however, ample evidence that second language learners progress to perceiving more and more contrasts as they expand their knowledge. To test predictions about perception of loanword adapters, one would ideally want adaptations from borrowing language speakers who are not familiar with the source language, since familiarity with the source language may lead to the perception of more non-native contrasts. If it does,

[7] Perceptual influence on adaptation is further tested through a preliminary cross-linguistic comparison study on categorical perception of well- and ill-formed clusters.

then results will be skewed towards more universal perception, i.e. showing more universal tendencies than are true for monolingual speakers. To exclude skewed results for the perception of non-native forms, experimental situations with elicited adapted forms are therefore necessary.[8]

What speaks further in favour of on-line loanwords is that in the course of the history of a loanword in a language, especially when the original word of the source language is not continually present in the speakers' environment, there will be a process of what is called lexicon optimisation in optimality theory. Speakers will store as an underlying form what is most harmonic, or put simply, least different to their output, i.e. the adapted form, and will thus "lose" features of the original form. For example, English loanwords from Greek with initial <pn> are pronounced with initial [p] in Greek, but this sound is deleted in the English form, although in spelling the <p> is still present. Thus, English children have no possibility of storing an underlying /p/ for this word, later orthographic knowledge notwithstanding. This means that, for a fully assimilated loanword whose stored form contains no non-native structures, the original form of the word in the source language is not relevant in production. Hence, there is no conflict between source form and native phonology which needs to be resolved. Consequently, the interesting cases from the loan phonologist's point of view are those where a loanword is being adapted for the first time - a first-generation loanword, as it were.

Furthermore, with established loanword data it is often problematical to determine how a word has entered a language, and whether orthography and previous adaptations of a word in other languages may have played a role. Additionally, with loanwords from such a universal lingua franca as English, there is a scarcity of languages that do not have a set of loanwords from English in their vocabulary, and therefore a convention or routine of adapting them (such as /θ/ being adapted as /s/ in Japanese). To determine the origin of mechanisms of adaptation and the role of perception in their formation, it is advisable to avoid existing loanwords. The conclusion was to investigate hypothetical rather than actual loanwords, and pairs of languages for which such an adaptation mechanism is not established.

Finally, as noted previously, on-line adaptations are likely to display a greater degree of variation as opposed to existing loanword corpora, where in most cases only one form of the loanword is recorded, namely the one which prevailed in the

[8] This issue may also have caused some of the differences between Silverman's (and Yip's) and more universal or feature-based views: while Silverman included elicited forms, and dismissed any influence a knowledge of English might have, most other researchers work with established loanwords and assume the origin of these loanwords with bilinguals and an effect of their knowledge on the output form. That these two kinds of adaptations yield the same result, is clearly not obvious, but must be tested experimentally.

borrowing language's lexicon over time. This will provide insight into the true nature of variability in adaptation, and show whether the choice of the optimal adapted form is as unambiguous as it seems from existing loanwords.

Segmental vs. phonotactic conflicts

As described in 1.3.4, there is a large body of research concerning the perception and production of foreign phonemes by second language learners. Furthermore, Silverman (1992) and Yip (1993) agree that perception also changes the segmental input to the phonological adaptation of loanwords. Thus there is a broad consensus about the interaction between a native system and foreign phonemes. In terms of phonotactic restrictions of the borrowing language, there is no such consensus. Whereas Dupoux and Peperkamp's viewpoint is that adaptations are solely perceptual, claiming that "there is no loanword phonology" (Peperkamp 2003), Silverman, Paradis and optimality theory phonologists maintain that phonotactic (if not segmental) adaptations are phonological.

The overwhelming majority of phonotactic loanword adaptations that have been analysed have, as reported in 1.3.2, those of English or French terms into CV languages. Unless the adaptations are not complete, this leaves only epenthesis and deletion as means to break up consonant clusters. The discussion about the application of either of these two repairs may thus be skewed because it is an artefact of this constellation. To remove the possibility of such an artefact in an experiment, a borrowing and donor language must be chosen that both contain consonant clusters, with the set that is well-formed in the borrowing language being a smaller or subset of the clusters in the source language. This setup provides the possibility of a third kind of adaptation, namely the change of one consonant cluster to another. If this possibility is used as an adaptation strategy, then the focus on epenthesis and deletion as the only methods of phonotactic adaptation (e.g. Paradis & LaCharité 1997) is mistaken.

Russian to English adaptations

Keeping in mind the prerequisites of a source and recipient language constellation that will enable us to investigate on-line adaptations without a predetermined adaptation routine, and to observe potential cluster to cluster adaptations, Russian and English were chosen as source and recipient language, respectively.

Loanwords from Russian are few and far between in English, and those that exist may not have been adapted by English speakers directly, for example *bistro*, via French, from Russian /bɨstrʌ/, which means that English speakers are naïve to ways of adapting Russian words. Furthermore, Russian has a wealth of initial consonant clusters, such as /dv, fp, vl, zb/, which contain segments that are legal in

English, but are not themselves well-formed, since the set of English initial clusters is much more restricted (further differences between the languages will be described in chapter 2). Consequently, the adaptation scenario from Russian to English provides phonotactic conflicts which can be resolved through the change of an ill-formed cluster to a well-formed cluster, and can therefore be used to test claims about strategies of phonotactic adaptation (cf. previous section).

1.4.2 Main predictions of models

In the same way as not all models of loanword adaptation pertained to all of the debates reviewed above, not all models make clear predictions on the outcomes of the experiments. This section gives an overview of instances where the models discussed do have a specific viewpoint on the outcome of the experiments.

The production experiment

Paradis and LaCharité's Theory of Constraints and Repair Strategies (TCRS) predicts epenthesis across the board for foreign structures with phonotactic conflicts only, i.e. without segments which are not in the borrowing language's inventory. This is the case for all target clusters in the production experiment which are not well-formed in English. Furthermore, the model does not allow for any perceptual adaptations, and would thus predict no adaptations in a perception-oriented task.

Similarly, while in contrast to Paradis' theory on the issue of the existence of perceptual adaptations in general, in this scenario Silverman's model conforms to its predicting no perceptual adaptations, since phonotactic problems are adapted in Silverman's Operational Level, i.e. in production.

As Paradis' theory, Steriade's approach to loanwords is based only on speech production, and makes the same prediction as TCRS in terms of the perception-oriented task. However, in contrast to TCRS she predicts different outcomes in production depending on the similarity between targets and potential adaptations: for each target, the most similar adaptation will be chosen. According to Steriade's P-map, the most similar change possible is a voicing change. For example, the target cluster /vr/ is predicted to be repaired by devoicing, resulting in legal /fr/, rather than by vowel epenthesis, creating /fər/. On the other hand, devoicing /zr/ does not make it legal; here the most similar change is not an option, so /zr/ is not devoiced, but a more drastic repair, such as deletion or epenthesis, is necessary. In short, wherever it creates a legal (or marginal) cluster, i.e. for /vr, vl, sv/ and possibly /zv, zb, ʒb/ (where double devoicing is necessary) only devoicing will be used in adaptation. Furthermore, place of articulation (POA) changes are claimed to be more similar than stricture changes, hence in a case where devoicing is not viable such as for /sr/, the POA change to /ʃr/ will be employed rather than the adaptation /tr/ which changes a fricative to a stop.

Both Boersma's and Peperkamp's loanword models are situated firmly within perception, and hence predict that there will be no difference between the two kinds of tasks. Boersma specifically states that the result of the perceptual process is well-formed in the listener's native language, which would result in complete adaptation of all tokens. In Peperkamp's case, since in her view all adaptations are phonetic approximations to the nearest native structure, the adaptations are expected to have a clear phonetic motivation and a certain uniformity. For example, stop-initial clusters like /pt/, where the stop is released in Russian, are predicted to be epenthesised, since the release can be reinterpreted as a vowel. For fricative-initial clusters such as /sv/, no such phonetic motivation exists, hence no epenthesis is predicted.

Frequency, gradient grammaticality and similarity

A role of the lexicon and frequency has not been previously claimed in loanword phonology specifically (although Johnson 2004 refers to influence of the lexicon on speech perception); however, usage-based theories such as Bybee's (2001) have claimed an influence on phonology in general. For this reason, direct relevance of type or token frequency on the choice of strategy in loanword adaptation would provide evidence in favour of those theories, and indirectly against the existing loanword models.

Variability and gradience is also rarely discussed in loanword adaptation, in favour of modelling the adaptation process resulting in a single predominant adapted form per borrowed word. Evidence for gradience in judgments of grammaticality, and for this gradience affecting how consistently and with which strategy a foreign word is adapted, would enforce the need to incorporate this gradience into accounts of loanwords. Furthermore, if grammaticality is relevant, and phonological parameters determining it can be found, the case of a phonological (such as Boersma's or Steriade's) rather than a phonetic (such as Peperkamp's) model of adaptation is strengthened.

Finally, the concept of minimal changes in a loanword, or, conversely, maximal similarity between the source form and the adapted form, has manifested itself in one way or another in all approaches to the topic. This means that all models would support the hypothesis that maximally similar adaptations will also be most frequently used. The difference lies in the views on how similarity is defined and measured. For example, Kenstowicz and Yip do not make definite general claims in this respect, but instead provide case-by-case arguments that certain adaptations in their data are chosen due to their similarity. Peperkamp and Boersma on the other hand assume that similarity is determined by the native language: its phonetics and phonology, respectively. Cross-linguistic differences in the perception of similarity would support this idea. Specifically, tieing in with Peperkamp's predictions on production, the comparison /pt - pət/ should be rated as very similar, whereas /sv - səv/ would be less similar. Boersma's predictions, on the other hand, would depend on his constraint ranking for the English perception phonology. For both Peperkamp and Boersma it follows from the idea of perceptual adaptation that, in the case studied in this thesis, English listeners perceptually adapt the foreign sequences and hence give very high similarity values to target-adaptation pairs, while Russian listeners do not perceptually adapt one form to another and give lower similarity judgments.

Steriade's approach to similarity predicts that it is determined by perception, but cross-linguistically so: for example, the deletion of the first consonant of a cluster is predicted to be perceived as more similar than deletion of the second cluster (e.g. /fp-p/ more similar than /fp-f/), since the acoustic cues are more salient before a vowel. Any similarity rankings counter to cue-based predictions, and any cross-linguistic differences in similarity judgments provide evidence against Steriade's model.

1.4.3 Overview of chapters

Chapter 2 compares the phonology and phonetics (to the extent relevant for the experiments and their analysis) of English and Russian.

Three experimental chapters follow. Firstly, chapter 3 describes the initial adaptation experiment to determine the output of the adaptation process, and the variability thereof. English listeners are presented with auditory Russian-sounding stimuli, which are phonotactically ill-formed in their L1, and which they are asked to reproduce in different ways. The results are analysed with respect to phonological parameters for evaluating clusters such as sonority distance (see chapter 2), and a pilot study explores the role of misperception for the adaptation of one of the Russian target clusters.

The results of these experiments necessitated further investigation into influences on the adaptation process. The factors studied in chapter 4 are:

- the frequency of potential adapted onsets, as measured in the CELEX database of spoken English
- the degree of (un-)grammaticality of the Russian targets with respect to English, as drawn from native speaker judgments

- the similarity between targets and adaptations, also judged by native speakers of English

The chapter contains an analysis of what determines similarity and grammaticality, and of the effect these variables have on the choice of an adaptation.

Chapter 5 studies in more detail the similarity between foreign forms and adaptations, and the perception of minimal or more grave alternations, since the concept of this minimality pervades the recent work on loanword adaptations. The experiment compares Russian and English similarity judgments of the same stimulus pairs, differentiated by phonotactics status in the two languages; results are used to test hypotheses based on the P-map (regarding the relative similarity of different contrasts), and other models of similarity, and to evaluate the relative contributions of cross-linguistic similarities caused by the human perceptual apparatus, and of the native phonology of a listener.

Finally, chapter 6 summarises the findings, and discusses their implications for loanword research specifically, and the significance of the results for the broader issues reviewed in general.

Chapter 2

Russian and English phonology

2.1 Introduction

In this chapter, some main points of Russian phonology in comparison with English phonology are outlined; the aim is to provide an overview over prominent aspects of Russian phonology, while looking in more detail at the characteristics which are relevant in the experiments in chapters 3, 4 and 5. The first part compares the segment inventories; the next section gives an overview of phonological processes in Russian, and finally some phonetic details of Russian and English are compared. For a detailed overview of the literature on Russian phonology see Chew (1999).

2.2 Phoneme inventory

2.2.1 Consonants

Since the consonants and the possibilities of combining them in a syllable onset are the main focus of the following chapters, initially the consonant inventories of the two languages will be compared. Table 2.2.1 shows the consonants of English.

	labial	dental	alveolar	post-alveolar	palatal	velar	glottal
oral stop	p b		t d			k g	
affricate				tʃ dʒ			
fricative	f v	θ ð	s z	ʃ ʒ			h
approximant				r	j	w	
nasal	m		n			ŋ	
lateral			l				

Table 2.1: The English consonant inventory, following Giegerich (1992). Segments occurring only in some varieties, such as /ʍ/ and /x/, are omitted. The bilabial and labiodental places of articulation are combined in one column, as are post- and palato-alveolar.

Table 2.2.1 shows Chew's (1999) version of the Russian inventory. The most striking difference between the inventories is the existence of two contrasting sets of consonants in Russian, namely the palatalised (or soft, or [-back]) set, and the

	labial		dental		alveolar		palatal		velar	
	hard	soft	hard	soft	hard	soft	hard	soft	hard	soft
stop	p b	p^j b^j	t d	t^j d^j					k g	
affricate			ts			$tʃ^j$				
fricative	f v	f^j v^j	s z	s^j z^j	ʃ ʒ				x	
approximant								j		
nasal	m	m^j	n	n^j						
lateral			l	l^j						
trill					r	r^j				

Table 2.2: The Russian consonant inventory, following Chew (1999). The bilabial and labiodental places of articulation are combined in one column.

non-palatalised (variously termed hard, plain or [+back]). Similarly obvious is the lack of dental fricatives in the Russian inventory, and of the labial approximant /w/, the glottal fricative /h/, and the velar nasal /ŋ/. On the other hand, the Russian velar fricative /x/ is not a phoneme in most varieties of English. Since the experiments in the following chapters are concerned with phonotactic rather than segmental conflicts, /x/ was not used in the stimuli.

A further difference between the two languages lies in the anterior coronal consonants: while these are alveolar in English, the Russian equivalents are dental. This will be described in more detail in the section *Phonetic Characteristics*.

There are some disagreements in the literature regarding the phonemic status of some segments in Russian, as discussed for example in Avanesov (1956) or Chew (1999). Most notably, it is not clear whether the velars [k, g, x] and their palatalised counterparts belong to the same phonemes or are separate. The same is in question for the post-alveolar fricatives. However, since this issue does not bear on the present thesis, it is not discussed further.

2.2.2 Vowels

Russian is usually described as having a five phoneme vowel system (Chew 1999, Jones & Ward 1969), with the vowels /a, e[1], i, o, u/, with allophonic variation dependent on context and stress.

As with the velar consonants, there is conflicting evidence on whether [ɨ] is an independent phoneme (Bulanin 1970), or an allophone of /i/ (e.g. Jakobson 1929, Avanesov 1956). Again, this issue does not pertain to the content of this thesis and is therefore not discussed.

[1]This vowel is sometimes transcribed as /ɛ/

Additionally, Kučera & Monroe (1968) also list unstressed *i, a, u* as separate phonemes. These correspond to somewhat laxer versions of their stressed counterparts, realised as [ɪ, ʊ], as well as [ʌ] and schwa. This will be explained in more detail in the section *Stress-related phenomena*.

For English, vowel systems differ widely between varieties, which is why it is not appropriate here to describe in detail the inventory of a single variety. Prominent features across varieties of English are for example tense-lax distinctions which are not part of the Russian system such as between /i/ an /ɪ/, /u/ and /ʊ/ and /e/ and /ɛ/, three closing diphthongs, and a larger number of distinctive low vowels compared to the single Russian /a/ phoneme. These features result in a considerable larger English than Russian vowel inventory.

Most importantly, while there are phonetic differences between Russian and English vowels, such as a diphthongal realisation of Southern British English /o/ and /e/, or a fronted realisation of Scottish /u/, the realisations of the vowels in the Russian inventory correspond to realisations of English vowels, including [ɨ], with F1 and F2 at 285 Hz and 1,655 Hz, respectively (Bolla 1981, p.66), overlapping with the phonetic space for English i-vowels to some extent (e.g. Ladefoged 2003, p.212, Peter Ladefoged's [ɪ] at about 275/1500 Hz, [i] at 300/1650 Hz). This makes the choice of vowels unproblematic for Russian stimuli presented to English speakers in the experiments.

2.3 Phonological differences

2.3.1 Voicing: final devoicing and voicing assimilation

In contrast to English, Russian has no voiced/voiceless contrast word-finally. All codas of the final syllable consisting of obstruents are devoiced (Avanesov 1956). For example, nouns with word-final voiced obstruents have a voicing alternation between the different forms of the declension, such as /drug/ [druk] (friend-NOM) and /ˈdrug+a/ [druga] (friend-GEN). This word-final devoicing can be viewed (Lightner 1972) as a voicing assimilation process, namely to the word boundary (assuming a word boundary is specified as [-voice]; see Chew, 1999).

This view connects devoicing to other voicing assimilation processes in Russian (discussed e.g. in Jakobson 1956, Avanesov 1956, Halle 1959): all consonants in an obstruent cluster undergo regressive voicing assimilation. This assimilation also crosses morpheme boundaries, such that the prefix {s-} forming the perfective of /kruˈtitʲ/ (to twist) is voiceless in [skruˈtitʲ], but the perfective of /ˈdʲelatʲ/ (to make) formed with the same prefix is [ˈzdʲelatʲ].

While this generalisation is true for most obstruents, the behaviour of /v/ is split, depending on its neighbours (e.g. Jakobson 1956, Halle 1959): when it is not followed by an obstruent, it behaves like a sonorant, in that it neither assimilates nor triggers assimilation. Preceding voiceless obstruents remain voiceless, as in /svʲet/ (light). However, if /v/ is followed by an obstruent, it behaves like an obstruent. It devoices in front of voiceless obstruents (e.g. /v + komnatʲe/ (in the room): [ˈfkomnatʲe]), and it triggers assimilation of preceding obstruents, as in /bʲez + fstupʲlʲenija/ (without introduction), [bʲesfstuˈpʲlʲenija].

With respect to the experiments described in the following chapters, voicing assimilation puts a restriction on the available set of Russian onset clusters: obstruent clusters differing in voicing are impossible except with /v/ as the last consonant. Further restrictions on possible clusters are discussed in the section *Constraints on syllables*.

2.3.2 Palatalisation, fronting and retraction

A further assimilation process is often discussed in the literature on Russian phonology; in contrast to the spreading of voicing specifications, this section is concerned with place feature dependencies between consonants and following vowels. Generally speaking, palatalised, or [-back], consonants occur followed by the front vowels /i, e/, while [+back] consonants occur before other vowels.

Rubach (2000) discusses the relevant processes in detail. The term *palatalisation* refers to consonants being palatalised before front vowels and /j/, which occurs for example in derivations such as /ʒeˈnʲitʲ/ (to marry) from /ʒeˈna/ (wife). Secondly, Rubach describes the process of fronting, which relates to the realisation of /ɨ/, as in the nominative plural suffix, as either front [i] or back [ɨ] depending on the preceding consonant. Hence, a stem ending in a soft consonant like /gostʲ/ triggers [i], while the plural of /stol/ is realised as [stoˈlɨ]. Conversely, retraction of /i/ also occurs, for /i/-initial stems, if they are preceded by prefixes or prepositions with a hard final consonant, or even across word-boundaries, such as in /ˈgolos ɨˈvana/ (Ivan's voice; example taken from Rubach 2000).

As a result of the realisation of a consonant in Russian being closely connected with the following vowel, an allophonic difference dependent on syllable position, as is the case for English dark and light /l/ in coda and onset, respectively, is not found in Russian.

2.3.3 Constraints on syllables

The concept of the syllable has been connected to notions of sonority since the 19th century, for example in Whitney (1865). According to Trask (1996, p. 327), sonority is "a particular sort of prominence associated with a segment by virtue of the way in which that segment is intrinsically articulated. Sonority is an elusive notion." Bloomfield (1933) connects sonority to the loudness of a segment; other accounts work with vocal tract aperture (being inversely related to sonority) or mostly with major class features and voice. Various sonority hierarchies have been proposed (e.g. Ladefoged 2003, Zwicky 1972, Steriade 1982, Clements 1990, Selkirk 1984, Hume & Odden 1996): they vary in detail, but agree in stop consonants being the least sonorous, followed by or equal to fricatives, then nasals, liquids, and finally glides and vowels. Blevins (1995) summarises the literature in a tree showing sonority relations between feature values (p.211): consonantal segments are less sonorous than consonantal ones, obstruents less sonorous than sonorants, nasals less sonorous than non-nasals, non-continuants less than continuants, and voiceless segments less sonorous than voiced ones. For the calculation of the sonority distance of the sonorant clusters used in this thesis, I am using Steriade's (1982) 6-step scale of:

Vowels > Glides > Liquids > Nasals > Fricatives > Stops

This scale is based on the manner of sounds and does not include the factor voicing. While it would be desirable to integrate the notion of the higher sonority of voiced segments into the scale, it is not clear how to quantify voicing, as compared to the whole steps of manner differences. However, for example the sonority relations in Blevins (1995) show the effect of voicing as subordinate, so that a voiceless fricative ranks higher on the sonority scale than a voiced stop; for example, [s] is more sonorous than [d]. Hence, the basis for calculation is the above scale, with the relevance of voicing within a group of clusters which have the same sonority distance discussed as appropriate in the experiment chapters.

With regard to syllables, sonority is decisive in that syllables tend to have their sonority peak at the nucleus, with the sonority of the surrounding elements declining towards the syllable edges. This tendency has also been noted since the 19th century, for example in Sievers (1881), and has been formulated in several ways, such as in Selkirk (1984), as the Sonority Sequencing Generalisation (SSG): "In any syllable, there is a segment constituting a sonority peak that is preceded and/or followed by a sequence of segments with progressively decreasing sonority values". In reference to syllable onsets (the focus of this thesis), the SSG demands that starting from the first consonant, each following segment must have the same or higher sonority, precluding for example sonorant-obstruent clusters, or fricative-stop clusters. Exceptions to the SSG can to some extent be explained through the extrasyllabicity of the consonant at the edge; nevertheless, exceptions remain. This has led to the view of the SSG as a preference, which, however, can be violated.

A further constraint on syllable onsets, which varies cross-linguistically, is the parameter of minimal sonority distance (MSD), which determines how many steps further up the sonority hierarchy the following consonant must at least be to create a well-formed cluster (e.g. Harris 1983, Steriade 1982).

For English, the MSD value is set at 2 (e.g. Archibald 1998, p. 158). This means that no obstruent-obstruent or fricative-nasal clusters are possible. Additionally, plosive-nasal clusters are also prohibited. The largest group of well-formed clusters is plosive plus liquid, for which all combinations are allowed except homorganic /dl/ and /tl/. The fricative-liquid group is more limited in that the fricative must be voiceless (i.e. less sonorous). For this group, the only full set of clusters exists in the labial place of articulations with /fl, fr/, while /θ/ is subject to the same restrictions as /t/ and only combines with /r/. /s/ and /ʃ/ behave to some extent complementary, in forming /sl/ and /ʃr/, while /sr/ and /ʃl/ are prohibited (or possibly marginal). Furthermore, /s/ behaves exceptionally (with respect to the SSG rather than cross-linguistically) in that it combines freely with voiceless plosives and nasals (/sp, st, sk, sm, sn/), as well as combining with well-formed clusters to form CCC onsets (e.g. /sprus/), which are only possible with /s/ as the first segment. Due to the status of these combinations violating of the SSG, but being cross-linguistically very frequent, initial /s/ has received special attention in the literature, and been treated as being outside of the syllable structure proper, variously seen as extrasyllabic, an adjunct to the syllable, or as forming a complex segment with the following consonant (Selkirk 1984). For this reason, /s/-stop clusters may been seen as a different type of cluster to others, and could be expected to be have differently in adaptation. However, this is of limited relevance, since, being legal in English, they would not be expected to be changed. The relevance of the special status may lie in /s/-stop clusters as possible adaptations themselves, specifically for other fricative-stop clusters, such that sonority distance is not increased. The role of /s/-stop clusters as adaptations is addressed in the results section of the production tasks.

Furthermore, English permits CC onsets with glides (e.g. /kwin/), again restricted to some extent by avoidance of homorganicity (e.g. */pw/). English also allows single onsets, as well as onsetless syllables. For further literature on English onsets, see for example Hammond (1999), Giegerich (1992), and also Algeo (1978) for a discussion of the inconsistency in the literature as to which sequences are well-formed. An overview of clusters regarded as possible clusters of English can be found in appendix A.

In Russian, single and empty onsets are also permitted. Furthermore, onset clusters are much less restricted than in English. Russian allows up to four consonants in a cluster, whose initial segments are anterior fricatives and combinations thereof, as in /ˈfstrʲetʃa/. In terms of CC onset clusters, all combinations but glide (i.e. /j/)-initial and liquid-liquid combinations are attested (Halle 1959). Not only does Russian not have a minimal sonority distance, but it also allows clusters which substantially violate the Sonority Sequencing Generalisation; for example, the onset cluster in /rtut^j/ (mercury) has a sonority distance of -3, as well as a voiceless second consonant after a voiced first one. Only approximately a third of CC onsets conforms to the SSG (Chew 1999). The most frequent initial consonants are again the labial and dental fricatives /f, v, s, z/ (see Zalizniak 1977, and Chew 1999 for a frequency table derived from the former). What plays a large role in this are the prefixes {vz-, v-, s-} with their voiced and voiceless allomorphs /vz-, fs-, v-, f-, s-, z-/. Additionally, the prepositions {k-, v-, s-} also combine with the following noun, forming a prosodic word, which results in further /v, f, s, z, k, g/-initial onsets (cf. [ˈfkomnat^{j}e] in the room). The reader is referred to appendix A for a table of clusters occurring in Russian.

Since only onsets will be the subject of experiments for this thesis, restrictions on codas are not described here. For a detailed discussion of Russian syllables and overview of the relevant literature, refer to Chew (1999).

2.3.4 Alternations with zero

A common feature of Slavic languages is the existence of so-called yers, defined as underlying vowels, which may or may not be realised in different environments (e.g. Lightner 1965, 1972). For this reason they are also called "floating vowels". Russian has two yers, a back and a front mid vowel, for example transcribed as E and O. The basis for the claim of yers lies in vowel-zero alternations, for example through a declension paradigm, such as in /ˈn^{j}em^{j}ets/ (German-NOM) as opposed to /ˈn^{j}emtsa/ (German-GEN), and /b^{j}eˈlok/ (protein-NOM) as opposed to /b^{j}elˈka/ (protein-GEN).

Since no conditions for insertion or deletion rules can be found, the proposal is that these are underlying vowels which may be vocalised (i.e. assigned a mora and realised as vowels) in certain environments, for example depending on the inflectional suffix. The concrete formulations of triggers for yer realisation differ. Lightner (1972) and Rubach (1986) claim that yers are realised if the following vowel is also a yer; in contrast, Yearley (1995) argues in an optimality-theoretic account that yers are realised to avoid a tautosyllabic consonant cluster. These theoretical disputes, as discussed for example in Hermans (2001), are not central to the present thesis. However, what is relevant to the comparison of Russian and English listeners (see chapter 5) is the commonness of vowel-zero alternations in Russian.

2.3.5 Stress-related phenomena

While predictions about stress can be made for English depending on word class and syllable weight, such predictions are not straightforward in Russian. Chew (1999) writes that "stress may be attached to any syllable of the word, and may surface on virtually any type of morpheme." Russian has therefore been claimed to have free word stress (Avanesov 1956), although this is not undisputed. For example, Halle (1997) argues for a Russian stress system, as derived from Indo-European.

In English, unstressed vowels are reduced, and in most cases realised as schwa, with the exception of high front vowels. Consequently, in unstressed position, there are only minimal phonemic distinctions. In Russian the set of distinctions is also reduced by lack of stress - while in stressed syllables there are five (or six, if /ɨ/ is regarded as separate) different phonemes, the number decreases to three in unstressed position: the distinctions between /i, e/ is neutralised and realised as a lax [ɪ]-like vowel, unstressed /u/ is realised as the equally lax [ʊ], and the /a-o/ distinction is also neutralised. However, in the latter case there are two possible realisations, depending on proximity to stress and position (Avanesov 1956, Bondarko 1977), namely [ʌ] and schwa.

Jones & Ward (1969) describe the conditions for [ʌ]: "it occurs only before the stress: *immediately* before the stress ('pretonic' position), or as the first sound in a word, or as either one of a sequence of two vowels where the letter a or the letter o is written.". Schwa, on the other hand, is found in the following environment: "Before the stress it cannot be preceded by a soft consonant, nor can it occur immediately before the stress or as the first sound in a word or as one of a sequence of two vowels before the stress [...] It occurs therefore in pre-pretonic syllables after a hard consonant, but not as one of a sequence of two vowels."

Examples given by Jones & Ward are [gəlʌˈva] (head) and [məlʌˈko] (milk). In reference to the experiment chapters to follow, this means that when an initial cluster of a Russian word is to be compared with the same word with a schwa inserted in the cluster, stress is relevant. In Russian orthography, <o> or <a> are inserted. In order for this vowel to be pronounced as schwa rather than lower [ʌ], it must be two syllables away from the stress. This means that the corresponding word with the CC cluster must be (at least) bisyllabic and stressed on the last syllable.

2.4 Phonetic characteristics

The labial hard consonants /p, b, f, v, m/ have a secondary velar articulation (e.g. Kochetov, to appear). Alternatively, the hard consonants are described as pharyngealised. Bolla (1981) reports varying degrees of pharyngealisation for the [-back] consonants, most strongly for /f/.

2.4.1 Plosive release

A further contrast between English and Russian consonants is described for example in Jones & Ward (1969): the release of plosive consonants. In English, voiceless plosives in the onset of a syllable are aspirated (less so in an unstressed syllable), unless they are preceded by /s/ (e.g. [pʰaɪ] vs. [spaɪ]). In Russian, on the other hand, no aspiration occurs.

A further difference lies in the presence or absence of the burst itself. In English in a cluster of two consecutive heterorganic obstruents such as in the word *act*, there is no burst for the first consonants, i.e "the tongue does not leave the roof of the mouth in passing from the k to the t" (Jones & Ward 1969), also see Ladefoged 2003. Conversely, in Russian stop-stop clusters, which, in contrast to English, also occur in onsets, there is a burst also in the first consonant, as can be seen in figure 2.1.

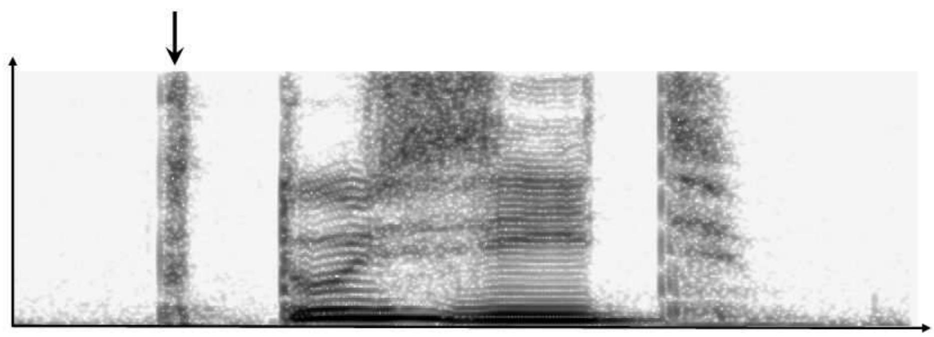

Figure 2.1: The sequence [ptuˈzʲit], as pronounced by a Russian native speaker. The x axis shows time, the y axis shows frequency, ranging up to 8000Hz. The burst of [p] is clearly visible, and indicated with an arrow.

Instead of an overlap between neighbouring consonants, there is a lag between consonant closures. Kochetov & Goldstein (2001) and Zsiga (2000) found that this lag is longest if the first consonant is articulated further back in the mouth than the second one (as in /kp/), with Zsiga claiming that this difference occurs for perceptual reasons, to ensure the perception of the C_1 burst.

2.4.2 Coronal consonants

As seen from the consonant inventories of English and Russian, the anterior coronal consonants in English are alveolar, while in Russian they are dental. Since all coronals used in the clusters of the experiments will be non-palatalised, only those are described.

The dental/alveolar difference implies that the spectral characteristics of, for example, Russian and English [s] (or the [t] burst) differ; since the tongue is further forward in Russian, energy is expected to be concentrated above a higher frequency.

However, while there is a large body of research and descriptions of the acoustic characteristics of different varieties of English (Lehiste 1964, Fant 1973, Bush 1964, Halle et al. 1957, Blumstein & Stevens 1981, Olive et al. 1993, Stevens 1998, to name only a few), and also of Russian (e.g. Bolla 1981, Fant 1970, Zinder et al. 1964, Skalozub 1963, Padgett & Zygis 2003, Zsiga 2003, Kochetov 2006), I am not aware of any studies directly comparing the coronals of both languages directly. Therefore, detailed comparisons could not be drawn from the specifications in the studies concentrating on one of the languages. Nevertheless, the realisations of Russian [t, d, s, z, n] are articulatorily (and impressionistically) clearly different from their English counterparts.

In the group of the liquids, Russian /l/ is classified as dental by Chew (1999), but described as alveolar by Bolla (1981). Non-palatalised Russian [l] is similar to English dark [l]. As the point of interest in the experiment will be Russian onset clusters, it must be noted that speakers of most English varieties have the light /l/-allophone in this position, while before Russian back vowels [l] is dark also in onset position. However, a number of experiment participants are speakers of Scottish varieties, in which there is no allophonic distinction, and onset [l] is also dark.

Russian /r/ is classified as alveolar, as is the English phoneme, so there is no dental/alveolar difference for this phoneme. However, the two realisations are distinguished by manner: the Russian realisation of /r/ is a trill, while in most varieties of English, the sound is realised as an approximant. For Scottish speakers, the realisation may be a trill, a tap or an approximant.

2.5 Summary

While English and Russian exhibit number of phonetic and phonological differences, their inventories and syllable structure make them ideal for a phonotactic comparison. There is a large overlap between the consonant inventories, which allows the use of segment combinations where only the phonotactic status differs for English and Russian listeners. Moreover, with the exception of clusters with Russian-illegal consonants such as /θr/, English CC clusters are a subset of the Russian well-formed clusters, which makes the comparison of both legal and illegal (as well as marginal) clusters possible. This widens the adaptation possibilities of

Russian sequences compared to other languages which do not allow any clusters at all (see next chapter).

Chapter 3

Production and perception

3.1 Introduction

Segmental adaptations in loanwords and second language acquisition are well researched; therefore, the focus of this project is on a different layer of the phonological hierarchy, namely phonotactic conflicts - more specifically, syllables which are allowed in the source language, but not the borrowing language. Previous work in this area has been limited by the nature of the borrowing languages, as described below. The specific case under investigation here is that of word-initial onset clusters which are well-formed in Russian, but not in English.

This constellation seems at first glance unlikely for a study of loanwords. There are only few loanwords which have been borrowed from Russian into the English lexicon, such as *glasnost, vodka, balalaika*, or *bistro* /bɪstɹo/ from Russian /bɨstrʌ/, which has, however, been imported through French /bistʁo/. Of the kind which will be under investigation here, i.e. those with word-initial illegal clusters, there is only a handful to be found in the The Oxford English Dictionary (2007), such as *dvornik* (house servant), *knout* (whip) or *tvorog* (soft cheese), which are virtually unknown to English speakers - with the possible exception of the increasingly well-known *kvass* (fermented cereal drink). This scarcity was in fact one of the reasons why Russian was chosen as the source language for this project, with English as the borrowing language: since there is no previously existing corpus of loanwords with a similar problem (and the situation with words from other Slavic languages is comparable), there is no existing strategy in place of dealing with these conflicts. In contrast, there is an abundance of previously existing loanwords for the constellation of English to Japanese adaptation (e.g. used in Dupoux 1999), and the strategy could have developed through factors other than phonology. This issue was one of the major criticisms of Dupoux' study, and has been avoided here.

A further reason why Russian and English were chosen for this project is that, while previous work on loanwords was often based on corpora of English or French to CV languages only, as commented on in chapter 1, the languages involved here both allow onset clusters. However, the set of well-formed clusters in Russian is much larger and more varied than the English set, thus leaving a sizeable set of Russian clusters which are not well-formed in English. This setup predicts frequent adaptation of ill-formed clusters, but opens up a third possibility besides vowel epenthesis and consonant deletion to "repair" a foreign cluster, namely adapting it to a native one, for example ill-formed /ml/ to well-formed /bl/.

The choice of Russian had a further advantage: besides the fact that no established strategy exists, which may have been influenced by hidden factors, the lack of an existing corpus of loanwords also means that there are no loanword data which have been shaped by orthography or adaptations of words in third languages, and thus distorting the role of phonology. Thus, this lack, which may appear to be a disadvantage at first glance, is in fact an advantage of this project. As a result, the first step of the project was the production of a set of adaptation data tailored to the questions to be answered. To this end, a production experiment was carried out in which English speakers were presented with the task of reproducing a Russian target online, in various ways described below. This experiment setting allowed more control by far over the factors involved; for example, in real-life adaptation the adapters may range from monolingual in the borrowing language to bilingual with the source language; these different groups may use and adapt different subsets of the loanwords. Since greater proficiency is associated with greater faithfulness to the original form of the word, this may result in a heterogeneous loanword corpus within a borrowing language. In contrast to this, the experiment setup allows for control of knowledge of the source language (in this case reduced to zero), which homogenises the group and precludes distortion of the data.

The production experiment consisted of a shadowing task, in which participants were instructed to repeat the Russian words they heard. The results of this can reflect the influence of a number of factors, including articulation, and phonotactic well-formedness in production as well as in perception. Therefore an orthographic task, in which the participants had to give orthographic representations of what they heard, was added. The aim of this experiment was to exclude articulation-based adaptations based on potentially unaltered phonological output, and to take a first step towards determining whether there is a role for perception in the field of loanword study: since the completion of this task does not require speech production, the effects of the production phonology can be avoided[1].

To concentrate directly on perception, a categorical perception experiment modelled on Dupoux et al.'s (1999) study was piloted, yet the full experiment was not

[1]However, see cautionary note in section 3.2.2.

completed for reasons explained below. Nevertheless, the results of the pilot merit mentioning in this chapter, as - in spite of the small amount of data - it shows clearly that speakers of different languages perceive input differently and thus gives a first indication of the importance of perception, which will be further investigated in chapter 5 with a different methodological approach.

3.2 Production

3.2.1 The shadowing task

Introduction

The shadowing task is similar to a study by Davidson (2001), which investigated English speakers' adaptation of Polish word-initial clusters by asking them to reproduce Polish-sounding stimuli in English carrier sentences. The shadowing experiment carried out here used a larger number of English-illegal clusters, and two experiment conditions rather than one. The decisive difference, however, is that in Davidson's study, participants were presented with an orthographic representation of the stimuli, as well as an auditory one. This precluded any potential effects of perception on the performance. In contrast, here there was no visual aid, thus allowing all perception effects which may apply in loanword adaptation.

Method

Materials.

The stimuli consisted of Russian pseudo-words, i.e. non-existent words conforming to Russian phonology. For these, 22 initial consonant clusters of various types (for example stop-stop, fricative-stop, fricative-fricative) were selected. Not all of these clusters appear in the Russian lexicon, however all of them appear in Russian prosodic words, for example through prefixation (cf. chapter 2). Of these clusters, two are legal in English (/sp, fr/), three are marginal in English (/sf, ʃp, ʃm/), and 17 are illegal in English (/pn, sr, zr, vr, vl, tv, dv, ps, ʃv, sv, zv, vz, pt, fp, vb, zb, ʒb/). Only consonants which are also in the English segment inventory were used. For each of the clusters, five bi- or trisyllabic target words were designed. The vowel following the cluster was one of /u, ʌ, ɛ, ɨ/ to avoid palatalisation of the cluster consonants, and the stimuli contained no further conflicts with English consonant phonology, although differences in vowel formants were unavoidable. A list of the target words can be found in appendix B.

These words were transcribed in the Cyrillic alphabet according to Russian orthography, which allows for unambiguous pronunciation if the stress is marked, and the list was recorded twice by a female native speaker of Russian. The recordings were made in a sound-proof recording studio at the Department of Theoretical &

Applied Linguistics, University of Edinburgh, using an AKG CK98 half rifle cartridge on an AKG SE 300B power unit and a SONY PCM-2700A DAT recorder. Recordings were digitised using a Townsend Datlink+ at 16khz mono, multiplexed to the SONY PCM-2700A and to a Sun Sparc Ultra running Solaris 2.4 for archiving to disk.

Following the recording, one target word (/vzʌˈdajə/) was excluded because in neither version could the initial /v/ be detected auditorily or on the spectrogram, which resulted in a list of 109 target words. This list was tested by two further Russian native speakers who were asked to write down the words they heard in Cyrillic. Neither of them had any difficulty in performing this task correctly, and transcribed all onsets targetlike.

Since sonority distance has been used to evaluate cluster well-formedness and varying acquisition difficulties for different clusters (see chapter 2), it will also be relevant in assessing differences between the productions of clusters. Specifically, it is studied whether sonority distance is connected to the success in reproducing the Russian clusters targetlike, or to the preferred adaptation strategy. Therefore, table 3.1 shows the target clusters used in the experiment with their sonority distance.

Participants.

The experiment participants were 10 native speakers of British English varieties, between 18-24 years old, recruited at the University of Edinburgh. None of the participants reported any hearing or speech production difficulties, and none of them had any knowledge of Russian (or any other Slavic languages). This was ascertained after the experiment, to avoid influencing the participants' expectations of the stimuli.

Procedure.

The participants were not informed about the identity of the language the words they were to hear belonged to. However, they were warned not to expect English words. The participants' task was to pronounce these words in two different conditions. In Condition 1 the target word was pronounced within an English sentence to ensure an English setting, whereas in Condition 2 the words had to be repeated without a carrier sentence.

[2]These are calculated by assuming values of O to 5 along stops, fricatives, nasals, liquids, semivowels and vowels. If C_1 is voiceless, and C_2 is voiced, then a "+" is added to the number to indicate that, since voiced segments are more sonorous than voiceless ones, sonority is somewhat larger than the value given.

Sonority Distance of Illegal Clusters [2]	
sp	-1
fr	2
ʃm	1
sf	0
ʃp	-1
pn	2+
sr	2+
zr	2
vr	2
vl	2
tv	1+
dv	1
ps	1
ʃv	0+
sv	0+
zv	0
vz	0
pt	0
fp	-1
vb	-1
zb	-1
ʒb	-1

Table 3.1: The sonority distance of the target clusters used in the production experiment. The clusters are divided into groups of different well-formedness (legal, marginal and illegal), which are separated by double lines.

The stimuli were presented auditorily via Sennheiser HD 457 headphones, in randomised orders. These orders were different for each participant. In the first part participants were shown a computer screen with a carrier sentence with a gap marked for the target word, and were asked to read out the carrier sentence with the target after hearing the target word twice. The carrier sentence remained on screen until the response was complete. The sentences were designed such that it was impossible to syllabify the first consonant (C_1) of the target word with the last syllable of the word preceding the gap, for example *Look at this fantastic ____ I bought!* or *Look at this amazing ____ I bought!*, , where a C_1 of the target word other than /s, t/ cannot syllabify with the final /k/ of *fantastic*, while /s,t/ cannot syllabify with ŋ/ of *amazing*.

The second part of the shadowing task consisted of repeating the stimulus without a carrier sentence after hearing it once. The same stimuli were used, but in a different randomised order. The experimenter activated the next target word after the response was complete.

Analysis.

The participants' responses were recorded in a soundproof studio and digitised in the same way as the stimuli, and analysed by the author, auditorily as well as with the aid of wide-band spectrograms, in *Entropic*'s *waves+* software.

The adaptations were classified into three main groups: vowel epenthesis, deletion of a consonant, and segment change, i.e. the change of a cluster into a different, legal cluster.

Within the latter group, the responses were classified as either resulting in a legal or marginal English cluster or not, and the non-legal class was subdivided into three subcategories, according to whether the resulting cluster was higher, equal, or lower in sonority distance. In the vast majority of cases this is achieved through changing one of the two consonants, for example /vl/ to /fl/. Only in a few cases a double consonant change resulted in a legal or marginal cluster in considerable numbers, such as for /zv/ to /sf/. These will be discussed below. All frequent adaptations are shown in tables in the result section.

For most tokens, data classification was straightforward: for example, the presence or absence of an entire segment was clearly visible except in some cases of vowel epenthesis, and the frequent change from target /v/ to [w] was indicated by a slow formant rise and no frication. Similarly, place of articulation in fricatives, which was frequently changed, was easily discerned, such as a change from target /ʃ/ to [s] from the spectrum above 4000Hz.

In a case of for example /zb/, the output was classified as /sb/ if the fricative was devoiced, rather than as /sp/, unless the plosive was aspirated. This is because in fricative-stop clusters in English the voicing of the stop cannot be determined, since there are no contrastive pairings, and for example in English /sp/, the realisation of the /p/ is phonetically closer to a single onset [b] than [p]. Therefore in the absence of phonetic evidence to the contrary a minimal modification was assumed. Similarly, in the reverse situation of a stop+fricative cluster, the voicing of the plosive at the edge of the word was difficult to determine. For this reason, in the case of target /dv/, where the fricative was devoiced, the token was labelled as [df]. For the target /tv/, when the fricative was voiced, the token was classified as [dv] only where a voice bar during closure was clearly visible, as in figure 3.1.

In contrast to stop voicing, the voicing of fricatives is much easier to determine. Even at word edges, clear differences such as the presence or absence of a voice bar were visible from the spectrograms. Figure 3.2 shows two tokens of target /ʒb/ from the same speaker, one voiced, one devoiced.

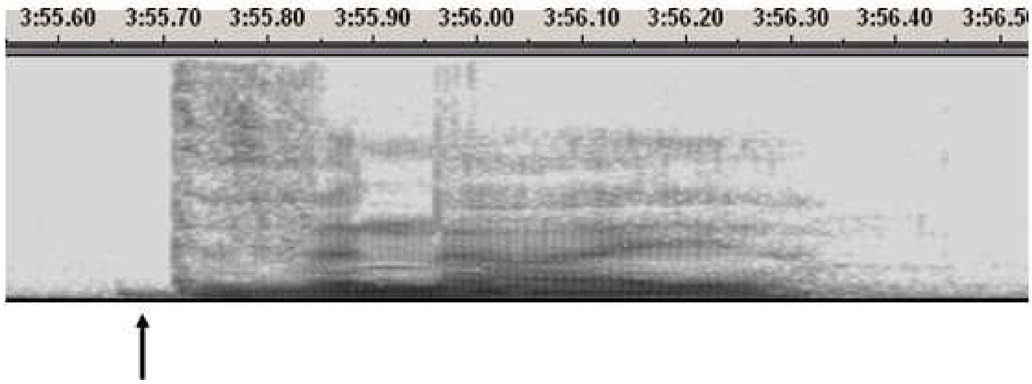

Figure 3.1: The target sequence [tvɛˈnajə], with voiced initial stop. The voice bar of [d] is indicated with an arrow. In this as in the following figures, the frequency on the y-axis ranges up to 8000Hz.

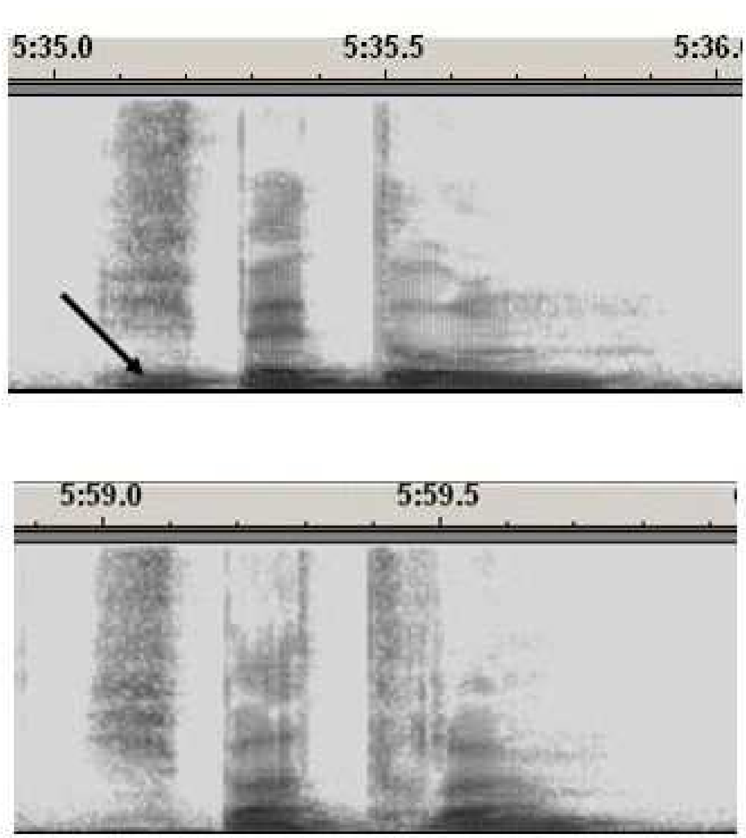

Figure 3.2: The target sequence [ʒbɛˈdom] and [ʒbʌˈtra]; in the first spectrogram, the arrow indicates the voice bar during the fricative, which is not present in the second spectrogram.

Where an epenthetic vowel was produced by the speakers, it varied considerably in length. Whereas in some cases a full vowel was easily distinguished both auditorily and in the acoustic information visible in the spectrogram (figure 3.3), very short vowel-like portions with glottal pulses also occurred (figure 3.4). For these tokens, a cutoff point of 20ms, or at least three glottal pulses in the waveform, was used for labelling these portions as epenthetic vowels.

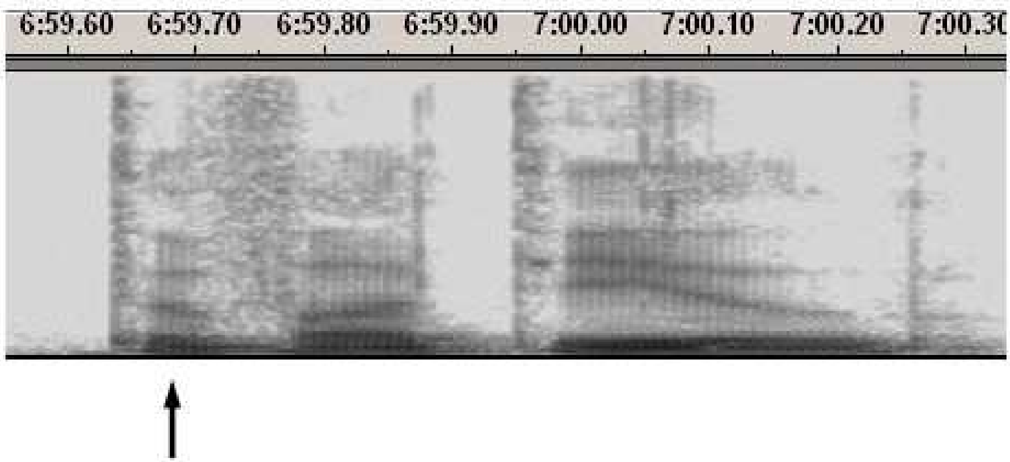

Figure 3.3: The target sequence [dvʌˈdʲel], with an epenthesised vowel marked by an arrow.

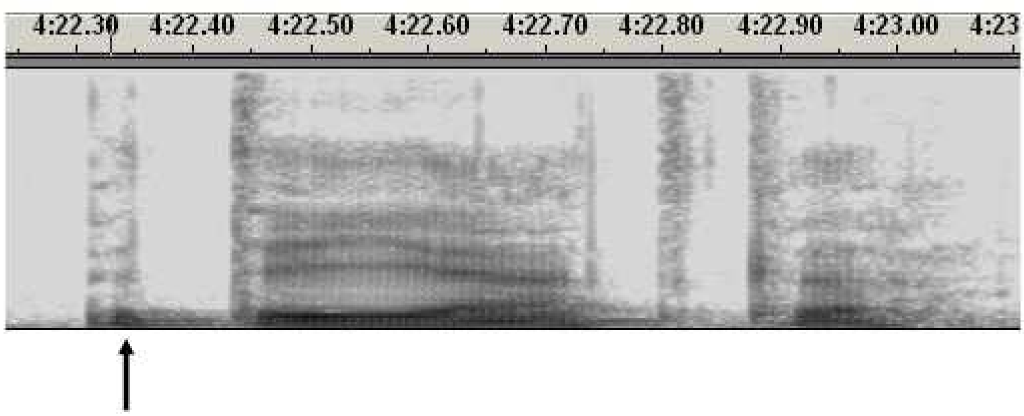

Figure 3.4: The sequence [ptɛˈjatkə], with a short epenthesised vowel marked by an arrow.

If a reliable classification of a token could not be achieved, it was labelled as "other". "Hyper-corrections", i.e. more changes than necessary to achieve a legal English sequence, such as /ʒb/ ⇒ [ʃəb] or /fp/ > [b] were classified as changes on the higher prosodic level (since those always result in legal sequences, whereas the feature changes alone do not), i.e. the former was labelled as epenthesis, and the latter as deletion rather than segment change.

Predictions

With respect to the outcome of this experiments, predictions can be derived from two sources: firstly, from the most prominent factor in discussions of the English syllable and in the acquisition of clusters, i.e. the sonority distance between the onset consonants; secondly, from previous work on clusters in loanword adaptation and various models of loanword adaptation. Not all of the models make specific predictions for this experiment: for example, Silverman's split of the adaptation process into the Perceptual Level (dealing with segmental conflicts) and the Operational Level is not relevant at this stage for two reasons: the stimuli contain no ill-formed segments, and the results of the experiment only give the form which is eventually produced, with no direct indication of what is perceived. For those models from which specific predictions for the case at hand can be derived, the expected outcomes are described in this section.

Firstly, independently of the models, better performance was expected in the word as opposed to the sentence condition, since the target word was embedded in a series of words conforming to English phonology, and the possible distraction of the carrier sentence. However, since each participant completed the sentence condition before the word condition to ensure linguistic processing, this expected result could also be explained with a learning effect.

Furthermore, a better performance was also expected for legal clusters than marginal clusters, which were in turn expected to be produced correctly more often than illegal onsets. It is within the group of the illegal clusters like /vl/ that the predictions diverge.

If specific phonotactic tendencies in English are relevant for adaptation, then those clusters conforming better or to more of these tendencies were predicted to fare better in the experiment. For example, if the /s/-initial clusters in English reflect a preference for voiceless and coronal C_1 obstruents, and these phonotactic generalisations guide English native speakers, then target clusters with initial voiceless or initial coronal consonants are expected to be produced more successfully than those with voiced or non-coronal C_1 segments.

The most prominent variable associated with consonant cluster well-formedness is that of sonority distance (SD), as described in chapter 2. A syllable contour typically rises and falls, which extends to the sonority within onsets and codas. Except for /s/-initial clusters, all onsets in English must have a minimal sonority distance of two, precluding stop-fricative or nasal-liquid clusters, or those with homogeneous stricture. The relevance of this variable has also been studied in second language acquisition (Eckman & Iverson 1993, Broselow & Finer 1991), to the extent that lower sonority distance between cluster consonants increased the difficulty learners had in acquiring a cluster. It follows from this that, within the illegal group, English speakers should have more problems with clusters with a smaller or negative sonority distance (e.g. /zv/, SD=0; /zb/, SD=-1), and produce those with a larger one (e.g. /pn/, SD=2) correctly more often.

In those cases where correct reproduction is unsuccessful, the outcomes were predicted to have higher acceptability with respect to English phonology; this could either be achieved through being completely well-formed, or through increasing sonority distance at least to some extent and thereby producing an output onset that is not part of the English lexicon, but does not have the added disadvantage of violating preferences about sonority distance.

For adaptations leading to a well-formed sequence, there are three main categories of possible strategy: the epenthesis of a vowel, the deletion of a consonant, and the change to a different cluster which occurs in English. Of these, traditional loanword research shows a clear preference of epenthesis over deletion in most cases (for example Paradis & LaCharité 1997, Brasington 1981). Segment change in these studies is not a viable option, since consonant clusters are not well-formed in the relevant borrowing languages. If, as Paradis and LaCharité claim, the only relevant factor is the type of ill-formedness, i.e. the illegal combination of two well-formed phonemes, then the strategy is predicted to be homogeneous across the ill-formed clusters: vowel epenthesis.

Furthermore, in Davidson's experiment with Polish target clusters, where segment change was theoretically possible, its occurrence was negligible at 2%. This means that the outcome of the shadowing task anticipates to some extent the question about the role of perception which will be discussed more extensively in the second part of the experiment: if perception, whose effects were excluded in Davidson's experiment, is irrelevant for the choice of adaptation, this percentage is predicted to remain very low here, but it may rise if perception did indeed have an effect on adaptation.

A model which relies solely on perceptual adaptation is Boersma's account. It does not predict a prevailing strategy, like vowel epenthesis for all clusters; instead, claiming that the result of perceiving the foreign input is well-formed in the L1, his model would predict an adaptation to a an English-legal form for any input (unless the perception grammar is gradient). This means that there should be neither targetlike productions nor any other ill-formed sequences: for example, a target like /zr/ is predicted not to be produced as either /zr/ or /sr/. The same prediction comes from Peperkamp's model of phonetic adaptation in perception; additionally, this model claims that there be phonetic motivations for the adaptations; for example, the insertion of an epenthetic vowel may be triggered by a plosive burst (/pt/ to /pət/), whereas the same motivation for epenthesis is not given for a fricative-initial cluster like /sv/.

Along with Kenstowicz, Steriade's model of loanword adaptation, i.e. adaptation in production only, guided by the knowledge of similarity as detailed in the P-map, predicts that the adapted forms will be maximally similar to the target forms. In the case of the P-map, there are very clear statements about relative similarity: for example, if the problem posed by the target can be solved through either devoicing or vowel epenthesis, then the chosen strategy will be devoicing; for example, /vl/ is predicted to be produced as /fl/ rather than /vəl/.

A summary of the predictions to be tested is given in the following list:

a) correct production: higher in

- word condition than sentence condition
- legal than marginal and illegal clusters
- clusters with higher sonority distance

b) adaptations:

- improvement with respect to English
- epenthesis across the board (Paradis and LaCharité)
- very little segment change (derived from Davidson)
- complete adaptation to English phonology (Boersma)
- complete adaptation with acoustic-phonetic motivation (Peperkamp)
- devoicing (rather than epenthesis or deletion) if it results in a legal cluster (Steriade)
- maximally similar to targets (Kenstowicz)[3]

Results

Results for condition 1

The graph in figure 3.5, as well as the following graphs in this chapter, show the different target clusters, or the experiment condition (this is specified in the labels), on the x axis, and the outcome for all the tokens for a cluster or condition. The following categories of outcomes are recorded in the graphs:

[3]Since the role of similarity will be further studied in the following chapters, the discussion of this approach is postponed to chapter 5.

- targetlike production
- vowel epenthesis between the two consonants, for example /zəv/ for /zv/
- deletion of a consonant, for example /z/ for /vz/
- segment change resulting in a different cluster which is legal in English, for example /sp/ for /fp/
- segment change resulting in a different cluster which is not legal in English, but has increased sonority distance compared to the target cluster, for example /sv/ for /zv/

These categories are overall those with the largest shares in production; all of them are improvements with respect to English, resulting in legal sequences in all but the last point, where the improvement consists of an increase of sonority distance. Since the improvements do not cover all cases, the percentages do not add up to 100%; the breakdown of the remaining categories, split into segment changes with same and reduced sonority distance as well as unclassifiable productions can be found in appendix C.

As can be seen from figure 3.5, the prediction that legality is reflected in the rate of correct responses is borne out. Legal clusters, as predicted, have high ratios of correct responses and a negligible percentage of various possible adaptations. For marginal clusters there is a sharp decline in correct responses, and the largest group of adaptations by far are those which change the target cluster to an English legal cluster. The picture is more spread out for the illegal targets: a success rate of only 24% is coupled with four types of adaptations which are used in over 10% of cases each, of which epenthesis is the largest one with 26%. This reflects the variation of both the percentages of correct responses and predominant adaptations within this group, and shows that there is no such clearly preferred overall adaptation as segment change is in the marginal group.

In a comparison to Davidson's experiment, the amount of segment changes is markedly increased: instead of 2%, in this experiment in 31% of illegal clusters the targets were adapted into a different consonant cluster, slightly less so for the marginal targets (28%). Since the main difference between the two studies is that in Davidson's experiment the participants were provided with an orthographic representation of the auditory stimuli, it follows that the higher frequency of segment change is caused by adaptations before production.

It can be further observed from figure 3.5 that, consistently with the prediction, the vast majority of adaptations result in an improvement of the cluster with regard to English, be it through the production of a legal sequence, or an increase of

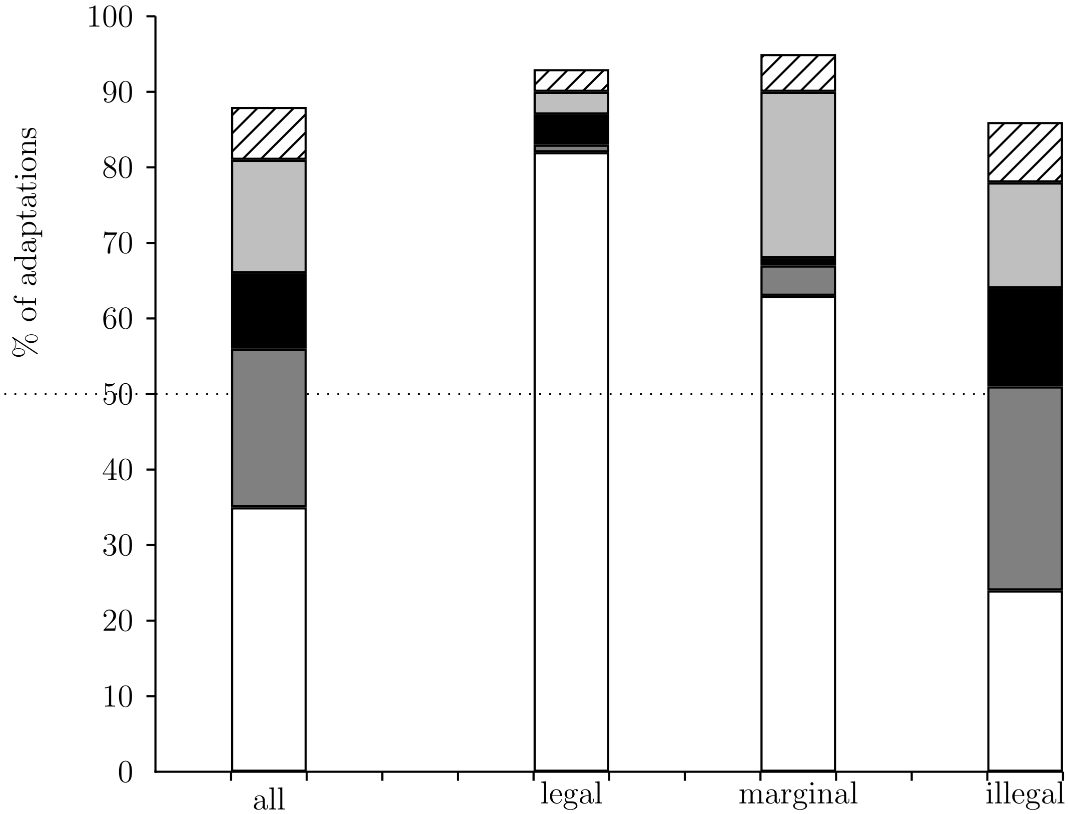

Figure 3.5: The summary for all clusters in the sentence condition. The colours symbolise the following adaptations and adaptation types: no adaptation (i.e. correct pronunciation of target): white; epenthesis: gray; deletion: black. The light gray and hatched bars show segment changes: English legal results are light gray, hatched are those where the produced cluster is not legal in English, but has increased sonority distance compared to the target cluster.

sonority distance. For an evaluation of the further predictions, the results for the individual clusters must be studied.

Legal/Marginal clusters. In a cluster by cluster comparison of legal and marginal clusters, figure 3.6 shows that the rate of correct responses is high with 80% or over for the English legal clusters /sp/ and /fr/, as expected. No clear pattern of changes can be established. In the group of the marginal clusters /ʃm, sf, ʃp/ somewhat more errors can be observed: for the latter two clusters the ratio of correct productions is below 60%. For /ʃm, ʃp/ the main alterations are those intuitively expected, namely changing the place of articulation from postalveolar to alveolar, resulting in a well-formed s-cluster. Predictably, this is the case more frequently for /ʃp/, since it reflects that the differentiation in English from /sp/ is nowhere near as clear as for the /ʃm-sm/ pair. For example, the pronunciation of English *spiel* varies between /sp/ and /ʃp/,

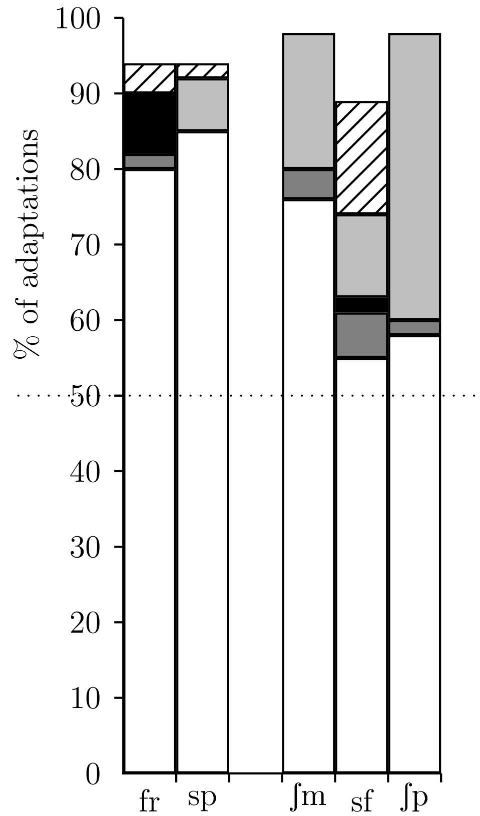

Figure 3.6: The productions of the legal and marginal clusters in the sentence condition.

> whereas /ʃm/-initial words in English, as in *schmaltz, schmendrik, schmuck* the first sound is invariably /ʃ/.

In figure 3.7, the illegal clusters are ordered by sonority distance, with clusters with a voicing difference[4] placed at the left edge of each group. /pn/ on the left has the highest SD of 2, whereas the last group of fricative-stop clusters has SD=-1, with none of them being an s-cluster with a possible different status.

As discussed above, a case could be made for either expecting an equally unsuccessful performance for all clusters, since all are illegal in English, or a graded outcome in that production of clusters with higher sonority distances should be easier to accomplish and therefore successful more often. Clearly, neither of these predictions holds: the ratio of correct responses ranges from 0% to over 50%, providing evidence against the TCRS model. Moreover, there is no downward trend along the graph's x axis. Sonority distance fails as a predictor for success in producing illegal clusters.

[4]For Russian clusters, a voicing difference normally means that the first consonant is voiceless, and the second voiced (for details, refer to chapter 2). Consequently, clusters like /sv/ have a higher sonority distance than those which agree in voicing, like /zv/ (see table 3.1.)

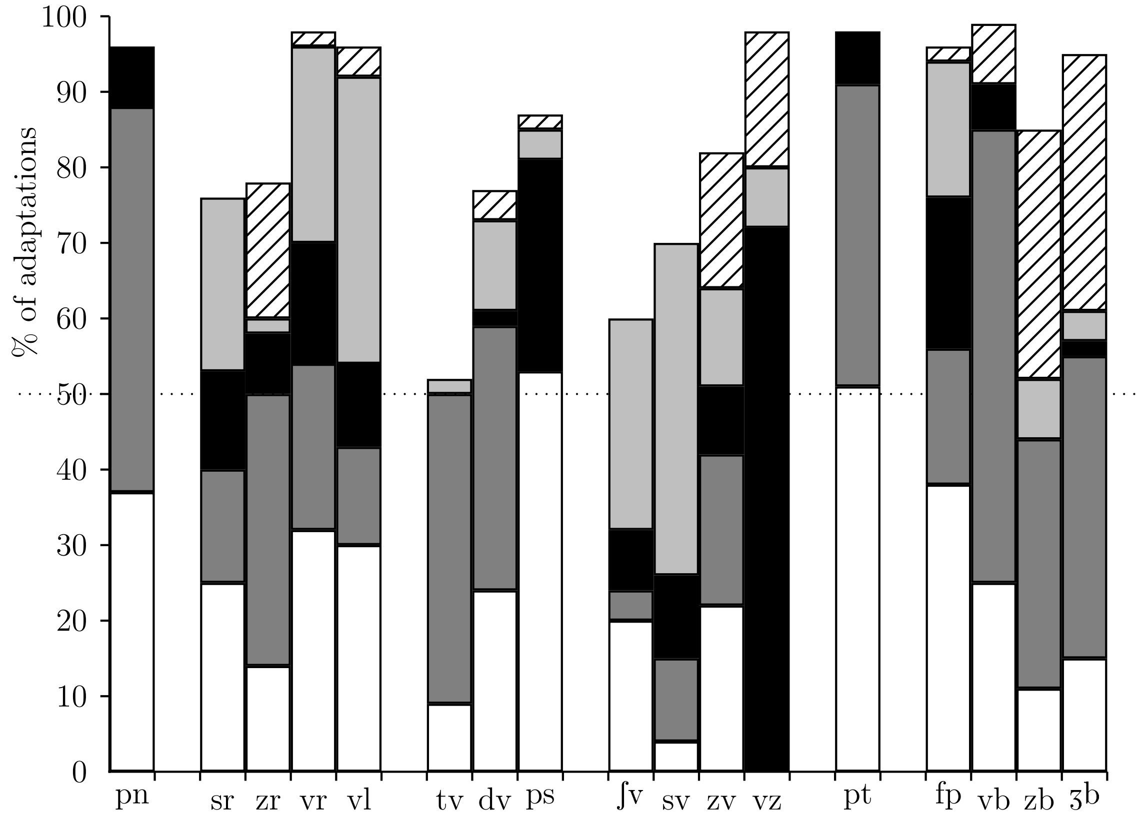

Figure 3.7: The productions of the illegal clusters in the sentence condition. The clusters are grouped by sonority distance in this graph as well as in the corresponding figures showing performance in the other conditions.

Nevertheless, sonority distance appears to be of influence in the cases where the targets are not correctly reproduced, and are not adapted to a legal sequence: for example, for all voice-initial target clusters (except /dv/, see specific note) the devoicing of the first consonant, and other changes which increases sonority distance, outweigh the sonority-neutral or sonority-negative adaptations (for a breakdown of percentages, see appendix C).

Paradis' prediction that all clusters will be adapted in the same way since all are combinations of legal segments in ill-formed combination, is shown to be incorrect through the wide variety in adaptations used, as well as the differences between clusters in terms of their most frequent repairs.

Furthermore, Boersma's and Peperkamp's notion that the result of the adaptation process conforms to the native language system, is contradicted by a targetlike, i.e. non-native, production which is as frequent as over 50% of tokens for several target clusters.

The most successful prediction is Steriade's claim that the option of devoicing to resolve a conflict will be taken if it exists. Indeed, for the target clusters /vr, vl, sv, zb, ʒb/, where the devoicing of one or both segments results in a legal or marginal cluster, it is indeed a frequent or even the most frequent repair. However, this preference is not as exclusive as it is in Steriade's prediction, and epenthesis and deletion are used almost as often.

Illegal clusters by sonority distance. Since for illegal clusters, none of the predictions for adaptations are fully supported, the different types of clusters are analysed in detail. Within groups of the same sonority distance, there are some weak tendencies to similar outcomes, although these tendencies are restricted by further differentiation within the groups.

For example, the fricative-liquid group shows a rough similarity between the clusters in that the ratios for correct shadowing, deletion and segment changes are comparable; however, a look at the most frequent adaptations shows that segment changes are most common in the cases where the necessary change to achieve a legal English sequence is devoicing of the fricative, for /vr, vl/. The share of segment changes (along with correct pronunciation) decreases for /sr/, where alteration involves a different place of articulation, to /ʃr/, and finally becomes lowest within the group for /zr/, where a one-feature modification resulting in a legal cluster is impossible.

The group of fricative+plosive clusters (e.g. /zb/) has a sonority distance of -1. In English, there is one kind of cluster like this, i.e. the s+stop clusters. Since /s/ is a coronal fricative, one might expect a different outcome for /fp, vb/ as opposed to the coronal-initial /zb, ʒb/. In fact, the latter appear to pose more difficulty to the English listeners, and there is also a difference in the preferred alterations; the largest share of alterations for the coronal group besides vowel epenthesis falls on the devoicing of the fricative, which not only improves sonority distance, but in fact brings the result closer to the legal (but violating the Sonority Sequencing Generalisation, cf. chapter 2) English cluster /sp/ by creating initial /s/. Support for this interpretation comes from the results for /fp/ and /vb/, where a comparable trend to devoicing cannot be observed. Instead, as above, where a single feature change can result in a legal cluster (/fp/ > /sp/), this method is chosen more often, whereas in the other case vowel epenthesis is by far more frequent.

The group of plosive+fricative clusters (e.g. /ps/), of which there are none at all in English, is rather heterogeneous: on the one hand, /dv/ and /tv/ are fairly similar in that the most frequent improving adaptation is vowel epenthesis, and deletion plays no role at all. On the other hand, /ps/ is completely different: in over 50% of tokens, participants succeed in reproducing this cluster. This is the best performance for any of the illegal clusters. Furthermore, no epenthesis takes place. Instead, the most frequent change is a deletion of /p/, which shows that /s/ has a special role - likely linked to the acoustic salience of [s] as compared to [v] - both in first and second position. The fricative-fricative clusters (e.g. /sv/), which equally do not exist in English at all and are the second-lowest in terms of sonority distance at 0, show on average the worst performance. In this group also, the differences between clusters are more striking than the similarities: the combination of coronal in first place and labiodental in second place is still produced with a limited amount of success up to the 20% range, and a variety of adaptations. The reverse case /vz/ is never produced correctly at all, and instead of a combination of alterations, the deletion of /v/ in C_1 position is by far the most frequent adaptation, and is used in 70% of the tokens. In the coronal-first clusters, on the other hand, in case of deletions C_2 is affected, i.e. again /v/. Participants are relatively successful in producing /pn/ with the highest SD=2 and /pt/ with SD=0. Notably, both of these are stop-stop clusters, where the first stop is released.

Overall, the comparison of individual clusters shows that what is more important both in success of reproduction and in the choice of adaptation than the sonority distance in a cluster is makeup of a cluster in terms especially of place and voicing, in combination with the salience of segments and the possibility of creating a legal or marginal cluster: for example, a voicing change is preferred for the targets /zb, vr/, but not for the targets /vb, zr/ where the result is not well-formed in English.

Results for Condition 2

As predicted, in the word condition (Condition 2), without the carrier sentence, the rates of correct responses were somewhat higher, as compared to the sentence condition. However, as noted above, this may have been caused by a practice effect. The performance in both conditions was nevertheless very similar. In order to test how similar the performances were, a Pearson correlation was calculated between the percentages of correct responses the two tasks, for the set of illegal clusters. The correlation was found to be highly significant (r_p=.85, N=17, $p<.01$, 95% CI: $.63< r<.95$), showing no evidence of the condition influencing performance.

Comparing the summary of adaptations by legality to the Condition 1 summary, hardly any change can be found. Again, the prediction that the rates of targetlike production are highest for legal and lowest for illegal clusters is confirmed. The largest difference lies in the reduced number of segment changes in the marginal clusters. In the illegal targets, however, aside from the slightly higher success rate,

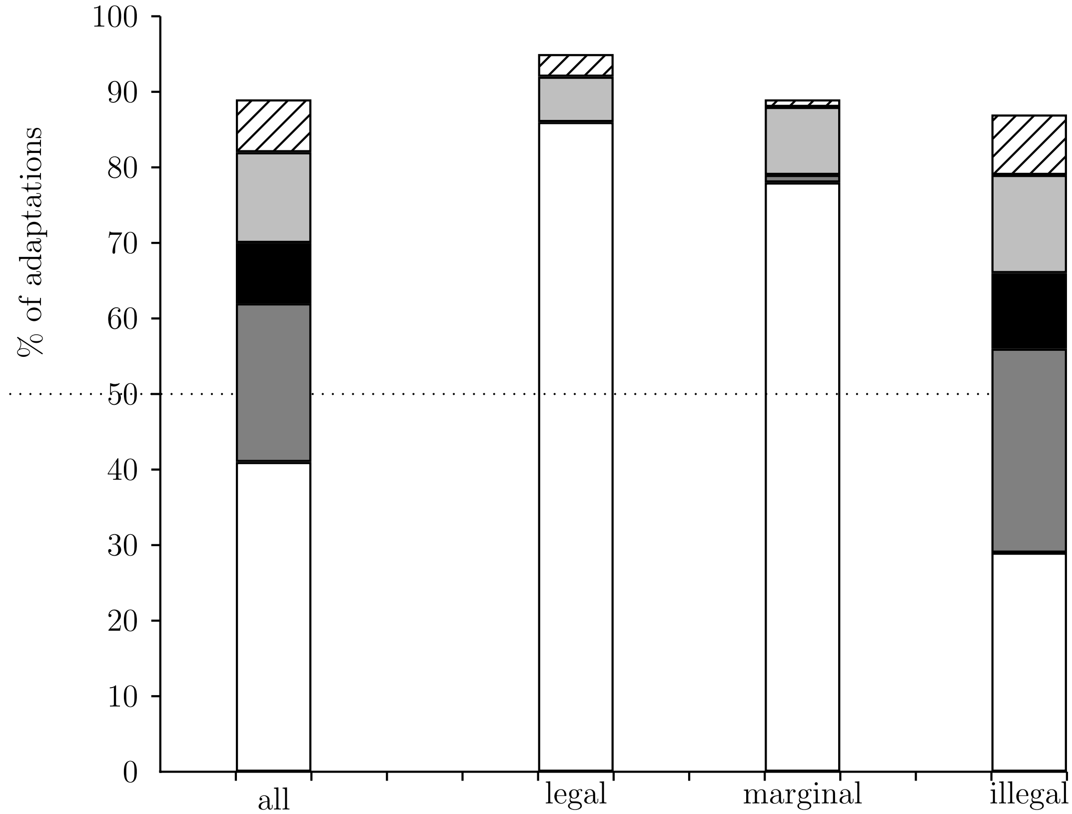

Figure 3.8: The summary for all clusters in the word condition.

the distribution of adaptation mechanisms is virtually the same, with the largest difference lying in a reduction of deletion cases from 12% to 10%.

Legal/Marginal clusters. A comparison to the sentence condition for the legal and marginal clusters shows that, within the legal clusters, the /fr/ performance is somewhat improved in the word condition, while /sp/ remains at the same value for correct production. For the marginal clusters, however, the improvement is at least 10% for each cluster, and consequently all changes including those creating a legal cluster are reduced. However, the cluster where the sonority-increasing adaptation is most frequent is still /sp/ due to the /sp-ʃp/ interchangeability of some English <sp> words, as explained above.

For the group of the English-illegal clusters, the picture looks much the same, while the rate of correct responses has risen to 29% as opposed to the average 24% in Condition 1. The range of correct response ratios for individual clusters again ranges from 0% to over 50%.

With respect to the predictions, the results overall confirm the conclusions of the sentence condition, and the lack of support for any predictions on the nature of the adaptations, except to some extent for Steriade's preference of voicing changes.

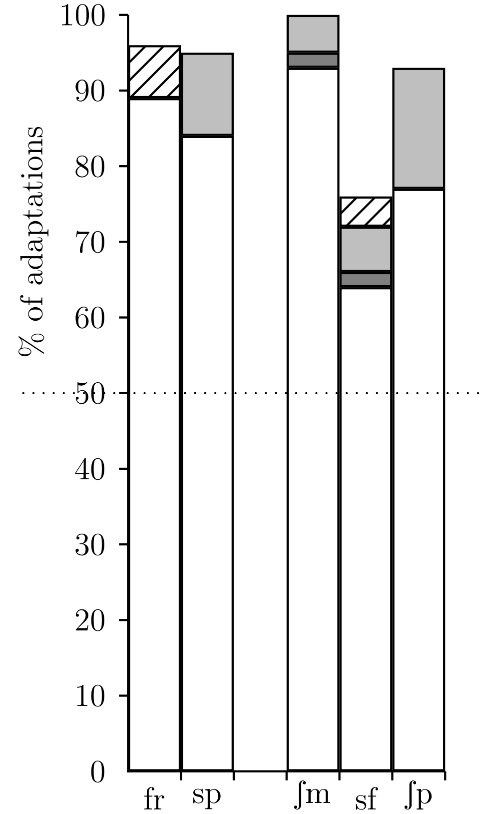

Figure 3.9: The productions of the legal and marginal clusters in the word condition.

Illegal clusters by sonority distance. Looking at the groups divided by sonority distance, the fricative-liquid clusters display the same preference for segment change for the targets which can be made legal with a voicing change, a less pronounced preference where a POA (point of articulation) change is necessary, and hardly any responses in this category for /zr/ without a possible one-feature change to make it legal.

For the fricative-stop targets, all clusters aside from /vb/ show marked improvement in condition 2, and higher success rates, taking /fp/ to over 50%. Consequently, the high/low success division for this condition is between the voiceless and voiced targets rather than labial- versus coronal-initial clusters. Nevertheless, in terms of adaptations, the distinct preference for devoicing remains for /zb/ and /ʒb/, but not for /vb/, while only /fp/ and /vb/ show any deletion percentage to speak of. This underlines the relevance of both the voicing and the coronal factors.

In the group with reverse configuration, i.e. with a plosive before a fricative, only /tv/ and /dv/ show any improvement in this condition. The adaptations which do occur, however, show much the same distribution as in Condition

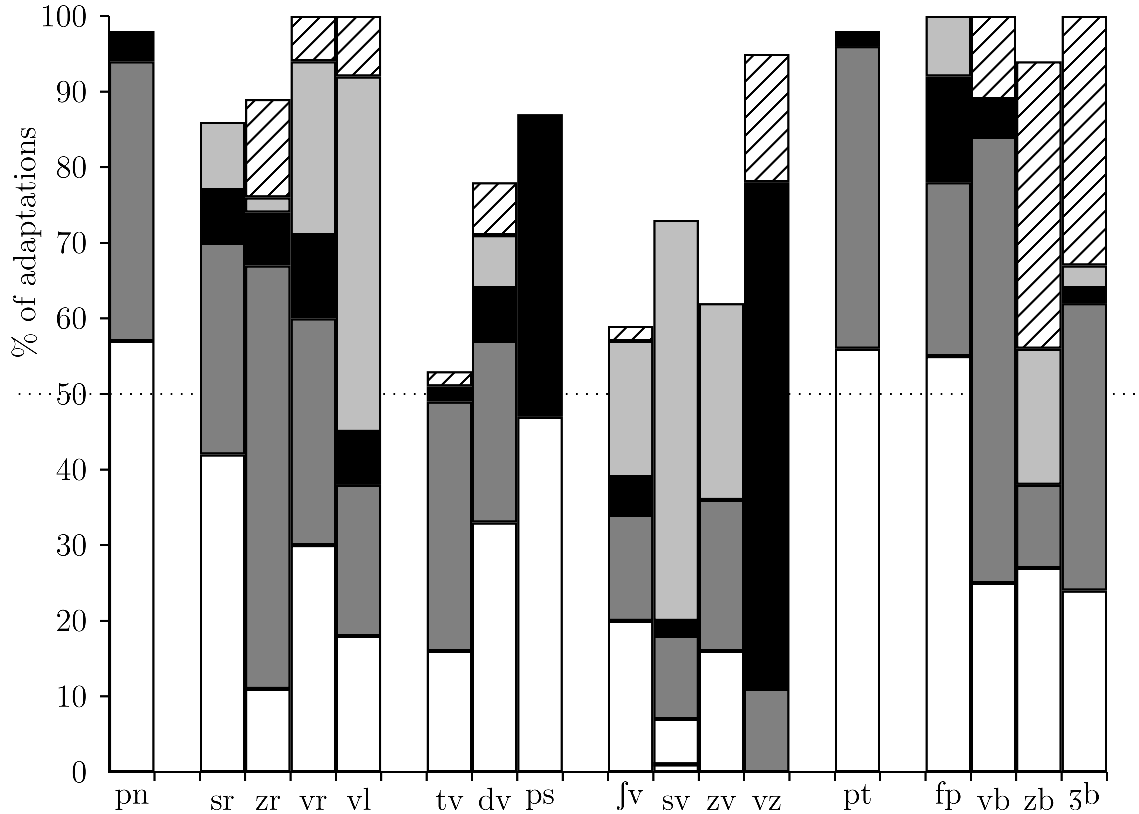

Figure 3.10: The productions of the illegal clusters in the word condition.

1: epenthesis is only used in the coronal-first clusters /dv/ and /tv/, whereas /ps/ is again repaired almost exclusively by deletion of the initial [p], while preserving the more salient coronal fricative.

The set with the lowest percentage of successful production in the sentence condition, i.e. the fricative-fricative clusters, is also the set with the lowest score here, with no improvement to the first task, in contrast to most other groups. The distribution of adaptations is mostly the same, except that the role of deletion is slightly diminished; for example, /vz/, which was previously only improved by /v/-deletion, is now also repaired by vowel epenthesis in 10% of productions.

/pt/ and /pn/, which have a similar stop-stop configuration (albeit a difference of two between their sonority distance values), are again the most successfully reproduced clusters.

As in the sentence condition, the adaptations are not determined by sonority distance or another general feature of the target cluster, but more by salience and the individual possibilities of changing an ill-formed cluster into a well-formed one. Again it can be observed that if a voicing change is a possibility for achieving this goal, it is used very frequently (e.g. for target /vl/), more so than a place of articulation change which creates a well-formed cluster (e.g. for target /fp/).

Frequently used adaptation strategies

Table 3.2 lists the main adaptations, i.e. those used in at least 10% of cases, in order of decreasing frequency within each cluster. This table reinforces what is shown in the graphs, in more detail: the adaptation strategies are very similar between the two conditions, and while the precise percentages of adaptations change in both directions (for example, there is more epenthesis in the sentence condition for /pn/, but more in the word condition for /zr/), the *ranking* of the most favoured adaptations rarely changes.

Frequent adaptations of marginal and illegal clusters (≥10%)[5] in the sentence and word conditions		
Cluster	Sentence	Word
ʃm	sm	
sf		tsf
ʃp	**sp**	sp
pn	***pən***	**pən**
sr	sər, s	**sər**
zr	**zər**	***zər***, sr
vr	**fr, vər**, v	**vər, fr**, v
vl	**fl, vəl**, v	**fl**, vəl, v
tv	***təv***, **tf**, dv	**təv, tf**, dv
dv	**dəv**, df, dw	**dəv**, df
ps	**s**	***s***
ʃv	**ʃf**, sf, sw	**ʃf**, səv
sv	**sf**, səv, s, ʃf, tsf	***sf***, **səv**
zv	**zəv**, sf	**sf, zəv**
vz	***z***, ts	***z***, ts
pt	***pət***	***pət***
fp	**p**, fəp, sp	**fəp**, p
vb	***vəb***	***vəb***, fb
zb	**zəb, sb**	**sb**, zəb
ʒb	***ʒəb***, **ʃb**	**ʒəb, ʃb**

Table 3.2: Main adaptations for the shadowing task, divided into the sentence and the word only conditions.

What is noticeable in a comparison of the two conditions is that, in Condition 2, along with the improvement in producing the clusters comes a slight uniformisation:

[5] Adaptations occurring over 20% of the time are printed in bold, those in over 30% are printed in bold italics.

for most clusters the number of different adaptations over 10% is less than in the sentence condition. Besides this, there is not a clear trend towards an increase or reduction for any given strategy.

Vowel epenthesis. In both conditions, the strategy of epenthesising vowels is most often the most frequent, as the graphs above showed as well. However, its occurrence is not connected to the target clusters' sonority distance; for example, both /pn/ with the highest sonority distance and /vb/ and /ʒb/ show epenthesis as the preferred repair. Within groups with the same sonority, it is also prominent for the coronal-initial fricative-liquid clusters, and also the coronal-initial stop-fricatives. The dominance of vowel epenthesis for the fricative-stop clusters does not stretch over both conditions and all clusters, and the most heterogeneous group is that of the fricative-fricative targets.

Deletion For the deletions, the most striking aspect about their application is that for those clusters where it is the most frequent repair, they dominate the production of these clusters, and other adaptation possibilities occur only marginally.

Two further observations can be made about consonant deletions: one, coronal stops and fricatives are not deleted; two, there is a preference for the acoustically motivated prediction that the consonants at the word edge - rather than those adjacent to a vowel - will be deleted, presumably since their acoustic cues are less salient. These two factors point to the same result for /ps/, resulting in /p/-deletion being the dominant repair, and /vz/, where /v/ is deleted in over 70% of productions. However, for /sr/ these trends pull in opposite directions, with the avoidance of coronal deletions winning out.

A third factor involved in deletion appears to be voicing; a more sonorous /v/ in /vb/ (up against a less salient rival than /s/) is preserved, and repaired by epenthesis, while the voiceless /f/ in /fp/ is deleted more often. Similarly, /v/ preceding liquids is preserved and repaired through epenthesis, and any deletions remove the liquid rather than the fricative, hence opposing the *delete C_1* trend. However, in many productions there was a reflex of the liquid in the shape of a back rounded vowel, which may speak against a classification of these productions as deletions, particularly given the darkness of Russian [l].

Segment changes While in deletions either the first or second consonant of a cluster can be removed, the group of segment changes is more variable still: either consonant can be changed, and each consonant can be altered with respect to one or more of a set of defining features. The most frequent of all of these changes was a voicing change, specifically the devoicing of C_1. This was especially common where it led to a legal cluster, i.e. for /vl/ and /vr/ adapted to /fl/ and /fr/. For a cluster of the same sonority distance, /zr/, where the same change does also increase sonority distance, but results in still illegal /sr/, C_1 devoicing still occurs, but to a far lesser degree. Similarly, in the voiced fricative-stop clusters C_1 devoicing is used, yet the dominant repair is epenthesis, since only to change the voicing of the first consonant would not produce a legal cluster. Notably, it is least frequent for /vb/, where devoicing does not result in a voiceless coronal first fricative, i.e. those allowed in English legal or marginal clusters, either.

A double devoicing or devoicing of C_2, i.e. those which cannot be motivated by sonority distance, occurs where the factor sonority distance can be overridden by the motivation to produce a legal cluster or marginal cluster, for example through the adaptation of both /zv/ and /sv/ as /sf/. Note that in the reverse cluster, /vz/, devoicing is not used since it would not result in an improvement. By analogy to /sf/, /ʃv/ is also produced as /ʃf/ as most frequent adaptation. Further voicing changes, not all of which can be explained as improvements with regard to English, are produced for /tv/ and /dv/. This may be due to voicing confusion for the initial stop due to lack of aspiration in Russian voiceless plosives.

The second most frequent segment changes after voicing changes are alterations of place of articulation, the most common of which is from palato-alveolar to alveolar. It is applied most extensively for the marginal clusters /ʃp/ and /ʃm/, as discussed above, and also in a mixed strategy together with devoicing, producing /ʃv/ as the marginal /sf/.

A further place of articulation change occurs in /fp/, resulting in /sp/. Notably, this place change is not combined with devoicing for /vb/. A possible explanation for this is the somewhat fuzzy boundary between /s/ and /ʃ/ in marginal clusters in English, whereas a similar fuzziness involving /f/ and /s/ does not exist.

In several cases manner changes occur, namely changing C_2 from /v/ to /w/, again producing a legal output, for /ʃv/ changing to /sw/, and /dv/ changing to /dw/. Phonetically, the velar secondary articulation of /v/ may contribute to this change, suggesting [w] as a close English alternative to Russian [v]. However, a manner change is never the most frequent repair, possibly due to this change being perceived to be a more drastic alteration than devoicing or POA changes.

Summary

The following list shows the predictions which were borne out in the shadowing experiment in boldface:

a) correct production: higher in

- **word condition than sentence condition**
- **legal than marginal and illegal clusters**
- clusters with higher sonority distance

b) adaptations:

- **improvement with respect to English**
- epenthesis across the board (Paradis and LaCharité)
- very little segment change (derived from Davidson)
- complete adaptation to English phonology (Boersma)
- complete adaptation with acoustic-phonetic motivation (Peperkamp)
- **devoicing (rather than epenthesis or deletion) if it results in a legal cluster (Steriade)**

Participants' performance was more successful for legal and marginal clusters, as predicted. Within the illegal targets, contrary to Paradis' approach, a differentiation was to be found, both in terms of correct response rates and with respect to preferred adaptations.

Overall, the most frequent repair strategies in both the sentence and word conditions were found to be vowel epenthesis (/ʒbedom/ ⇒ [ʒəbedom]) and segment changes of one or both consonants (/fpurot/ ⇒ [spurot]), followed by deletion (/vzana/ ⇒ [zana]). Most of the segment change repairs altered voicing or place of articulation, and roughly half of them resulted in a legal English cluster. For the other half there was often an improvement in that the sonority distance between the two consonants was increased, for example /ʒbedom/ ⇒ [ʃbedom].

These sonority increases as well as all changes resulting in a well-formed English cluster were regarded as improvements. Overall, 80% of adaptations improved the cluster in both the sentence and word conditions. However, the results showed large differences between the target clusters in terms of percentage of correct, i.e. unadapted, responses, as well as in terms of preferred adaptation. Sonority distance does not influence the rate of targetlike productions, nor does it determine the preferred adaptation: while /zv/ with a sonority distance of 0 is most often adapted by epenthesis, /vz/ (with the same sonority distance) is repaired by deletion almost without exception.

None of the predictions of adaptation following from theories of loanword adaptation were supported, with the exception of Steriade's claim of a preference for voicing changes, however not absolute. The adaptations vary with the possibility of adapting the target to a legal or marginal English cluster, while keeping the alterations to the original minimal. This means that in some cases segment changes are viable (e.g. /vl/ to /fl/), whereas in other cases the necessary modifications may be deemed too severe to be considered an appropriate adaptation.

3.2.2 The orthographic task

Introduction and predictions

In the shadowing experiment, participants were instructed to repeat words they heard. The adaptation process, however, not only consists of mapping an underlying form to an output form; as well as potential problems in articulating non-native forms, the target forms must first be perceived before being produced; i.e., an underlying form must first be created before a phonological production process can take place. Therefore a different task was designed to exclude articulation-based adaptations, and to gain insight into the role of perception.

This task was designed to create a variation on the shadowing experiment, but to avoid actual speech production. The method to achieve this and at the same time to approach the representation which is formed in perception, was for participants to produce an orthographic interpretation of what they heard, rather than pronouncing it.

If the perceptual factor is decisive for the adaptations observed in the shadowing task, then the same adaptations are expected in this experiment, since the perceptual process is a part of both experiments. If, on the other hand, perception does not filter out any foreign sequences, and all adaptations occur in production, then a large discrepancy is predicted because there is no speech production involved in the orthographic task. However, since speech production consists of the phonological production process *as well as* articulation, and it is not known how much of the phonological process is passed through before writing[6], this discrepancy is not necessarily complete. Additionally, the existence of subvocalisation, i.e. internal speech, during the reading process, might suggest that a similar sub-production process would take place during the reverse process of producing rather than perceiving a

[6]The connection between phonology and orthography is unclear: there is a close connection of learning an alphabetic writing system and phonological awareness (e.g. Morais & Kolinsky 1994). However, awareness of phonemes and their correspondence to graphemes does not entail an application (or awareness!) of phonological production and phonotactics.

written form. However, reading subvocalisation is associated with registering visual information in a "phonological store" (e.g. Baddeley 1992), i.e. with creating a phonological form, while in the experiment at hand the phonological form is created from an auditory input, and is then followed by the production of the orthographic representation. In sum, the orthographic task is not claimed to represent exactly the results of the perceptual process, but to unveil outcomes which are closer to those than the results of the shadowing tasks are.

An increase of targetlike responses as compared to the shadowing tasks would indicate that there is indeed a separation between the production and the perception process, such that adaptations occur in both parts. In contrast, if the performance in this task is not noticeably different from the previous task, this would amount to the same performance whether or not speech production is involved, and hence strengthen the role of perception in loanword adaptation, as Peperkamp and Boersma suggest[7]. On the other hand, the position of Paradis and Steriade, i.e. that all adaptation occurs within the realm of speech production, would be strengthened if the number of targetlike responses rose to a nativelike level. This would also be predicted by Silverman, since all segments in the target cluster are part of the English sound inventory.

The following predictions with regard to the orthographic task will be considered.

- The rate of targetlike responses is higher than in the previous tasks
- The rate of targetlike responses the same as in the previous tasks (Peperkamp, Boersma)
- The rate of targetlike responses close to 100% (Paradis, Steriade, Silverman)

Method

Materials.

The same materials as in the shadowing task were used and recorded onto DAT tape in randomised order, with two repetitions of each stimulus directly following each other.

Participants.

A different group of 11 native speakers of British English varieties, between 19-26 years old, also recruited at the University of Edinburgh. None of the participants reported any hearing or speech production difficulties, and none had any knowledge

[7]Note, however, that their models go further, and Peperkamp claims a phonetic motivation for all adaptations, which goes counter to the large amount of variation found in this experiment.

of Russian (or another Slavic language). As in the previous experiment, this was ascertained by a questionnaire which filled out after the experiment, to avoid creating expectations about the stimuli in advance.

Procedure.

Participants were asked to listen to the two repetitions of each word from the DAT player via Sennheiser HD 457 headphones and then to pause and to write down an orthographic representation that best fit the stimuli. The experiment was self-timed.

There was no instruction of a specific notation system. The conventions of English orthography were deemed to be sufficient for this task, since only the representations of the consonants at the word onsets were relevant for the experiment, which were unambiguous. For example, both <sch> (as in *school*) and <sk> would represent /sk/, and <thr> would represent /θr/.

Analysis.

The same classification system as in the shadowing experiment was applied. However, while previously for fricative-stop target clusters like /zb/, the stop was classified as having the same voicing value as the target if the fricative was devoiced, in the orthographic task either a voiced or voiceless plosive were possible.

Results

Overall, it can be seen from figure 3.11 that the rate of correct responses is only slightly higher in the orthographic task than the shadowing task: overall, the proportion is 44%, as opposed to 35% and 41% respectively. To test the correspondence between the to tasks, a Pearson correlation was calculated between the rate of correct responses in the shadowing tasks (with the average value of both conditions for each cluster) and those in the orthographic task. The rate of correct responses for the two kinds of tasks is significantly correlated (r_p=.47, N=17, p=.03; 95% CI: .00$<$r$<$.78), with the r-value, i.e. the degree of correlation, being noticeably lower between tasks than within the shadowing task (r_p=.85). Additionally, the distribution of adaptation strategies is also similar across tasks.

Compared to the overall similar rate of correct responses, the difference between the tasks becomes more pronounced if one looks specifically at the illegal clusters, where the success rate in this task is at 41% (figure 3.12), as opposed to 24% and 29% in the shadowing experiment. This result supports the conclusion that there are indeed fewer adaptations in the orthographic than in the shadowing task.

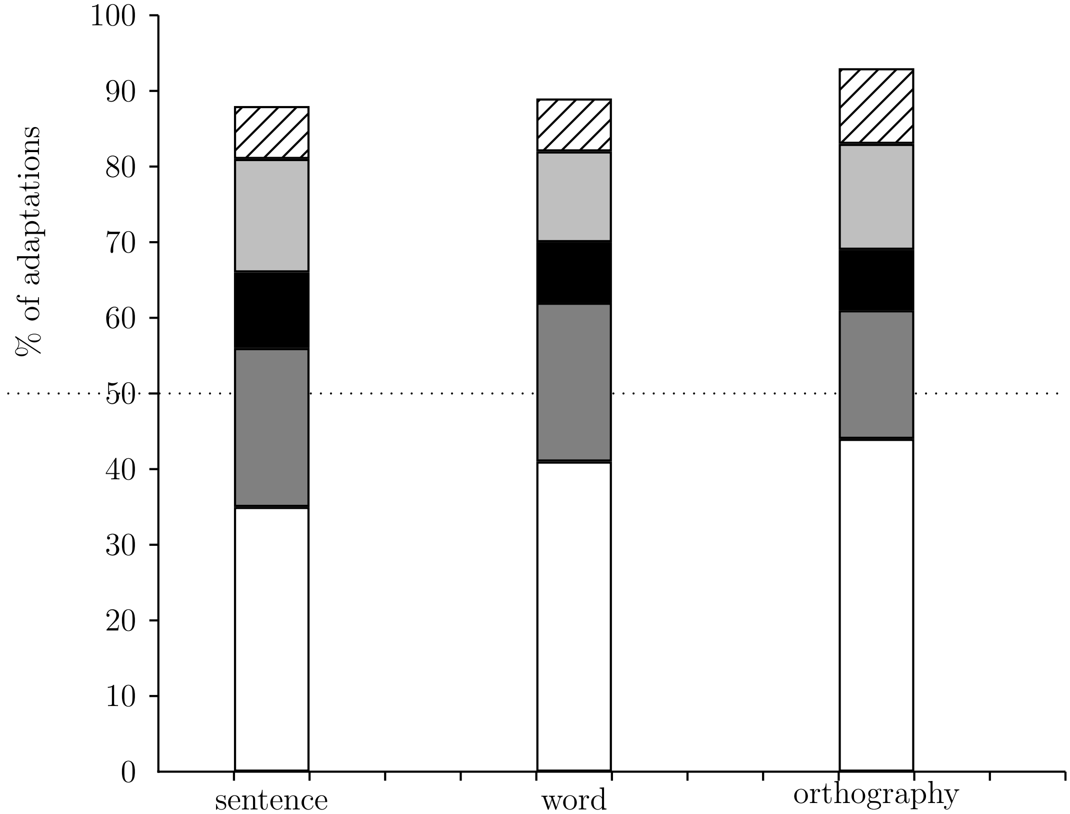

Figure 3.11: The summary for all clusters in all conditions.

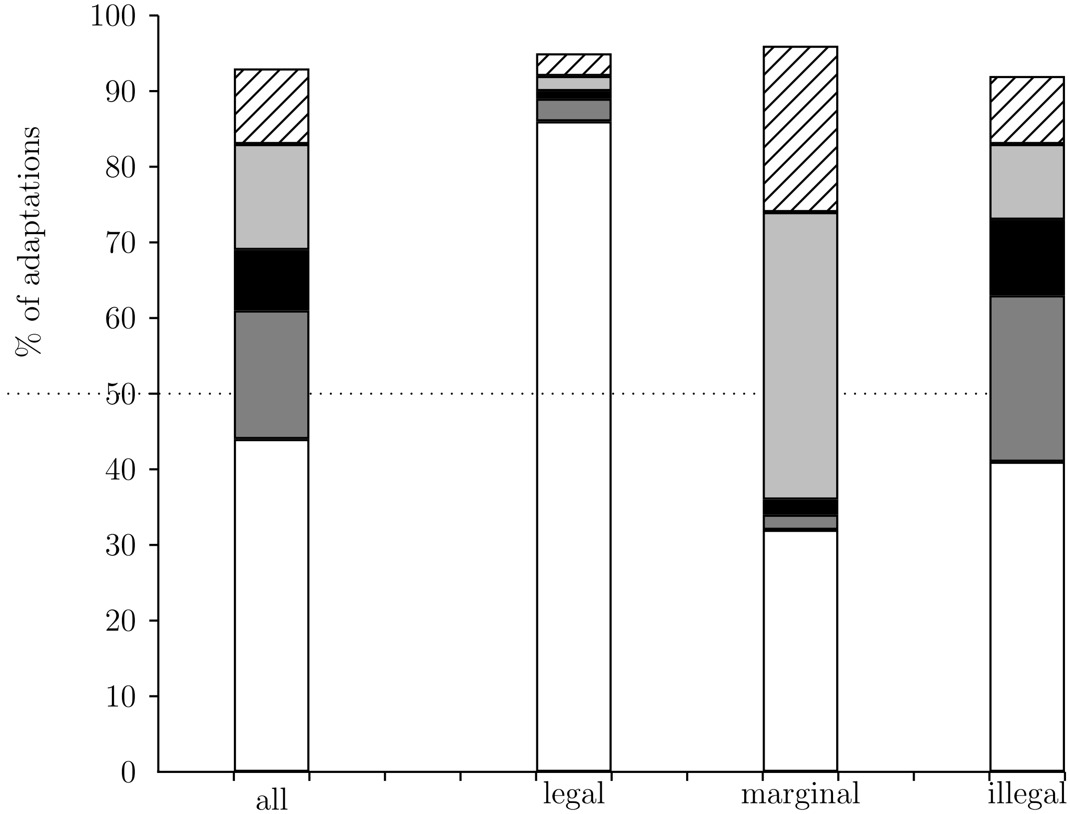

Figure 3.12: The summary for all clusters in the orthographic task.

Nonetheless, it is clear that they have not been reduced to zero, as proponents of production-only adaptations predict.

In a comparison of the summary for the orthographic task only (figure 3.12), divided by the legality of the target clusters, the most striking difference to the shadowing task is the lowering of the correct response rates for the marginal clusters by about half, and the relatively smaller role of epenthesis and larger role of segment changes. Within the illegal clusters, however, an increase of correct response rates by over 10% compared to the shadowing tasks can be observed. Therefore each kind of adaptation is reduced in frequency; nevertheless, the distribution of methods remains almost the same as in the first experiment. Despite this similarity, the slight shift in preference - the insertion of an epenthetic vowel is used slightly less often here (21% rather than 26% and 27%), as is segment change (26% vs. 30% and 28%) - means that the most frequent adaptation type in the orthographic task is a segment change, whereas epenthesis was favoured in the previous task.

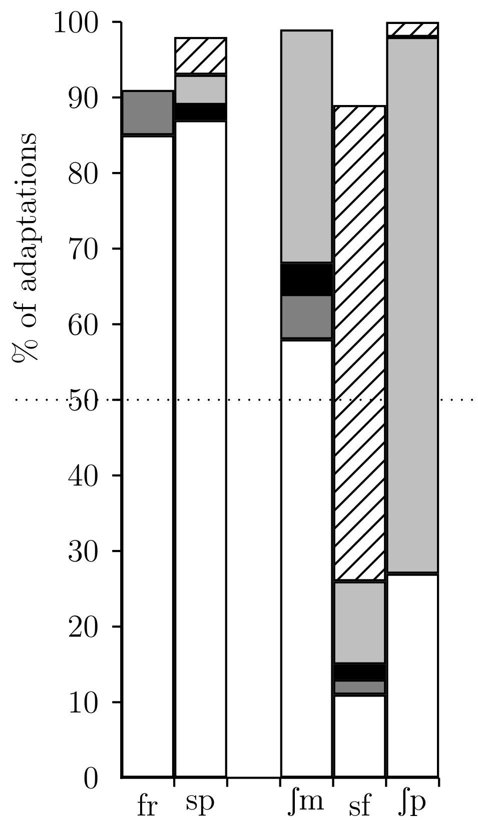

Figure 3.13: The productions of the legal and marginal clusters in the orthographic task.

Marginal/Legal clusters. The performance regarding the legal clusters is very similar in this task as compared to the shadowing. As commented above, a striking difference can be found in the marginal group: contrary to expectation, participants produce the correct representation for /sf, ʃm/ in significantly fewer cases when writing them down than when pronouncing them. For /sf/, over 60% of the tokens are represented as *sv*, and /sm/ is produced in over 30%, whereas for /sp/ the results are similar (figure 3.13). An explanation can once again be found in English orthographic conventions, since a perceived /ʃp/ can - and is in fact likely to - be represented as *sp*. A perceived /ʃm/, on the other hand, is represented as *shm* because *sm* does not allow the interpretation of the fricative as post-alveolar.

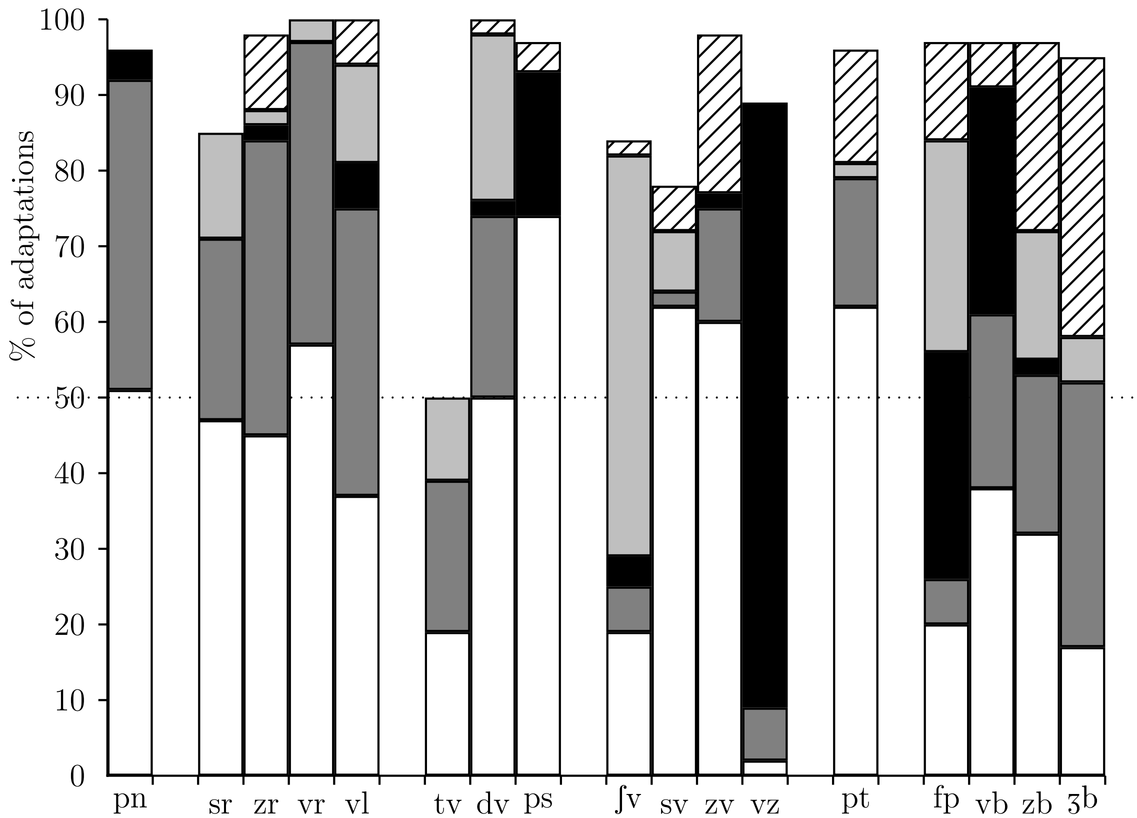

Figure 3.14: The productions of the illegal clusters in the orthographic task.

For the illegal clusters, the results, again divided by sonority distance groups and individual clusters (figure 3.14), shows that the rise in targetlike responses is not distributed evenly through the target clusters, and is concentrated on groups or subsets of groups (for ease of comparison, figure 3.15 compares only rate of correct responses across tasks).

As in the previous tasks, sonority distance does not determine the rate of correct responses or the extent of its increase in this task. For this reason, again the subsets of clusters were considered separately.

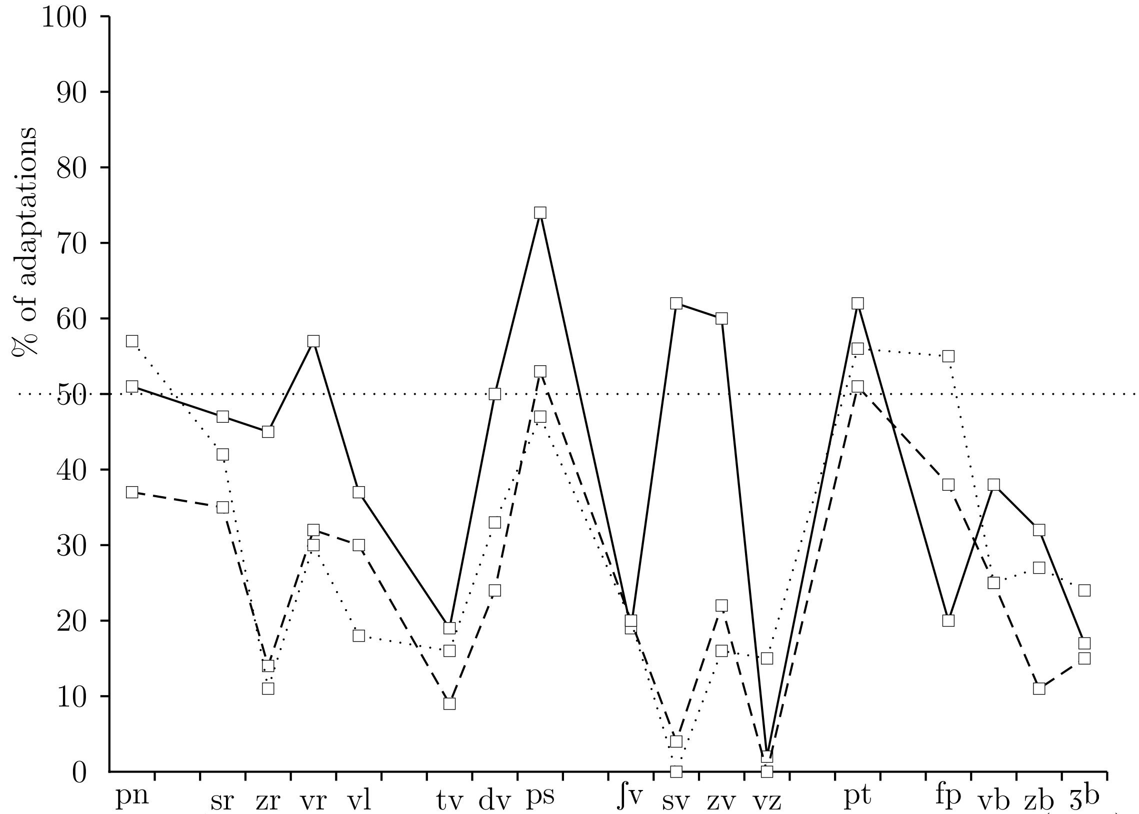

Figure 3.15: A comparison of the correct response rates for orthography (solid), shadowing/sentence (dashed) and shadowing/word (dotted).

Illegal clusters by sonority distance. For example, there is a marked increase for the fricative-liquid clusters like /vl/; they are on average produced correctly almost 50% of the time. This suggests that for fricative-liquid clusters perception does not appear to be the biggest obstacle to correct production. This conclusion is supported by the role of vowel epenthesis being increased and deletion almost disappearing: if a vowel is epenthesised, the production of two consonants necessitates also the perception of two consonants. When one consonant is deleted in production, no such conclusions can be drawn: the sound may have been deleted in perception or during the production process. For the fricative-stop clusters, no similar boost of correct responses can be observed. On the contrary, /fp/ production is lower than in the shadowing task. No clear trend regarding the distribution of adaptations could be observed for this group.

The stop-fricative clusters also show a noticeable rise in correct productions, however this rise is higher, the more correct productions there were in the shadowing task. Furthermore, there is also an increase in segment changes to a legal cluster for both /dv/ and /tv/, the majority being produced as /dw/. Again, this may be due to voicing confusion as a consequence of lack of aspiration.

Within fricative-fricative clusters, there is a clear divide: while in the alveolar-initial clusters there is the largest jump in correct productions of all targets, /ʃv/ and /vz/ remain at the bottom of correct production rates, with no increase. This may be connected to the predominance of alveolar-initial clusters in English. While the distribution of adaptations of /vz/ remain unchanged, there is a sharp increase of /ʃv/ to /ʃw/ segment changes.

The percentage of successful productions of /pt/ and /pn/, which were highest in the shadowing task, remain over 50%. While in the case of the single plosive the voicing appears to be perceived correctly, in the plosive-plosive combination voicing is misperceived more often.

Overall, neither extreme prediction regarding the difference between the two tasks is correct. The prediction supported is given in boldface:

- **The rate of targetlike responses is higher than in the previous tasks**
- The rate of targetlike responses the same as in the previous tasks (Peperkamp, Boersma)
- The rate of targetlike responses close to 100% (Paradis, Steriade, Silverman)

While the trend towards a higher percentage of correct responses than in the shadowing tasks is evident in the majority of clusters, the extent of this increase differs. In the same way as for the rate of correct responses for shadowing and for the choice of adaptation, no factor was found which explains all of the variation.

Frequently used adaptation strategies

Table 3.3 adds the preferred adaptations in this task to the shadowing results given in table 3.2. The main adaptations are, as above, very similar to each other, yet the table shows a larger difference in preferred adaptations between the two tasks than between the two conditions in the shadowing task. The shift in the /v/+liquid clusters from segment change to epenthesis is further advanced - more so for /vr/, where epenthesis was more frequent even in the word condition.

For /tv/, there is a sharp rise of /dv/ productions. At first glance these are classified as segment changes with a counterintuitive reduction of sonority distance. However, they can be analysed as correct productions except for the aforementioned voicing confusion.

[8]Any case of vowel insertion is represented with schwa

[9]Adaptations occurring over 20% of the time are printed in bold, those occurring over 30% of the time are in bold italics.

Frequent adaptations of marginal and illegal clusters (≥10%)[8] in the shadowing and the orthographic task			
Cluster	Sentence	Word	Orthography[9]
ʃm	sm		**sm**
sf		tsf	***sv***
ʃp	**sp**	sp	***sp***
pn	***pən***	**pən**	***pən***
sr	sər, s	**sər**	**sər**
zr	**zər**	***zər***, sr	**zər**, sr
vr	**fr, vər**, v	**vər, fr**, v	***vər***
vl	**fl, vəl**, v	**fl**, vəl, v	**vəl**, vV
tv	***təv*, tf**, dv	**təv, tf**, dv	***dv*, təv**
dv	**dəv**, df, dw	**dəv**, df	**dəv, dw**
ps	**s**	***s***	s
ʃv	**ʃf**, sf, sw	**ʃf**, səv	**ʃw**, sw, sv
sv	**sf**, səv, s, ʃf, tsf	**sf, səv**	zv
zv	**zəv**, sf	***sf*, zəv**	zəv, sv
vz	***z***, ts	***z***, ts	***z***
pt	***pət***	***pət***	pət, pd
fp	**p**, fəp, sp	**fəp**, p	**p, sp**
vb	***vəb***	***vəb***, fb	b, vəb
zb	**zəb, sb**	**sb**, zəb	**sb, zəb**, sp
ʒb	***ʒəb*, ʃb**	**ʒəb, ʃb**	**ʒəb, ʃb**

Table 3.3: Main adaptations in all production tasks, divided into the two shadowing tasks and the orthographic task.

A further noticeable difference between the two tasks can be seen in /ʃv/, for which the manner change from /v/ to /w/ occurs only in the orthographic task as a main adaptation, alongside the double segment change /sw/, which was frequent in the shadowing task as well.

In the case of /zv/, there is little change in the role of epenthesis, which remains frequent. However, while the second main adaptation is /sf/, an English marginal cluster, in the shadowing results, a slightly different adapted cluster comes to the fore, namely /sv/. This suggests that the voicing of /v/, or perhaps both fricatives, may have been correctly perceived more often, but devoiced in production.

As reported above, for the target /pt/ vowel epenthesis is once again the most frequent adaptation. Additionally, the orthographic task shows the new adaptation /pd/. This cluster with two plosives differing in their voicing specification is difficult to produce; nevertheless, the perception of the target as /pd/ is possible due to lack of aspiration, but a release even in C_1, which is the case in Russian. This shows again that targets may be correctly or near-correctly perceived, but not necessarily produced.

Finally, there is an apparent difference between the tasks for /zb/, namely the introduction of /sp/ as a repair. However, as explained above, fricative-stop cluster productions were analysed as having the same voicing value in the stop for target and production, since a distinction could not be drawn. In the orthographic task, on the other hand, the voicing value was decided by participants choosing <p> or <b> to represent the sound. Consequently, this difference does not necessarily point to a difference in perception *or* production.

Subject variability. On the topic of subject variability, the numbers in the production tasks were too small for statistical evaluation. However, several observations can be made: speakers are not equally successful in producing the target clusters in either task. This means that extensive variability can be mostly found in the productions of the speakers with lower targetlike performance. With only five tokens per clusters, not each speaker can display exactly the whole range of adaptations for those targets where many different adaptations are used. There are cases where individual speakers display less variation than others, for example /vl/ being adapted four times through devoicing to /fl/ by one speaker, and another choosing vowel epenthesis in three out of four adaptations. However,the remaining speakers show mixed adaptations, and overall the individual productions reflect the variation that can be observed in most cases. For example, for target /vr/ in the shadowing task one speaker produced the cluster targetlike once, epenthesised once, and three times devoiced as /fr/, another four times with epenthesis and once devoiced, so that each displays the same adaptations with differing frequency.

Summary

The legality of target clusters corresponds to the success in reproducing them, with a high variability found within the group of illegal clusters, both in the percentage of cases in which participants reproduced the cluster successfully and in terms of the preferred method of adaptation. Performance was better in the word as opposed to the sentence condition in the shadowing task.

Compared to a similar experiment used in a study by Davidson (2000) where stimuli were presented orthographically as well as auditorily, the percentage of segment changes were noticeably higher, while the rate of vowel epenthesis was lower.

The factor sonority distance was not found to be of relevance predicting the variability of either correct response rate or choice of adaptation. However, within the changes which did not result in a legal sequence, those which increased sonority difference were the most frequent.

The shadowing and the orthographic tasks show similar results, with slightly improved performance in the latter, and therefore a slightly decreased amount of all types of adaptations. Since the difference between tasks is relatively small, it suggests that perception plays an important role in adaptation. Of the adaptations, around 82% (orthographic task: 88%) resulted in phonotactic improvement, i.e. a legal sequence or one with higher sonority distance. This shows that in all three tasks the vast majority of changes is clearly motivated by phonotactic considerations.

The choice of adaptations for a given cluster was shown to be relatively similar across tasks, but varying widely across clusters; epenthesis, deletion and segment changes are possible as preferred adaptations. The choice of segment change depends to some extent on the availability of a legal cluster as a result of the change.

3.2.3 Discussion

The most striking finding of this experiment is that illegal clusters vary considerably in terms of production success and adaptation strategies. Different targets showed different adaptations. Clusters with initial /dv/ were mostly epenthesised, /zb/ was frequently altered to /sb/ or /sp/, whereas /vz/ clusters were reduced to /z/ almost without exception. This provides evidence against the *Theory of Constraints and Repair Strategies* (TCRS) described in Paradis & LaCharité (1997): all illegal clusters in this experiment were of the same type - the two segments were legal in English, but their combination in syllable onset was not. For this type TCRS predicts epenthesis in all cases. TCRS therefore does not provide correct predictions for an adaptation scenario where the borrowing language allows consonant clusters.

Furthermore, the sonority distance between the two consonants in a target cluster is not adequate for predicting the correct response rate, nor does it have the power to predict the differences in adaptation strategies. For example, consider again the targets /vz/ and /zv/: since both segments are voiced fricatives, the sonority distance between them is zero for both targets. Nevertheless, /vz/ is the most difficult cluster to produce for English speakers with 0% correct. /zv/ on the other hand, was successfully repeated in 20% of cases in the shadowing task. And not only do the error rates differ dramatically, the strategies of adaptation are also divergent: the main adaptation for the former target is a deletion of the initial /v/ to produce a singleton /z/ as onset, whereas for /zv/ there is epenthesis, or the initial segment becomes devoiced to yield the cluster /sv/.

No global predictor of correct response rates could be found, although specific consonant combinations and manner/place combinations lend themselves to explanations of high or low rate of successful productions, which have been given above. Chapter 4 explores the issue of factors influencing correct response rate further.

Similarly, an overall factor which determines the choice of adaptation, or the distribution of adaptations within the productions of a target cluster, was not determined. As with the rate of correct productions, individual explanations of these choices have been provided in the results sections, and are based on various influences. For example, English grapheme-phoneme correspondences are relevant for the adaptations of the marginal clusters.

For the illegal clusters, the decision which strategy to use, or to use foremost, is influenced by the proximity of an available, acceptable alternative. The vast majority shows a tendency towards greater acceptability, but the options of achieving this goal without altering the input to an unacceptable extent are different for each target. For some of these, changing one of the consonants in one feature suffices to make the output legal, as with /vl/ to /fl/. For other targets, no such straightforward alternative exists. For example, neither /z/ as C_1 nor /v/ as C_2 are acceptable in English onset clusters, so more than one segment change would be necessary to achieve legality with /zv/ as starting point.

Within the group of segment changes, some preferences emerged: for example, voicing changes are the most frequent alterations, even applied to both consonants of the cluster, or in cases where the result is not a legal cluster. Less frequent are place of articulation changes (mostly between alveolar and post-alveolar), and the rarest are manner changes, the majority of which only occur where the change produces a legal output. The preference for voicing changes where the resulting output is well-formed in English supports Steriade's claim that this is the least perceptible and therefore preferred resolution of a phonological conflict. Chapter 5 will investigate of the source of these preferences and judgments of perceptibility further.

Segment changes were overall very frequent, with percentages ranging around the same level as vowel epenthesis, i.e. the adaptation which in previous research was claimed to be the most prominent by far. Consonant deletions were very frequent for certain target clusters, but rare for others, with rates comparable to other studies of loanword adaptation.

An important point to note is that the percentage of segment changes is substantially higher in this experiment compared not only to much of previous loanword research, but also to Davidson's study (2%), which has a very similar design.

Epenthesis and deletion rates are comparatively more similar between the two experiments. There are two main differences between Davidson's experiment and the one carried out for this project. Firstly, participants received different instructions. Instead of being asked to reproduce the input as faithfully as possible, they were instructed not to pronounce the targets too perfectly, in other words they were prompted to make adaptations. Secondly, the presentation of the stimuli differed in that they were not only presented auditorily, but there was also an orthographic representation made available to the participants. Thus only the productive phonology was accessed, rather than the English-speaking listeners' perception. Hence, the difference of results between Davidson's (where the perceptual step was circumvented) and this experiment suggests that perception is relevant for even the phonotactic - not only the segmental - adaptation of foreign words.

Support for this claim comes also from a comparison of the two tasks: while successful production (i.e. correct transcription in this case) was increased in the orthographic task, adaptations nevertheless occurred in over 50% of the cases in a setup where participants did not have to produce speech. This suggests that neither a loanword model placing adaptations in the perception process only, such as Boersma's or Peperkamp's, nor one placing them in the production process only, such as Steriade's or Paradis', are entirely accurate.

However, since orthographic representation is also linguistic output of some kind, it may have passed through the productive phonology mapping to some extent. Additionally, the favourite adaptations for any given cluster do not always correlate between both experiments. Hence, while this experiment does give an indication that perception is relevant for adapting phonotactically illegal structures, an experiment studying perception more directly, i.e. without the mediating factor of orthography, is necessary.

3.3 Perception

3.3.1 Motivation

By showing that the differences between repetition and orthographic representation are small, the production results suggest that perception does have some influence on how foreign segments are adapted. Nevertheless, there might still be an effect of the grammar that comes into play between perception and the forming of an orthographic representation of the auditory stimulus. Therefore a method is needed that tests perception directly, excluding any effects of the native phonology that might take place after a representation of the input has been formed.

One such method is the categorical perception task, which tests identification and discrimination abilities for points along an acoustic continuum, the end points of which are minimal pairs, for example /b/-/p/ (see e.g. Repp 1984). Moreover, a categorical perception experiment compares the Russian/English setup directly to Dupoux et al.'s study about perceptual vowel epenthesis by Japanese speakers.

The pilot for a categorical perception experiment was designed. The results suggested problems with the design, which resulted in the preference of alternative research strategies laid out in chapters 4 and 5. Nevertheless, they are enlightening with respect to the role of perception; for this reason, the results of the pilot study are included in this chapter.

3.3.2 Method and results

Method

Materials.

The comparison chosen for the pilot study was between the two clusters /fp/, one of the illegal target clusters in the production experiment, and /sp/, an English legal cluster frequently used as an adaptation for /fp/. The SenSyn formant synthesiser (for a description see Klatt 1979) was used to synthesise continua of eleven steps between the syllables /fpo/ and /spo/ as well as between the sounds /f/ and /s/. [s] was modelled after Russian characteristics, specifically examples of [s] from target clusters of the production experiment. The fricative only continuum was used as a control, as these sounds are part of the English as well as the Russian inventory. The two syllables were designed without a coda to minimise distractions from the onset. This restricted the choice of possible vowels: the vowel needed to be in both the Russian and English inventories, and long, to make a coda-less syllable possible. Furthermore, to prevent palatalisation of the consonants, a front vowel could not be chosen. /a/ was not used to avoid to the lexical bias with *spa* for English listeners. The diphthongal nature of English [u] made it too different from Russian [u], and so /o/ was chosen, in its monophthongal realisation, which is used in Scottish varieties of English. The specifications for the synthesiser parameters of the end points are given in appendix G, with the values for the intermediate steps de- or increased in equal intervals for each step.

For the discrimination task pairings with a difference of three steps were used (e.g. points 2 & 5), which means that there were 8 pairs of stimuli along one continuum. For both tasks, there were five different pre-randomised orders of the stimuli, of which each participant listened to two.

Participants.

There were three participants with Scottish varieties of English (ages 21-26), and three Russian participants (ages 23-27) as controls, all recruited at the University of

Edinburgh. Scottish listeners were chosen as a result of the necessary choice of /o/ as the vowel in the target syllables. The Scottish participants had no knowledge of Russian (or other Slavic languages), and none of the participants reported hearing or speech production difficulties.

Procedure.

Participants performed an ABX discrimination task for both the /f-s/ continuum and the /fpo-spo/ continuum, and an identification task between /fpo-spo/, in different orders, and with 2 different randomised orders for each participant. There were 128 tokens for the ABX discrimination (8 pairings x 2 orders x 2 outcomes x 2 repetitions x 2 randomisations) and 176 tokens for the identification task (11 points x 8 repetitions x 2 randomisations), i.e. 16 data points for each point or pair.

Predictions

The hypothesis was that, if perception is influenced by phonotactics, the responses of Russian and English participants would differ. Since the end points of the continuum are two phonotactically possible structures in Russian, but belonging to different categories, Russian listeners' responses are expected to show a more categorical perception along each continuum. This would predict that there is a specific point along the continuum where listeners' perception of the onset switches from /fp/ to /sp/; at this point they should be able to discriminate best between the corresponding stimulus pair. However, if the response of the Russian listeners is gradual along the continuum (similar to the responses of the French listeners in Dupoux's experiment) because of the differing element being a fricative, the responses are nevertheless expected to be close to 0% identification as /spo/ at the /fpo/ end and close to 100% at the /spo/ end of the continuum. In the discrimination function, Russians are more likely to have a clear peak, at the point of the sharpest rise in the identification function. Scottish speakers, on the other hand, were expected to respond in a less categorical manner and be biased towards the acceptable (i.e. the adapted, here: /sp/) form. This was predicted to show as a lack of a peak in the discrimination function and, in identification, a high /spo/ response even near the /fpo/ end of the continuum.[10] For the fricative continuum, Russian and Scottish responses are expected to be the same.

[10]In the full experiment, continua between each target and all its adaptations in the production task would have been compared. The bias towards the adapted form would have been predicted to be strongest for the most frequent adaptation. Consequently, the lowest performance of English speakers would have been predicted for the continuum between each cluster and its preferred adaptation.

Results

The results showed, contrary to expectation, little difference between languages in the identification task, but differences between languages as well as between tasks in the discrimination test.

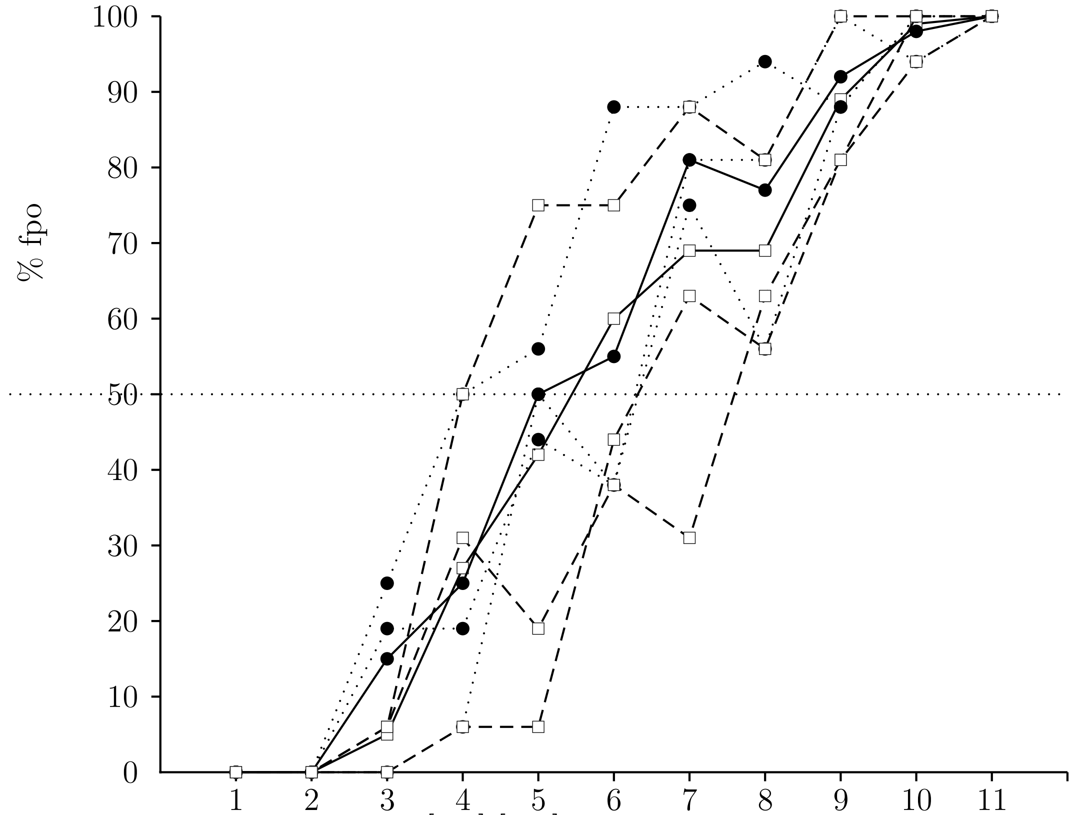

Figure 3.16: Identification task [fpo]-[spo]. The x axis shows the steps of the continuum, the y axis shows percentage of *fpo* responses. The left end of the continuum is [spo]. The squares show Russian results, the solid dots Scottish ones. The solid lines are the summaries of the two L1s, the dashed and dotted lines are the results of single listeners, dashed for Russian, dotted for Scottish.

Figure 3.16 shows the descriptives for the identification task which shows little difference between the performances of the Russian and the Scottish group. The appropriate statistical test to evaluate influence of the native language and the point along the continuum on the ratio variable *rate of identification as fpo* was a mixed ANOVA (N=66), with L1 as between-group factor and point of continuum as a within-group factor. It showed that, contrary to expectation, no effect of the native language could be found for the identification part ($F\ (1,44)=2.10$, $p=.16$). There was the expected main effect for point of continuum such that /fpo/ identification rose with approximation to the [fpo] end point ($F\ (10,44)=35.70$, $p<.01$), and no interaction ($F\ (10,44)=0.38$, $p=.95$) between L1 and point of continuum.

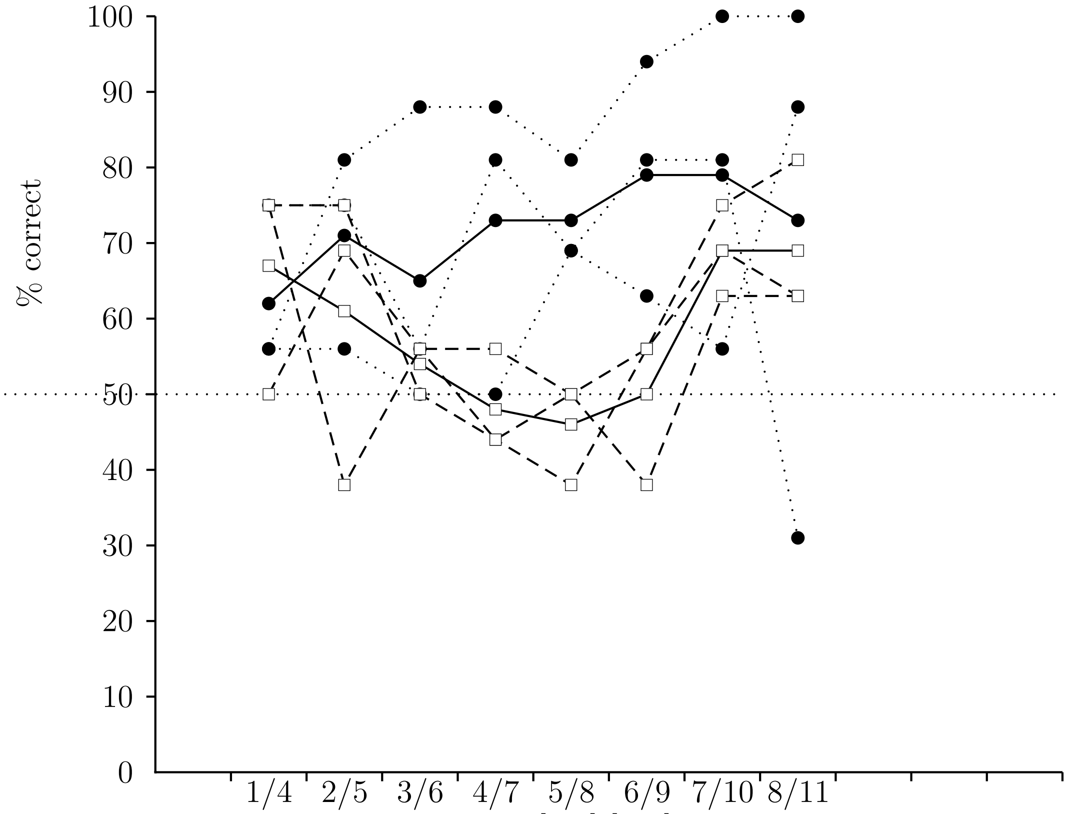

Figure 3.17: ABX discrimination task [fpo]-[spo]. The x axis shows the steps of the continuum used as A and B, the y axis shows percentage of correct responses. In the same way as for identification, the squares show Russian results, the solid dots Scottish ones. The solid lines are the summaries of the two L1s, the dashed and dotted lines are the results of single listeners, dashed for Russian, dotted for Scottish. The left end of the continuum is [spo].

Factor / Interaction	F(df)	p
L1**	F(1,64)=16.75	<.01
test type**	F(1,64)=36.35	<.01
point	F(7,64)=1.78	.11
point x L1	F(7,64)=2.11	.06
point x test	F(7,64)=0.42	.89
L1 x test	F(1,64)=0.95	.33
point x test x L1	F(1,64)=0.12	1.00

Table 3.4: The statistics for the factors L1, test type and point of continuum on discrimination performance (N=96). Significant effects are marked with ** at p<.01.

In the discrimination task, however, the picture changes. Figures 3.17 and 3.18 show the results for the syllable and the fricative continuum, respectively. Again, a mixed ANOVA was run with the between-group factor L1, and the point of continuum

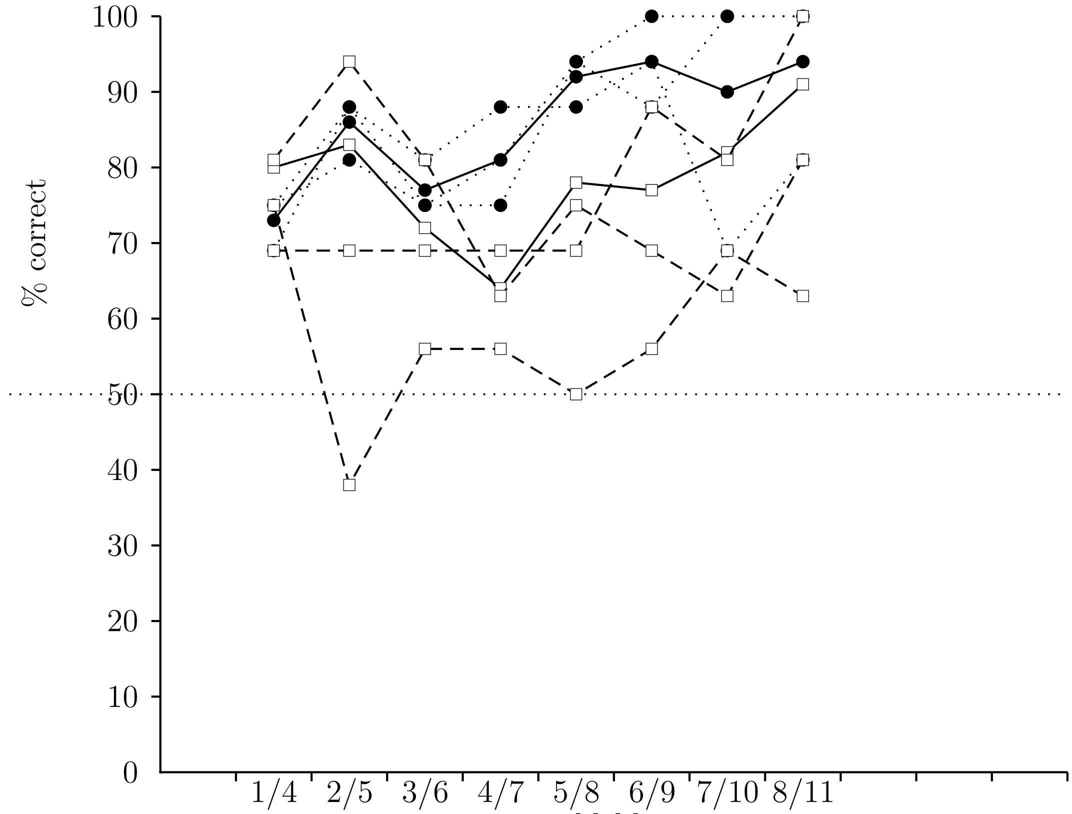

Figure 3.18: ABX discrimination task [f]-[s]. Descriptions as for figure 3.17.

and the test type (i.e. syllable or fricative only)as within-group factors. Table 3.4 shows the results. There is a significant main effect of L1 and test type. No effect was found for point of continuum, where a grouping of lower performance near the end points and higher performance in the middle area might have been expected. No significant interactions were found. There is a near-significant interaction between L1 and point of continuum, which indicates a greater difference between Russian and English responses near the middle of the continuum.

3.3.3 Summary and discussion

With a first glance at identification results, the null hypothesis could not be rejected: Russian and Scottish speakers identify initial /fp/ and /sp/ very much the same, although Russian phonotactics allow both, whereas in English only /sp/ is possible. However, the discrimination results contradict the apparent language independence seen for identification. Two effects can be observed. Firstly, discrimination of the syllables is generally worse than the discrimination of the fricatives only. Since this applies to both groups equally, the effect may be caused by cues in the -[po] section conflicting with those in the fricative section. Secondly, a clear L1 effect emerges: Russian listeners have higher error rates for both syllables (the opposite of what would be expected based on Dupoux et al.'s views) and fricatives only, which - independently of phonotactic considerations - shows that the cues perceived from the synthesised forms suit Scottish listeners better. Thus, the phonetic detail of a language makes a significant difference to the discrimination of sounds, which supports the viewpoint that a listener's native language influences their perception.

This also supports one of the findings from the production experiment, i.e. the improved performance in the orthographic task, where perception would have had a relatively greater weight for the outcome. However, the cue weighting of Russian and Scottish listeners is not known; to investigate which aspects of the phonetic form are most relevant in identifying fricatives for each language group would have necessitated a large number of further pilot studies. Since this was not within the scope of the present thesis, the categorical perception experiment was discontinued.

Instead, the question of perception is taken up again within the study of further factors beyond sonority influencing loanword adaptation; specifically, it is relevant in the issue of perceived similarity, which is the topic of chapter 5.

Chapter 4

Alternative factors in production

4.1 Introduction: motivation and hypotheses

The previous chapter described a production experiment of English participants producing Russian onset clusters and changing or adapting the clusters in the process. The experiment showed that not all illegal clusters are treated equally by English speakers: some clusters are reproduced with greater success than others, and the types of adaptation vary to a great extent.

The hypothesis that either of these variations was connected to or caused by the different sonority distances between the two consonants of each cluster could not be confirmed. Instead, the cluster-specific combination of manners of articulation or what was termed the availability of a possible adaptation were suggested as better explanations for rates of adaptations, and for particular choices of adapted forms.

For this reason, a more systematic look at possible other factors influencing the production tasks, and the extent to which they exert influence, was undertaken. While previously the focus was on a language-specific grammatical value of minimal sonority distance - a variable which is suggested to be of cross-linguistic relevance -, the results of chapter 3 suggest that the perspective needs to be widened. Three further factors will be studied, the nature of which is not necessarily grammatical or even language-specific.

Firstly, gradient grammaticality is investigated, since a varying degree of grammaticality of source language forms with respect to the donor language can conceivably result in varying adaptation. Naturally, gradient grammaticality is both a language-specific and a grammatical concept. However, the a priori choice of the established phonological variable sonority distance to determine a ranking of target clusters, and the following predictions about the role of sonority-based differences in adaptation was shown to be unsuccessful in the previous chapter, and needs to be revised.

Traditionally, grammar has been seen as an all-or-nothing phenomenon: a structure is either well-formed or not. This strict picture of grammar prevailing in syntax and phonology has become more fuzzy in recent years, and the concept of gradience has been incorporated into models of both syntax (e.g. Sorace & Keller 2005) and phonology (e.g. Hayes 2000). When applied to research of loanword adaptation, this concept may have implications on both rate and method of adaptation. If there is a gradience of structures within the native grammar, then this gradient measure of acceptability, the levels of which decline towards the periphery of possible structures, may extend also to structures which are beyond this periphery, i.e. those which would be classified as ill-formed in a language. When foreign words are adapted as loanwords, the perceived position of an ill-formed structure with respect to those which are well-formed may have an effect in several ways: it may determine how likely an ill-formed onset is to be changed at all. Secondly, it may have an effect on which method of adaptation is preferred; for example, an onset cluster perceived to be nearer the core may be more likely to be adapted by epenthesis, and a cluster further away from acceptable onsets may have a higher probability to undergo consonant deletion.

Since sonority distance failed as a predictor, and no numerical measure of phonological grammaticality is available, a different approach, namely the judgment of native speakers on grammaticality, is used. Subsequently, three questions are studied:

- Is there a variation in (un-)grammaticality of illegal clusters akin to the variation in correct production?
- Can we find determinants of grammaticality judgments which may be part of the English grammar, and determine a ranking of ill-formed structures?
- Does perceived grammaticality have any bearing on whether and how illegal clusters are adapted?

A second factor under investigation, which may be connected to grammaticality, is the influence of the lexicon. The role of frequency in phonological adaptation is examined. As described in chapter 1, frequency in phonology has received much attention in recent years, most famously in Bybee (2001); furthermore, Frisch et al. (2000) found an influence of segment probabilities in judgments of wordlikeness, Berkley (2001*a*,*b*) describe pressure from the lexicon on the formation of derived words. Johnson (2004) claims that differences in speech perception are caused by cross-linguistically differing lexica, and Broe (1993) stresses that phonotactics is equal to a structured specification of the lexicon. If phonotactics is indeed contained in the lexicon, then the frequency of a structure in the lexicon should be related to its perceived grammaticality. But frequency in the lexicon can conceivably be reflected also in the adaptations: those sequences which occur most often in the lexicon may also be those most readily available or preferred when a foreign structure is substituted with a native one, and hence occur more often than their less frequent alternatives.

Finally, the notion of similarity and its role especially in loanword adaptation have been discussed, with the definitions and the details of how and where in the process it is applied varying widely. Most research employs a concept of minimal adaptations (for a more extensive discussion of previous studies see chapter 5); also, Fleischhacker (2001) presents a study of alterations to a stimulus word in which preferred changes were also considered to be the most similar to the stimulus. An image not unlike the core-periphery idea in grammaticality may illustrate the possible effect of similarity on adaptation. In the space of all possible onsets[1], some pairs or groups may be perceived to be more similar or closer together, while others are heard as more different from each other and hence having a longer distance between them. This distance may serve as a predictor for how likely a conversion of one of these clusters into another is, i.e. in loanword adaptation, how likely an onset will be adapted as another, if it is adapted at all.

However, similarity remains ill-defined in three ways: first, it remains unclear whether similarity is perceived differently by listeners with different language backgrounds; second, there is no accepted measure of similarity; third, it is not known at which stage in the adaptation process similarity is relevant, i.e. whether high similarity leads to misperception, or is applied in production. For this reason, English native speaker judgments regarding the similarity of word pairs are used here to determine the influence of similarity on adaptation. It is tested whether different types of adaptation differ significantly in how similar they are perceived to be to the target, and whether more similar adaptations are more frequently used.

This chapter has three aims: to determine whether any of the factors described have any relevance on the adaptation process; to study if it is possible to predict how likely a cluster will be produced correctly and how likely a specific adaptation is to be used; and to determine whether the influences on adaptation are or can be of a grammatical nature, or whether they are independently resident in a language's lexicon or the human speech perception apparatus. The last issue will be taken up again, and the role of similarity in phonology in general and in loanword phonology specifically will be discussed in chapter 5. The strategy for this chapter is to measure the *perceived* grammaticality of a set of onsets (including the targets of the production experiment) and the *perceived* similarity between as set of pairs of targets and potential onsets, as judged by native speakers of English. After obtaining

[1]or initial sequences, since CəC is not an onset, but a syllable plus another onset

values for these measurements, the factors' relevance as predictors of adaptation is tested.

The hypotheses regarding the factors grammaticality, similarity and frequency which are tested in this chapter are the following:

1. The less well-formed a foreign sequence is perceived to be by speakers of the borrowing language, the less likely it is to be reproduced in a target-like way, i.e. to remain unadapted
2. The degree of acceptability of a target sequence influences the choice of adaptation type
3. The more frequent a potential adapted form is in the native lexicon of the borrowing language, the more likely it is to be used
4. The higher the perceived similarity is between a target form and a potential adaptation, the more often this adaptation will be used

4.2 Methods of measuring possible factors

While the measurement of frequency from corpora of spoken language is theoretically straightforward, measuring the perceived grammaticality and similarity is a more demanding objective, since it aims to put numbers to degrees of native speaker intuition which are not otherwise directly manifested. One method of solving this problem is with the use of scales of, for example, from one to five, or one to seven, on which participants are asked to mark their judgment. However, this is problematic in that participants may use the whole range of the given scale before encountering a stimulus which they would intuitively place off the scale. Secondly, such a predetermined scale can only ever give rankings of preferences, i.e. ordinal data for analysis. A method to deal with these issues is the use of *Magnitude Estimation*. This technique was first used in the area of psychophysics (Stevens 1956, see also Lodge 1981) to measure impressions of physical stimuli such as loudness or brightness, as expressed proportionally for example by lines of differing lengths, or numbers. An initial stimulus is matched with a line of a certain length, or a certain number. The participant is then asked to give a judgment of the strength of the following stimuli by lines of different length (or numbers) relative to the initial *modulus*: a stimulus with a perceived double strength is supposed to result in double the number or a line with double the length. Since participants are free to choose any length or number that they deem representative of their impression, they can each create their own scale, as fine-tuned and as extensive as necessary. Bard et al. (1996) have introduced this method of measuring impressions to linguistics, to elicit intuitions and judgments of syntactic acceptability, i.e. perceived grammaticality, from native speakers.

For this section, *Magnitude Estimation* will be extended to judgments of phonological acceptability, and to intuitions about phonetic similarity. After measuring frequency, grammaticality and similarity, the results are studied with respect to the determinants of the values. Subsequently, the relative influence of each of these factors on the adaptations which occurred in the production experiment is analysed with statistical tests of the four hypotheses listed above.

4.2.1 Grammaticality

As there is no numeric measure of the actual grammaticality, or the extent of the ill-formedness of a phonological structure, the notion of grammaticality used here was one of *perceived* grammaticality, i.e. a value of how acceptable a group of native speakers of the borrowing language, i.e. English, judged the stimuli to be, using the described method of magnitude estimation.

Materials.

The stimuli consisted of 102 initial sequences (clusters, single onsets and epenthesised /CəC/, see appendix D) combined with 5 endings, namely (/-ʌvoj/, /-ʌdiʃ/, /-ʌpat/, /-ʌsal/, /-ʌla/), with stress on the final syllable. As in the previous experiment, the vowel following the initial sequence ensured that no palatalisation of the preceding consonants occurred. The set of initial sequences included the target clusters of the production experiment as well as the same clusters epenthesised, corresponding single onsets, further clusters related to the target clusters of the shadowing experiment through feature changes like devoicing, and clusters designed to extend the patterns, such as the nasal-nasal cluster /mn/. All stimuli were in keeping with Russian phonology, but aside from the subset of illegal onsets, the trial stimuli did not contain any further phonological difficulties for English speakers. The list was transcribed into Cyrillic, and recorded by a male native speaker of Russian in a soundproof recording studio. The recordings were made in a sound-proof recording studio at the Department of Theoretical & Applied Linguistics, University of Edinburgh, using an AKG CK98 half rifle cartridge on an AKG SE 300B power unit and a SONY PCM-2700A DAT recorder. Recordings were digitised using a Townsend Datlink+ at 16khz mono, multiplexed to the SONY PCM-2700A and to a Sun Sparc Ultra running Solaris 2.4 for archiving to disk.

The resulting list of 510 stimuli words was divided into two lists of 255, to which 235 distractor pairs each were added to give two trial lists of 490 pairs. Each trial list was judged by five subjects. Examples of distractor stimuli used are /kuda/, /mʌros/ and /tʃʌdit/.

Participants.

10 native speakers of British English varieties (ages 18-23), recruited at the University of Edinburgh. They were paid 5 pounds for taking part. None of the participants reported any hearing or speech production difficulties, and none of them had any knowledge of Russian (or another Slavic language). This was ascertained after the experiment, to avoid influencing the participants' expectations of the stimuli.

Procedure.

Before introducing the participants to the experiment itself, the idea of Magnitude Estimation was explained to the participants. The method was illustrated with a set of lines of differing length, with which participants practised assigning numbers proportionally: they were shown a line and asked to assign the length of the line a number. Subsequently, they were shown a sequence of shorter and longer lines, and asked, for each line, to express its length proportionally to the number used for the length of the first line. When the participants felt comfortable with the concept, the nature of the task was explained: to judge how acceptable a word was with respect to English, i.e. how possible or impossible it was for a word they heard to be a word of English.

The stimuli were presented auditorily via Sennheiser HD 457 headphones, within the Eprime experiment program, in a different randomised order for each participant. The participants typed the responses, followed by the ENTER key, on the keyboard. The participants were asked to give higher values for words they considered more grammatical. A practice block of 10 stimuli could be repeated until the participants were comfortable with the task and felt that they had calibrated their scale. The trial stimuli were grouped into 24 blocks of 20 and a final block of 10. Participants started each block in their own time and were free to take short breaks between blocks.

Before the responses were analysed, they were checked for typing mistakes to avoid erroneous skewing of the results. If one response of a participant was an outlier by a margin larger than three times the range of the remainder of the scale, this response was removed, or corrected where unambiguously possible. For example, in one case the scale ranged from 3 to 17, but one response value was 555, and was changed to the value 5. Since in a Magnitude Estimation experiment each participant creates their own scale, these scales were normalised by a transformation to z scores.

4.2.2 Similarity

As with grammaticality and acceptability, there is no available measure of similarity which reliably reflects the intuitions of listeners. Disadvantages of measures devised so far (which will be further discussed in chapter 5), are for example

- a calculation of phoneme similarities only rather than sequence similarity
- a calculation of phoneme similarity without taking the feature makeup into account
- a disregard for the possible effect of context
- a disregard for the possible effects of a listener's native language

Therefore, any suggested measurements of similarity are inappropriate when trying to determine how the specific similarity of two sequences of sounds, as perceived by English native speakers, may bear on the adaptations English native speakers carry out. Consequently, the solution is to conduct a similarity judgment study testing the perceived similarity between the Russian targets and their adaptations in the production study.

Materials.

The stimuli consisted of 102 pairs of initial sequences (e.g. /fp-sp/, /vz-z/, /dv-dəv/, see appendix E), each combined with 5 endings (/-ʌvoj/, /-ʌdiʃ/, /-ʌpat/, /-ʌsal/, /-ʌla/), with stress on the final syllable. The pairs of initial sequences were combinations of a target and one of its adaptations, as found in the production experiment. The materials were (analogous to the production experiment) in keeping with Russian phonology. Therefore a subset of the target-adaptation pairs had to be excluded from this experiment, most prominently the manner changes, where C_2 /v/ was changed to /w/. The Russian inventory does not contain /w/, and so, for example, /dv/-/dw/ and /sv/-/sw/ could not be compared. Similarly, a lack of /θ/ in Russian precluded the comparison /sr/-/θr/ or /fr/-/θr/, which were less frequent in the previous experiment.

As in the grammaticality experiment, the vowel following the initial sequence ensured that no palatalisation of the preceding consonants occurred, and all trial stimuli conformed to Russian phonology, but excluded further difficulties for English speakers aside from the subset of illegal onsets. The list was transcribed into Cyrillic, and recorded by the same male native speaker of Russian in the soundproof recording studio.

The resulting list of 510 word pairs was divided into 2 lists of 255, to which 235 distractor pairs each were added to give two trial lists of 490 pairs. Each list was judged by five participants. The distractors of the grammaticality judgment were used, each paired with a word in which other constituents were changed, or single onsets exchanged for each other. Examples of distractor pairs used are /kuda/-/kuʒa/, /mʌros/-/mʌres/ and /tʃʌdit/-/tʃʌdut/.

Participants.

10 further native speakers of British English varieties (ages 18-25), recruited at the University of Edinburgh. They were paid 5 pounds for taking part. None of the participants reported any hearing or speech production difficulties, and none of them had any knowledge of Russian (or another Slavic language). This was ascertained after the experiment, to avoid influencing the participants' expectations of the stimuli.

Procedure.

As in the grammaticality judgment, participants were at first familiarised with magnitude estimation. The procedure including blocks and timing were the same as used in the previous experiment, and the participants were asked to give higher values to pairs of words with higher similarity. The responses were checked and errors corrected, and the scales were normalised in the same z score transformation.

4.2.3 Frequency

The frequencies of the target clusters and their adaptations, such as single onsets, other clusters or epenthesised /CəC/ sequences, were examined in the CELEX spoken corpus for English. Only word-initial onsets (as in the experiment stimuli) were counted. For example, for the target cluster /sv/ (with adaptations /sf, səv/) the frequencies of initial /sf/ and /səv/ were compared, and showed around double frequency for the latter, in words such as *severe* and (in non-rhotic varieties of English) *survive*.

Both type and token frequencies were calculated. Of these, the logarithmic values were used for the analysis in the following section. As expected, there was a highly significant correlation between type and token frequencies, with a very high correlation coefficient of .96 (r_p=.96, N=75, $p<.01$, 95% CI: $.94<r<.98$).

4.3 Analysis and results: the influence of alternative factors on production

4.3.1 Determinants of grammaticality and similarity

Grammaticality

The main aim of collecting judgments of grammaticality for ill-formed clusters is to study the relevance of a possible gradience in grammaticality for the variation observed in the adaptations of these clusters. However, these judgments are also useful to study what determines these judgments - for example, whether there are certain phonological parameters influencing grammaticality, or statistical extrapolations from the lexicon. Therefore, this section tests the influence of several factors on the judgments. Firstly, it is tested whether the judgments are different depending on what type the initial sequence is - a single onset, a cluster, or a cluster with an epenthesised schwa. For the clusters, the influence of three variables connected to cluster goodness in English is tested, namely voicing and coronality of the first consonant, and the sonority distance between the two consonants. Finally, the connection between grammaticality judgments and frequency in the English lexicon is studied.

To test the correspondence between grammaticality and frequency, the correlation between the frequency of an onset sequence in the English lexicon and the mean perceived grammaticality of the same sequence is calculated. The influence of the other factors (i.e. onset type, and sonority distance and voicing/coronality of C_1 in clusters) on the ratio variable perceived grammaticality is tested with ANOVAs.

Subject variability. The tests were run by-subjects and by-items, i.e. with participant and word ending (e.g. /-ʌvoj/) as random factors to exclude the possibility that the hypothesised main effects were caused by variability between listeners or different endings. All reported statistical values are taken from tests with both subject and item as variables, and therefore the reported effects are valid across participants and endings. However, for reasons of conciseness, the descriptives and inferential statistics are reported only for the fixed factors.

Considering the different types of onset sequences, an ANOVA was performed to test whether listeners would indeed assign higher grammaticality to legal sequences; for this test, the onset type was chosen as the independent variable, i.e. a cluster, an epenthesised cluster, or a single onset. The result shows an effect of onset type ($F(2,2349)=32.36$, $p<.01$).

The descriptives are given in table 4.1. They show that, while the legal single onsets are judged most grammatical, clusters are the least acceptable, with

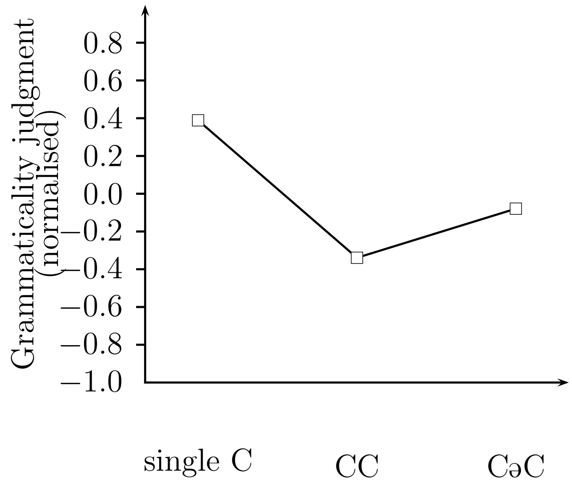

Figure 4.1: Grammaticality judgments of single onsets, CC clusters and CəC sequences.

Onset type	Mean	Std. Dev.	N
C	.39	1.26	344
CəC	-.40	.97	543
CC	-.80	.92	1617
Total	0.01	.99	2504

Table 4.1: This table shows the descriptive statistics for grammaticality, as influenced by onset type.

epenthesised clusters, which show no phonotactic problems with regard to English, ranked only marginally higher (also see figure 4.1). Post-hoc comparisons showed that the difference between clusters and epenthesised clusters was not significant (p=.68), while the difference between either of them and single onsets was significant at p<.01. This throws up further questions about the ability of English native speakers to perceive differences between illegal clusters and their epenthesised adaptations, which will be further discussed in chapter 5.

In chapter 3, a target cluster's sonority distance was not shown to be relevant for the rate of correct responses or for the choice of adaptation. Nevertheless, sonority distance may influence the notion of acceptability in native speakers. This was tested along with the influence of two further factors, i.e the coronality of the first consonant in the cluster, and the voicing status of the first consonant. These are motivated by the special status of /s/-initial clusters, both in English and cross-linguistically, and the fact that, independently of manner-based sonority distance, the C_1-voiced counterparts of legal fricative-initial clusters are ill-formed in English, for example /sw/, but */zw/, /fl/, but */vl/, and /θr/, but */ðr/.

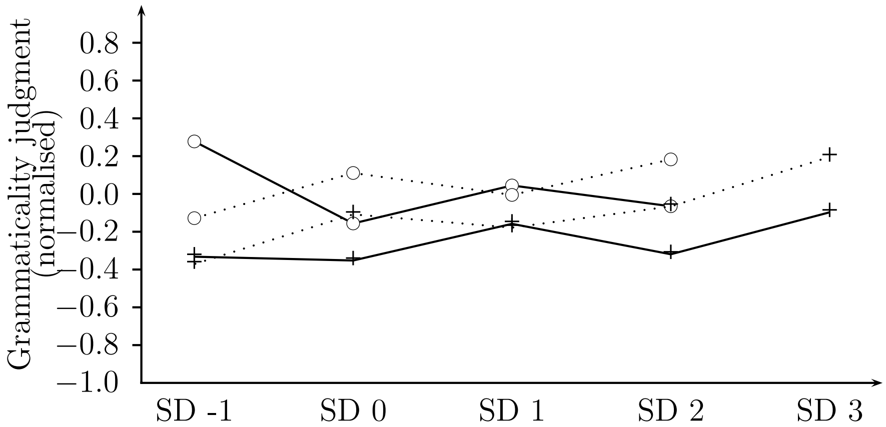

Figure 4.2: Grammaticality judgments by sonority distance (SD), and coronality and voicing of C_1: solid lines show coronals, dotted lines non-coronals, + and o dots show voiced and unvoiced, respectively. The highest value can be found for coronal, voiceless clusters with SD=-1, i.e. the English legal and marginal /s/ and /ʃ/ plus stop clusters.

An overview of the descriptive statistics can be found in table 4.2. A look at the mean values for grammaticality shows that sonority distance is of limited importance: while the highest value is indeed found for SD=3 at 0.10, there is no discernible upward trend as sonority distance increases. Another striking point is the highest mean value in the table, for clusters with negative sonority distance, and a voiceless coronal first consonant. While SD=-1 would suggest low grammaticality, the high rank of this group is reflective of the existence of legal and marginal coronal-initial fricative-stop clusters such as /sp, ʃp/. This special status can also be observed in figure 4.2.

An ANOVA with the factors sonority distance, C_1VOI and C_1COR was calculated for all CC stimuli (see table 4.3), and shows a main and an interaction effect. There is a main effect for voicing of C_1 only. Additionally, the test confirms the observations from the descriptive statistics: there is a significant interaction between sonority distance and the coronality of C_1. Non-coronal-initial clusters have their lowest value for the lowest sonority distance and the highest for the highest sonority distance (SD); coronal-initial clusters, on the other hand, had the highest ranking for SD=-1, which is the only constellation of English clusters where a sonority distance of a value lower than two is permitted: the subset raising this value here is, the voiceless one, i.e. precisely the legal and marginal /s/ and /ʃ/ clusters.

This result hence only reflects what is possible and impossible in English, rather than giving a real indication of the factors influencing the goodness of a cluster.

Sonority distance	C_1VOI	C_1COR	Mean	Std. Dev.	N
-1	no	no	-.13	.73	75
		yes	.28	1.04	150
		total	.14	.97	225
	yes	no	-.37	.79	74
		yes	-.33	.82	125
		total	-.35	.81	199
	total	no	-.25	.77	149
		yes	.00	1.00	275
		total	-.09	.93	424
0	no	no	.11	.95	75
		yes	-.15	.87	100
		total	-.04	.91	175
	yes	no	-.11	1.00	100
		yes	-.35	.80	49
		total	-.19	.95	149
	total	no	-.01	.99	175
		yes	-.22	.85	149
		total	-.11	.93	324
1	no	no	-.01	1.16	75
		yes	.04	.99	150
		total	.03	1.05	225
	yes	no	-.18	.85	124
		yes	-.16	.89	123
		total	-.17	.87	247
	total	no	-.11	.98	199
		yes	-.05	.95	273
		total	-.08	.96	472
2	no	no	.18	1.08	98
		yes	-.07	.80	98
		total	.06	.96	196
	yes	no	-.07	.80	99
		yes	-.32	.68	50
		total	-.15	.77	149
	total	no	.06	.96	197
		yes	-.15	.77	148
		total	-.03	.88	345

Table 4.2: The descriptive statistics for grammaticality (by sonority distance, C_1VOI and C_1COR). For part 2 of this table, please refer to the next page.

The only independent factor influencing acceptability is therefore the voicing of the initial consonant, which lowers the perceived grammaticality of a cluster.

Sonority distance	C_1VOI	C_1COR	Mean	Std. Dev.	N
3	yes	no	.19	.91	50
		yes	-.10	.74	25
		total	.10	.86	75
	total	no	.20	.91	50
		yes	-.10	.74	25
		total	.10	.86	75
total	no	no	.05	1.00	323
		yes	.05	.96	498
		total	.05	.98	821
	yes	no	-.13	.88	447
		yes	-.26	.82	372
		total	-.19	.86	819
	total	no	-.05	.94	770
		yes	-.08	.92	870
		total	-.07	.93	1640

Factor / Interaction	F(df)	p
Sonority distance (SD)	F(4,1622)=1.90	.108
C_1VOI**	F(1,1622)=30.54	<.01
C_1COR	F(1,1622)=2.94	.09
SD x C_1VOI	F(3,1622)=1.37	.25
SD x C_1COR**	F(4,1622)=4.51	<.01
C_1VOI x C_1VOI	F(1,1622)=0.95	.33
SD x C_1VOI x C_1VOI	F(3,1622)=0.93	.43

Table 4.3: The statistics for the factors sonority distance (SD), C_1VOI and C_1COR and interactions. In this and the following statistics tables, significance is marked with * at p<.05, and with ** at p<.01.

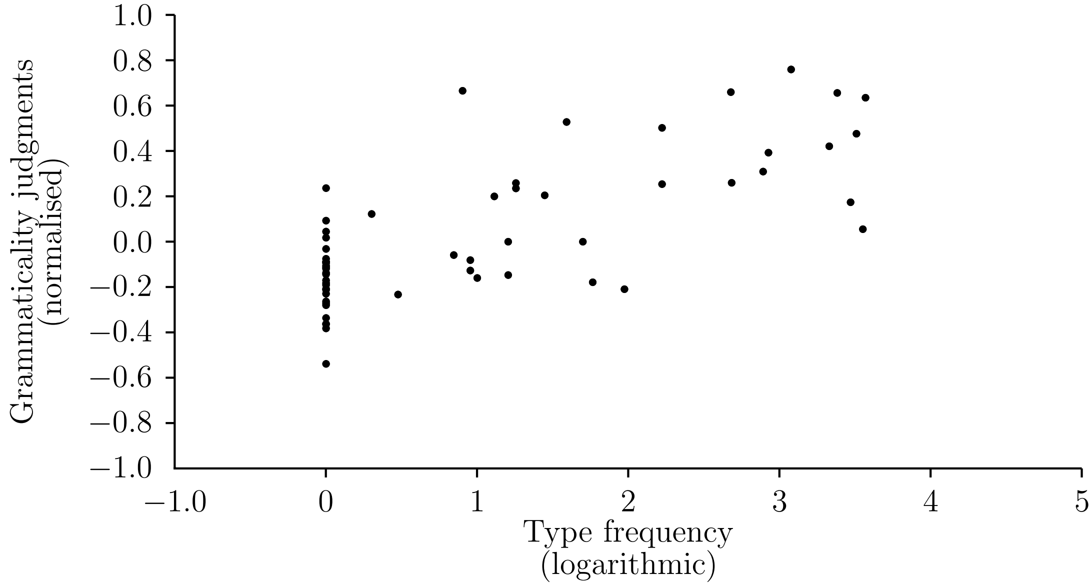

Figure 4.3: Grammaticality judgments and frequency in English. For this scattergram, the frequency is type frequency in English.

Finally, the correspondence between the frequency of a cluster in English and its perceived grammaticality was tested to explore the influence of the lexicon on perceived grammaticality. The Spearman's correlation was found to be significant for clusters which do occur in the English lexicon, for both type and token frequencies (type: N=27, r_s=.61, p<.01, 95% CI: .34<r<.82; token: N=27, r_s=.54, p<.01, 95% CI: .23<r<.78). When clusters with zero frequency were added, the correlation was still significant (type: N=58, r_s=.69, p<.01, 95% CI: .53<r<.81; token: N=58, r_s=.68, p<.01, 95% CI: .63<r<.80).

While figure 4.3 shows that zero frequency items tend to have lower grammaticality ratings than those which do occur in the lexicon, the wide spread of the judgment values for non-legal clusters (between -.54 and 0.24) makes clear that the influence of the lexicon does not reach as far as assigning to all structures outwith it the same low grammaticality value, namely one below the values for all structures within the lexicon.

The exploration of factors determining how grammatical or ungrammatical stimuli are perceived to be by native speakers of English shows that frequency clearly plays a role in determining the goodness of clusters. However, this influence is mostly constrained to well-formed structures, while a frequency of zero does not automatically result in a low grammaticality rating. Phonological generalisations are relevant for ratings of grammaticality only in terms of the voicing of C_1.

Similarity

In the shadowing experiment it was observed that some types of adaptations appeared to be more frequent. Before ascertaining whether this preference is due to these adaptations being perceived to be more similar to the targets, a first step towards exploring nature of how similarity is perceived was taken. The influence of the type of adaptation on the notion of similarity was tested.

Subject variability. In the same way as for the grammaticality judgments, the appropriate test for the influence of the relevant factors on the ratio variable similarity judgment was an ANOVA. Again, the tests were run both by-subject and by-ending (e.g. /-ʌvoj/) as random factors to exclude the possibility that the hypothesised main effects were caused by variability between listeners or different endings. All reported statistical values are taken from tests with all variables, and therefore the reported effects are valid across participants and endings. However, for reasons of conciseness, only the descriptives and inferential statistics for the fixed factors are reported.

In terms of the type of adaptation, the values of the fixed factor *type of change* were:

- deletion of C_1
- deletion of C_2
- vowel epenthesis
- voicing change C_1
- voicing change C_2
- voicing change of both consonants
- place of articulation (POA) change
- manner change
- segment change combinations

This list was reduced to deletion, epenthesis, voicing, place of articulation, manner and combinations for a second ANOVA. Table 4.4 gives the descriptive statistics, figure 4.4 shows the means for each of the types.

Within the reduced list of change types, the mean similarity values suggest a division into two groups: consonant deletion, manner changes and combinations of changes were assigned low similarity. In the other group there were place of articulation, epenthesis and voicing change.

The test for the extended list shows a main effect for the type of change (F (8,2129)=41.49, p<.01). A post hoc test showed five overlapping homogeneous

Change type	Mean	SD	N
deletion	-.30	.87	1100
epenthesis	.38	1.10	550
manner	-.45	.92	50
combinations	-.44	.74	75
POA	.16	.96	300
voicing	.26	.96	474
total	.00	1.00	2549

Table 4.4: The descriptive statistics for similarity, with the factor change type (reduced).

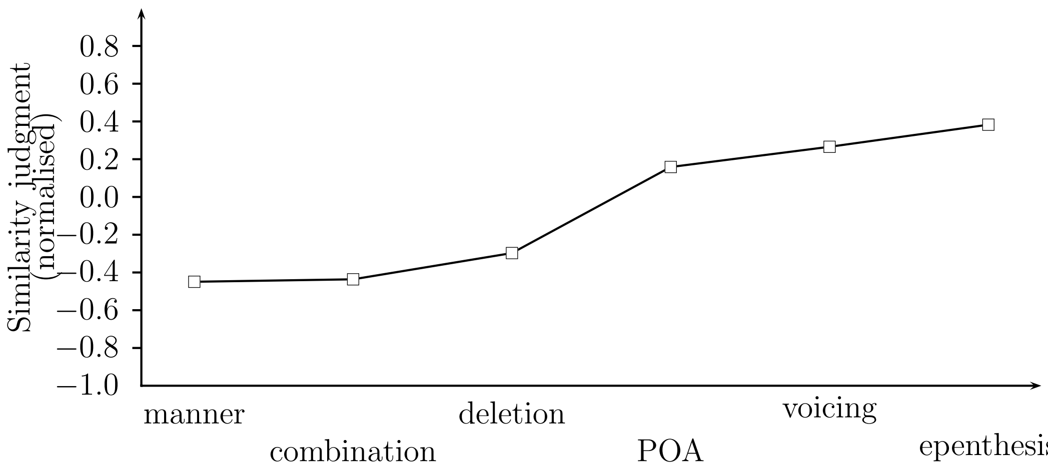

Figure 4.4: Similarity judgments depending on the type of change: segment changes of manner, combinations of more than one segment change, C deletions, POA segment changes, voicing segment changes and vowel epentheses

subgroups of change types[2]. When the ANOVA was re-run with the reduced list of changes, the results were similar (F(5,1958)=56.89, p<.01). Additionally, a clear division in homogeneous subgroups crystallised between deletions, manner changes and segment changes (with low similarity values) on the one side and highly ranked epenthesis, voicing changes and place of articulation changes on the other. This result taken together with epenthesis and voicing changes being the most frequent adaptations in the production experiment indicates that perceived similarity is a decisive factor in adaptation. This and the remaining three hypotheses are tested in the next section.

[2]In ascending order of similarity: {manner, combinations, deletion C_2, deletion C_1}, {deletion C_1, voicing of both Cs}, {voicing of both Cs, POA}, {POA, epenthesis, voicing of C_1}, {epenthesis, voicing of C_1, voicing of C_2}

4.3.2 Grammaticality, similarity and frequency in adaptation

To determine the relative influence of frequency, similarity and perceived grammaticality on the adaptation strategies that English speakers use, the frequency values and the overall means for the similarity and grammaticality judgments were compared to the results of the production tasks. The percentage of how frequently an adaptation is used was measured and converted to relative percent, i.e. the percentage of non-correct responses only. This is a better measure of preference for a strategy than percentage of all responses, since the targets differed markedly in the rate of correct responses, and comparing the actual percentages of an adaptation between a target with 5% correct and a target with 50% correct would skew the picture. The four hypotheses tested are:

1. The less well-formed a foreign sequence is perceived to be by speakers of the borrowing language, the more likely it is not to be reproduced in a target-like way, i.e. to be adapted
2. The degree of grammaticality of a target sequence influences the choice of adaptation type
3. The more frequent a potential adapted form is in the native lexicon of the borrowing language, the more likely it is to be used
4. The higher the perceived similarity is between a target form and a potential adaptation, the more often this adaptation will be used

Hypothesis 1

To test the association between grammaticality and successful reproduction, the mean perceived grammaticality of a target cluster was correlated with the percentage of correct responses in the production experiment. There was a significant positive correlation for the 22 clusters in all 3 conditions (N=66, r_p=.68, p<.01, 95% CI: .52<r<.79), showing that clusters with higher grammaticality were indeed produced targetlike more often than lower-rated clusters. The individual correlations for each condition separately are also significant, although it is noteworthy that this correlation is strongest in the sentence condition (N=22, r_p=.78, p<.01, 95% CI: .73<r<.90), weaker in the word condition (N=22, r_p=.65, p<.01, 95% CI: .34<r<.84), and weakest in the orthographic task (N=22, r_p=.63, p<.01, 95% CI: .30<r<.82). This is inversely related to the rate of correct responses overall; i.e. the grammaticality of the target becomes less decisive (even if it is still a strong factor) in how well a specific target cluster will be reproduced in a condition which favours better reproduction success overall. The scattergram in figure 4.5 shows all three conditions together.

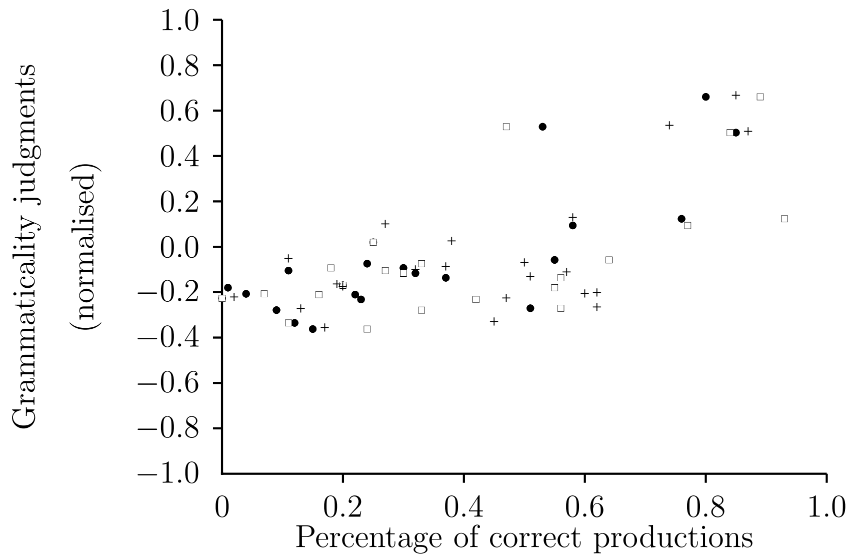

Figure 4.5: The perceived grammaticality of a target cluster and the percentage of its correct production, as divided by experiment condition: shadowing/sentence (o), shadowing/word (square) and orthography (+).

Hypothesis 2

This hypothesis is concerned with whether English speakers show a tendency to use a specific type of adaptation, depending on how ungrammatical they perceive the target stimulus to be. A somewhat similar idea can be found in Paradis' Theory of Constraints and Repair Strategies, claiming that deletion is used only where phonotactic and segmental problems coincide, i.e. in case of double ill-formedness. This was shown not to be correct in the production experiment; however, applying the idea to this case, deletion may be chosen predominantly for target clusters which are perceived to be at the lower end of the grammaticality scale.

Grammaticality and adaptation type				
Target	Grammaticality	Sentence	Word	Orthography
fr	0.666	**deletion**	combinations	EPENTHESIS
ps	0.529	**deletion**	**deletion**	**deletion**
sp	0.503	*segment change*	*segment change*	*segment change*
ʃm	0.123	*segment change*	*segment change*	*segment change*
ʃp	0.093	*segment change*	*segment change*	*segment change*
vb	0.019	EPENTHESIS	EPENTHESIS	**deletion**
sf	-0.058	combinations	combinations	*segment change*
dv	-0.075	EPENTHESIS	EPENTHESIS	EPENTHESIS
vl	-0.093	*segment change*	*segment change*	EPENTHESIS
zb	-0.105	EPENTHESIS	*segment change*	*segment change*
vr	-0.118	*segment change*	EPENTHESIS	EPENTHESIS
pn	-0.137	EPENTHESIS	EPENTHESIS	EPENTHESIS
ʃv	-0.170	*segment change*	*segment change*	combinations
fp	-0.181	**deletion**	EPENTHESIS	**deletion**
sv	-0.208	*segment change*	*segment change*	*segment change*
zv	-0.211	EPENTHESIS	*segment change*	EPENTHESIS
vz	-0.228	**deletion**	**deletion**	**deletion**
sr	-0.232	EPENTHESIS	EPENTHESIS	EPENTHESIS
pt	-0.272	EPENTHESIS	EPENTHESIS	EPENTHESIS
tv	-0.280	EPENTHESIS	EPENTHESIS	*segment change*
zr	-0.336	EPENTHESIS	EPENTHESIS	EPENTHESIS
ʒb	-0.363	EPENTHESIS	EPENTHESIS	EPENTHESIS

Table 4.5: This table lists the targets in descending order of judged grammaticality, and pairs them with the most frequent type of adaptation in each production task.

Table 4.5 lists the most frequent adaptations for all targets in descending order of perceived grammaticality. It shows that, while being the most frequent main adaptation overall, epenthesis shows a concentration at the lower end of the grammaticality scale, whilst none of the other adaptation types have a similar concentration. Figures 4.6 to 4.8 illustrate the connection of target grammaticality and each of the subsets.

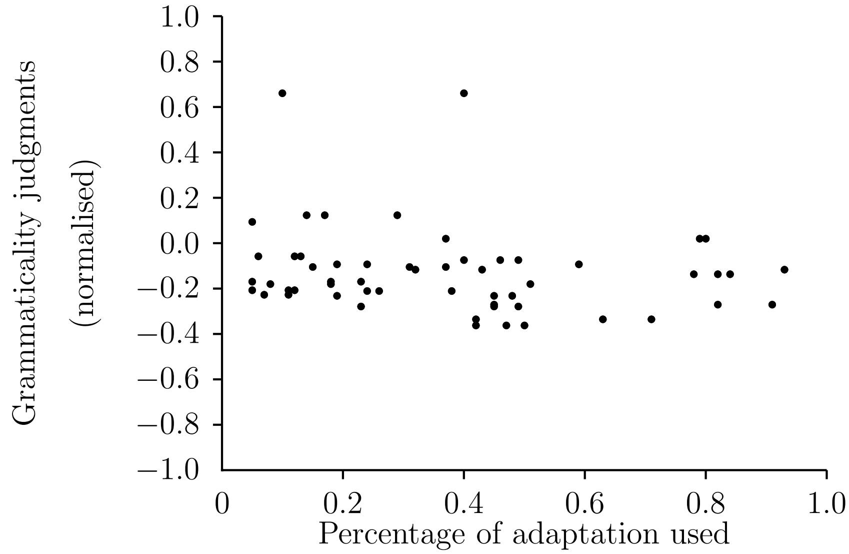

Figure 4.6: The grammaticality of a target and the proportional percentage of adaptation for all target-adaptation pairs in the epenthesis subset

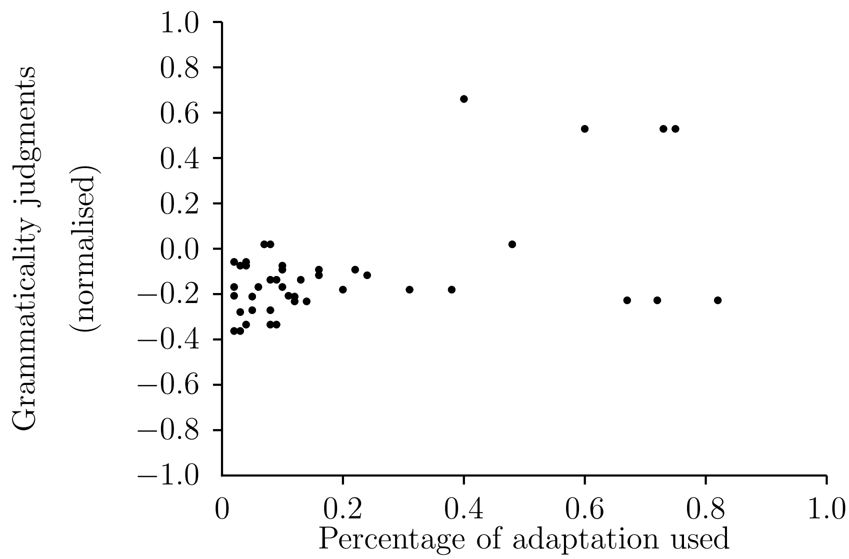

Figure 4.7: The grammaticality of a target and the proportional percentage of adaptation for all target-adaptation pairs in the deletion subset

To test statistically whether the grammaticality rating of a target cluster influences the choice of adaptation, a multinomial regression: this is to test the influence of a ratio variable, i.e. the grammaticality judgment in this case, on the choice of a

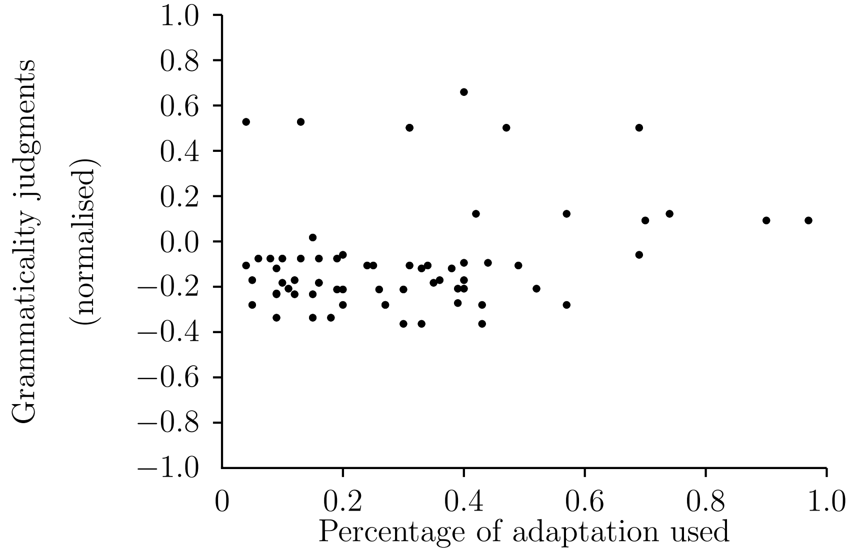

Figure 4.8: The grammaticality of a target and the proportional percentage of adaptation for all target-adaptation pairs in the segment change subset.

category, i.e. the method of adaptation. The levels of the category parameter, *main adaptation type*, are deletion, epenthesis, and segment change, with the latter as reference category (see table 4.6). This shows that the grammaticality of a given target cluster influences the choice of the preferred adaptation to some extent (Pseudo-R^2 (Cox and Snell) =.18). However, this influence can only be demonstrated to be significant for epenthesis, which is (at p<.01) significantly more associated with lower target grammaticality than segment changes, while there is no significant difference between the grammaticality of targets which associate with segment change and deletion.

Adaptation type	epenthesis			deletion		
	B	exp(B)	p	B	exp(B)	p
grammaticality	-4.09	.02	.01	1.00	2.73	.38
intercept	-.23		.46	-1.02		.01
N	30			10		

Table 4.6: The results of the multinomial regression testing the factor grammaticality in the assignment of the main adaptation type. Total N=66; the reference category is segment change.

Hypothesis 3

For the test of the influence of the frequency of a potential adapted form on its frequency in production, all target-adaptation pairs were used. It was calculated whether there was a significant correlation between type or token frequency of an adaptation in the lexicon and the frequency of its use in adaptation. As described above, rather than the absolute percentage of an adaptation for a specific target, the percentage within the adapted responses was used as a variable, reflecting the preference within the set of adaptations for each target.

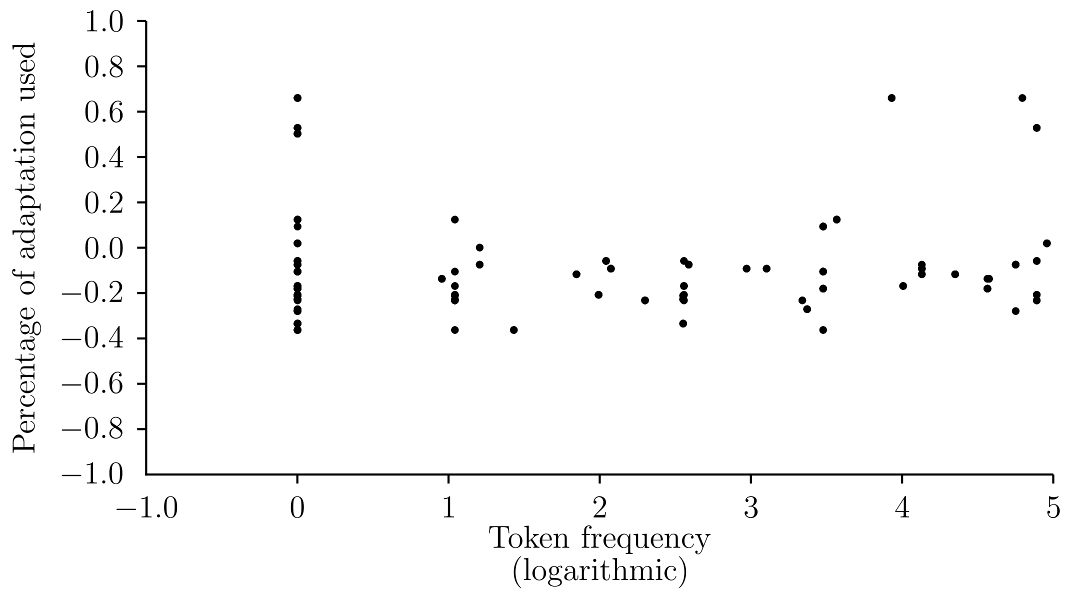

Figure 4.9: The token frequency of a potential adaptation and the proportional percentage of its use in the production task (not significantly correlated).

The distribution of values in the scattergram in figure 4.9 (for token frequency) does not suggest that higher frequency of a potential adapted onset goes along with higher use of this adaptation. In the statistical test of the correlation, neither with type nor token frequency, and neither overall nor in any of the conditions a significant correlation was found (overall: type: N=204, r_p=-.02, p=.72; token: N=204, r_p=-.05, p=.45). Hence, there is no support for the hypothesis that the frequency of a potential adaptation plays a role in the choice of adaptation. This suggests that factors outwith the lexicon are decisive.

Hypothesis 4

This hypothesis is to test the proposal that the preferred adaptation to a foreign word is the one most similar to the form in the source language. Table 4.7 clearly indicates that in the vast majority of cases the adaptation used most frequently by participants in the production experiment is fact also the one considered the most similar by English listeners.

Most frequent and most similar adaptations			
Target	Most frequent[3]	Most similar	Notes
ʃm	sm	sm	
sf	sw, sv	sv	sw could not be tested
ʃp	sp	sp	
pn	pən	pən	
sr	sər	zr, sər	zr not an improvement
zr	zər	zər	
vr	fr/vər, vər	vər	
vl	fl, vəl	vəl	
tv	təv, dv	təv	
dv	dəv	dəv	
ps	s	ts, s	
ʃv	ʃf, ʃw	ʃf	ʃw could not be tested
sv	sf, zv	sf	
zv	zəv/sf, zəv/sv	sv	
vz	z	z	
pt	pət	pət	
fp	p/fəp, p	sp	
vb	vəb, b	vəb	
zb	zəb/sb, sb	zəb	
ʒb	ʒəb	ʒəb	

Table 4.7: The most similar and the most frequent adaptations for marginal and illegal target clusters.

To support this observation statistically, the hypothesis was tested through a correlation of the mean perceived similarity of the target and adaptation pairs with (as above) the proportional percentage of one adaptation. N is smaller for this test than for the frequency correlations for hypothesis 2, since not all adaptations are legal in Russian, and hence eluded comparison to the target, as described in the method section above. From the scattergram it can be seen that percentage of adaptation rises with the perceived similarity between target and adaptation.

[3]The most frequent adaptations are given for both the shadowing and the orthographic tasks. If they coincide, the adaptation is only listed once.

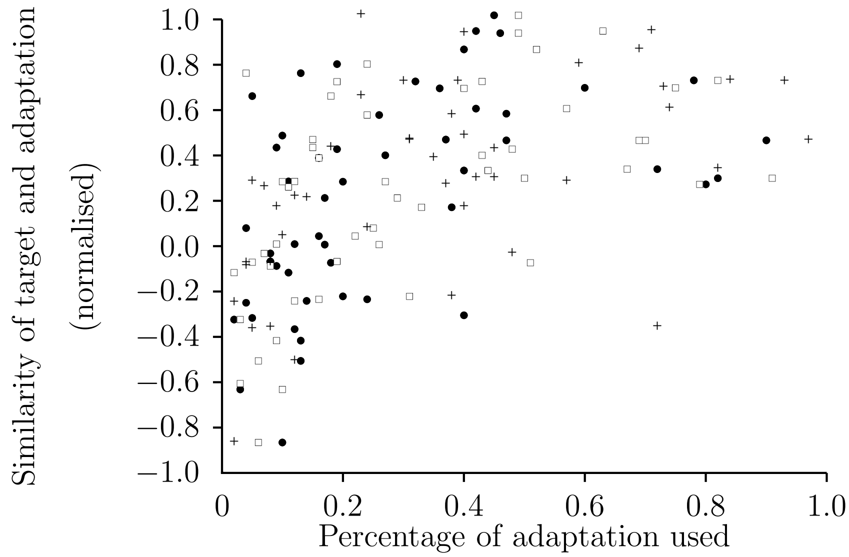

Figure 4.10: The similarity between a target and an adaptation for all target-adaptation pairs, and the proportional percentage of an adaptation being used. The figure shows all three production tasks: shadowing/sentence (o), shadowing/word (square) and orthography (+).

The correlations proved to be significant both overall (N=147, r_p=.52, p<.01, 95% CI: .36<r<.57) and for the different experiment conditions (sentence condition: N=52, r_p=.49, p<.01, 95% CI: .26<r<.67; word condition: N=51, r_p=.52, p<.01, 95% CI: .29<r<.70; orthographic task: N=44, r_p=.52, p<.01, 95% CI: .28<r<.71). Note that no reduction of r_p, i.e. the degree of correlation, through the conditions takes place, like for target grammaticality influencing success in reproducing a cluster. The role of similarity does not appear to change depending on the condition. Overall, similarity is the most decisive factor in which of the possible adaptations to a cluster will be chosen.

4.3.3 Summary for hypotheses

The frequency of potential adapted onsets in the English lexicon does not correlate with the strategy of adaptation. Judgments about the grammaticality of words containing illegal initial clusters, and about the similarity between pairs of words partially containing illegal onsets, were obtained from English native speakers. Similarity of a target to an adaptation was shown to be a predictor of its rate of use. The perceived grammaticality of a target cluster influenced performance in two ways: high-grammaticality target clusters were modified less often, and low-grammaticality clusters were mostly associated with vowel epenthesis.

4.4 Overall summary and discussion

The production experiments in the previous chapter showed that the way in which English speakers deal with phonotactically illegal foreign words cannot be predicted on the basis of a dichotomy between grammatical and ungrammatical, or on grammatical factors such as sonority distance alone. For this reason further factors, which are, in contrast to the focus of loanword research, not purely grammar-driven, and their influence on adaptation strategies have been examined in this chapter.

Of these factors, the frequency of a sequence in the L1 lexicon is shown to be associated with its perceived grammaticality, and a better predictor for how a structure's grammaticality will be perceived than grammatical factors such as sonority distance.

However, frequency is not a factor in the choice of one potential adaptation over another when speakers are faced with the task of dealing with an illegal sequence of phonemes in a foreign word. The perceived ungrammaticality of the target cluster, on the other hand, does influence this choice. Whereas the data do not allow a conclusion as to whether deletion and segment change have a specific target group, the strategy of vowel epenthesis is clearly associated with clusters having a lower grammaticality rating.

The third factor, the perceived similarity between a target cluster and an adaptation, is a good predictor of which changes are made to target clusters. The more similar an adaptation is perceived to be, the more likely it is to be used. The form of adaptation which English speakers consider to alter the sound of a sequence least is vowel epenthesis, followed by a single kind of feature change, whereas deletion and multiple feature changes are seen as more drastic.

This perception is interesting in the light of the fact that both epenthesis and deletion change the sequence by a whole segment, and epenthesis even creates an extra syllable, whereas segment changes only constitute the alteration of a single feature, i.e. a much smaller unit. It must be noted, however, that this segment is taken from a varied set for the consonants, but the vowel is always schwa. Nevertheless, the adaptation resulting in a change on the highest prosodic level is regarded as less drastic by English listeners than lower level changes.

Together with the fact that the grammaticality ratings for CC clusters and CəC sequences are only marginally different (while single onsets rank much more highly), this once again throws up the question of perception, and whether the English speakers are in fact capable of perceiving the difference. Such an inability would suggest, as claimed in Dupoux et al. (1999) for epenthetic vowels in Japanese, that the English speakers perceive a non-existent vowel between the two consonants of an illegal cluster; this would explain the high similarity rating between clusters and epenthesised clusters, as well as the prevalence of epenthesis as an adaptation. Problematic for this proposal is the low grammaticality rating of CC and CəC. If a vowel is perceived between the two consonants, then this creates a single-onset syllable. It is not clear why this should be rated lower than other single onsets. Alternatively, if the schwa is not perceived, or rather, if it is interpreted as a variant of the C_1 realisation instead of a syllable peak (as listeners did in a study by Donselaar et al. 1999 for epenthesised Dutch words), these ratings would appear more explicable. However, the frequent occurrence of epenthesis in the orthographic task of chapter 3 speaks against this hypothesis.

The next question which must be addressed is therefore the perception of differences between words, and thus the nature of similarity. Several issues must be addressed:

- how it can be measured
- which prosodic levels must be taken into account
- whether it is influenced by one's native language and the well-formedness of the comparanda in it
- whether its role in adaptation is situated in the perceptual process, i.e. misperceptions lead to adaptations, or in the production process, such that intuitions of similarity are part of the productive phonology

These questions will be addressed by comparing Russian speakers' similarity judgments on the materials presented to English speakers. Such a comparison will shed light on cross-linguistic differences in the perception of similarity for various objective acoustic differences, and determine whether the status in terms of legality has an impact on how a structure is perceived.

Chapter 5

Perceived Similarity and L1 grammar

5.1 Similarity and research

Chapter 4 described how judgments were collected from native speakers of English about the similarity of pairs of words differing in their onset. These judgments were used as a basis for testing the prediction that the relative similarity of conceivable adaptations to an ill-formed target sequence influences the tendency to use them. After this prediction was confirmed, a further investigation into the workings of what determines perceived similarity was necessary, since - in contrast to grammaticality and frequency - it is least obviously a language-specific criterion, as the adjective *perceived* suggests: it is not at all clear that a speaker's first language is the only or even the main determinant of similarity. In fact, as the overview of the field will show, both the idea that perceived similarity is determined by universal, perceptual principles, and the idea that L1 influences the perception of similarity, have been put forward in an extensive body of recent work.

As Trask (1996) notes in his *Dictionary of Phonetics and Phonology*, phonetic similarity is "an important but elusive criterion for phonological analysis". It is important in that a notion of similarity, more or less well defined, has been the basis of a large number of accounts of phonological phenomena, in various areas. It is also elusive because it is not at all straightforward to devise a measure of similarity that can be used as the basis for statistical analyses to determine its role in phonological processes. Not only is it not agreed whether there is any influence of one's native language on how similar one perceives sounds to be, it is also not clear whether or not (and if so, how) a sound's context should be taken into account, or on the basis of which phonological or phonetic units similarity should be calculated. Nevertheless, the notion of similarity has often found its way into phonological research, as the next section shows.

5.1.1 Similarity as a tool in analysis

Similarity as a concept is used in linguistics in many different ways, and often intuitive ideas about similarity are used rather than explicit models. This section describes research leaning on the concept of similarity - the models of similarity they are based on will be explained in the next section.

One of the areas of research that has employed the similarity between words is that of the lexicon. The density of similarity neighbourhoods in the lexicon has been shown to influence word recognition, as in Vitevitch & Luce (1998) and Vitevitch et al. (1999). They compared the effect of phonotactic probability, i.e. the frequency with which a given sequence occurs in a language's lexicon, with that of neighbourhood density, i.e. the number of phonologically most similar words. They found that phonotactic probability facilitates recognition in non-words (i.e. on a sublexical level), whereas a dense similarity neighbourhood in the lexicon inhibits recognition on the lexical level. Thus they demonstrate processing on two different linguistic levels.

Lukatela & Turvey (1990) examine the effects of context words on the processing of phonemically and graphemically similar words in Serbo-Croatian (visually presented in either Cyrillic or Roman alphabets) in various tasks, such as naming and lexical decision. They find phonemic similarity effects, such as an inhibitory effect on lexical decision (dependent on the position of the sounds not shared by the word pair), and a facilitating effect on the naming of both words and pseudo-words.

Mueller et al. (2003) use phonological similarity as a factor in their study of verbal working memory, testing predictions of Baddeley's phonological-loop model (Baddeley 1986) such as that sequences of phonologically dissimilar words should be remembered better than those which are similar. Mueller et al.'s results suggest that the prediction is indeed correct for syllable onsets, but not necessarily so for rhymes.

A further large field of research which has employed a notion of similarity as an explanans is that of second language acquisition, specifically in the area of differences between two languages' segment inventories. Flege's speech learning model (SLM) (Flege 1987*b*, 1995) is based on the idea that the more similar a new phoneme is to one of the native language inventory, the more difficult it will be to form a new, independent category for this sound. Hence for an English speaker the acquisition of French /u/ would be more problematic than French /y/, since the latter is entirely new, whereas the phoneme /u/ exists in English, however with different phonetic properties, for example F2 differences. English native speakers will therefore substitute the French phonetic surface form with the specifications from their native sound. This process is termed "equivalence classification".

In a similar vein, Best's (e.g. 1995) Perceptual Assimilation Model (PAM) concentrates on discrimination abilities for foreign sounds. The model predicts that foreign sounds cannot be discriminated if they are perceptually assimilated to the same native articulatory category. Discrimination is predicted to be unimpaired if two foreign sounds are assimilated to different categories, or if only one of them is similar enough to a native category to be perceptually assimilated.

A further model of perceptual assimilation is Kuhl's (Kuhl & Iverson 1995) Perceptual Magnet Effect, which claims a perceptual warping of the auditory space as a consequence of exposure to one's native language. This warping, or shrinking, results in a magnetic effect of the prototypical native categories to attract any sounds similar to them, but becomes weaker with increasing distance. Any foreign sounds similar to native prototypes will therefore not be discriminated from them, and harder to acquire.

A relatively new field working with the notion of similarity is that of historical linguistics: the aim is to develop a new approach to language classification, to explore the relationship between dialects, and to calculate the similarity between any given pair of languages or dialects. This is achieved by computing phonetic similarity between cognate words and feeding the values for those into tree-drawing programs (e.g. PHYLIP, Felsenstein 2001), to produce a representation similar to traditional language family trees, but also representing the closeness or distance between the branches in terms of similarity (e.g. Dyen et al. 1992, McMahon & McMahon 2003, Ellison & Kirby 2006).

In the realm of loanword research, the notions of similarity and the importance of perception were introduced by Silverman's (1992) work on Cantonese loanwords, as described in chapter 1. Since then, the debate it triggered regarding the role of perception in (loanword) phonology has gained increasing momentum, and has been dominating the discourse in this area. Yip (1993) introduced Silverman's idea of language-specific salience into optimality theory, Jacobs and Gussenhoven's (2000) rejected this notion as well as any relevance of perception, along with Paradis (e.g. Paradis 2005, Paradis & LaCharité 1997).

However, what is dominant in Paradis' work as well as elsewhere in loanword research (e.g. Kenstowicz 2001, 2005, in press, Adler 2006) is the idea of minimality: any and all adaptations of loanwords are intended to be as minimal as possible, while resulting in a form that respects the phonology of the borrowing language. This principle underlies most current work in loanword research; Steriade (2001, p.5) claims that "for at least some phonological properties and perhaps for all, there appears to exist a cross-linguistically constant notion of minimal modification", which is precisely the modification that is preferred in loanword adaptations. Kenstowicz (in press) writes, "The speaker will tend to preserve features whose absence would be most noticeable; and when a repair must be made, like a good tailor he will make his alterations as unobtrusive as possible by substituting a sound that most closely resembles the original", and claims that decisions are based on "some notion of auditory similarity." Kang (2003) proposes for her analysis of Korean loanwords from English that "the motivation of vowel insertion is the maximisation of the perceptual similarity between the English input and the Korean output, as well as to obey a morphophonemic restriction in Korean", and Adler (2004) states that "loanword adaptation is grounded in the perceptibility of modifications from input to output", and that perceptual or articulatory features can be the target of preservation.

A further direction in loanword research is represented by Peperkamp (e.g. Peperkamp 2003), who also strives to explain loanword adaptation through similarity, but takes a different view of where it takes its effect: since according to her model all adaptations take place through phonetic mapping to the closest available native output during *perception*, no loanword phonology is necessary.

Minimality of change, i.e. an input and output form that are maximally similar to each other, plays an important role in analysing loanwords. Unfortunately, the answers to the question "What is minimal?", and the measurements of similarity are often intuitive, and thus applied differently in different interpretations.

5.1.2 Definitions of similarity

As noted above, since there is no commonly accepted measurement for phonological similarity, scientists often rely on their intuitions, and on their audience's intuitions coinciding with their own, or on subset relations, such as between the sets of features shared by two phonemes. For example, there will be little disagreement that /b/ and /m/ are more similar to each other than /p/ and /m/, since the former share all the features the latter share, plus the addition of voicing. However, since in most cases the comparisons are not as straightforward as this, and "plus the added feature voicing" does not answer the question "How *much* more similar?", this section will provide an overview of different viewpoints of how to calculate or define similarity. Some of these viewpoints have been the main objective of the works cited, some arise from standpoints on the relationship between perception and phonology, or on loanword adaptation, some have been created or used for other purposes.

The measure of word similarity, as used used by Vitevitch and colleagues (e.g. Vitevitch et al. 1999) in work with lexical neighbourhoods, is rather coarsely meshed and based on phonemes: the similarity neighbourhood of a word is comprised of all words with a distance of one step, where a step is any one-phoneme change, be it the addition or deletion or exchange of a segment. This is clearly not adequate for our purposes, since, among other reasons, the first similarity judgment showed that not all of the comparisons (which were all one-segment changes) were rated the same.

A more fine-grained analysis looks at a different, lower level of representation, namely the feature. Thus it avoids equating the loss or gain of a whole segment, which can be represented as a bundle of features, with changing the value of a single feature, for example /s/ - /z/.

One such measure of similarity is PSIMETRICA developed by Mueller et al. (2003), who are interested in verbal working memory and the *dis*similarity of whole words, for which they take into account syllable structure and the number of shared features. In their method the comparison between segments is extended to whole words by aligning syllables and syllable constituents, and then assigning within-coda and within-onset positions. After each segment has been allocated its counterpart (which, of course, can be zero), the pair is compared, and the dissimilarity values are averaged over syllable constituents and then whole words. Mueller et al. determine the similarity between segments with the use of a set of 13 features, as taken from Chomsky and Halle's SPE (Chomsky & Halle 1968). Therefore only those features are used which are necessary to differentiate between all segments of the English inventory, which means that Mueller et al.'s measurement depends heavily on both the English language and on SPE. A criticism is that PSIMETRICA's only way of taking into account positional effects is in an abstract phonological way of tying a segment to a specific place in the prosodic structure, rather than by evaluating the influence neighbouring segments (or boundaries) may have on the perception of the sound, and the assignment of consonants to specific onset slots has an undue influence on the result.

For example, the assignment of segments to positions in a syllable representation should have no effect on perceived similarity at all. For example, in the case of /splæʃ/ compared to /θɹaɪv/ for a cue-based model it does not matter whether the fricative and the rhotic of the latter occupy positions I and II, or II and III, of the syllable onset. For PSIMETRICA, however, this makes the difference between comparing either

1. /s, θ/, /p,r/, /l,∅/ or
2. /s, ∅/, /p,θ/, /l,r/ (used in PSIMETRICA) or
3. /s, θ/, /p,∅/, /l,r/

More seriously, PSIMETRICA contains no method of weighting features - all features have the same importance. The results of this chapter will demonstrate that this is not the case. Previously, Treiman et al. (1998) found in a phoneme recognition study that children did not, as was previously assumed, have fewer problems recognising fricatives than plosives. Voicing, on the other hand, was not very easily distinguished. Treiman et al. conclude from this that oral features may be more salient than laryngeal features. Whether that is correct or not, the experiment shows that equating features is not appropriate in studying phonetic similarity, and that some features appear to be more salient than others.

A more theory-driven motivation for weighting features is that their functional loads differ from language to language. In most languages, nasals are always voiced, hence the specification [+nasal] predicts the feature [+voice]. A specification of a value for voicing in for a nasal is redundant, and the underlying representation is therefore underspecified. The implication of underspecification theory (e.g. Archangeli 1988) for similarity is that redundant features are weighted less in determining how similar two sounds are to each other than contrastive features. A direct application of this for loanwords is found in Weinberger (1990), who uses underspecification in his account of differential substitution such that the most underspecified segment of a language will be substituted for a sound that does not occur in the borrowing language.

Frisch (1996) uses a model of similarity based on weighting different features, and specific to any given language. This is used as a basis of his account of effects of the Obligatory Contour Principle (OCP) in Arabic. His approach is based on Broe's theory of structured specification (Broe 1993) as opposed to underspecification. Structured specification is a representation of the segment inventory of a language in a lattice showing natural classes of sounds, i.e groups of phonemes sharing one or more features (as seen for English nasals in figure 5.1).

In this model the similarity of two sounds is determined in a very straightforward way. From the complete lattice of the segment inventory of a language, the number of shared nodes, i.e. natural classes, is divided by the overall number of nodes, which therefore results in a number between 0 (no nodes shared) and 1 (all nodes shared). Broe (1993) extends this approach to a language's lexicon, where lexical entries can share segments in certain positions (as opposed to the features that two phonemes may share), and thus allows for a calculation of the similarities between words (only those which occur in the lexicon) as well as phonemes.

Frisch's model of similarity is based entirely on the phonemic inventory of a language, and thus cannot expressly account for effects of similarity between native sounds and those from a foreign language, nor does it take context into account. The prediction *derived* from Frisch's account applied to the case at hand, which contains phonotactic, but not segmental adaptation problems, is as follows: the similarity of two words which differ only in one segment is proportional to the similarity between the two relevant phonemes. Thus any judgments of similarity

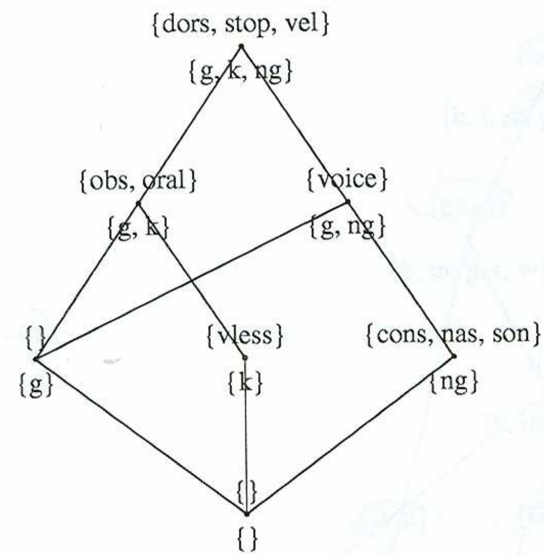

Figure 5.1: Lattice of English nasals, taken from Frisch (1996), p.25.

should correlate with the relevant phoneme similarity values, as derived from an inventory's lattice. Another prediction of this approach that, whichever context two phonemes are compared in, their similarity will be the same, since similarity is only determined by a phoneme's features and not by its context.

All work described so far has been based on phonological units and made no reference either to phonotactics or phonetic detail, and has not incorporated perception (with the exception of Treiman et al. 1998). The research described below broadens the horizons.

As noted above, in the area of loanwords, Silverman (1992) was the pioneer with regard to perception. While this is meritorious, his account of similarity remained vague and was not developed further. His approach, as detailed in chapter 1, includes a perceptual level, during which foreign segments are filtered by the native perceptual system, and allocated a "similar" feature matrix in the native inventory. What remains unclear is whether there exists any kind of threshold of similarity. How similar is similar enough? Or how far is too far? If foreign segments are allocated a similar feature matrix (and the question is also how speakers make the connection between a mess of known and unknown acoustic features and the phonological feature matrices they have stored), is there a point at which the segment is too dissimilar, for example when maybe only a single or two common features can be found with any native segment? Silverman does not elaborate on this process, and detects possible problems himself, such as the apparent perception of /v/ as /w/ rather than /f/, although in other cases the voiceless counterpart of a non-native voiced phoneme is chosen and /f/ is a native segment of Cantonese.

Yip's (1993) optimality-theoretic account of Silverman's data suffers from its own problems in terms of defining what is salient (and therefore preserved in perception), since a constraint such as PARSE(salient) should on the one hand, being a constraint in optimality theory, be universal, and on the other hand specific to Cantonese, since it eliminates for Cantonese speakers what would be salient for English speakers. The account also does not sit well with Silverman's perceptual level being pre-grammatical. This issue could be dealt with by Boersma's model, which will be described later in this section.

The problems with Silverman's perceptual level also exist for the SLA models, such as Best's PAM (Perceptual Assimilation Model; described in section 1.3.4); the overriding theme there is that there exist a threshold of similarity beyond which a foreign segment will not be equated with a native one. Kuhl's NLM (Native Language Magnet, described in section 1.3.4) predominantly deals with vowels, where a relationship to objective acoustic similarity is relatively easy to come by, i.e. the distance between them in the F1/F2 plane, since these formants are the relevant differences.[1] For consonants, the story is more complicated, and more complicated still if context is taken into account.

The role of context is the one of the points on which Steriade's work concentrates: in contrast to a language-specific, but context-independent concept of similarity, Steriade's licensing-by-cue concept determines similarity on the basis of available acoustic cues to a given contrast. This availability firstly depends on relative distance, which can be illustrated by three vowels with the same height, but the distance between the front and back vowel being greater (as shown by F2) than that between either and a central vowel. Further relevant factors are the subset effect described earlier and cue duration. How perceptible acoustic cues are also depends largely on context: for example, the place of articulation for a stop is more perceptible before a vowel than before another consonant, since this environment shows the transition cues for place. In her work on the P-map (Steriade 2001), Steriade cites evidence from half-rhymes[2] (see Steriade 2003), cross-linguistic tendencies

[1]However, it is conceivable that the two formants are not weighted equally. As Lindblom 1975 argues, from a perceptual point of view, F1 differences and contrasts should carry more weight than those in other formants.

[2]Half-rhymes are defined as incomplete rhymes, i.e. codas which do not completely coincide. Differences include for example place of articulation, as in *time-nine*, or consonant metathesis, as in *raised-raids*. In Steriade's data, half-rhymes with high similarity according to the P-map are preferred and significantly more frequent.

in the resolution of conflicts (such as the repair of prohibited voiced final obstruents), and previous studies of similarity or confusion rates to support the idea that perceived similarity is determined by cross-linguistically applicable, perceptual principles. This approach has been developed for example in Hayes & Steriade (2003), and also been used by Jun (1995) and Flemming (2001, 2003, 2005). The knowledge of these differences in similarity and of the ranking of contrast perceptibility - which speakers know intuitively - is then claimed to be part of a phonological production grammar. This grammar applies in loan phonology and is responsible for minimal adaptations. Specific claims of perceptibility rankings are described in the method section of this chapter, to provide the rationales and predictions for the tested hypotheses in direct comparison to the results.

Boersma and colleagues have developed yet another way of thinking of similarity in phonology, as described in chapter 1. With perception itself regarded as a grammatical process, and the output of this process argued to be a well-formed structure in the borrowing language, similarity must be a part of phonology; the the constraint hierarchy (Boersma's being an optimality-theoretic account) sees to it that the output of the perceptual grammar is on the one hand well-formed, and on the other as similar to the input as possible. This is achieved by using acoustic characteristics such as duration and formant values in classificatory constraints, such as "F1 of 349 Hz is not [ɪ]" or "74ms duration is not /i/" (Escudero & Boersma 2004). In Boersma's *functional phonology*, therefore, similarity is both a perceptual and grammatical parameter.

In one respect this resembles the view that Peperkamp and Dupoux argue for (e.g. Dupoux et al. 1999, Peperkamp 2003). Like Boersma, they do not explicitly attempt to define or measure similarity, but their approach to loanword adaptation makes the viewpoint on similarity clear: it is perceived of as a factor that plays its role in perception. However, in the other respect, their claim is diametrically opposed to Boersma's: the adaptation to the most similar forms in perception is, in Peperkamp's view, not a phonological, but a phonetic process. The influence of phonology is relegated to the provision of possible forms, while the important role is played by the phonetic similarity of the incoming signal to possible native realisations. For example, for the cluster adaptations at hand, Peperkamp's model predicts high rates of epenthesis for first consonants with a burst. While this is true for most of these clusters, epenthesis is no less frequent for several fricative-stop and, especially in the orthographic task, fricative-liquid clusters, where no phonetic motivation is obvious.[3]

[3]Moreover, not only do the clusters /pt/ and /pn/, where the burst is the most obvious, have high rates of epenthesis, but they are also the target clusters with the highest rates of correct

As becomes clear from the descriptions of models represented by Steriade, Boersma and Peperkamp, the issue at hand is not which acoustic cues exist (which is often obvious), and in many cases not even which cues are more salient than others. The core of the matter is how we use the cues that are available, and whether there are some situations (for example, certain phonological units or certain environments) or other factors, such as our native language, that cause us to use these cues differently, or to be less perceptive to these cues than general auditory principles would suggest.

To split this issue into smaller sub-issues, the following questions about the perception of sound similarity are in need of answering before we can confidently use it beyond restricted niches (such as cases where subset relations apply) in investigating loanwords:

- Are there cross-linguistically identical intuitions of similarity, or does a person's native language account for differences in these notions?
- Which level of representation is relevant for similarity: phonetics (acoustic cues)? phonology (features, segment inventory or phonotactics)? Is there a single decisive unit or level?
- Is similarity determined by (L1-influenced) perception? On which level of representation?
- Is context relevant for the perceived similarity between two units of sound?

Summarising the viewpoints of some of the researchers discussed above, table 5.1 gives an overview of their positions on some of the relevant aspects. The first column is concerned with whether similarity is assumed to be perceived in the same way cross-linguistically, or conversely, influenced by L1. Most models except for Steriade's[4] assume some form of L1 influence. Note, however, that this influence can take different shapes; specifically, Frisch's account contrasts with others by assuming L1 influence only through the segment inventory. The second and third column refer to the level of representation being involved in determining similarity. For Steriade, knowledge of phonetic similarity is reflected in the constraint hierarchy of the production phonology, for Boersma, phonetic detail is part of perception phonology. The next column considers the importance of perception for similarity, which all models but Frisch claim, however in different manifestations. The last column gives the standpoints regarding the relevance of context. This is advocated

responses. This suggests that in the majority of cases the burst is, in fact, used as an aid to correct perception by English speakers. On the topic of English and Russian production of consonant releases and default patterns thereof, also see Zsiga 2003.

[4]While she explicitly does not consider the P-map to be universal (pers.comm.), there is no description of a possible L1 influence in her work.

most strongly by Steriade, as a determinant of cue availability, and by Dupoux, through the (il)legality of sound combinations.

	cross-ling/L1	*Level of representation*		*Perception*	*Context*
		phonetics	phonology		
Frisch	L1	-	+	-	-
Best, Kuhl, Flege	L1	+	-	+	?
Dupoux	L1	-	+	+	+
Peperkamp	L1	+	-	+	+
Steriade	cross-ling	+	+	+	+
Boersma	L1	+	+	+	+

Table 5.1: The various models of similarity, and their positions regarding issues in determining similarity. A plus sign symbolises that the issue in question is relevant, a minus sign symbolises that it is not, and a question mark shows that no clear position has been taken.

5.2 Russian and English compared

As the table in the previous section shows, the issues are whether similarity is language-dependent or not, if yes, what level of L1 exerts the influence, and whether it is context-dependent or not. These issues were investigated by comparing Russian speakers' similarity judgments on the materials presented to those of English speakers - if there is indeed an effect of L1 phonotactics on perception as claimed in Dupoux et al. (1999), there is the possibility that the similarity judgments collected in the previous experiment might have been compromised (i.e. influenced by L1) by a failure to perceive the stimuli correctly. Support for the cue-based model would be garnered through good correlations between English an Russian listeners, and significance for predicted effects, such as a different similarity depending on position and context. L1 influence would be demonstrated through L1 effects showing significant differences between the groups, and interaction effects, specifically where the sequences in question have a different legality status depending on background language, or different phonetic details. Segmental differences are not part of this experiment, as in the previous chapter.

5.2.1 Method

Hypotheses that were either explicitly stated or could be derived from the models were selected to investigate the relative merit of the different approaches. All hypotheses were tested using a cross-linguistic comparison of judgments from Russian or English native listeners.

Participants

The participants were students recruited at the University of Edinburgh, and were reimbursed at a rate of £5/hour. There were 10 native speakers each of Russian (ages 20-28) and of British varieties of English (ages 18-26). None of the participants reported any hearing or speech production difficulties.

After finishing the experiment, participants filled in a questionnaire about their language background, second languages and impressions of the experiment. These revealed that among the participants there was one Russian speaker trained in phonetics and phonology, and one English speaker with knowledge of Russian, and so their results were disregarded. No other English participants had any knowledge of Russian or another Slavic language. One further Russian dataset had do be discarded after completion due to technical problems. The data for analysis therefore consisted of 9 English and 8 Russian sets.

Materials

The shortcoming - for a general appraisal of perceived similarity - of the first set of similarity judgments carried out was twofold: firstly, since it was English listeners performing the adaptations, only English listeners' judgments were collected as relevant for the initial study, and there was no cross-linguistic control. Secondly, the kinds of pairs compared were derived from the targets and their actual and potential adaptations only. In this experiment, however, in a step towards a more systematic insight into the perception of differences, the materials were tailored to testing specific hypotheses. Nevertheless, this was restricted by the available onsets in Russian. Furthermore, it was necessary in cases to combine English legal and illegal clusters to increase the number of data points. The comparisons used are given for each hypothesis, and the implications are discussed.

The stimuli were designed to conform to Russian phonology and phonotactics. In this respect, they are very similar to those in the similarity task of the previous chapter. However, in that experiment, only English listeners were participating who were oblivious to Russian phonology (although some participants speculated after the experiment that the stimuli were Russian). In this experiment, Russian participants were taking part, and so a balance had to be struck between designing the stimuli such that on the one hand they sounded natural enough to be recognised and accepted as Russian by the Russian listeners, but that on the other hand only phonological, and not semantic considerations, influenced their decisions. This was achieved by using common Russian verbal and nominal suffixes in combinations with desired sound sequences to improve naturalness, while at the same time avoiding actual words of the Russian lexicon. Furthermore, verbal prefixes were used, as well as cliticising prepositions forming (in prosodic words) clusters which are not found in Russian *lexical* items[5]. The disadvantage of this procedure is a possible influence of the Russian participants' morphological knowledge influencing their judgments. The disadvantage of forgoing this avenue of using initial clusters which occur in Russian prosodic words would have been either to severely reduce the number of possible comparisons and/or resort to clusters very rare in the Russian lexicon. To combat potential morphological effects, it was impressed on the participants that they were asked to rate the similarities solely on the basis of how the words sounded.

As in chapter 4, the stimulus materials consisted of pairs of words which differ in their initial sequence[6]. These were either, as in most cases, syllable onsets consisting of one or more consonants, or CəC sequences. Therefore, the pairwise combinations of these resulted in ±V and ±C comparisons, as well as segment changes, such as /dv-gv/.

To determine the onsets for comparison, initially an analysis of possible initial two-consonant clusters (including those generated by cliticised prepositions) was performed. Out of over 3000 resulting possible comparisons (including all possible segment changes, and deletions and vowel epentheses), 134 pairs of onsets were chosen for best fit to the hypotheses (see appendix F). For example, to compare changes of nasality and place of articulation, sets of comparisons like {/zd-zg/, /zd-zn/} were added. These onsets were combined with one of {ʌ, u} and a further CVC syllable, of the set {moj, dɛk, ʃok}. The syllables were chosen for "inoffensiveness" with respect to English phonology, and, with the help of a Russian colleague, for naturalness in Russian. They were stressed to enable the production of schwa in the CəC cases, since stress must be at least two syllables distant from an orthographic <a> or <o> for vowel reduction.

The stimuli were transcribed in Cyrillic, and recorded by a native Russian phonetician onto DAT tape in the soundproof recording studio in the phonetics department at the University of St.Petersburg onto DAT tape. The recordings were digitised at the University of Edinburgh using a Townsend Datlink+ at 16khz mono,

[5]For a fuller description, please refer back to chapter 2.
[6]Again, I will refer to these as onsets, although in the case of a CəC sequences this is in fact a whole syllable plus an onset.

multiplexed to the SONY PCM-2700A and to a Sun Sparc Ultra running Solaris 2.4 for archiving to disk.

The stimuli also follow Russian phonetics, as did the stimuli in the previous chapter. Since the auditory and spectrographic analysis of the recordings revealed that not all recorded stimuli were usable, two out of the three final syllables were chosen for most comparisons from the remaining, such that the endings were the same for the stimuli of each comparison. This resulted in four data points per participant per comparison.

There were 550 test items to be rated by each participant, interspersed with 42% distractor pairs, which differed in later parts of the words in order to distract participants from the onsets (/gabər - gabət/), and very often in the vowels, to distract from the consonant comparisons (e.g. /jenʲist - jenost/).

Procedure

The word pairs were presented auditorily via Sennheiser HD 457 headphones, in randomised order. The tokens for each comparison were split such that each onset was presented first in half the cases. *Psychology Software Tools*'s experiment program *Eprime* was used to present stimuli and record responses and response times. As in the previous experiment, participants were instructed to provide a rating of the similarity of the words in a pair based on "how they sounded", with magnitude estimation. Keying in the response followed by the ENTER key moved the experiment on to the next word pair, thus allowing the participants to proceed at a comfortable pace. Before beginning the experiment, the informants went through a practice round of 10 pairs at least once, but potentially more often, until they felt confident about their rating system. Since *Magnitude Estimation*[7] was used for the judgments, each participant developed their own scale. These scales were normalised to a single scale with a mean of zero by a transformation to z scores.

Altogether 950 word pairs were rated by each participant, divided into blocks of 25. Participants had the option of pausing after each block. The stimulus set was divided into two, with a break of at least 10 minutes inserted between them, to avoid a lowering of concentration due to the large number of comparisons.

[7]For a discussion of Magnitude Estimation, refer to the method section of chapter 4

5.2.2 Hypotheses and results

This section describes the hypotheses that were tested, their source or the background that they were derived from, and the motivations for them, as well as the results that were found. The rationales do not in all cases pertain to the exact conditions tested here (for example, in some cases they apply to word-final rather than initial cases) or to similarity per se, yet they are applicable and relevant to the conception of the hypotheses. Additionally, this section lists the comparisons from this experiment that were used to test the claims. There are three groups of hypotheses, each with a different approach to the similarity question:

1. Cue- and context-based (universal) predictions, as found for example in Steriade (2001), Jun (1995), Hayes & Steriade (2003), which are expected to apply cross-linguistically.
    - A voicing change is more similar than a nasality change
    - A place of articulation (POA) change is more similar than a nasality change
    - A voicing change is more similar for the word-initial than the prevocalic consonant
    - A POA change is more similar when preceding a non-coronal consonant
    - A POA change is more similar for a stop than a fricative
    - A change in anteriority is more similar than a stridency change
    - A voicing change is more similar than the addition or deletion of a whole segment
    - The similarity of a consonant to zero depends on stricture
    - Presence or absence of a consonant is more obvious pre-vocalically than word-initially
2. L1 inventory-based (language-specific and context-independent) predictions, derived from Broe (1993), Frisch (1996)
    - Inventory-derived segment similarity is applicable to perceived word similarity
3. L1 phonotactics-based (language-specific) predictions, following Dupoux et al. (1999)
    - Judgments differ cross-linguistically where phonotactic status differs
    - Acoustic cues are only used fully for phonotactically legal sequences

For purposes of conciseness, the hypotheses are given in the following notation hereafter: $\gg$ signifies "is perceived to be more similar than", $[\pm feature]$ signifies a change in the feature. For example, $[\pm voice(C_1)] \gg [\pm voice(C_2)]$ stands for "A change in voicing in the first element of the cluster is perceived to be more similar than a change in voicing in the second element".

Subject variability. Before the individual hypotheses were tested, the complete dataset was tested for subject and item (i.e. word ending) variability. To this end, ANOVAS were run for both the English and Russian data, for the dependent variable similarity, with the independent variable change type, with the levels voicing change, place of articulation change, manner change, epenthesis, and deletion. Further independent variables were the factors subject and word ending. A significant effect of change type was the result in a by-subject and by-item (i.e. word ending) test, confirming that the type of change effect is not due to subject or word ending variability. For the Russian set (change type: $F(4,3947)=37.30$, $p<.01$), the post-hoc Tukey tests showed the homogeneous groups to be voicing changes as most similar, followed by epenthesis, followed by the remaining changes. In the English group (change type: $F(4,4409)=15.84$, $p<.01$), two of the listeners rated voicing changes significantly lower. Nevertheless, overall the post-hoc tests showed voicing to be perceived as the most similar change, followed by the group of place of articulation and manner changes and epenthesis, and deletion as the least similar.

Cue- and context-based predictions (Hypotheses 1-9)

For this group of hypotheses, the perceived similarity will be compared for various changes, such as changing the place of articulation of a segment as compared to changing its nasality. The means for perceived similarity will be tested for influence from the type of change and from participants' native languages. In terms of the statistics, the dependent variable will in all cases be the ratio variable *similarity rating*, and the participants' native language and the type of change are the factors. Since the former is a between-group factor and the latter a within-group factor, the appropriate test is a mixed ANOVA.

For evidence in favour of the cue-based hypotheses, similarity must depend solely on the type of change and be independent of the native language. Hence, the outcome is expected to be an main effect for the type of change, and neither an L1 effect nor an interaction. This is not the case for all hypotheses. While the expected main effects are often significant, L1 effects and interactions also occur in most cases, and thus show that similarity is not entirely independent of one's native language system. The details of the potential points of influence, for example phonetic differences or phonotactic well-formedness, are discussed - along with motivations and results - separately for each hypothesis.

- Hypothesis 1: $[\pm voice] \gg [\pm nasality]$

This claim, stating that a voicing change is more similar than a change in nasality, is one of the central points in the P-map (Steriade 2001): in the case of a voiced stop occurring underlyingly in a word-final position where voiced obstruents are disallowed, according to Steriade (2001), a number of repairs are possible, such as C deletion, final V insertion, nasalisation of the obstruent, lenition of the obstruent to a glide, and devoicing. The disallowed /lib/ could for example be changed to /li/, /libə/, /lim/, /liw/, or /lip/. However, attested cases of the resolution of this conflict show that the sole solution employed is devoicing. More evidence for the high similarity of [± voice] comes from an analysis of half-rhymes, i.e. rhymes with a featural difference in the final C, in English (Hanson 1999), and in Romanian poetry (Steriade 2003), where voicing was the overwhelmingly most frequently used.

In support of voicing as the preferred repair Steriade also reports studies of similarity judgments or similarity-related tasks (Walden & Montgomery 1975, Broecke 1976, Vitz & Winkler 1972, Greenberg & Jenkins 1964), which support the claim that a voicing change is perceived less readily than a change in nasality. Note that, while concerned with word-edges, much of this work is concerned with word-final than initial consonants or consonant clusters. More detail will be provided below under hypothesis 7.

Hypothesis 1 comparisons		
voice	*nasality*	*environments (e)*
v-f	v-m	_l,r
b-p	b-m	_l,r
z-s	z-n	_r
d-t	d-n	_r

The graph in figure 5.2, like the following ones in this chapter, shows the means of English and Russian listeners' judgments for all comparisons listed. This means that the evaluations of all Russian listeners of all tokens with the same initial sequence have been used to calculate this average value for Russian. The y-axis shows perceived similarity (higher numbers mean a higher perceived similarity) on a normalised scale. For this reason, the numerical values are of no consequence - what is relevant are the values relative to each other. The x-axis shows the different L1s of the participants.

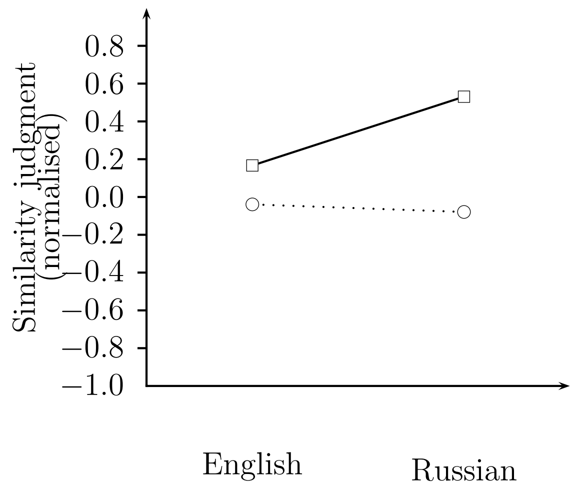

Figure 5.2: Similarity judgments of voicing (solid) and nasality (dotted) changes

Figure 5.2 shows that, for both language groups, the mean similarity values are higher for voicing changes than for nasality changes, and that the difference is smaller for the English than the Russian group.

As detailed in the first part of table 5.2, a mixed ANOVA (i.e. testing within-group and between-group factors), with L1 as a between-group factor and the type of change as a within-group factor, showed a main effect of type of change, as well as for the native language and the interaction between them: Russian listeners assigned higher similarity, and especially so for voicing. This was unexpected: since voice onset time is different in English and Russian, as explained in chapter 2, the expected L1 effect was that English listeners would be more likely to confuse voicing in Russian stops and therefore give higher similarity judgements for those, resulting in higher overall ratings by English listeners.

To test whether the L1 effect is different depending on clear acoustic differences between Russian and English, the data were divided into stop and fricative sets, expecting a relatively higher similarity values from English listeners for stops. The descriptive statistics for this hypothesis are given in table 5.3. They show that voicing changes are rated higher than nasality changes in both sets, but that the

Factor / Interaction	F(df)	p
V/N**	F(1,389)=34.60	<.01
L1**	F(1,389)=6.11	.01
V/N x L1**	F(1,389)=8.45	<.01
V/N(stop)**	F(1, 201)=11.84	<.01
L1(stop)	F(1,201)=0.03	.88
V/N x L1(stop)*	F(1,201)=5.21	.03
V/N(fricative)**	F(1,186)=25.39	<.01
L1(fricative)**	F(1,186)=16.75	<.01
V/N x L1(fricative)	F(1,186)=3.13	.08

Table 5.2: The statistics for the factors type of change (voicing or nasality change, V/N), L1 and interactions. In this as in the following statistics tables, significance is marked with * at p<.05, and with ** at p<.01.

difference between them in English is very small compared to the Russian group for the fricatives.

Data	Change	L1	Mean	Std. Dev.	N
all	nas	E	-.04	.93	204
		R	-.08	.86	187
		total	-.06	.90	391
	voi	E	.17	1.07	204
		R	.53	.88	187
		total	.34	1.00	391
stop	nas	E	.11	.93	108
		R	-.14	.85	95
		total	.00	.90	203
	voi	E	.23	1.17	108
		R	.45	1.01	95
		total	.33	1.10	203
fricative	nas	E	-.21	.91	96
		R	-.02	.87	92
		total	-.12	.89	108
	voi	E	.09	.96	96
		R	.61	.73	92
		total	.35	.89	188

Table 5.3: The descriptive statistics for voice vs. nasality change. In this as in the following tables, E stands for mean values of English participants' judgments, R for those of Russian participants.

In the statistical analysis for both sets (table 5.2), there was still the predicted main effect of change type; for stops,an interaction effect was found, but no L1 effect. The reverse was true, however, for the fricatives, where no interaction was found, but the L1 effect of Russian higher than English judgments.

Steriade's hypothesis that a voicing change is perceived to be more similar than a nasality change is clearly supported through the main effect both overall and for the subgroups. However, the interactions and L1 main effects are not expected under this hypothesis. On the other hand, if one assumes L1 influence on this judgment, English speakers would be expected to give higher similarity values, firstly due to English-illegal clusters being part of the comparison. Secondly, they are predicted to rate especially stop voicing differences higher, i.e. more similar, than Russian speakers, due to Russian voiceless stops lacking aspiration. Instead, Russian listeners judge voicing changes to be more similar than their English counterparts, which may be caused by the commonness of voicing alternations in Russian (see also hypothesis 7).

In terms of manner, no higher English perceived similarity can be observed for stops than for fricatives. A look at the comparisons used here shows that for stops, both [± voice] words of each pair are phonotactically legal in English (bl-pl, br-pr, dr-tr), in contrast to all other comparisons, which contain at least one illegal cluster for each comparison, which may have improved English listeners' discrimination. This provides support for the view that the legality of a sequence influences its perception. Nevertheless, a further point to note is that for Russian listeners, a voicing change appears to more similar in fricatives, a trend which cannot be found for English listeners. This may point to the realisational differences between English and Russian indeed affecting their similarity judgments.

Overall, while the higher similarity of the voicing change, as predicted by Steriade, has been shown, there is an unpredicted influence of L1 on the judgments, which may be caused either by an L1 phonological process, or the legality of a cluster in L1.

- Hypothesis 2: $POAchange \gg [\pm nasality]$

The similarity studies cited for hypothesis 1 provide evidence not only for the comparison between voicing and nasality, but also that any stricture differences, i.e. most changes in manner, are perceived more readily than changes not only in voicing, but also in point of articulation (POA). In addition, Steriade provides phonological support for this claim: one example is the Latin manifestation of the widely attested repair of ill-formed consonant combinations by place assimilation, and thus POA change of one of the consonants (such as ad-kelerare ≫ at-kelerare, [akkelerare]), rather than equally possible stricture changes such as [as-kelerare].

Hypothesis 2 comparisons		
poa	*nasality*	*e*
d-b	b-m	_l,r
b-d	d-n	_r
g-d	d-n	z_

Figure 5.3 shows the means for both types of changes for both language groups. English ratings are higher for both types of changes. Place of articulation changes are rated higher according to prediction; however, the difference is small.

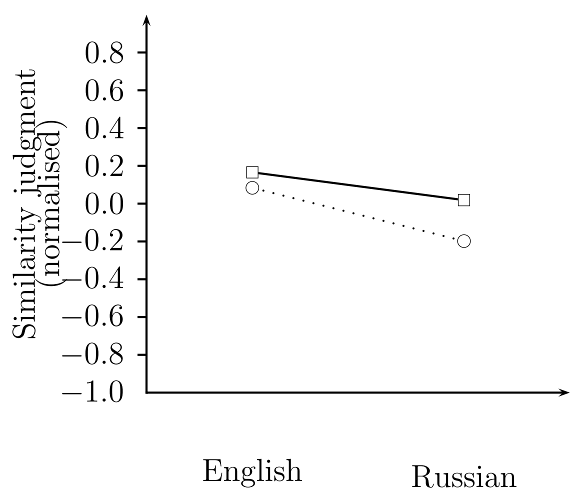

Figure 5.3: Similarity judgments of place of articulation (solid) and nasality (dotted) changes

To test the hypothesis, again a mixed ANOVA was calculated, with language as between-group and type of change as within-group factor (row 1 of table 5.4) and perceived similarity as the dependent variable. There is a main effect of of language, showing that English listeners are less sensitive to both manner and place of articulation changes. A possible independent reason for this effect applying in this experiment is the different articulation of English and Russian coronals: while English coronals are alveolar, Russian coronals are dental. All place of articulation changes used here are either labial-coronal or coronal-velar, i.e. involve coronal sounds, so the dental spectral characteristics are likely to have impaired this change in the English listeners' perception. The predicted main effect for a higher similarity of place of articulation changes over a nasality change was not found, however there was a trend in this direction.

Factor / Interaction	F(df)	p
P/N(*)	F(1,270)=3.54	.06
L1**	F(1,270)=6.95	.01
P/N x L1	F(1,270)=0.72	.40
P/N(C_1)	F(1,202)=2.67	.10
L1(C_1)	F(1,202)=0.50	.48
P/N x L1(C_1)(*)	F(1,202)=3.52	.06
P/N(C_2)	F(1,66)=0.95	.33
L1(C_2)**	F(1,66)=17.52	<.01
P/N x L1(C_2)(*)	F(1,66)=3.85	.05

Table 5.4: The statistics for the factors type of change (place of articulation or nasality change, P/N), L1 and interactions.

Since the acoustic cues to place of articulation in stops are better in pre-vocalic than pre-consonantal position, the similarity values for changing place of articulation are expected to be higher for C_1, resulting in a larger difference between the two types of changes for C_1. However, when the data are divided into the first and second element of the cluster, the positional comparison shows (table 5.5) that C_1 position shows higher values for both types of change. The comparison shows further that the prime source for the difference between the language groups is to be found in the C_2 data, where *all* clusters are illegal for English, while for C_1 the discrepancy is much smaller.

The statistical analysis (rows 2/3 of table 5.4) supports this split. The main effect for L1 is only significant for C_2. This can be taken to suggest the relevance of the status of clusters in this respect for perceiving similarity. Note, however, that the C_1 vs. C_2 division equals the labial-coronal vs. velar-coronal division in this set, so it cannot be pinpointed whether position or L1 are responsible for the difference.

data	Change	L1	Mean	Std. Dev.	N
all	POA	E	.17	.91	144
		R	.02	1.06	128
		total	.10	.99	272
	nas	E	.09	.91	144
		R	-.29	.86	128
		total	-.05	.89	272
C_1	POA	E	.09	.97	108
		R	.20	1.02	96
		total	.14	.99	204
	nas	E	.11	.93	108
		R	-.13	.85	96
		total	.00	.90	204
C_2	POA	E	.39	.91	36
		R	-.54	.87	32
		total	-.04	.89	68
	nas	E	.00	.84	36
		R	-.40	.88	32
		total	-.19	.88	68

Table 5.5: The descriptive statistics for place of articulation (POA) vs. nasality change

While there is a near-significant effect of the type of change, as predicted, the stronger influence on the judgment of the changes here is the participants' native language, either through phonetic detail or phonotactic status: English listeners assign higher similarity when unfamiliar phonetic detail or ill-formed clusters are involved in a comparison.

- Hypothesis 3: $[\pm voice(C_1) \gg [\pm voice(C_2)]$

C_1 and C_2 in this case are consonants in a #_C and a C_V environment, respectively. Since any cues to voicing are known to be more salient in the latter context than in the former, similarity judgments - if influenced by cue availability - should reflect the fact that the difference between a voiced and unvoiced segment are going to be smaller in the former than the latter environment. On the other hand, if context does not influence similarity judgments, as advocated for example by Frisch, no difference between the two positions should be found.

Hypothesis 3 comparisons		
voice	$e_1 : C_1$	$e_2 : C_2$
d-t	_r	r_

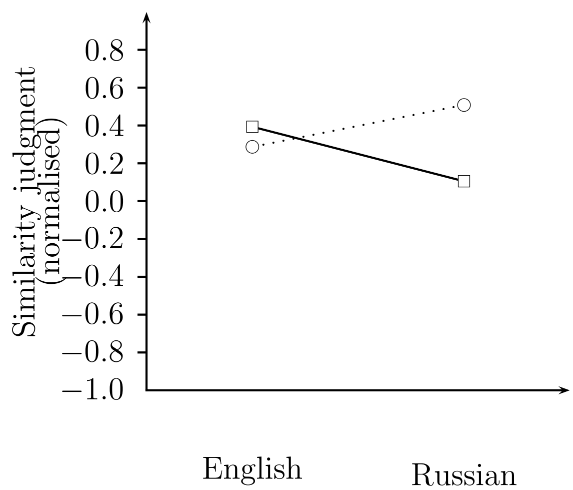

Figure 5.4: Similarity judgments of C_1 (solid) and C_2 (dotted) voicing changes

Figure 5.4 and table 5.6 compare the means for both changes, in both language groups. While the means for both changes hardly differ in the English group (0.1), there is a greater discrepancy for the Russian listeners (0.4). Contrary to the prediction, they perceive the change to be more similar in C_2 position.

position	L1	Mean	Std. Dev.	N
C_1	E	.39	1.25	36
	R	.11	1.18	32
	total	.26	1.21	68
C_2	E	.29	1.29	36
	R	.51	1.06	32
	total	.39	1.19	68

Table 5.6: The descriptive statistics for voicing change depending on position.

Table 5.7 demonstrates that in the mixed ANOVA no significant effects were found. Even for Russian listeners, for whom all clusters are legal, there is no indication that the judgments conform to the hypothesis. This outcome may have been influenced by the fact that, due to regressive voicing assimilation triggered by obstruents in Russian (as described in chapter 2), it was only possible to find a single onset pairing for this comparison. This pairing involves voicing change in plosives, where in one case two legal clusters are compared, in the other two illegal clusters. Consequently, due to the small sample size, the power of the statistics is also reduced.

Factor / Interaction	F(df)	p
position	F(1,66)=0.63	.43
L1	F(1,66)=1.85	.18
position x L1	F(1,66)=0.02	.88

Table 5.7: The statistics for the factors position of voicing change, L1 and interactions.

The results neither support the position dependency claimed in the hypothesis, nor do they indicate any influence of the native language; however, they may have been compromised by the small sample size.

- Hypothesis 4: POA change: $\_C[-cor] \gg \_C[+cor] \gg \_V$

This proposal is derived from Jun (1995): he compares the gestures for coronal and non-coronal consonants and concludes that, given that coronal gestures are shorter and faster than non-coronals, the latter kind are more prone to obscuring the gesture of a preceding consonant. This leads him to derive a higher salience of, for example, place cues before coronal consonants. These cues should be most obvious if there are added transitional cues, i.e. if the segment in question is followed by a vowel rather than a consonant. Overall, this results in the prediction that similarity judgments will be highest for a place of articulation (POA) change before a non-coronal consonant, intermediate before a coronal consonant, and lowest before a vowel.

Hypothesis 4 comparisons				
POA change	e_1: _C[-cor]	e_2: _C[+cor]	e_3: C[-cor]_V	e_4: C[+cor]_V
d-g	_v	_z	v_	z_
v-z	_g	_d	g_	d_
f-s	_k	_t	k_	
f-ʃ	_k	_t	k_	
s-ʃ	_k	_t	k_	

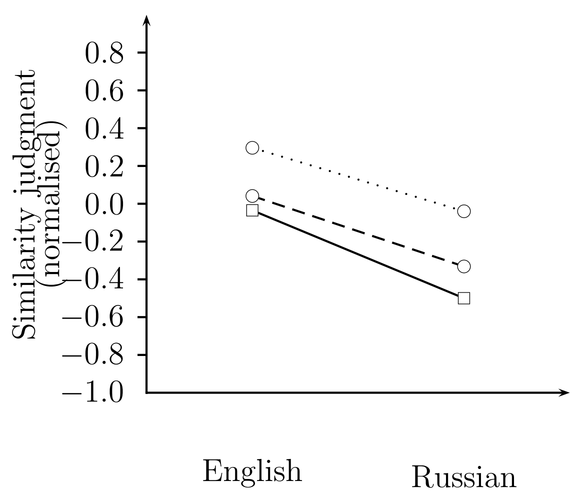

Figure 5.5: Similarity judgments of place of articulation changes before non-coronals (solid), coronals (dotted) and vowels (dashed)

Figure 5.5 shows that, for all of these changes in place of articulation, the English group's mean values are higher than the Russian group's means. Furthermore, the order of the types of changes is the same for both language groups, and the similarity values are highest in the non-coronal context. There is no difference between the two vowel contexts for the English group, but for the Russian group similarity with a preceding coronal consonant is noticeably lower.

Data	Change	L1	Mean	Std. Dev.	N
4 levels	C_1, [-cor]	E	.19	.90	72
		R	-.30	.91	62
		total	-.04	.94	134
	C_1, [+cor]	E	.47	.93	72
		R	.08	1.03	62
		total	.29	.99	134
	C_2, [-cor]	E	.06	1.00	72
		R	-.62	1.01	62
		total	-.26	1.06	134
	C_2, [+cor]	E	.04	.85	72
		R	-.90	1.06	62
		total	-.39	1.06	134
3 levels	C_1, [-cor]	E	.13	.95	180
		R	-.19	.94	157
		total	-.02	.96	337
	C_1, [+cor]	E	.31	.95	180
		R	.02	.93	157
		total	.17	.95	337
	C_2, [-cor]	E	-.10	1.03	180
		R	-.39	.99	157
		total	-.24	1.02	337

Table 5.8: The descriptive statistics for place of articulation change depending on position and coronality of neighbouring segments; C_1 stands for the changed consonant being in initial position, *[+cor]* for the preceding or following consonant being coronal.

Table 5.8 gives the descriptives for different subsets of the data: one, the comparison of all four levels to include a test for the influence of coronality of a preceding consonant; two, the comparison of three levels only, with the aim of achieving a greater sample size. The mean values for both subsets are similar with highest values for coronal context, and higher values for English listeners (as in hypothesis 2).

Significance was calculated for all four levels as well as for a three-level distinction (see table 5.9). The results were similar. The mixed ANOVA with context as the between-group factor shows a clear main effect, albeit this effect is not as predicted: the contrast is weakest, i.e. similarity is perceived to be highest, in front of coronal

Factor / Interaction	F(df)	p
context(4)**	F(3,132)=14.66	<.01
L1(4)**	F(1,132)=45.31	<.01
context x L1(4)	F(3,132)=2.21	.09
context(3)**	F(2,335)=16.66	<.01
L1(3)**	F(1,335)=20.68	<.01
context x L1(3)	F(2,335)=0.03	.97

Table 5.9: The statistics for the factors context (position and coronality of neighbour), L1 and interactions, for three and four levels of the variable.

consonants for both language groups, with figure 5.5 showing that the contrast for English listeners is somewhat smaller than for Russian listeners.

The observation of consistently higher means in the English group is confirmed by a main effect of L1. This may be triggered by native phonotactics, since the bulk of the clusters involved are illegal in English. Another possible reason is again the influence of the difference in the precise point of articulation between English and Russian coronals, i.e. the greater confusion of English listeners if coronals with unexpected formants are part of the comparison they are to assess.

Overall, these results fail to provide much support for Jun's hypothesis or the P-map in general, and once again the native language is shown to influence the perception of similarity, in the form of native phonotactics or phonetic detail.

- Hypothesis 5: Place of articulation change: *stop* ≫ *fricative*

Also derived from Jun (1995), this proposal claims that place of articulation changes in fricatives are more perceptible than in plosives; this is due to the cues to place features for plosives being mostly located in the transitions, whereas for fricatives there are stronger segment-internal cues. In C_V contexts, i.e. in C_2 position, the vocalic environment makes the plosive transition cues available, whereas in #_C contexts this is not the case. Therefore, the perceived similarity of two clusters with a place of articulation change should be higher for stops, especially in pre-consonantal position.

Hypothesis 5 comparisons		
POA (S)	*POA (F)*	*e*
p-t	f-s	_r
b-d	v-z	_l,r
p-t	f-s	k_
m-n	f-s	k_

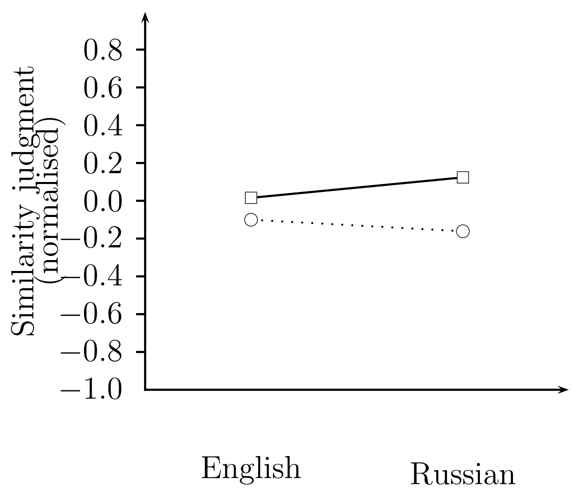

Figure 5.6: Similarity judgments of POA changes in stops (solid) and fricatives (dotted)

Figure 5.6 confirms that overall, i.e. for both positions, the similarity judgments for plosives are higher than those for fricatives, while there is little difference between the English and Russian groups. In the statistical test, The mixed ANOVA analysis reveals a main effect for manner in the predicted way (F(1,338)=7.29, p<.01). No main effect of L1 was found (F(1,338)=0.10, p=.75). Neither was there a significant interaction between the factors (F(1,338)=1.31, p=.25).

Data	manner	L1	Mean	Std. Dev.	N
all	stop	E	.15	.90	180
		R	.12	.93	160
		total	.07	.91	340
	fricative	E	-.10	1.05	180
		R	-.16	1.01	160
		total	-.13	1.03	340
C_1	stop	E	.16	.94	108
		R	.26	.97	96
		total	.21	.95	204
	fricative	E	-.02	.96	108
		R	-.11	1.05	96
		total	-.06	.99	204
C_2	stop	E	-.20	.81	72
		R	-.08	.84	64
		total	-.15	.82	136
	fricative	E	-.22	1.17	72
		R	-.24	.95	64
		total	-.23	1.07	136

Table 5.10: The descriptive statistics for place of articulation change depending on manner.

After the division of the dataset into C_1 and C_2 position, the similarity values in table 5.10 show that, firstly, the prevocalic environment does lead to lower similarity values, for both stops and fricatives. They also support the predictions that similarity values are higher for stops, and that the difference is greater in C_1 position. This is true overall as well as for both language groups separately.

Factor / Interaction	F(df)	p
manner(C_1)**	F(1,202)=8.34	<.01
L1(C_1)	F(1,202)<0.01	.96
manner x L1(C_1)	F(3,202)=0.98	.33
manner(C_2)	F(1,134)=0.56	.46
L1(C_2)	F(1,134)=0.20	.66
manner x L1(C_2)	F(1,134)=0.36	.55

Table 5.11: The statistics for the factors manner, L1 and interactions, split by position.

With the dataset divided in by position, the analysis showed that the manner effect remained only in the C_1 set, while it disappeared for the prevocalic set (see table 5.11). No further L1 effects or interaction effects could be found.

Thus all results gained for this hypothesis support the universal cue model: the hypothesis is confirmed, independently of language background.

- Hypothesis 6: [$\pm$*anterior*] $\gg$ [$\pm$*strident*]

One of Steriade's claims is that a difference in whether or not a consonant is changed with respect to anteriority is more similar than a change in stridency. She cites studies concerning the confusability (Miller & Nicely 1955) and perceived similarity Walden & Montgomery (1975) of the pairs [f]-[θ], [θ]-[s] and [f]-[s], which find that pairs distinguished by stridency such as [f]-[s] are less confusable and less similar, and thus that stridency is inherently more salient than other features. Since the Russian inventory lacks the dental fricatives, the comparisons used here were between [f], [s] and [ʃ], where [f]-[s] is differentiated by stridency, [s]-[ʃ] by anteriority, and [f]-[ʃ] by both. One debatable issue is the question of the stridency of /f/, and I am following Steriade here in regarding it as [-strident]. A further problem to keep in mind is the fact that a change in stridency always involves a change in POA also, although it is not necessary to *invoke* the place feature to establish the phonological contrast if stridency is used. In this case, any alleged effects of stridency in judgment may also be caused (or enhanced) by a difference between the salience of an alveolar/labiodental and an alveolar/post-alveolar change. This possibility could not be otherwise tested in this experiment, since there are no combinations with this place of articulation change only (i.e. without the stridency change) available in Russian.

Hypothesis 6 comparisons			
$\pm$*ant*	$\pm$*str*	$\pm$*ant&str*	*e*
s-ʃ	f-s	f-ʃ	ɹ,r,t,k

As can be seen from figure 5.7 and table 5.12, English listeners' judgments are higher than those of the Russian group. The change of anteriority only is judged to be the most similar by both groups. However, there is a difference in the perception of the other changes: while the Russian group judges them to be almost the same, for the English group the stridency-only change is more similar.

The mixed ANOVA found a main effect in accord with the prediction for anteriority/stridency in the predicted order. Additionally, a main L1 effect was found such that English participants rated these place of articulation changes to be significantly more similar than their Russian counterparts did. No significant interaction was found (see table 5.13).

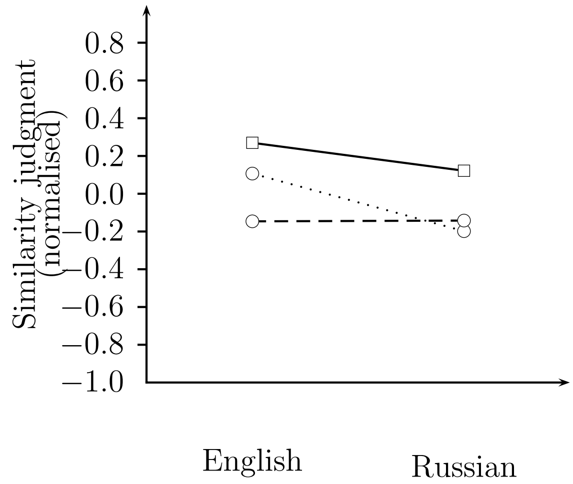

Figure 5.7: Similarity judgments of changes in anteriority (solid), stridency (dotted) and both (dashed)

A/S	L1	Mean	Std. Dev.	N
[±ant]	E	.27	1.03	132
	R	.12	.88	122
	total	.20	.96	254
[±str]	E	.10	.97	132
	R	-.20	1.09	122
	total	-.04	1.04	254
[±ant],[±str]	E	-.15	.92	132
	R	-.14	.90	122
	total	-.15	.91	254

Table 5.12: The descriptive statistics for anteriority and stridency changes (A/S).

Factor / Interaction	F(df)	p
change**	F(2,252)=8.40	<.01
L1*	F(1,252)=4.45	.04
change x L1	F(1,252)=1.64	.19

Table 5.13: The statistics for the factors change type, with the three levels anteriority change only, stridency change only, and both.

Pairwise comparisons showed a significant difference only between anteriority change and stridency (p<.01) as well as the combination (p<.01), but not between the two comparisons where stridency was changed in both comparanda (p=.25).

Notably, the L1 effect only shows where [s] is involved (note the flat line for the /f-ʃ/ comparison in figure 5.7), which differs most in its spectral specifications from the English sound, and not the [f]-[ʃ] comparison. Since uneven distribution of legality cannot be responsible in this case, the result supports the idea that this difference has impaired English listeners' discrimination abilities, as suggested above, in that Russian [s] is a dental rather than an alveolar coronal.

The results for this hypothesis support Steriade's prediction, as well as suggesting that L1 specific phonetics influences perception.

- Hypothesis 7: $[\pm voice] \gg \pm C \gg \pm V$

The similarity of a voicing change being higher than that of adding or deleting a segment is a further hypothesis taken from the P-map and its claim of similarity being based on cue availability.

In Steriade's P-map, the similarity of a voicing change is compared to other possibilities of solving a grammatical conflict, such as a prohibition of voiced final obstruents. Two of these solutions are the deletion of the offending consonant, and the insertion of a final vowel and thus moving the voiced obstruent to non-final position. Like a nasality change, according to Steriade these solutions are, in contrast to devoicing, not attested.

In a study mirroring the conditions in this thesis, Wingstedt & Schulman (1988) compared modifications of word-*final* and hence also word-marginal consonant clusters, and found a preference for deletion of the final C, i.e. the one between a C and a word boundary, followed by epenthesis between the consonants and then deletion of the first, and thus vowel-neighbouring, C of the coda. Note, however, that these were judgments of preference rather than of perceived similarity.

Further support for the suggested ranking of similarity comes from Fleischhacker (1999) who collected English listeners' judgments about the similarity of English target words and possible modifications. Like Wingstedt and Schulman, Fleischhacker used word-final clusters in her study, which revealed a higher ranking for voicing changes of the final segment than other alterations (e.g. [prɪnt] - [prɪnd] ≫ [prɪnt] - [prɪn]) and a higher raking of final C deletion than schwa insertion ([hɛft] - [hɛf] ≫ [hɛft] - [hɛftə]). The fact that these results were true for both similarity and preference ratings may add relevance in terms of perceived similarity to Wingstedt and Schulman's results. On the other hand, it must be noted that the vowel epenthesis here occurs after a word-final cluster, rather than between the consonants, and the ratings may be influenced by whether or not the final consonant was released.

In a further paper cited by Steriade, Magen (1998) analysed English listeners' attitudes to Spanish-accented English, and found that voicing differences were judged to be less grave than vowel insertions (e.g. *closed*: /klosəd/) and deletions of final sibilants, again in word-final clusters. The consonant deletions were, however, of morphological relevance, which might have influenced the results.

Overall, these studies support the view that voicing changes are less salient than either vowel (schwa) epenthesis or consonant deletion, yet the relationship between the latter two remains somewhat unclear, especially since the perception of C deletion is also hypothesised to be position- and manner-dependent. Both in Fleischhacker's and Wingstedt and Schulman's results word-marginal consonant deletion was preferred over schwa epenthesis, which was in turn preferred over deletion of the vowel-adjacent consonant. Steriade (2001)'s hierarchy of distinctiveness ranks schwa epenthesis higher than consonant deletion in V_# context.

Hypothesis 7 comparisons					
±voi	*epenthesis (ep) [+voi]*	*ep [-voi]*	*deletion (del) [+voi]*	*del [-voi]*	*e*
b-p	b-bə	p-pə	b-∅	p-∅	_r
d-t	d-də	t-tə	d-∅	t-∅	_r
g-k	g-gə	k-kə	g-∅	k-∅	_r
v-f	v-və	f-fə	v-∅	f-∅	_r
z-s	z-zə	s-sə	z-∅	s-∅	_r
d-t	d-əd	t-ət	d-∅	t-∅	r_
v-f	v-əv	f-əf	v-∅	f-∅	s_

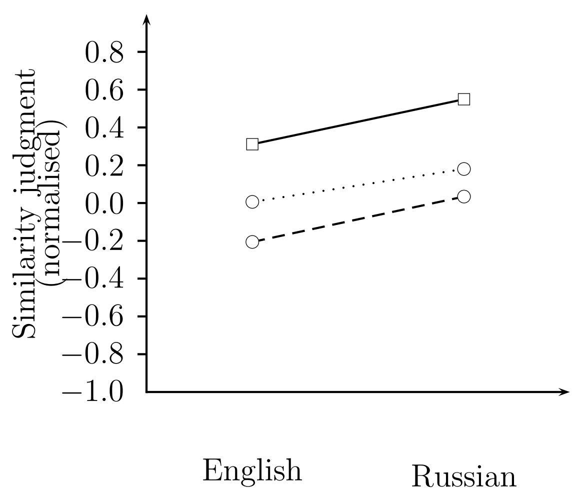

Figure 5.8: Similarity judgments of voicing change (solid), C deletion (dashed) and V epenthesis (dotted). For ease of comparison, only the values for epenthesis and deletion [-voice] are shown.

As figure 5.8 demonstrates, the Russian similarity values are higher than those from the English group for each type of change. Both language rank the types of change in the same order, with the voicing change as the most similar, and the presence or absence of a consonant the least similar.

Change	L1	Mean	Std. Dev.	N
voicing	E	.31	1.16	250
	R	.55	.92	223
	total	.42	1.06	473
ep[+voi]	E	.00	.90	250
	R	.18	.93	223
	total	.09	.92	473
ep[-voi]	E	.01	.95	250
	R	.18	.92	223
	total	.09	.94	473
del[+voi]	E	-.11	.95	250
	R	-.09	.97	223
	total	-.10	.96	473
del[-voi]	E	-.21	.89	250
	R	.04	.84	223
	total	-.09	.88	473

Table 5.14: The descriptive statistics for voicing, epenthesis and deletion.

For completeness, table 5.14 gives the similarity means for the voicing change, as well as for consonant deletion and vowel epenthesis of both the voiced and the voiceless version of the cluster. The means for both epenthesis groups are higher than the deletion means. This is the case overall, as well as for both the C_1 and C_2 subsets.

Factor / Interaction	F(df)	p
change**	F(4,471)=23.79	<.01
L1**	F(1,471)=19.25	<.01
change x L1	F(4,471)=0.96	.43

Table 5.15: The statistics for the factor change, with the levels voicing change, epenthesis and deletion, with the latter two split into the relevant consonant being voiced or not.

A mixed ANOVA test was run with the within-group factor *type of change* with all five levels. The results can be seen in table 5.15. The consistently higher similarity values from the Russian group are confirmed by a significant main effect of L1. Also, a clear main effect of change type was found, repeating the English listeners' judgments from the previous chapter.

Pairwise comparisons show that all levels are significantly different at p<.01, except for the two levels of epenthesis (p=1.00) and deletion (p=.82), showing that voicing in a cluster does not influence the perception of the similarity to a CəC or C sequence.

A voicing change is perceived to be the least drastic type of change, followed by vowel epenthesis and deletion. This means that the predicted order of vowel epenthesis and consonant deletion is reversed, also coinciding with the results for English speakers found in the previous chapter. It must be noted again, however, that there is of course a wide range of consonants and vowels that can be added or deleted, and that in this case the vowel was schwa without exception, which acoustically makes a smaller difference than a full vowel. The relevant deleted consonants here are always obstruents.

In reference to Wingstedt and Schulman's results, no difference between the first and second consonant of the cluster being affected was found. Of course, aside from the /d-t/ pair before and after /r/, the C_1 and C_2 conditions could not be matched, so the results are less meaningful in this respect than they would otherwise be. For a direct comparison, see hypothesis 9.

There is also a main effect of L1, i.e. that Russian participants perceive all these changes to be more similar than English participants, confirming the L1 effect shown for hypothesis 1 for voicing changes.

As described above, the illegality of the clusters and consequent confusion of English listeners between them and the alternative forms would predict the opposite result, i.e. higher similarity values for the English judgments. However, this effect would not be predicted to show very strongly for this set of comparisons, since most of the clusters are not illegal for English speakers. A tentative explanation I am offering for this L1 effect is the high frequency of both voicing alternations (through assimilation) and vowel/zero alternations (for example within declension paradigms) in Russian, which may result in Russian listeners assigning less importance and thus a higher perceived similarity than expected. This depends, however, on whether the proposed diminished sensitivity is general or context-dependent, as it is a necessary factor of the contexts used here that they do *not* trigger assimilation.

Once again, the results for this hypothesis are mixed: while there is strong support for Steriade's cue-based hypothesis demonstrated by a main effect of type of change, there is also a clear role for L1, possibly through common phonological processes in the native language.

- Hypothesis 8: Manner and position influence deletion ratings

This claim is based on the idea that in terms of insertions or deletions, the presence or absence of a sound that is very similar to its neighbourhood or, if it is at the edge of a word, to zero is going to be the least conspicuous. For example, Steriade (2001) attributes liquid vocalisation or loss in postvocalic position to the greater similarity of liquids to the preceding vowels, compared to lower sonority consonants. While this argument sounds somewhat circular by using similarity to determine similarity, in relative terms the situation is often clear: for example, schwa is cross-linguistically preferred as an epenthetic vowel, since, in the words of Steriade (p.46), “it is in both duration, and relative absence of invariant articulatory properties, the closest thing in a vowel system to no segment at all, i.e. to zero”.

In terms of consonant deletion, both manner or sonority and position are therefore predicted to play a role: while at the left edge of the word, i.e. in C_1 position, one of the neighbours of the segment in question is silence, lower sonority consonant are predicted to be less conspicuous when deleted. On the other hand, in C_2 position the relevant consonant is pre-vocalic, and hence higher sonority resulting in a higher similarity to a vowel is predicted to make its absence more noticeable.

Therefore, the order of similarity values for deletion in initial position is expected to be: plosive ≫ fricative ≫ nasal ≫ liquid. In second position, the order is predicted to be reversed.

Hypothesis 8 comparisons				
stop	*fricative*	*nasal*	*liquid*	*env*
b-∅	v-∅	m-∅		_r
d-∅	z-∅	n-∅		_r
p-∅	f-∅			_r
t-∅	s-∅			_r
p-∅	f-∅			k_
		n-∅	l-∅	m_
b-∅		m-∅		v_
d-∅	z-∅	n-∅	l-∅	v_
d-∅		n-∅	l-∅	z_

Since, unfortunately, the set of available comparisons was somewhat patchy, the descriptive statistics (table 5.16) are given of subsets of the data, to get comparative values for all levels of manner. In initial position, the similarity judgments are not in

Data	Manner	L1	Mean	Std. Dev.	N
C_1	stop	E	-.11	1.00	72
		R	-.14	.83	64
		total	-.12	.92	136
	fricative	E	.00	.97	72
		R	.13	1.06	64
		total	.06	1.01	136
	nasal	E	.07	.73	72
		R	.13	.78	64
		total	.10	.75	136
C_2	stop	E	-.19	.92	36
		R	-.70	1.06	31
		total	-.43	1.01	67
	fricative	E	-.10	.99	36
		R	-.69	1.16	31
		total	-.37	1.10	67
	nasal	E	.14	.80	36
		R	-.06	.86	31
		total	.0	.83	67
	liquid	E	-.07	.85	36
		R	.09	.83	31
		total	.00	.84	67

Table 5.16: The descriptive statistics for consonant deletion, by manner and position. These are values for those subsets of the comparisons where each level is represented, so that values of matching data for each condition can be presented: in the C_1 subset, the statistics are given for the first two comparisons, in the C_2 subset for the second but last comparison.

the predicted order, and no clear difference between the language groups is evident. For C_2 position, the mean similarity values conform better to the expectations: stops and fricatives have the lowest similarity values for both English and Russian listeners.

The statistical analysis of these sets (table 5.17, rows 1 and 3) gives no significant effects at all for initial position. In second position, there is a significant main effect of L1 such that the Russian groups similarity values are lower. Additionally, there is a significant effect of manner, with obstruent deletions judged to be less similar than sonorant deletions, as predicted.

As the dataset covering all manner levels resulted in a small sample size, further tests were run with larger datasets, comparing fewer manner levels. For example, row 2 of table 5.17 shows the results for a comparison of stop vs. fricative (in

Factor / Interaction	F(df)	p
S/F/N(C_1)	F(2,134)=2.60	.08
L1	F(1,134)=0.29	.59
S/F/N(C_1) x L1	F(2,134)=0.33	.72
S/F(C_1)(*)	F(1,271)=3.58	.06
L1*	F(1,271)=4.49	.04
S/F(C_1) x L1	F(1,271)=1.58	.21
S/F/N/L(C_2)**	F(3,65)=5.49	<.01
L1*	F(1,65)=4.95	.03
S/F/N/L(C_2) x L1	F(3,65)=2.37	.08
S/F(C_2)	F(1,133)=1.03	.31
L1**	F(1,133)=12.87	<.01
S/F(C_2) x L1	F(1,133)=0.18	.68
S/N(C_2)**	F(1,199)=8.06	<.01
L1*	F(1,199)=5.88	.01
S/N(C_2) x L1	F(1,199)=1.14	.29
S/L(C_2)*	F(1,134)=6.54	.01
L1	F(1,134)=1.95	.16
S/L(C_2) x L1*	F(1,136)=5.10	.03
F/N(C_2)*		.05
F/L(C_2)		.09
N/L(C_2)	F(1,202)=2.11	.15
L1	F(1,202)=0.94	.33
N/L(C_2) x L1	F(1,202)<.01	.96

Table 5.17: The statistics for the factors position and manner, the latter with the levels stop (S), fricative (F), nasal (N), liquid (L). For each position, the first data subset is the smallest one, with the most levels of manner. The following values are given for subsets, as indicated in the column *Factor / Interaction*. The two comparisons where only p is listed are those where a test with a larger N than in the first test was impossible. The p values are therefore taken from the pairwise comparisons of that test.

initial position) only. A significant L1 effect with higher Russian values, and a near-significant manner effect not in the predicted order, were found.

For the prevocalic position, more tests on larger samples were possible. No consistency is observed in terms of the relevance of L1 background, although in some cases the English similarity values are significantly higher; this may be caused by the illegality of the clusters involved. However, the predicted manner effect is found. The differentiation does not appear to be as fine-grained as expected, as is evident from the lack of significant differences between stops and fricatives, and between nasals and liquids. However, there are consistent significant effects, dividing the set into obstruents and sonorants: the deletion of a sonorant is perceived to be more similar in this position, as predicted by the cue-model.

The results for this hypothesis do give some support to the cue model. However, the sample sizes are too small to provide certainty regarding both the lack of a manner effect for C_1, and the positionally dependent L1 effects.

- Hypothesis 9: $\pm C_1 \gg \pm C_2$

While the previous hypothesis was concerned with the relevance of manner (and of manner with respect to position) on the perception of the presence or absence of a consonant, hypothesis 9 concentrates on the influence of position only. It is related to the findings of Wingstedt & Schulman (1988) and Fleischhacker (1999), as well as being based on separate acoustic considerations: the consonant at the edge of the word is situated in a #_C environment, and hence has no good transition cues showing its identity (and existence). C_2 on the other hand is followed by a vowel and so shows transition cues on its right edge. This means a deletion of this consonant is predicted to be more salient than that of C_1.

Hypothesis 9 comparisons		
del	e_1	e_2
d-∅	_r	r_
t-∅	_r	r_
p-∅	_s	s_
s-∅	_p	p_
z-∅	_v	v_
v-∅	_z	z_

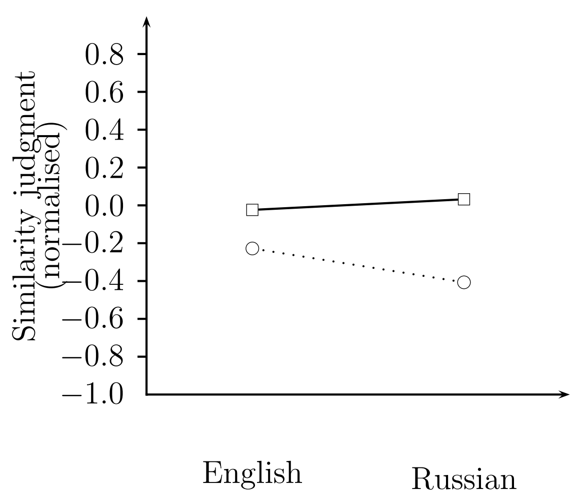

Figure 5.9: Similarity judgments of deletion of C_1 (solid) and C_2 (dotted)

As can be seen in figure 5.9 and table 5.18, the similarity values for initial C are indeed higher than for C_2 for both language groups. The difference is perceived to be slightly larger for Russian than for English listeners.

Position	L1	Mean	Std. Dev.	N
C_1	E	-.02	1.00	215
	R	.03	.93	189
	total	.00	.96	404
C_2	E	-.23	1.04	215
	R	-.41	.99	189
	total	.31	1.02	404

Table 5.18: The descriptive statistics for deletion, depending on position.

The mixed ANOVA confirms that there is a clear difference, as predicted, between a deletion of first and second consonant, as shown in the main effect for position. The deletion of the prevocalic consonant is perceived to be a more severe alteration than that of the word-initial consonant. The difference between English and Russian listeners is not significant: no other effects were observed.

Factor / Interaction	F(df)	p
position**	F(1,402)=23.03	<.01
L1	F(1,402)=0.72	.40
position x L1	F(1,402)=3.05	.08

Table 5.19: The statistics for the factor position, as applied to the deletion of a consonant.

This result comes up very clearly in support of the universal cue model. The hypothesis can be confirmed, and no L1 influence is evident.

Summary for cue-based hypotheses The evaluation for cue- and context-based claims is not straightforward. Most hypotheses from the P-map were confirmed; yet, for a number of hypotheses no support could be found, or the predicted order of similarity judgments was incorrect. Additionally, in most cases effects of the native language were found, which are not predicted under the P-map. For example, the higher similarity values from the Russian group for voicing changes, and the higher values from the English group for place or articulation changes can be attributed to L1 phonology and phonetics.

L1 inventory-based predictions (Hypothesis 10)

- Hypothesis 10: Inventory-derived segment similarity correlates with similarity judgments of English listeners

This hypothesis tests whether inventory-based measures of segment similarity such as the one devised by Frisch (1996) are useful for word similarity predictions. Since Frisch makes no claims about the similarity of any segment to zero, all deletion or epenthesis cases were disregarded. For all remaining cluster to cluster changes, i.e. all pairs where one and only one consonant of the cluster was exchanged for another consonant, it was calculated that word similarity judgments should be proportional to the segment similarity derived from the inventory model, since the clusters were the only difference between word pairs.

The predictions following from this are, firstly, that there should be a significant correlation of Frisch's segment similarity with the English listeners' judgments, but not with the Russian ones, since Russian has a different phoneme inventory. Secondly, the similarity values for comparisons where the same segment is exchanged should cluster together independently of context.

The comparisons used for this hypothesis were all pairs that changed segments rather than deleting or adding them, i.e. cluster to cluster comparisons. These are the 57 items marked as segment changes in appendix F.

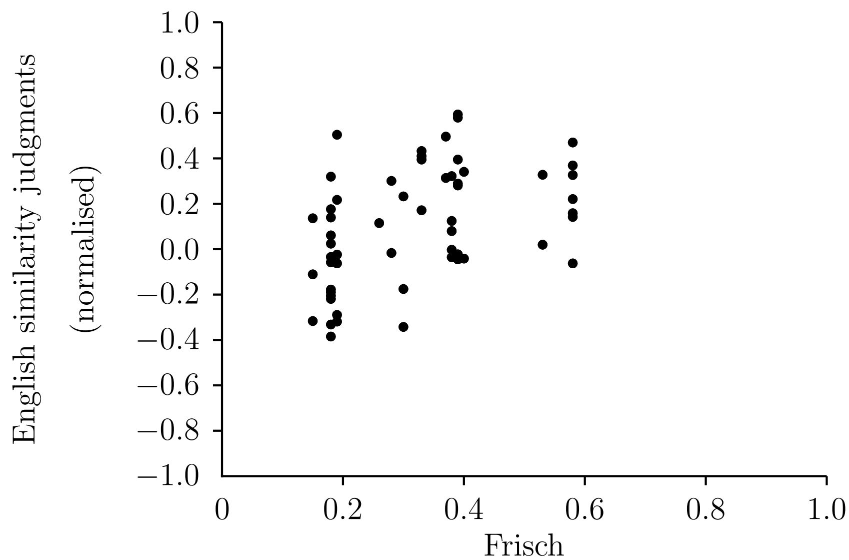

Figure 5.10: Correlation of English listeners' similarity judgments and Frisch-derived predictions

Figure 5.10 suggests that English listeners' similarity judgments are higher for higher segment similarity, as calculated by Frisch. The statistical test shows a significant positive correlation between them was found as predicted (r_p=.45, N=57, p<.01, 95% CI: .22<r< .64). Following from this, combining Frisch's segment similarity to calculate word similarity seems to work for English.

However, the prediction also claimed that there should be no correlation between Frisch's segment similarity and the Russian similarity judgments. The scattergram in figure 5.11 is not as clear as for the English group. However, in contrast to the prediction, the correlation between the segment similarities and the Russian judgments was almost as good (r_p=.40, N=57, p<.01, 95% CI: .15<r<.60) as for the English group.

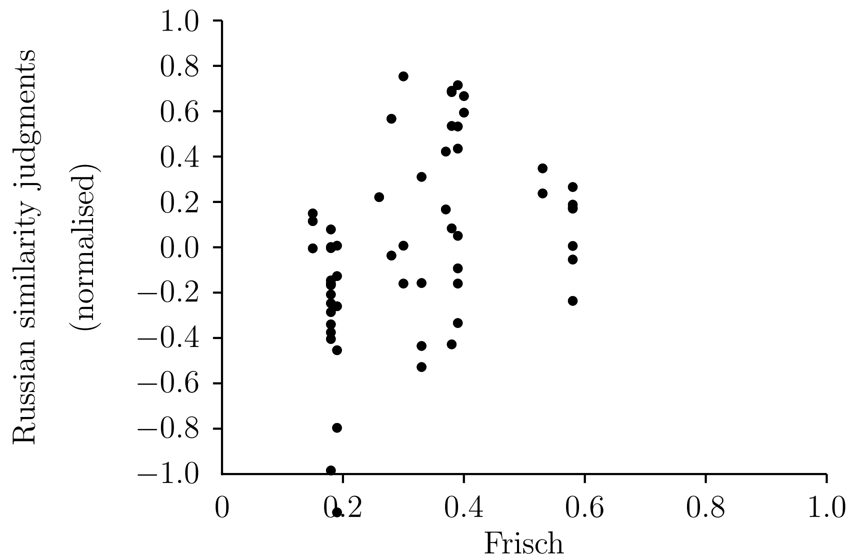

Figure 5.11: Correlation of Russian listeners' similarity judgments and Frisch-derived predictions.

This suggests that Frisch's model may actually reflect a more cross-linguistic pattern in terms of segment similarity. A further argument that speaks against the Frisch-derived approach to L1 dependent word similarity is that, when the results are split up into sets of comparisons such that within each set the same segment is changed, for example /gv-gz/, /vb-zb/, the predicted outcome, that each set should cluster around one similarity value, is not found. Instead, as figure 5.12 shows, there is a wide within-set variety of values[8].

In summary, Frisch's inventory-based segment similarity appears to work well to predict English listeners' similarity judgments. However, a closer look at individual

[8]Thanks to Donca Steriade for discussing this issue with me.

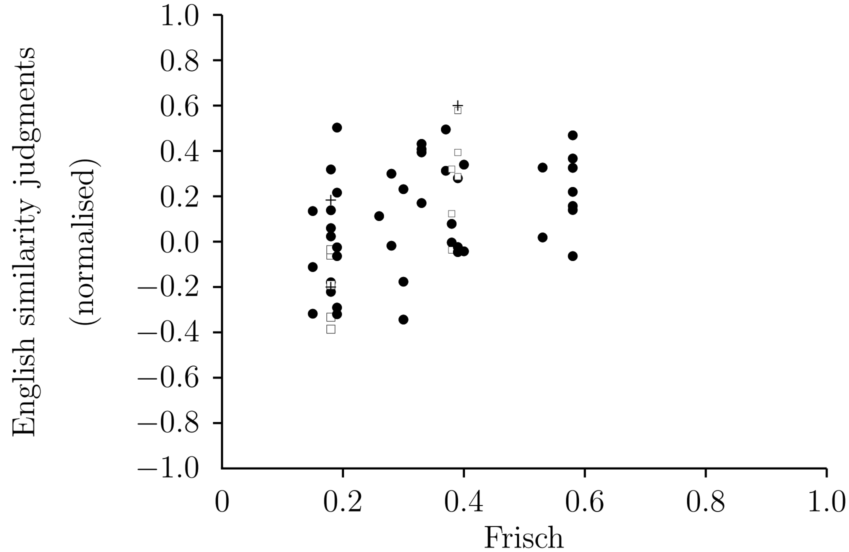

Figure 5.12: Correlation of English listeners' similarity judgments and Frisch-derived predictions. With the Frisch values on the x axis, all comparisons where the same segment is changed lie on the line of the same x value. Where more than one pair of segments has the same Frisch value, the pairs are distinguished by dot style.

clusters shows that the claim of segment similarity being independent of context cannot be supported.

L1 phonotactics-based predictions (Hypotheses 11-12)

Both the hypotheses in this section are based on the idea that the perception of similarity is dependent on whether or not the sequences of sounds compared are well-formed in the listener's native language. The claim is that perception is compromised when phonotactically ill-formed sequences are encountered, and that therefore similarity is perceived to be higher when ill-formed sequences are part of the comparison.

There has been some support for this claim through the phonotactically motivated L1 effects which were detected when testing the cue-based hypothesis. Here, evidence for phonotactics-based predictions can be found through a discrepancy between Russian and English similarity judgments where the clusters involved are illegal in English. The results support this approach partially.

- Hypothesis 11: English and Russian judgments concur, but not in English-illegal cases

This hypothesis is deduced from Dupoux et al.'s (1999) study of French and Japanese performance in categorical perception tasks, which concludes that Japanese vowel epenthesis in illegal clusters has its basis in perception skewed by L1 phonotactics. Following this line of argument, English listeners should be expected to perform similarly to their Russian counterparts unless they encounter English-illegal clusters. In this experiment, in one group of comparisons illegal targets are paired with potential adaptations (as seen in chapters 3 and 4), in one group the illegal clusters are paired with equally illegal onsets. For these, English listeners should be more prone to perceptually equate the pairs and thus provide higher ratings of similarity than the Russian listeners. There is also a group comparing pairs where both clusters are legal in English, where this effect should not appear.

There were 126 comparisons used, which were split into the set of English legal pairs (e.g. /bl/-/pl/, /kət/-/kəp/), illegal pairs (e.g. (/dv/-/tv/), and mixed pairs (e.g. /br/-/mr/, /vd/-/d/). The aim was to test if Russian and English listeners judged similarity the same, depending on legality in English. Therefore, correlations between the Russian and English means were calculated for each of the three subsets.

The scattergram in figure 5.13 for the English-legal subset displays that, as Russian judgments of similarity become higher, so do the English values. The

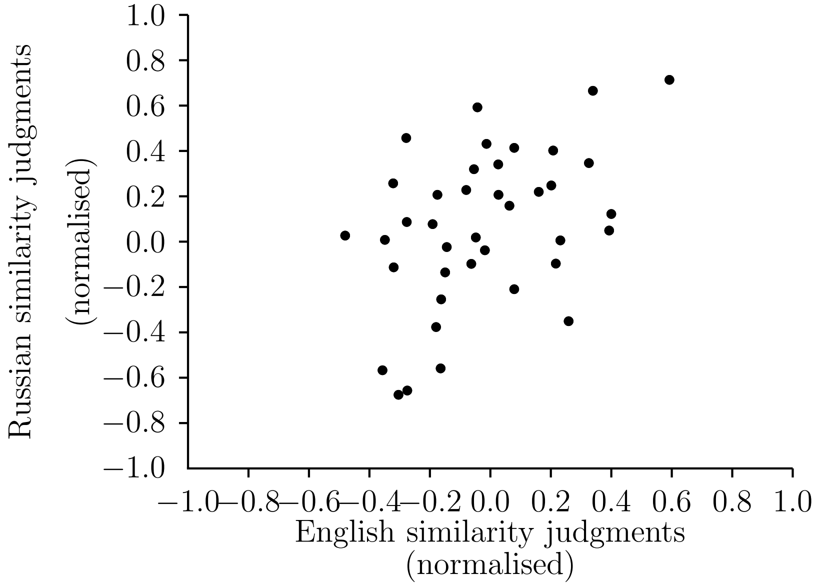

Figure 5.13: Correlation of English and Russian listeners' similarity judgments for word pairs where both onsets are legal in English

statistical analysis confirms this: a significant positive correlation was found for the legal onsets (r_p=.45, N=39, p<.01, 95% CI: .15<r<.67).

However, the scattergrams for the illegal (figure 5.14) and mixed subset (figure 5.15) show that both groups conform well even where the English group is judging ill-formed structures.

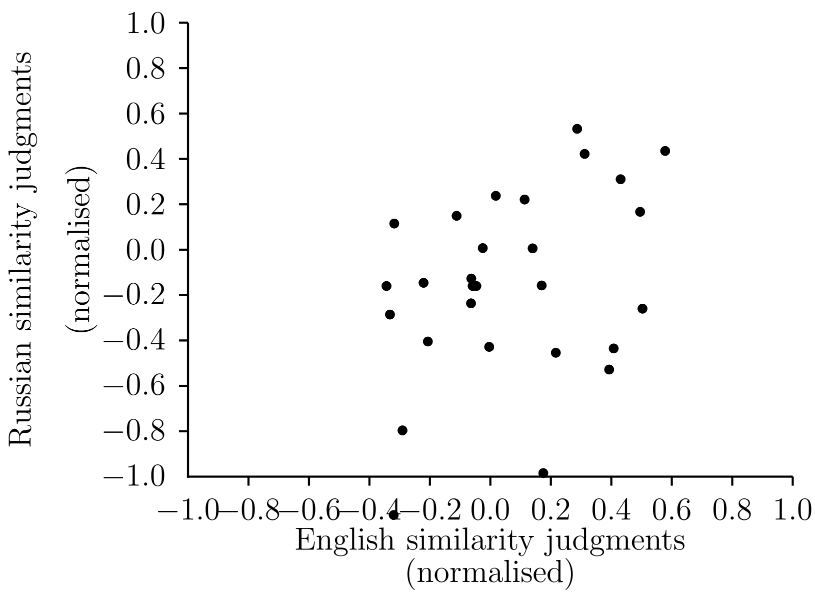

Figure 5.14: Correlation of English and Russian listeners' similarity judgments for word pairs where both onsets are illegal in English

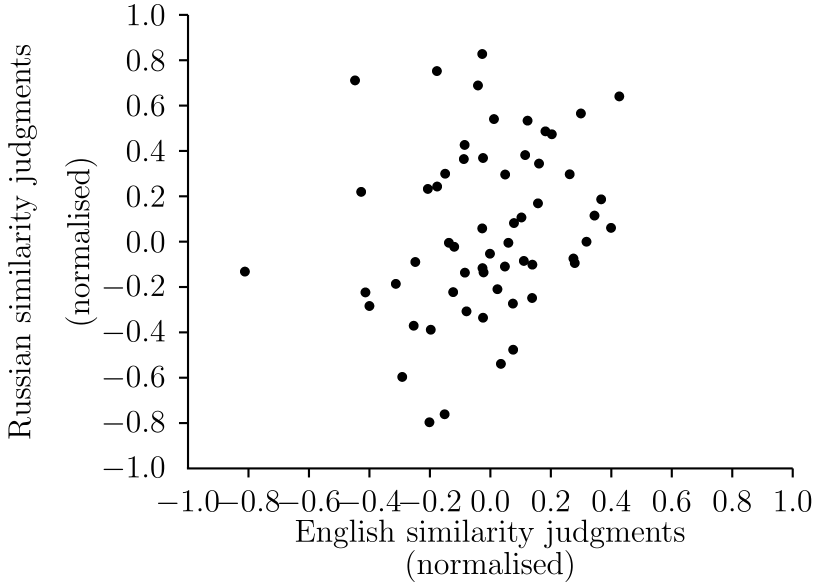

Figure 5.15: Correlation of English and Russian listeners' similarity judgments for word pairs where one onset is legal, one illegal in English

When the correlations for the mixed and illegals sets were calculated, neither was significant, as predicted. Yet, a closer look at the r- and p-values for the mixed (r_p=.25, N=59, p=.06, 95% CI: -.01<r<.47) and illegal set (r_p=.35, N=28, p=.07, 95% CI: -.03<r<.64) reveals that even for those a trend towards significance is evident.

This means that there is some support to be found here for a role of legality in the perception of similarity. Nevertheless, this role must be supplemented by other factors, since the contribution that it makes is relatively small.[9]

[9] I must add a caveat for the interpretation of these results, however: dividing the comparisons by the legal status of onsets harbours dangers in that at first glance the reader may overlook that the sets are not and cannot be matching. For example, the set of illegal comparisons cannot contain ±C or ±V comparisons, and there is a disproportionate amount of ±C in the mixed set. This means that these sets should be viewed separately rather than in direct comparison. Nevertheless, the observation that the correlation is significant in one case and marginally not significant in the others remains.

- Hypothesis 12: English, but not Russian judgments are equal for CC and CəC

Since CəC is overall the strategy most employed in adapting an illegal CC sequence, English speakers are expected to show more confusion between any given illegal CC cluster and its epenthesised counterpart. Thus, a comparison such as C_1C_2 to C_3C_2 should yield much the same rating as C_1əC_2 to C_3əC_2. Russian listeners, on the other hand - and English listeners, too, according to cue availability - are expected to utilise the additional information about the first consonant available through the V transitions to give lower similarity ratings for the epenthesised pairs.

To test this hypothesis, the similarity values from the English and Russian groups will be compared, depending on the onset type. Therefore, the appropriate test is again a mixed ANOVA with L1 as between-group factor, and CC/CəC as within-group factor. If the hypothesis is correct, an interaction effect is expected such that the similarity values are higher for clusters in the Russian group, but the same as for the epenthesised clusters for English listeners.

Hypothesis 12 comparisons		
cluster	epenthesised	*e*
d-t	də-tə	_v
v-f	və-fə	_l,r
z-s	zə-sə	_v
b-m	bə-mə	_l
p-t	əp-ət	k_

In this comparison, particular attention must be paid to the legality of the CC pairs: the effect described should rest more, if not exclusively, on CC comparisons for with both onsets are illegal.

Figure 5.16 and the similarity values in table 5.20 suggest that, in line with the prediction, the English group treats both types of onsets the same, while Russian listeners utilise the added cues for the initial consonant and thus have lower similarity values in the CəC set.

In the statistical analysis, no significant result was found, but there is a trend towards the expected interaction effect. The statistical values can be seen in table 5.21.

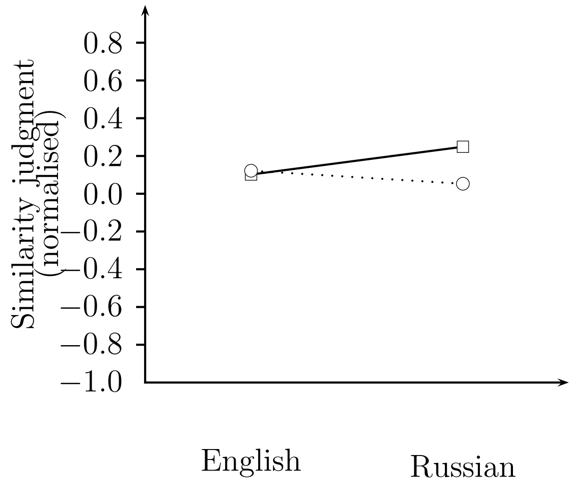

Figure 5.16: Similarity judgments of deletion of clusters (solid) and epenthesised clusters (dotted)

CC/CəC	L1	Mean	Std. Dev.	N
CC	E	.10	1.10	216
	R	.25	.84	191
	total	.17	.99	407
CəC	E	.12	.97	216
	R	.05	.94	191
	total	.09	.95	407

Table 5.20: The descriptive statistics for clusters as compared to epenthesised onsets.

Factor / Interaction	F(df)	p
onset	F(1,405)=2.29	.13
L1	F(1,405)=0.26	.61
onset x L1(*)	F(1,405)=3.44	.06

Table 5.21: The statistics for the factor onset type, with the levels consonant cluster and epenthesised onset.

The fact that in the Russian group, the similarity values for clusters are consistently higher than for the epenthesised clusters supports the idea that Russian listeners perform differently because all clusters are legal in Russian, but not in English. However, if the dataset is split by legality to determine the relative role of this factor, no interaction effects remain. However, in the Russian group the similarity values for clusters are again consistently higher.

On the other hand, under the acoustic cue model only a main effect for type (higher value for CC than CəC) is expected. No evidence for this effect was found either.

Overall, the results for the two hypotheses derived from claiming an influence of native phonotactics on perception provide only limited support in favour of this approach. While trends in the expected direction can be observed, no clearly significant effects are shown.

5.2.3 Summary and discussion

Summary of results

First of all, the list of hypotheses is presented again, with the supported predictions in bold:

1. Cue- and context-based (universal) predictions, as found for example in Steriade (2001), Jun (1995), Hayes & Steriade (2003), which are expected to apply cross-linguistically.
    - **A voicing change is more similar than a nasality change**
    - A place of articulation (POA) change is more similar than a nasality change
    - A voicing change is more similar for the word-initial than the prevocalic consonant
    - **A POA change is more similar when preceding a non-coronal consonant** (however, not in the predicted order)
    - **A POA change is more similar for a stop than a fricative**
    - **A change in anteriority is more similar than a stridency change**
    - **A voicing change is more similar than the addition or deletion of a whole segment**
    - **The similarity of a consonant to zero depends on stricture**
    - **Presence or absence of a consonant is more obvious pre-vocalically than word-initially**
2. L1 inventory-based (language-specific and context-independent) predictions, derived from Broe (1993), Frisch (1996)
    - Inventory-derived segment similarity is applicable to perceived word similarity
3. L1 phonotactics-based (language-specific) predictions, following Dupoux et al. (1999)
    - **Judgments differ cross-linguistically where phonotactic status differs**

- Acoustic cues are only used fully for phonotactically legal sequences

Since the evidence for or against a hypothesis often needs to be qualified, only applies to a subset of the data (e.g. fricatives, but not stops), or, in the case of cue-based hypotheses, is mixed with L1-effects, table 5.22 offers a more differentiated picture. It summarises which of the possible determinants of perceived similarity found support in the results for each of the hypotheses tested. Overall, strong support for the cue-based hypotheses could be found. For example, there are clear cross-linguistic tendencies for a voicing change (and, with reservations, a place of articulation change) to be perceived as more similar than manner/stricture changes such as [±nasality], and a change in anteriority more similar than a change in stridency.

	universal	L1 dependent		
hypothesis	*cue/context*	*legality*	*inventory*	*phonetics*
1	+	(+)	([10])	(+)
2	(+)	+?[11]		+
3	-	-		
4	(-)[12]	+?		+
5	++			
6	++			+
7	+		([10])	
8	(+)			
9	+			
10	(+)[13]		-	
11		(+)		
12	(+)	(+)		

Table 5.22: Summary of results for the similarity comparison. The factors have been split into universal and language-dependent. A + sign represents evidence in favour of a factor, a - sign evidence against a factor. Gaps in the table show that no conclusions relevant to a specific factor could be drawn, bracketed entries show the evidence is only partial, or one of more possible interpretations.

Besides these findings for certain features, the importance of cue availability manifests itself in other ways, too: for example, sonorant consonants are more inconspicuous pre-vocalically due to the higher sonority and therefore (relative) similarity[14] to the adjacent vowel than obstruents, when added or deleted. Addi-

[10] A possible effect of Russian phonology in terms of assimilations and alternations

[11] This is based on C_2 only, where all clusters are illegal. However, the positional division coincides with the one between different POA changes.

[12] A main effect was found, but not in the predicted order

[13] Cue availability is supported here through the clear role context plays in determining how similar two segments are to each other

[14] See comment about circularity in the description of hypothesis 8

tionally, since different manners show their place cues differently, for example mostly segment-internal as opposed to in transitions, a change in place is not perceived by listeners in the same way. The same comparison also shows the significance of context in this approach to similarity. The manner difference is only evident where the place cue difference is of concern, i.e. where transition cues are in fact possible: before vowels. In C_1 position the rating difference for manner is not found. A further positional effect due to cue availability is the clear distinction between C_1 and C_2 in terms of deletion. The comparison of any consonant deletion and the presence or absence of an intervening schwa, and the finding that the latter is cross-linguistically deemed more similar, speaks not only for this approach, but also for its relevance in loanword adaptation: as discussed in the overview of loanword research, vowel epenthesis is by far more frequent an adaptation than the deletion of an offending consonant.

Nevertheless, not all evidence is in favour of the cue-based approach: several hypotheses find no support in this experiment at all, or only for subsets which cue availability does not explain. In one case, the influence of adjacent coronal consonants on similarity, a significant effect does exist, but not in the way predicted.

Akin to these results for a more universal viewpoint of similarity, the outcome for the language-specific factors is also mixed. An influence of whether the clusters compared are legal was suggested as an explanation in various cases, where English speakers showed unpredicted high values for comparisons of illegal sequences, or made use of available cues less than the Russian counterparts. Furthermore, English and Russian judgments were better correlated for the group of English legal comparisons, and the correlation was not (quite) significant for the other groups. Nonetheless, the fact that even in the face of English-illegal clusters the English listeners often managed to use the available cues similarly to the Russian speakers, speaks strongly against a dominant role of legality.

The inventory-derived prediction did not fare well in experiment. While English judgments did correlate well with the derived similarity values, these values do not appear to be as language-specific as suggested by Frisch (1996), since the correlation with Russian judgments was only marginally worse. While the Russian values for lattice-based similarity could be coincidentally similar, there is a further, major problem. According to Frisch's model, the similarity between two segments should remain the same in any context, while in the experiment the judgments varied widely depending on the phonological environment of a comparison (cf. figure 5.12).

Discussion

The conclusion to be drawn from this experiment is overall quite simple: none of the tested approaches to measuring similarity is entirely without merit, but none tells the whole story.

Before discussing details, one point must be made with reference to seemingly "cross-linguistic" or "universal" factors: while coinciding judgments from Russian and English listeners do suggest influences at work which are not language-dependent, the reader must not forget that only two languages have been compared, and two languages which, for all their differences in phonetic detail and phonotactics, are also both Indo-European and have obvious correspondences in terms of the phonological features they use and the possibility of complex onsets and codas.

Keeping this in mind, there are some cases where principles of cue availability are clearly the overriding factor in determining similarity: a voicing change is the least conspicuous feature change, as Steriade has claimed repeatedly (e.g. Steriade 2001, 2003), and deletion is perceived to be more severe if the context is conducive to showing cues to the identity of the sound. But more than the "failing" cue-based hypotheses, which could be due to an error in proposing the relevant hypothesis, it is the frequent mixing of cross-linguistic cue-based effects with interactions or L1 main effects that diminishes the unique role of the cue-based factors in determining similarity. While they are evidently decisive in many cases, there is still much room for explanation beyond them.

The second point of influence on measuring similarity is of course the impact of L1 which often shows through. In the experiment described, the most obvious L1 influence was in the different phonetic detail in the two languages, and the question whether the sequences compared were allowed in a language; others on different levels of the prosodic hierarchy are equally conceivable.

A multitude of further experiments and analyses of phenomena such as half-rhymes are necessary to determine more details of measuring similarity: the extent of the influence of any given factor (in any given language); the points where universal and where L1-specific considerations win out; the question whether it is at all possible to use our knowledge of objective acoustic differences between two sounds, cross-linguistic tendencies, and language-specific idiosyncrasies to define an accepted measure of similarity; and the question whether this kind of fine-grained measure will improve our analyses. I hope to have shown, however, that we need to exercise care with a measurement on which our analysis rests, and that rough-and-ready or one-sided approaches are not sufficient, despite theoretical misgivings which supporters of a model might have. Trask's point, describing similarity as an "important but elusive criterion for phonological analysis" is still valid, and it will take a great deal of work to change this fact.

The implications of the outcome of this experiment for the role of similarity in loanword research, for models of the interplay between perception and phonology, and for the relationship between language-specificity and grammar, will be discussed in chapter 6.

Chapter 6

Discussion

6.1 Summary of results

The previous chapters described a series of experiments which were designed to shed light on the factors active in loanword adaptation. Given their nature as experiments, the output produced were not actual existing loanwords, but they mirrored a borrowing situation: individuals with the borrowing language as their native language (and with no knowledge of the donor language in this case) encountering foreign words. The relevance of these experiments to the issues in loanword adaptation has been laid out in chapter 1, and the implications of the results will be discussed in the following section. First of all, this section recapitulates the motivations and gives a summary of the results.

6.1.1 Production experiment

Experiment 1, as described in chapter 3, laid the ground for the subsequent experiments by investigating the *how* of adaptations. English speakers were confronted auditorily with illegal structures and were asked to reproduce these structures either through pronouncing them or giving an appropriate orthographic representation of them. These tasks were performed for a specific kind of phonological conflict scenario: a phonotactic conflict involving word-initial onset clusters, between two languages which both allow complex sub-syllabic constituents, yet to a differing degree of complexity and, as relevant here, with a different number of possible combinations in two consonant onsets. In this setup, an alternative is possible to the well-reported vowel epenthesis or deletion strategies, as used in borrowing languages which only allow CV syllables: the strategy of the transformation of an offending cluster into a different, L1 compatible one. In this way, the restriction of possible phonotactic adaptations to epenthesis and deletion (e.g. Paradis & LaCharité 1997) can be put to the test.

Indeed, the alternative of changing one cluster to another was used frequently; the most favoured was the devoicing of a voiced C_1 segment, followed by a change of place of articulation, and the smallest group, manner changes. The overwhelming majority of the devoicing cases were those where the adaptation yielded a legal or marginal English onset cluster, such as /fl/ from /vl/ or /sf/ from /sv/. However, another subgroup did not produce a legal English cluster, but only an improvement of the target in terms of sonority distance, by devoicing the first consonant of the cluster to give a steeper rise in sonority, for example for the target /zv/ to /sv/ and /zr/ to /sr/. A less frequent, but also employed strategy was the change of a segment's place of articulation. The most prominent of these changes was from a palato-alveolar fricative to an alveolar one, as in /ʃm/ changed to /sm/, or in /ʃv/ to /sf/, but also from labio-velar to alveolar, in /fp/ to /sp/. These changes were in most cases moving the targets to English legality, however there were also neutral cases such as the change from /sv/ to /ʃf/. A final group of alterations from one cluster to another were manner changes, in this case the use of /w/ instead of /v/ after /s/ or /d/ to produce a legal English output.

These results show a contrast to previous reports of cluster adaptations which reported an overwhelming majority of vowel epenthesis to solve the problem, with very few cases of deletion. The proportions of the strategies here are very different: while epenthesis is overall indeed the largest group of changes, segment changes are almost as common, and deletion is used in about 10% of cases. The vast majority of deletions are those of the first consonant, i.e. the one which is word-marginal and not adjacent to a vowel.

Furthermore, a high degree of variation on a number of levels is evident here: Firstly, while all target clusters are designed such that they pose a phonotactic, but not a segmental conflict with English phonology, in the very same position in word and syllable, they are not treated equally in terms of adaptation: the ratio of correct responses ranges between 0% (for /vz/) and over 50% (for example for /pt/, which has the same sonority distance as /vz/). The number of adaptation strategies employed per cluster ranges from two important changes (where important is defined as used over 10% of the time) to five different ones. The main adaptation varies between clusters; while for many targets it is indeed an epenthesised vowel, for others deletion is favoured (e.g. /ps/ to /s/, /vz/ to /z/), and yet in others devoicing (e.g. /vr/ to /fr/). The main adaptation has no connection with how successfully a target cluster may be produced correctly: for example, deletion of a consonant is used most often for both /ps/ and /vz/, the latter having the lowest success rate, the former ranging among the highest. Additionally, the choice of main adaptation does not allow conclusions with respect to how diverse the remainder of adaptations are, or indeed what they are: while for /pt/ the primary adaptation strategy is vowel epenthesis to an extent that leaves no room for other options, in the case of /fp/ epenthesis is the most frequent, but the possibilities of deletion and segment change are also used.

Finally, this production experiment highlights that concepts which have traditionally been employed to explain the existence or non-existence of clusters in English have little relevance to the performance of English speakers on these illegal target clusters. For example, /fp/, violating the OCP condition, which is used to account for the absence of clusters such as /dl, tl/ in English (e.g. Moreton 2002), enjoys a ratio of nearly 50% correct in the shadowing tasks. More importantly, the sonority distance between the two initial consonants, which explains the prohibition of all consonant clusters with a lower sonority rise than two steps in English, has remarkably little influence on whether the target will be pronounced correctly by English speakers or not. Although it is true that the cluster with the highest sonority distance, /pn/, is indeed produced rather successfully, around 50% of the time, the group of zero sonority distance, i.e. the fricative-fricative and plosive-plosive clusters, range from 0% to over 50%. Additionally, sonority distance cannot predict the preferred adaptations. While there are some commonalities for groups of clusters with the same sonority distance, these do not combine to any overall influence of sonority distance, and may instead be attributable to the manner of segments involved.

In terms of perception, this experiment gives a first indication that one of the problems preventing the correct production of illegal clusters may be be incorrect perception: in the shadowing task the participants had to produce a spoken output of what they had perceived and hence processed the target through the whole speech perception and production mechanism. The second task, on the other hand, did not involve speech production, but only perception and the orthographic representation of the perceived input. Nevertheless the results for the two different task were remarkably similar, and there was a significant correlation of the correct response ratios. A further indication of cross-linguistic differences in perception was the pilot study described in chapter 3, showing significant differences in Scottish and Russian discrimination in ABX categorical perception tasks.

6.1.2 Factors in adaptation

The set of experiments and investigations described in chapter 4 dealt with a number of factors which were conceivably relevant for the production of foreign structures, and the adaptation of these structures, respectively. The first of these factors was the frequency of potential adaptations in the English lexicon, and the possible primacy of frequent English onsets over infrequent one as adaptations. The CELEX database was used as a basis for determining the spoken frequency of the adaptations that had been used by the participants in chapter 1. These ranged from the very frequent, for example of singleton /t/, to the very infrequent such as /sf/, and indeed the absent, like /tf/. The statistical analysis showed no influence of either token or type frequency on the commonness of a potential adaptation in experiment 1.

The second factor investigated was the concept of gradience in the acceptability of the Russian target clusters, as judged in a magnitude estimation task by English speakers. The results showed that there is clearly a gradience effect in phonotactics, which pertains even to structures with zero frequency, i.e. those which never occur in a speaker's language. With regard to the role of the lexicon in this respect, the perceived grammaticality of an onset which does occur in the English lexicon correlates with its frequency therein. However, since the group of zero frequency onsets shows a wide variety of grammaticality judgments, the lexicon has no direct bearing on the perception of how well a foreign structure would fit into it, but may be of indirect influence via a generalisation of preferences which can be gleaned from statistical analysis of frequencies in the lexicon. Variables relevant for English, and tested for influence on the rating of grammaticality were sonority distance and voicing as well as coronality of the first consonant of the cluster. A voiceless first consonant clearly indicated goodness, whereas other factors showed influence mainly in interactions, such as coronal first consonants indicating acceptability, but only when voiceless, i.e. demonstrating an unsurprising preference for s-initial clusters, which exist in the English lexicon.

The question what effect grammaticality has when those zero frequency structures are adapted could not be answered conclusively. One clear effect is that, the more unacceptable a target is for English speakers, the lower the successful correct production ratios in the adaptation task. This means, the worse something sounds with respect to English, the less likely it is to be pronounced correctly. Less clear is the effect of acceptability on the choice of adaptation strategies. The analysis showed that neither deletion nor segment change were associated with a particular band on the grammaticality spectrum, but that there was a significant pairing of low-acceptability clusters with vowel epenthesis: clusters which were perceived to be very ungrammatical were likely to be epenthesised.

Finally, the effect that the perceived closeness of a potential adaptation to the prohibited target cluster has on adaptations was studied. To this end, judgments of similarity between target/adaptation onset pairs were collected, again using a magnitude estimation experiment, with native English participants. For the majority of target clusters, the most frequently used adaptation was also the one judged most similar to the target. The hypothesis that the perceived similarity correlates with how often an adaptation is used, was confirmed. This shows that there is a strong motivation to keep the alteration to a target minimal and to pick the closest possible adaptation. It was shown that overall, a change in voicing was regarded as the minimal change, followed by vowel epenthesis. However, this experiment did not provide any conclusions about the nature of similarity, which, given the prime role of similarity in adaptation, needed to be remedied.

6.1.3 Similarity

To this end, the next experiment was designed to study in more detail the determinants of similarity; hypotheses from the literature on rankings of similarity needed to be tested, as well as comparing cross-linguistic differences and commonalities to evaluate to what extent perceived similarity is dependent on a listeners native language.

Therefore the following experiment improved on the previous similarity judgment task by including a wider range of comparisons, and comparing Russian and English participants' judgments. The results were used to evaluate a number of hypotheses concerning the nature of similarity, the influences determining it and the possibility of measuring it. Issues of concern were whether the perception of similarity is universal or determined by one's native language, and if so, which aspect of the native language is relevant (i.e. whether this influence is of a phonological nature or takes place on a phonetic level), and whether context plays a role in determining perceived similarity.

By definition, influence of the native language is apparent in cross-linguistic differences. These were found frequently when analysing the results of the judgments; for example, a difference between Russian and English listeners was evident in terms of the legality of the compared sequences: while all onsets were legal for the Russian participants, this was only partially true for the English group. This group showed greater difficulty in distinguishing, or rather, a higher similarity rating for those consonant combinations which are not part of English. Furthermore, while comparisons of pairs with some alterations were perceived by both groups in the same way, others showed clear distinctions: the Russian group rated manner and POA changes less similar than English listeners, specifically between coronal and other places of articulation. Voicing changes, on the other hand, and the presence or absence of schwa or a consonant are perceived to be more similar from a Russian point of view. Of these differences, most are attributed to phonetic differences between Russian and English: the confusion arising from different phonetic properties, such as the dental nature of Russian coronals, causes a higher perceived similarity when clusters containing coronals are involved.

On the other hand, commonalities between the Russian and English group were also found abundantly. Whereas illegal clusters and unfamiliar spectral values appear to somewhat hinder English listeners' perception of differences, overall there is a good correlation between the two language groups. Furthermore, many of the predicted universal rankings of perceived similarity were evident cross-linguistically: voicing was perceived to be the least distinguishing change overall, followed by the presence vs. absence of schwa; the presence or absence of a consonant was perceived to make the largest difference. Within the group of segment changes, voicing was followed by changes in place of articulation, which were perceived more readily in fricatives than stops. Consonant deletions were perceived to be more drastic when the second, vowel-adjacent consonant was involved rather than the word-marginal first consonant.

Overall, the experiment showed both universal and language-specific forces at work in governing the perception of similarity, which will be discussed below.

6.2 General discussion

There are several issues which this thesis has investigated. These range from finding insights into the nature of similarity, via the role of perception (with respect to phonology in general and loanwords specifically), to the relationship between language-specificity and phonology. All these contribute to the original aim of garnering new conclusions about the workings of a loanword phonology - or indeed non-phonological, or universal loanword adaptations -, and about which aspects of loanword adaptation may be predictable.

Specific focus was placed on variability in adaptation, and its connection to frequency and gradient grammaticality. Furthermore, the dispute whether loanwords are adapted in perception or production was approached firstly through a comparison or perception- and production-oriented adaptation tasks, and secondly through studying to what extent the perception of similarity, which is the main driving force in adaptation, is influenced by the native language system.

Before moving on to these issues, however, I would like to remind the reader that indeed none of the experiments undertaken for this thesis deal with real loanwords that have been borrowed from one language and incorporated into another. While some may reason that this is a disadvantage or even disqualifies any conclusions drawn from this for real loanwords, it has, in fact, an important advantage over real loanwords in terms of giving answers about loanword phonology: the means by which the foreign words have entered the linguistic world of the speakers of the borrowing language is purely speech, unadulterated by orthographic input. This is precisely the situation that is needed when studying the *phonological* workings of incorporating foreign words. It ensures that, if adaptations do not take place during the perceptual process, then this is not only because perception was aided by other means such as written input. This prevents us from a priori assigning perception a smaller role than it may be due in the adaptation of loanwords. Yet another advantage is that, as explained above, this set-up precludes the production of loanwords from being merely an implementation of a pre-learned adaptation strategy which may have arisen from factors other than phonology. Moreover, since none of the speakers had any knowledge of the source language, there is no explicit instruction from the SLA classroom that might have distorted the nature of the adaptations. A further point to support the relevance of the adaptation experiments is the effort which was made to ensure that the production experiment did not result in mere parroting of the Russian forms: each participant did the sentence condition before the word condition. They were thus set up to process the auditory input they received in a linguistic fashion. In a similar vein, the orthographic production necessitated the forming of a representation preceding the performance of the task for each stimulus.

6.2.1 Gradience, grammaticality and variation

The results of the experiments conducted provide a great many examples of gradience and variation in phonology. Both the percentages of correct responses and the type and diversity of adaptations for each cluster varied widely between the members of the set of English-illegal clusters, despite their status being the same, as phonotactically, but not segmentally illegal. This is in sharp contrast to most work on loanwords, where only one native form for a given foreign form is analysed, and it complicates any grammars designed to account for it. One possibility of modelling this variation would be probabilistically, as for example Anttila (1997) and Hayes & MacEachern (1998) have done within Optimality Theory: constraints may have overlapping ranking intervals, which allows for variable ranking and thus variable output. This idea will be taken up in the section on loanword adaptation.

The variation in successful correct replications is connected to the gradience in how grammatical or acceptable, or indeed unacceptable, a given form is perceived to be with respect to the borrowing language. Again, although all illegal clusters are the same in that the segments contained, but not the combination of segments are allowed, not all of them are considered to be as unacceptable as each other. There is a wide range in grammaticality judgments, with a significant correlation with how successfully a cluster was produced in experiment 1. As described above, gradient grammaticality is connected to the lexicon in that there is a significant correlation between the frequency of a legal or marginal cluster and its judged grammaticality; this supports exemplar-based models of phonology such as Bybee's (e.g. Bybee 2001) and Pierrehumbert's (e.g. Coleman & Pierrehumbert 1997) in which frequencies and probabilities of structures in the set of perceptual categories or the lexicon of a language play a decisive role in phonology, for example in the occurrence or not of a phonological process.

However, the present thesis is not only concerned with structures of low and high occurrence probabilities within the native lexicon, but with sequences of sounds, and acoustic signals, of zero frequency, i.e. those that the native listener never encounters. All the illegal clusters used in the experiments have zero frequency, but diverging perceived grammaticality. Hence, the lexicon cannot influence the judgments directly. One possible way of deriving implications from the native lexicon for these is by an abstraction from the native exemplars and from the absence of other structures to generalisations and statistical tendencies, which may eventually be termed rules, or constraints. The grammaticality of foreign structures would then be judged by how large a part of these generalisations they conform with. In the present thesis, the influence which the different factors tested have on the judgments is mixed: while one generalisation (voiceless first consonant) proved to be a good predictor of perceived goodness, other factors were only influential in interactions, which made the combined factors so specific, for example indicating increased grammaticality for s-clusters, that the term "generalisation over the lexicon" is hardly appropriate. Hence the mechanism of what determines the perceived grammaticality of a foreign structure with respect to one's native language needs further investigation.

6.2.2 Similarity

In chapter 5, several models of similarity and their background were described. The issues in dispute are: firstly, the universality vs. language-specificity of perceived similarity; secondly, the relevance of acoustic cues and context; and thirdly, the level of phonetics or phonology relevant in determining similarity. Furthermore, the possibility of measuring similarity was evaluated. The summary of the results of the cross-linguistic similarity judgment suggests that for none of these issues a simple answer is appropriate.

The question whether the perceived similarity of two strings of sounds is language-specific, or universal and objective (to the point where the attribute *perceived* would be inappropriate due to its connotations of variation and subjectivity), can be answered with a decisive "both". On the one hand there is strong evidence that impressions of the degree of difference between two linguistic signals are to some extent common across speakers of different languages[1], that listeners from different linguistic backgrounds indeed have in many cases the same sense of similarity as researchers using their intuitions of what is similar. Listeners with different native languages often use the objectively present acoustic cues to contrast in the same way, such as finding stronger evidence for place information in consonants adjacent to vowels, with transition cues; and where the consonants are not next to a vowel, stronger intrinsic cues to place information in fricatives as compared to plosives. The English and Russian groups agree that some changes are always less drastic than others, and that a difference in voicing creates the smallest difference of all. In short, there is plenty of evidence for cross-linguistic tendencies in the

[1] However, one needs to be wary of drawing stronger than warranted conclusions since, although Russian and English exhibit a number of differences, they also belong to the same language family and use the phonetic features compared in a similar way; for example, they both have a two-way voice distinction and use the same major place categories.

perception of similarities in these experiments, even in the face of phonotactic (and phonetic) issues for one of the participating groups; it remains to be seen whether similar cross-linguistic tendencies can also be found when languages with larger differences between them than English and Russian are contrasted; for example, in a comparison of listeners from a Indo-European language with listeners whose native languages are Arabic (for place similarity), Thai (for the perception of voicing changes), Mandarin (as a tone language), or Xhosa (for the perception of similarity with clicks).

Despite such cross-linguistic tendencies, there is also, as reported, convincing evidence that perception of similarity is not the same cross-linguistically, and that the degree of difference that is assigned to two forms is influenced by the native language even in the face of strong acoustic cues available for a specific contrast. For example, in spite of the absence and presence respectively of a vowel, English listeners tend to assign CC pairs and the corresponding epenthesised pairs roughly the same ratings, whereas Russians give lower similarity values to the epenthesised pairs. English listeners also assign comparatively higher similarity to place of articulation changes in spite of the available cues which Russian speakers use. Russian speakers rate voicing changes as comparatively similar, whereas English speakers give these cues more importance.

Together with other predictions which are derived from the relative strength of cues for certain segments or contexts, and which are not supported (such as the influence of the coronality of a neighbouring sound) a mixed picture emerges: on the one hand, the availability of cues plays a major part in how the salience of a contrast is perceived, and indeed in a comparison of English and Russian similarity judgments a number of rankings are the same; on the other hand, factors of native language origin also play a role and cause cross-linguistic differences in perception of similarity. While segment similarity derived from the distribution of the native inventory does not extend to accounting for the similarity of sequences of sounds[2] since the experiments show that the similarity of two phonemes is judged very differently depending on context, both a lower level and a higher level are shown to have a bearing on the judgments: the low-level phonetic details of the realisations of sounds, and the possible phoneme combinations, in the listener's native language. Both the L1 phonetic detail and phonotactics influence similarity judgments in

[2]Note that this does not mean that the native inventory has no bearing at all on similarity judgments; of course it would be relevant in a different setup if sound strings containing non-native elements were compared.

that they can hamper listeners' recognition and use of acoustic cues, and lead to perceiving a contrast as weaker than the strength of the cues would suggest.

This complicates the matter of measuring similarity, or even devising a ranking of contrasts (in specific contexts) in terms of how large the difference between two forms will be perceived, as Steriade (2001) has done in the P-map. Such rankings may be the basis of the perceptibility of contrasts, but they must be modified and reshaped and, in the terminology of OT, re-ranked with respect to corresponding faithfulness constraints, for each language according to its own phonetics and phonological system. To develop a complete similarity ranking for a language, or the parts of it necessary for a specific phonological analysis, may be laborious and time-consuming; at its basis needs to be an extensive body of cross-linguistic comparisons, with segmental, phonotactic and phonetic variables. For example, if languages display the same segments with phonetic differences, the relevance of the phonetic characteristics can be tested through two variants of the same set of stimuli, with each variant conforming to the phonetic detail of one language.

The influence of the native phonology may change the relevant rankings only marginally, such that the analysis remains the same as when using only cue-based rankings. However, a detailed study of what determines similarity for listeners of a certain native language will in some cases resolve mismatches between similarity and adaptations that a more objective or universal approach may have been unable to explain.

Beyond the topic of loanword adaptation, a more detailed model of similarity, which takes language-specific differences into account, will also be able to put further similarity-related research on a firmer footing; for example, it is applicable to claims regarding the Obligatory Contour Principle, which prohibits neighbouring too similar to each other (e.g. Frisch et al. 2004), to claims for example of deletion due to low salience, to the new approaches to language classification described in chapter 5, and to cross-linguistic generalisations and language-specific trends in historical phonological change.

6.2.3 Perception and phonology

For several of the approaches to what determines similarity, the opinions about similarity are directly defined and derived from the researcher's model of perception, and of the relationship between perception and phonology. The term *perception* in this context was introduced in chapter 1 as referring to the complete process of transforming, or distilling, an acoustically detailed input into a linguistic representation. Since the perceptual and phonological models bear on the view of similarity, the results from experiments on judging similarity, and cross-linguistic comparisons thereof, are also suitable to evaluate some aspects of these models of perception, which were discussed in chapter 1.

To recapitulate, four kinds of models were discussed:

- perception and phonology are independent from each other, i.e. the null model, as it were: perception is equal for listeners from all linguistic backgrounds, and any acoustic input can be assigned its "correct" phonological representation without difficulty (e.g. Jacobs & Gussenhoven 2000).
- phonology influences perception: a person's native phonology affects their perception, and in the case of foreign input, can lead listeners to perceive and categorise it differently from listeners of the originating language. This influence can take various shapes, such as a warping of the acoustic space to be denser around native perceptual categories (e.g. Kuhl & Iverson 1995), or an assimilation of the input to native articulatory categories (e.g. Best 1995). The result of the influence is that much foreign input is allotted to these native categories in spite of belonging to different categories in the source language. Evidence supporting this has been gathered for segmental cross-linguistic differences in many SLA studies.
- perception influences phonology: perception is determined by the availability of acoustic cues to a specific contrast, in a specific context that may emphasise or diminish the cues. This perceptibility of contrasts influences inversely how much importance is placed on preserving a contrast when it collides with other preferences of the language, or, in optimality-theoretic terms, how the corresponding faithfulness constraints are ranked. Evidence for this position comes for example from cross-linguistic preferences for strategies of solving conflicts[3], and from half-rhymes preferring differences with poor acoustic cues.
- perception is phonology: in some respects this is an extreme case of phonology influencing perception; according to Boersma (e.g. Escudero & Boersma 2004) one's native phonology does not only affect the perception of foreign input, but indeed the perceptual process itself is phonology, and can modelled like the mapping of a representation to the output in standard phonology.

[3]For example, in case of a disallowed final voiced obstruent, a change of voicing, which has the poorest cues in this position, is claimed to be universally preferred over other repairs such as vowel epenthesis and deletion.

How can the studies in this thesis contribute to the discussion? Firstly, the initial production experiment has had one very interesting outcome: whether or not a foreign word that English speakers were confronted with had to be pronounced by them, i.e. whether they had to produce speech output or were only asked to convert their representation of what they had heard to orthographic form - in both cases the participants performed remarkably similarly. There were fewer adaptations, or errors, in the latter task; nevertheless, the percentages of producing the correct output showed a significant correlation between both tasks, suggesting that the English participants had difficulty not only in producing, but also in perceiving the sounds (while Russian listeners testing the stimuli had no such difficulty). On the other hand, in spite of the correlation, correct productions were consistently higher for illegal clusters in the orthographic representations. This already indicates that, as with similarity, a clear-cut separation of influence vs. non-influence of perception cannot be achieved. A further indication of perceptual differences depending on native languages was, although not further explored for reasons explained above, the pilot study comparing Scottish and Russian ABX categorical perception. The strength of both results is somewhat diminished, the former by the uncertainty of how much, if any, of the phonological production process a representation has to pass through before an orthographic output can be created, the latter through a small number of stimulus sets and participants, i.e. data.

More evidence for L1 phonology influencing speech perception comes from the cross-linguistic similarity comparison, which, as described, shows that perceived similarity is influenced both by L1 and by cue perceptibility, and that, while obviously the cues are objectively available to the listener, they are not always *used* in the same way, and that the (L1-)subjective salience of a contrast is not necessarily the same as what the spectrogram would indicate. The experiment shows that perception, or the judgment of what has been perceived, is not radically different between Russian and English, but there are clear discrepancies. As suggested in chapter 5, these discrepancies are caused by phonotactics and phonetic detail, to the effect that phonotactically illegal forms and phonetic detail which deviates from the native realisations of a segment impair listeners' perception of a contrast, and hence lead to higher similarity ratings.

One possible interpretation of this is that English listeners do perceive the Russian stimuli "correctly", i.e. utilising all the available cues in the same way the Russian participants do, and evaluate the similarity differently in a later step after creating the correct representation. What speaks against this is that on the one hand we know from second language acquisition that segmental perception is often unsuccessful for foreign input, and that on the other hand this merely shifts the problem to a different, later, undefined level: in contrast to loanword adaptation, where a straightforward split of processes enables the viewpoint that adaptations do not happen in perception, but in production, a similarity judgment task has no production process: similarity is determined by what is perceived. Moreover, in a case where phonetic detail has clearly been demonstrated to be the relevant factor, an explanation where similarity judgments are made *after* a phonological representation is formed cannot succeed for a representation which contains no phonetic information.

The alternative interpretation is that, although all humans have the same perceptual apparatus, and the perception of many contrasts takes place in the same way across languages, the perception of linguistic structures is not free of L1 influence, on a phonetic, segmental and phonotactic level, and that the claim of a perception independent of phonology is not tenable. This does not as such pertain to the idea that perceptual cues influence phonological constraints and their organisation, but it is relevant for Steriades model of loanword adaptation, which will be discussed in section 6.2.4.

The remaining issue concerns the nature of the influence of L1 on perception, and whether it is a phonological-grammatical one. In general, characteristics of one language's speech are not necessarily grammatical simply due to their being language-specific, if we consider for example the average pitch of English as compared to Dutch or German female speakers Mennen (to appear). Additionally, we have seen that L1 influence can apply on various levels, and it is not a given that it functions in the same way on each of these levels. A relevant point here is the question whether experimental results such as from similarity judgments can be useful in providing evidence either way: would listeners perceive differently depending on whether the perception process is a grammatical one or not? It would appear that, if the corresponding grammar includes phonetic detail, as Boersma's perception grammar does, then the phonological representation that is eventually formed, i.e. the output of perception, is the same whether or not the perceptual process is thought of as grammatical or not.

What is more interesting are the theoretical issues this question brings with it, if, as noted above, perception is thought of as the conversion of an acoustically rich input to an abstract phonological representation, i.e. in many ways the mirror image of production. Since at either end of both these processes is an abstract linguistic representation, there is a temptation to label both processes grammatical: if a process works with (production) or produces (perception) something grammatical, then some aspect of the process itself must be grammatical. Is, then, a perception grammar the mirror image of (at least part of) the corresponding production grammar? If so, the production grammar would, just like the perception grammar, have to incorporate (an excessively large number of) phonetically detailed constraints, such as assigning a certain vowel specific formant frequencies - an aspect of production which is normally called *phonetic implementation*, and would be placed outside of phonology proper. If not, it would have to be specified in which parts perception and production grammar overlap and in which ways they do not (and why). A further problematic aspect of a perception grammar with respect to loanwords is that it creates a duplication problem in adaptation, which will be discussed in section 6.2.4.

In summary, speech perception is shown to be similar, but differing across languages in detail, and hence subject to the influence of L1 on various levels. To quote Hume & Johnson (2001*a*), "[T]he perceptual influence on phonology is static because the pattern of perceptual salience in segmental context remains relatively constant across languages, but the perceptual influence on phonology is also dynamic because overall [...] salience differs from language to language." The language specificity is caused by L1 phonetics and phonology, and the question whether this a grammatical process remains disputed for the time being.

6.2.4 Loanword adaptation

The insights regarding the perception of similarity and perception in phonology can, together with the results of the production (i.e. adaptation) experiment be fed into an evaluation of the models of loanword adaptation which were described in chapter 1. Before entering this discussion, the main conclusions from the experiments, as relevant to loanword adaptation, are summarised as follows:

- Despite all illegal clusters having the same status in English, i.e. containing legal segments in illegal combinations, there is extensive variation with respect to both the success rates of reproducing them and the preferred adaptations
- Sonority distance does not determine production success, and is of very limited relevance for adaptation choice
- Voicing changes are preferred where they result in a legal cluster

- Adaptations are partially made in perception
- Despite all clusters having the same status in English, English listeners do not perceive them to be equally unacceptable
- Higher perceived grammaticality of a target results in higher targetlike production
- Potential adaptations which are perceived more similar to the target are preferred
- Perceived similarity is cross-linguistically similar, but depends to some extent on the native language

The following section will assess how well various models of loanword adaptation are compatible with these conclusions. Firstly, there is Paradis' loanword-specific Theory of Constraints and Repair Strategies (TCRS), which proposes a universal strategy of epenthesis over deletion for disallowed consonant clusters, with deletion only occurring in case of a combined phonotactic and segmental conflict, so that the change of a segment would have to occur in addition to epenthesising a vowel. The reason for the concentration of TCRS on epenthesis and deletion, while the third possibility of changing one cluster to another is not discussed, lies in the data on which it is based, namely adaptations of clusters in CV languages. In the present scenario, however, this third option plays a large role, in some cases larger by far than that of epenthesis. A further fact irreconcilable with TCRS is that deletion occurs at all in this experiment: since none of the clusters contain segments outwith the English inventory, all problematic clusters should have been repaired through vowel epenthesis, as well as being perceived to be equally ungrammatical. Moreover, while Paradis denies all influence of L1 perception on adaptation, the difference between shadowing and orthographic task - indicating adaptations in perception - provide a separate problem for TCRS. Furthermore, Paradis' universal phonological hierarchy as a basis for minimal change would predict cross-linguistic conformity in judging similarity, which has shown not to be the case. Overall, Paradis' proposed model of loanword adaptation fails to account both for the adaptation processes observed here and the outcome of the similarity comparison.

A model which is similar to TCRS in that it claims that all loanword adaptation takes place in production rather than perception is Steriade's, seeing the locus of adaptation in productive phonology and not in perception, while on the other hand there is a strong perceptual element incorporated into production: the perceptibility of contrasts results in rankings of constraints (the P-map), which then militate to keep the change from the original form minimal, i.e. minimally perceptive; for example, a voicing change is the minimally perceptive in word-final position and is used to remedy cases of ill-formed final voiced obstruents, so that /b/ is changed to /p/ rather than being deleted or followed by a vowel to form /pə/. This means that similarity plays a key role in adaptation. This aspect of the model works well with the present results to some extent.

A result speaking in favour of it is that, within the segment changes, a voicing change was the largest group, as opposed to place of articulation or manner changes, which are less similar according to Steriade. Specifically, for those target clusters in the production experiment where devoicing resulted in a well-formed cluster, such as /vl/ to /fl/, this strategy was indeed used very frequently by the English speakers (especially in the shadowing task). However, two problems arise with regard to Steriade's model. Firstly, if there is clear cue-based ordering of similarity resulting in a clear ranking of faithfulness constraints such that for example a constraint against a voicing change is lowest, then no variability in the outcome is expected. The productions are in fact rather varied, however: for example, for /vr, vl, sv/ not only vowel epenthesis, but also deletion are observed to a considerable extent. While in theory this problem could be solved through overlapping or unranked constraints, in Steriade's case this solution is not applicable, since the ranking has been pre-determined through perceptual considerations.

Secondly, an influence of perception on the adaptations in the experiment is also exhibited by the similarity between the shadowing and orthographic tasks; this goes counter to the claim that all adaptation occurs in production. Steriade (2005) raises important questions as to a possible perception/production split, i.e. part of the adaptations taking place in perception and part in production, namely: how would the task of adaptation be split between the two processes, and why should it be split in this fashion? These questions are as yet unanswered, and Steriade's loanword analyses strive to find explanatory solutions to loanword problems within optimality theoretic production phonology including the perceptually guided P-map.

What is problematic about the P-map influence in adaptation is the fact that, while Steriade explicitly denies that her approach to perceptibility is supposed to be universal, her hypotheses are formulated in an implicitly universal way, and lack counterexamples or explanations of admissible language-specificity. It remains therefore unclear how demonstrated influence of L1 on similarity judgments and on perception in general, as also shown by a wide variety of second language acquisition studies, could fit into Steriade's model. However, even if it is possible to provide a satisfactory analysis of loanword adaptations in this manner, it is not self-evident that it should be appropriate to ignore that adapters do misperceive input, and hence that adaptations do happen in perception.

The evident split of adaptation into perception and production conforms to positions taken by Silverman (the split into the Perceptual and the Operational Level), as well as proposals by Kenstowicz and Yip suggesting adaptations taking place in both processes. However, Silverman's Perceptual Level only repairs segmental problems, leaving phonotactic issues to the Operational Level, while for the - phonotactically problematic - clusters in this thesis adaptations were made in perception, too. In Yip's and Kenstowicz' work, the descriptions of the perceptual adaptations are as yet too general to evaluate with any precision; for example, "the adapter is not a passive recipient but exercises active control [...] as well as possibly calling on implicit knowledge of phonetic similarity" (Kenstowicz & Suchato 2006).

A different approach, which places all emphasis on perception instead of production, is Boersma's. As described in chapter 1, in this view the adaptation process is still considered to be a grammatical one, but at a different stage: perception is regarded as the process by which a rich acoustic form incompatible with L1 is mapped onto a representation which is well-formed in L1, which means the adaptation process is complete by the time the representation is formed. This would mean that, unless the perception grammar is not applied fully for an independent reason, no ill-formed representations should last beyond perception, which has, however, been the case in the adaptation experiment; in fact, for many instances, the ill-formed structures survived even beyond production. Moreover, there should not be a difference between the shadowing and the orthographic task, which also was the case: the rate of targetlike responses was higher in the orthographic task, and in some cases the preferred adaptation changed (e.g. /vl/ > /fl/ in shadowing, /vl/ > /vəl/ in orthography). If, on the other hand, the perception grammar does indeed not filter all ill-formed structures (perhaps modelled through a probabilistic constraint ranking in optimality theory, as in Boersma & Hayes 2001), then once again Steriade's questions apply, i.e. why and how adaptations would be split between perception and production. The problem is aggravated for this model because both the perception and the production processes are labelled as grammatical, and stem from the same L1. It is therefore not clear why and how aspects of the same grammar would apply at different stages, whether perceptual adaptations would be duplicated in production, and whether a grammar that models the variation of foreign word adaptation would also fit the native production data. On the other hand, one aspect of Boersma's theory is supported by the similarity results, namely the influence of a listener's native language on how similar to sequences are perceived to be, as shown by the frequent L1 effects in the comparison between Russian and English judgments.

Peperkamp et al. (2004) agrees with Boersma in that perception is the decisive stage in adaptation. Nevertheless, she is at odds with perceptual grammar as well as with P-map production phonology and TCRS: according to Peperkamp's claims, the term loanword phonology as a whole is inappropriate, since phonology or grammar are not involved in the process. Rather, all adaptations are viewed as phonetic approximations to the nearest existing forms of the native language (thus L1 being the cause, but not the means of adaptation). This accounts, for example, for why the same stops with and without a burst in the source language, respectively, are adapted with or without an epenthesised vowel (Peperkamp 2003). The arguments against perception-only adaptation listed above apply here, too; in fact, they apply even more strongly because Peperkamp's model claims explicitly that the adaptation process is complete before production, and there is no obvious potential solution, such as a partial ranking of constraints for Boersma's perception phonology. Moreover, what speaks against Boersma as well as against Peperkamp (more severely against Peperkamp) is that for many targets, instead of a single obvious perceptual adaptation, there are two or three frequent adaptations; this could be solved more easily in a perception grammar where not all constraints are ranked with respect to each other, or overlap, whereas in Peperkamp's case a clear phonetic motivation must be found for each adaptation. This fails for example in the cluster adaptations described in chapter 3, where Peperkamp's model predicts high rates of epenthesis for first consonants with a burst, where a motivation such as a plosive burst causing epenthesis was only found for a subset of the clusters with high epenthesis rates. For example, the release of /pt/ being reinterpreted as a sign of a vowel is possible explanation for the very frequent vowel epenthesis in production. However, epenthesis is equally frequent in /vb/, where a similar motivation is not obvious. What speaks further against Peperkamp's model is the influence of gradient grammaticality on production. If the forming of an adaptation consists of nothing but the perceptual system choosing the phonetically closest native form, then neither should the concept of grammaticality influence production, nor should English speakers judge the grammaticality of the ill-formed clusters differently to begin with.

This is not to say that phonetics does not play an important part in loanword adaptation; in fact, since perceived similarity is the most decisive factor in the choice of adaptation, the importance of phonetic detail in adaptation is corroborated by its relevance in determining perceived similarity. It is relevant in two ways: one, through differences between languages in the realisation of sounds, which may impede perception by foreign listeners, even if they have the relevant phonemes in their inventory; two, through the perceptual availability of acoustic cues to contrasts, which has been (in the spirit of the P-map) shown to play a major role cross-linguistically in determining how similar two structures are. Further support for the importance of phonetics comes from the failure of phonological factors such as sonority distance to account for the variation in loanword production. On the other hand, the involvement of phonological factors in adaptation can also be clearly seen: the success of reproducing illegal clusters changes with grammaticality (in the perception of which some phonological generalisations are evident), adaptations often do increase sonority distance, and the phonotactic illegality makes a cluster more prone to misperception and thus to adaptation. Therefore, the results of this thesis support a phonological model of loanword adaptation which makes reference to phonetic facts and their consequences for perception and similarity.

Models of loanword adaptation all agree that, for all loanword adaptation, the goal (or at least the outcome) is to incorporate the foreign word into the set of structures which are well-formed in L1 in such a way that the alteration to the original form is minimal. They also have in common that their approaches to determining minimality, and the precise site of adaptation, are of an all-or-nothing kind. This is more straightforward in theoretical terms, but hampers the understanding of what appear to be more complicated subdivisions and load-sharing involving phonetic and phonological, cross-linguistic and L1-specific, perceptual and production elements. None of the models in question is without its merits; on the other hand, none of the models is without its failings.

A complete model of loanword adaptation, or loanword phonology, must be one that can incorporate the fact that a split can occur into perception and production adaptations, where the *kinds* of adaptations (e.g. vowel epenthesis, or place of articulation change) may be the same in both stages. It must further allow a cooperation of language-specific and cross-linguistic preferences to determine the minimal modification for a target/L1 combination, and variability between two or more possible adaptations. That is, while such a model may not be able to predict the precise optimal modification for each case, it will be able to provide a ranking of possible modifications, and a prediction on the relative frequency of each.

An outline of such a model must contain two parts: the perceptual part and the production part, with a phonological representation resulting from the perception process and serving as the input to the production process.

The perceptual part must allow for variability, both in how frequently perception is "successful", i.e. in identifying the ill-formed sequence, and in the number and frequency of potential alterations. This can be achieved through a perceptual module in which the phonetic analysis of the acoustic input does indeed activate the targetlike representation: for example, frication spread over a wide range, but centring around 4000Hz, followed by silence and a burst of noise followed by rising vowel formants will activate the representation /fp/. However, since this is not of possible onset sequences, other, legal, representations will also be activated, which the acoustic form is compatible with - less so, but within a certain space of probability. So, in this case other possible representations are a legal fricative-stop onset, /sp/, an epenthesised version /fəp/ retaining both consonants, but placing them in different syllables, or a singleton /p/, ignoring the relatively weak frication at the beginning. This means, an ill-formed input creates a degree of uncertainty which results in a set of possible representations rather than a single unambiguous one. These representations must be "close enough" to make a connection to auditory input plausible. This is a stage in the process where the concept of similarity comes in, deciding which representations' realisation is similar enough to the phonetic input. The ill-formed and the alternative representations are in competition, and the former can override the uncertainty for example through high salience. For example, compare the targets /zv/ and /vz/: for /vz/, /v/ is in the weak pre-consonantal position and shows comparatively weaker frication than sibilant /z/. Therefore, the alternative interpretation of singleton /z/ is strongly activated, and /vz/ is virtually never perceived. In the reverse case, /v/ is in pre-vocalic position, and sibilant /z/ is salient even pre-consonantally. Hence, it overrides the problems of ill-formedness and is perceived correctly in over 50% of tokens. The winning competitor among the possible representations is clearly not unambiguous; individual activations may vary in strength, which leads to variable outcomes. However, in general it appears that certain rankings are possible: for example, the experiment results suggest that voicing differences are of lesser weight, and hence representations with voicing changes are frequently the largest group.

As a result of perception, the input to the speech production phonology may or may not be well-formed in the native language. If it is ill-formed, the phonology must be such that some adaptations are made, but in other cases the ill-formed sequence is actually produced, since there are more adaptations in the shadowing than the production task, but clusters are produced targetlike in up to 50% of tokens for a given cluster. A phonology which can achieve this is very similar to Steriade's P-map, which ranks faithfulness constraints in terms of similarity such that the change with the highest similarity is lowest among the constraints. However, two modifications are necessary. Firstly, the P-map, i.e. similarity, must be conceived of as influenced by the native language, such that certain rankings are the same cross-linguistically (e.g. the constraint against deleting a pre-consonantal consonant being lower than against deleting a pre-vocalic consonant) while others differ (e.g. a constraint against voicing changes being lower ranked in Russian compared to English). Secondly, the notion of variability must be incorporated in two ways. One, there must be a rise of faithfulness with respect to markedness constraints to an extent that constraints overlap, resulting in a mixed output of ill-formed sequences and adaptations. Two, overlapping constraints must be allowed within the P-map to allow for variation within the adaptations. For example, constraints against devoicing and insertion of a vowel may overlap such that a perceived input /zv/ may be adapted as either /sf/ or /zəv/.

The largest blank space in this outline is similarity, both in the perception and production parts of the process. For this reason, part of the spadework for such a model must be further studies of perceived similarity, and its determinants on phonetic, segmental and suprasegmental levels.

Appendix A

Possible clusters in English and Russian

The following table gives the possible two-consonant onset clusters in English and Russian. Note that the table only lists the clusters which are relevant to the present thesis; for this reason, the set is constrained to clusters formed from segments which are in both languages' inventories, and do not contain semivowels. Clusters arising from prepositions and nouns, i.e. onsets of prosodic rather than lexical Russian words (cf. chapter 2), are marked in the same way as lexical onset clusters.

The clusters are given as combinations of first (C_1) and second elements (C_2), such that each column shows the combinations of a given first element with all possible second elements; i.e. the crossing of column s and row p signifies the cluster /sp/.

The notation for the clusters is as follows:

+: wellformed in both languages

(+): wellformed in Russian, but marginal in English

R: only wellformed in Russian

(R): occur in Russian as palatalised or at the beginning of larger clusters (Zalizniak 1977)

The diagonal pattern within the obstruent-obstruent section (broken by /v/-second clusters) illustrates the principle of voicing assimilation in Russian described in chapter 2. The combinations of /t/ with /s, ʃ/, which are marked in the table, are affricates.

C_2 \ C_1	p	b	t	d	k	g	f	v	s	z	ʃ	ʒ	m	n	l	r
p			(R)		R		R		+		(+)					
b						R		R		R		R			(R)	
t	R				R		R		+		(+)					R
d		R				R		R		R		R				R
k			R		R		R		+		R				R	
g						R		R		R		R	R	R	R	
f					R		R		(+)				(R)			
v			R	R	R	R		R	R	R	R	R				R
s	R		R		R		R		R				(R)			
z		R		R		R		R		R			(R)			
ʃ	R		R		R		R		R				R			
ʒ				R		R		R		R		R			(R)	R
m			(R)		R	(R)	(R)	R	+	(R)	(+)	R				
n	R			R	R	R		R	+	R	(+)	(R)	R			
l	+	+	R	R	+	+	+	R	+	R	(+)		R			
r	+	+	+	+	+	+	+	R	R	R	+	R	R	R		

Appendix B

List of stimuli for production experiment

dvʌˈdʲel	dvɛˈlot	dviˈrja	dvʌˈnʲets	dvuˈrit
fpʌˈjavə	fpɛˈvoj	fpiˈlʲir	fpʌˈlagə	fpuˈrot
frʌˈnut	frɛˈtʃok	friˈʃije	frʌˈboj	fruˈdost
pnʌˈvʲest	pnɛˈbitʃ	pniˈdat	pnʌˈrʲist	pnuˈloʃ
psʌˈrʲest	psɛˈdʲinə	psiˈtok	psʌˈvod	psuˈlut
ptʌˈrʲinə	ptɛˈjatkə	ptiˈnomkə	ptʌˈlat	ptuˈzʲit
sfʌˈjot	sfɛˈmu	sfiˈbʲil	sfʌˈritʲe	sfuˈda
ʃmʌˈdoj	ʃmɛˈkanjə	ʃmiˈlotkə	ʃmʌˈvʲek	ʃmuˈnʲiskʲi
ʃpʌˈvotə	ʃpɛˈgorə	ʃpiˈrʲenːə	ʃpʌkˈno	ʃpuˈnʲir
ʃvʌˈsol	ʃvɛˈtro	ʃviˈtʃinkə	svʌnˈjota	ʃvurˈbʲina
spʌvˈno	spɛnˈkajə	spiˈmʲir	spʌˈtʃal	spuˈtʲor
srʌˈvotkə	srɛˈskol	sriˈnʲije	srʌˈjut	sruˈmʲot
svʌˈdom	svɛˈrʲit	sviˈloj	svʌˈglut	svuˈʃal
tvʌˈnut	tvɛˈnajə	tviˈgoj	tvʌnˈdaʃ	tvuˈrʲot
vbʌnʌˈja	vbɛˈkʲist	vbiˈnar	vbʌlˈto	vbuˈtʃist
vlʌˈskajə	vlɛˈdʲil	vliˈgost	vlʌˈsʲibə	vluˈbʲist
vrʌˈdjenka	vrɛˈbʲiskʲi	vriˈnotʃ	vrʌˈʃat	vruˈjos
vzʌˈna	vzɛˈjut	vziˈdoj	(vzʌˈdajə)	vzuˈgof
zbʌˈnor	zbɛˈroʃ	zbiˈnat	zbʌˈtʃjevə	zbuˈmʲa
ʒbʌˈtra	ʒbɛˈdom	ʒbiˈdat	ʒbʌˈmʲil	ʒbuˈsʲevə
zrʌˈtavə	zrɛˈnaskə	zriˈpal	zrʌˈkʲiva	zruˈbi
zvʌˈnʲetskə	zvɛˈvoje	zviˈtajə	zvʌˈrul	zvuˈʃat

Appendix C

Production experiment - further results

The following tables show the remaining results for the production experiment which were not recorded in the results graphs in chapter 3. These are the outcomes with the smallest overall shares in the productions. They are also not regarded as improvements with regard to English. The categories are:

- segment changes resulting in a cluster illegal in English, with the same sonority distance as the target cluster, for example /ʒb/ for /zb/
- segment changes resulting in a cluster illegal in English, with a lower sonority distance than the target cluster, for example /zr/ for /sr/
- further tokens which are unclassifiable, unanalysable or have added consonants, for example /tsf/ for /sf/

The tables give the percentages for each category as well as the sum of these non-improving outcomes. The values are given individually for each target cluster, as well as summaries by target legality and overall. There are three tables, one for every task of the production experiment.

Shadowing task - Sentence condition				
Cluster	Same sonority distance	Lower sonority distance	Other	Sum
fr	0	4	2	6
sp	0	0	6	6
legal	0	2	4	6
ʃm	0	2	0	2
sf	0	0	11	11
ʃp	0	0	2	2
marginal	0	1	4	5
pn	0	0	4	4
sr	0	17	7	24
zr	2	10	10	22
vr	2	0	0	2
vl	0	0	4	4
tv	5	43	0	48
dv	6	12	5	23
ps	6	0	7	13
ʃv	2	29	9	40
sv	0	12	18	30
zv	0	5	13	18
vz	0	0	2	2
pt	0	0	2	2
fp	0	0	4	4
vb	0	0	2	2
zb	4	0	11	15
ʒb	0	0	5	5
illegal	2	7	5	14
overall	2	6	4	12

Shadowing task - Word condition				
Cluster	Same sonority distance	Lower sonority distance	Other	Sum
fr	0	4	0	4
sp	0	0	5	5
legal	0	2	3	5
ʃm	0	0	0	0
sf	4	0	20	24
ʃp	0	0	7	7
marginal	1	0	9	10
pn	0	0	2	2
sr	0	5	9	14
zr	0	2	9	11
vr	0	0	0	0
vl	0	0	0	0
tv	0	47	0	47
dv	13	4	5	22
ps	4	0	9	13
ʃv	2	34	5	41
sv	5	5	17	27
zv	4	16	18	38
vz	3	0	2	5
pt	0	0	2	2
fp	0	0	0	0
vb	0	0	0	0
zb	2	0	4	6
ʒb	0	0	0	0
illegal	2	7	4	13
overall	2	5	4	11

Orthography task				
Cluster	Same sonority distance	Lower sonority distance	Other	Sum
fr	0	9	0	9
sp	0	2	0	2
legal	0	5	0	5
ʃm	0	1	0	1
sf	0	0	11	11
ʃp	0	0	0	0
marginal	0	1	3	4
pn	4	0	0	4
sr	0	6	9	15
zr	0	2	0	2
vr	0	0	0	0
vl	0	0	0	0
tv	0	50	0	50
dv	0	0	0	0
ps	0	2	1	3
ʃv	10	6	0	16
sv	0	15	7	22
zv	0	0	2	2
vz	2	9	0	11
pt	0	0	4	4
fp	0	3	0	3
vb	2	1	0	3
zb	0	0	3	3
ʒb	5	0	0	5
illegal	1	6	1	8
overall	1	5	1	7

Appendix D

List of onsets for English listeners' grammaticality judgments

CC			CəC	C
bd	mr	sv	dəv	b
bl	nr	tf	fəp	d
bn	pn	ts	fər	f
bz	ps	tv	pən	l
dl	pt	vb	pəs	m
dv	sf	vd	pət	n
dz	ʃf	vg	ʃəm	p
fk	ʃk	vl	ʃəp	r
fl	ʃl	vm	ʃəv	s
fn	ʃm	vn	səf	ʃ
fp	ʃn	vr	səp	t
fr	ʃp	vz	sər	v
fs	ʃr	zb	səv	z
ft	ʃt	zd	təv	ʒ
gd	ʃv	zg	vəb	
gl	sk	ʒb	vəl	
gn	sl	ʒd	vər	
gv	sm	ʒm	vəz	
kn	sn	ʒv	ʒəb	
kt	sp	zl	zəb	
kv	sr	zr	zər	
mn	st	zv	zəv	

Appendix E

List of comparisons for English listeners' similarity judgments

C_1C_2-C_2	C_1C_2-C_1	C_1C_2-C_1əC_2	C_1C_2-C_3C_4	
dv-v	dv-d	dv-dəv	pt-ft	ʒv-ʃv
fp-p	fp-f	fp-fəp	pt-ps	zr-sr
fr-r	fr-f	fr-fər	vl-sl	zv-sv
pn-n	pn-p	pn-pən	vr-sr	ʃv-ʃf
ps-s	ps-p	ps-pəs	ʒb-sp	sv-sf
pt-t	pt-t	pt-pət	fp-sp	tv-tf
sf-f	sf-s	sf-səf	ps-ts	dv-tf
ʃm-m	ʃm-ʃ	ʃm-ʃəm	ʃm-sm	ps-bz
ʃp-p	ʃp-ʃ	ʃp-ʃəp	ʃp-sp	pt-bd
ʃv-v	ʃv-ʃ	ʃv-ʃəv	ʃv-sv	vb-fp
sp-p	sp-s	sp-səp	sr-fr	vz-fs
sr-r	sr-s	sr-sər	sr-ʃr	zb-sp
sv-v	sv-s	sv-səv	vb-zb	ʒb-ʃp
tv-v	tv-t	tv-təv	vl-zl	zv-sf
vb-b	vb-v	vb-vəb	vr-zr	
vl-l	vl-v	vl-vəl	ʒb-zb	
vr-r	vr-v	vr-vər	dv-dz	
vz-z	vz-v	vz-vəz	dv-tv	
zb-b	zb-z	zb-zəb	pn-bn	
ʒb-b	ʒb-ʒ	ʒb-ʒəb	vl-fl	
zr-r	zr-z	zr-zər	vr-fr	
zv-v	zv-z	zv-zəv	ʒm-ʃm	

Appendix F

List of comparisons for cross-lingustic similarity judgments

C_1C_2-C[1]			C_1C_2-C_1əC_2
bl-l(LL)	rt-r(IL)	zn-z(IL)	bl-bəl(LL)
br-r(LL)	sf-s(M)	zr-r(IL)	br-bər(LL)
dr-r(LL)	sl-s(LL)	zr-z(IL)	dr-dər(LL)
fl-l(LL)	sm-s(LL)	zv-v(IL)	dv-dəv(IL)
fr-r(LL)	sn-s(LL)	zv-z(IL)	fl-fəl(LL)
gr-r(LL)	sp-p(LL)		fr-fər(LL)
kf-k(IL)	sp-s(LL)		gr-gər(LL)
km-k(IL)	sr-r(IL)		kr-kər(LL)
kp-k(IL)	sv-s(IL)		kt-kət(IL)
kr-r(LL)	tr-r(LL)		ml-məl(IL)
ml-l(IL)	vb-v(IL)		pr-pər(LL)
ml-m(IL)	vd-v(IL)		rd-rəd(IL)
mn-m(IL)	vl-l(IL)		rt-rət(IL)
mn-n(IL)	vl-v(IL)		sf-səf(M)
mr-r(IL)	vm-v(IL)		sr-sər(IL)
nr-r(IL)	vn-v(IL)		sv-səv(IL)
pl-l(LL)	vr-r(IL)		tr-tər(LL)
pn-n(IL)	vz-v(IL)		tv-təv(IL)
pr-r(LL)	vz-z(IL)		vl-vəl(IL)
ps-p(IL)	zd-z(IL)		vr-vər(IL)
ps-s(IL)	zl-z(IL)		zr-zər(IL)
rd-r(IL)	zm-z(IL)		zv-zəv(IL)

C_1C_2-C_3C_4			C_1əC_2-C_3əC_4
bl-dl(IL)	fr-sr(IL)	sf-sv(M)	bəl-məl(LL)
bl-ml(IL)	fr-vr(IL)	ʃk-sk(IL)	dəv-təv(LL)
bl-pl(LL)	ft-ʃt(M)	ʃl-sl(M)	fəl-vəl(LL)
br-dr(LL)	ft-st(IL)	ʃp-sp(M)	fər-vər(LL)
br-mr(IL)	gr-kr(LL)	ʃr-sr(IL)	kəp-kət(LL)
br-pr(LL)	gv-gz(II)	ʃt-st(M)	səv-zəv(LL)
dr-nr(IL)	kf-ks(II)	sl-sn(LL)	
dr-tr(LL)	kf-kʃ(II)	sm-sv(IL)	
dv-dz(II)	km-kn(II)	sr-zr(II)	
dv-gv(II)	kp-kt(II)	sv-zv(II)	
dv-tv(II)	ks-kʃ(II)	vb-vm(II)	
dz-gz(II)	ml-mn(II)	vd-vg(II)	
fk-ʃk(II)	ml-vl(II)	vd-zd(II)	
fk-sk(IL)	mr-vr(II)	vg-zg(II)	
fl-ʃl(M)	nr-zr(II)	vl-zl(II)	
fl-sl(LL)	pl-tl(IL)	vn-vz(II)	
fl-vl(IL)	pr-tr(LL)	vr-zr(II)	
fp-sp(IL)	ps-pʃ(II)	zd-zg(II)	
fr-ʃr(LL)	rd-rt(II)	zd-zn(II)	

[1] All pairs are marked for legality in English. LL stands for both initial sequences being legal in English, LL for both being illegal, and IL for a mixed pair. A pair marked with M contains one marginal English cluster.

Appendix G

Synthesiser specifications for [fpo] and [spo] in the categorical perception experiment

The following tables show the parameters changed for the end points of the [fpo]-[spo] continuum. The header row describes the segments or segment portions, as they correspond to the timeline (second row). The following columns describe the values of the manipulated parameters at any give time. The parameters are amplitude voicing (av), amplitude frication (af), open quotient (oq), fundamental frequency (f0), formants (f1-f6), amplitude of formants (a2f-a6f), and bypass path amplitude (ab). *d* stands for the default value of the synthesiser. The first table gives the values for [spo] (i.e. one end point of the continuum), the second one the values for [fpo] (the other end point).

par	fricative					closure		burst			vowel										
time	0	10	100	115	120	125	225	230	235	240	245	250	255	270	290	330	360	430	495	496	500
av	0									0	60				62			55	0		0
af	0	22	40	22		0	0	70	60		60	50	0								0
oq	50						50	60													60
f0	900								900								800				900
f1	d							d	200			300		350							350
f2	d							d	720					800			850				850
f3	2000				2000	2100	2300									2600					2600
f4	3800				3800	d															d
f5	5800				5800	4500					4500	d									d
f6	7500				7500	d															d
a2f	5				5	0															0
a3f	60				60	0															0
a4f	65				65	0															0
a5f	80				80	6						6	0								0
a6f	80				80	0															0
ab	5				5	50						50	0								0

par	fricative					closure		burst			vowel										
time	0	10	100	115	120	125	225	230	235	240	245	250	255	270	290	330	360	430	495	496	500
av	0									0	60				62			55	0		0
af	0	22	40	22		0	0	70	60		60	50	0								0
oq	50						50	60													60
f0	900								900								800				900
f1	d							d	200			300		350							350
f2	d							d	720					800			850				850
f3	2500				2500	2100	2300									2600					2600
f4	4000				4000	d															d
f5	6000				6000	4500					4500	d									d
f6	7500				7500	d															d
a2f	65				65	0															0
a3f	65				65	0															0
a4f	60				60	0															0
a5f	50				50	6						6	0								0
a6f	60				60	0															0
ab	5				5	50						50	0								0

References

Adler, A. (2004), 'Perception in loanword adaptation: a study of English loans into Hawaiian'. Paper presented at the Manchester Phonology Meeting.

Adler, A. (2006), 'Faithfulness and perception in loanword adaptation: A case study from Hawaiian', *Lingua* **116**(7), 1024–1045.

Algeo, J. (1978), 'What consonant clusters are possible?', *Word* **29**(3), 206–224.

Anttila, A. (1997), Deriving variation from grammar: a study of Finnish genitives, *in* F. Hinskens, R. v. Hout & W. L. Wetzels, eds, 'Variation, change and phonological theory', Amsterdam: John Benjamins.

Archangeli, D. (1988), 'Aspects of underspecification theory', *Phonology* **5**, 183–208.

Archibald, J., ed. (1998), *Second Language Phonology*, Amsterdam: John Benjamins.

Avanesov, R. (1956), *Fonetika sovremennogo literaturnogo russkogo jazyka*, Moscow: Izdatel'stvo Moskovskogo Unversitet.

Baddeley, A. (1992), 'Working memory', *Science* **255**, 556–559.

Baddeley, A. D. (1986), *Working memory*, New York: Oxford University Press.

Bard, E. G., Robertson, D. & Sorace, A. (1996), 'Magnitude estimation of linguistic acceptability', *Language* **72**(1), 32–68.

Berkley, D. M. (2001*a*), Gradient Obligatory Contour Principle Effects, PhD thesis, Northwestern University.

Berkley, D. M. (2001*b*), 'The OCP and gradient data', *Studies in the Linguistic Sciences* **24**(1-2), 59–72.

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-language research', Timonium, MD: York Press.

Best, C. T., McRoberts, G. W. & Sithole, N. M. (1988), 'Examination of perceptual reorganisation for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants', *Journal of Experimental Psychology: Human Perception and Performance* **14**(3), 345–360.

Best, C. T., Traill, A., Carter, A., Harrison, K. D. & Faber, A. (2003), !Xoo click perception by English, Isizulu, and Sesotho listeners, *in* M. J. Sole, D. Recasens & J. Romero, eds, 'Proceedings of the International Congress of Phonetic Sciences', Barcelona: UAB.

Blevins, J. (1995), The syllable in phonological theory, *in* J. A. Goldsmith, ed., 'The handbook of phonological theory', Oxford: Blackwell, pp. 206–244.

Bloomfield, L. (1933), *Language*, New York: H. Holt & Company.

Blumstein, S. E. & Stevens, K. N. (1981), 'Phonetic features and acoustic invariance in speech', *Cognition* **10**(1-3), 25–32.

Boersma, P. (1998), *Functional Phonology*, University of Amsterdam.

Boersma, P. (2000), *The OCP in perception grammar*. Ms., University of Amsterdam.

Boersma, P. (2004), 'Separating the production task and the filtering task of the grammar in a bidirectional model of phonology'. Talk presented at Linguistic Circle, Edinburgh.

Boersma, P. & Hayes, B. (2001), 'Empirical tests of the gradual learning algorithm', *Linguistic Inquiry* **32**(1), 45–86.

Bolla, K. (1981), *A conspectus of Russian speech sounds*, Cologne, Vienna: Boehlau.

Bondarko, L. V. (1977), *Zvukovoj stroj sovremennogo russkogo jazyka*, Moscow: Prosveshchenie.

Brasington, R. (1981), Epenthesis and deletion in loan phonology, *in* 'Work in Progress, University of Reading Phonetics Lab', pp. 97–103.

Broe, M. (1993), Specification theory: the treatment of redundancy in generative phonology, PhD thesis, University of Edinburgh.

Broecke, M. v. d. (1976), *Hierarchies and Rank Orders in Distinctive Features*, Assen, Amsterdam: Van Gorcum.

Broselow, E. (2005), 'Stress adaptation in loanword phonology: perception and learnability'. Talk presented at DGfS Workshop *Speech perception in or outside phonology?*

Broselow, E. & Finer, D. (1991), 'Parameter setting in second language phonology and syntax', *Second Language Research* **7**(1), 35–59.

Bulanin, L. L. (1970), *Fonetika sovremennogo russkogo jazyka*, Moscow: Visshaja Shkola.

Bush, C. N. (1964), *Phonetic variation and acoustic distinctive features: a study of four General American fricatives*, The Hague: Mouton.

Bybee, J. L. (2001), *Phonology and language use*, Cambridge: Cambridge University Press.

Chew, P. A. (1999), A computational phonology of Russian, PhD thesis, University of Oxford.

Chomsky, N. & Halle, M. (1968), *The sound pattern of English*, Cambridge: Cambridge University Press.

Clements, G. N. (1990), The role of the sonority cycle in core syllabification, *in* 'Papers in laboratory phonology I: between the grammar and physics of speech', Cambridge: Cambridge University Press, pp. 282–333.

Coleman, J. & Pierrehumbert, J. (1997), 'Stochastic phonological grammars and acceptability', *Computational Phonology* **3**, 49–56.

Côté, M.-H. (1997*a*), Phonetic salience and consonant cluster simplification, *in* 'MIT Working Papers in Linguistics', pp. 229–262.

Côté, M.-H. (1997*b*), Phonetic salience and the OCP in coda cluster reduction, *in* 'Papers from the Regional Meetings, Chicago Linguistic Society', pp. 57–71.

Côté, M.-H. (2001), Consonant Cluster Phonotactics: a Perceptual Approach, PhD thesis, MIT.

Davidson, L. (2000), 'The effects of hidden rankings in the acquisition of consonant clusters'. Paper presented at the New Sounds Conference, Amsterdam.

Davidson, L. (2001), Hidden rankings in the final state of the English grammar, *in* G. Horwood & K. Se-Kyung, eds, 'Rutgers Linguistics Working Papers 2', New Brunswick: Department of Linguistics, Rutgers University, pp. 21–48.

Donselaar, W. v., Kujpers, C. & Cutler, A. (1999), 'Facilitatory effects of vowel epenthesis on word processing in Dutch', *Journal of Memory and Language* **41**, 59–77.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C. & Mehler, J. (1999), 'Epenthetic vowels in Japanese: a perceptual illusion?', *Journal of Experimental Psychology: Human Perception and Performance* .

Dyen, I., Kruskal, J. B. & Black, P. (1992), 'An Indoeuropean classification: a lexicostatistical experiment', *Transactions of the American Philosophical Society* **82**.

Eckman, F. R. & Iverson, G. K. (1993), 'Sonority and markedness among onset clusters in the interlanguage of ESL learners', *Second Language Research* **9**(3), 234–252.

Ellison, T. M. & Kirby, S. (2006), 'Measuring language divergence by intra-lexical comparison', *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL* pp. 273–280.

Elman, J. L., Diehl, R. L. & Buchwald, S. E. (1977), 'Perceptual switching in bilinguals', *Journal of the Acoustical Society of America* **62**, 971–974.

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

Fant, G. (1970), *Acoustic Theory of Speech Production. With calculations based on X-Ray studies of Russian articulations*, The Hague: Mouton.

Fant, G. (1973), *Speech Sounds and Features*, Cambridge: MIT Press.

Felsenstein, J. (2001), *PHYLIP: phylogeny inference package, version 3.6 (alpha)*, Seattle: University of Washington.

Firth, J. R. (1948), 'Sounds and prosodies', *Transactions of the Philological Society* **46**, 127–152.

Flege, J. E. (1987*a*), Effects of equivalence classification on the production of foreign language speech sounds, *in* J. Allan & J. Leather, eds, 'Sound patterns in second language acquisition', Dordrecht, Providence: Foris.

Flege, J. E. (1987*b*), 'The production of new and similar phones in a foreign language: evidence for the effect of equivalence classification', *Journal of Phonetics* **15**, 47–57.

Flege, J. E. (1993), 'Production and perception of a novel, second-language phonetic contrast', *The Journal of the Acoustical Society of America* **93**(3), 1589–1608.

Flege, J. E. (1995), Second language speech learning: theory, findings, and problems, *in* W. Strange, ed., 'Speech perception and linguistic experience: issues in cross-language research', Timonium, MD: York Press.

Fleischhacker, H. (1999), 'The location of epenthetic vowels with respect to consonant clusters: An auditory similarity account'. Poster presented at the ICPhS Perception in Phonology Meeting.

Fleischhacker, H. (2001), Cluster-dependent epenthesis asymmetries, *in* A. Albright & T. Cho, eds, 'UCLA Working Papers in Linguistics 7, Papers in Phonology 5', pp. 71–116.

Flemming, E. (1997), Phonetic detail in phonology: towards a unified account of assimilation and coarticulation, *in* K. Suzuki & D. Elzinga, eds, 'Proceedings of the 1995 Southwestern Workshop in Optimality Theory (SWOT)', University of Arizona.

Flemming, E. (2001), 'Scalar and categorical phenomena in a unified model of phonetics and phonology', *Phonology* **18**(1), 7–44.

Flemming, E. (2003), Contrast and perceptual distinctiveness, *in* B. Hayes, D. Steriade & R. Kirchner, eds, 'Phonetically based phonology', Cambridge: Cambridge University Press.

Flemming, E. (2005), Speech perception and phonological contrast, *in* D. Pisoni & R. Remez, eds, 'The handbook of speech perception', Oxford: Blackwell.

Frisch, S. (1996), Similarity and Frequency in Phonology, PhD thesis, Northwestern University.

Frisch, S. A., Large, N. R. & Pisoni, D. B. (2000), 'Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords', *Journal of Memory and Language* **42**(4), 481–496.

Frisch, S., Pierrehumbert, J. B. & Broe, M. B. (2004), 'Similarity avoidance and the OCP', *Natural Language and Linguistic Theory* **22**(1), 179–228.

Giegerich, H. (1992), *English phonology: an introduction*, Cambridge: Cambridge University Press.

Greenberg, J. & Jenkins, J. J. (1964), 'Studies in the psychological correlate of the sound system of American English', *Word* **20**, 157–177.

Halle, M. (1959), *The sound patterns of Russian: a linguistic and acoustical investigation ; with an excursus on the contextual variants of the Russian vowels*, The Hague: Mouton.

Halle, M. (1997), 'On stress and accent in Indo-European', *Language* **73**, 275–313.

Halle, M., Hughes, G. W. & Radley, J. P. (1957), 'Acoustic properties of stop consonants', *Journal of the Acoustical Society of America* **29**, 107–116.

Hallé, P. A., Best, C. T. & Bachrach, A. (2003), Perception of /dl/ and /tl/ clusters: a cross-linguistic perceptual study with French and Israeli listeners, *in* M. J. Sole, D. Recasens & J. Romero, eds, 'Proceedings of the International Congress of Phonetic Sciences', Barcelona: UAB.

Hallé, P. A., Segui, J., Frauenfelder, U. & Meunier, C. (1998), 'Processing of illegal consonant clusters: a case of perceptual assimilation?', *Journal of Experimental Psychology: Human Perception and Performance* **24**(2), 592–608.

Hammond, M. (1999), *The phonology of English: a prosodic optimality theoretic approach*, Cambridge: Cambridge University Press.

Hanson, K. (1999), 'Formal variation in the rhymes of Robert Pinskys *The Inferno of Dante*'. Ms., University of Berkeley.

Harris, J. W. (1983), *Syllable structure and stress in Spanish: a nonlinear analysis*, Cambridge: MIT Press.

Haugen, E. (1950), 'The analysis of linguistic borrowing', *Language* **26**, 210–231.

Hayes, B. (2000), Gradient wellformedness in linguistic data, *in* J. Dekkers, F. v. d. Leeuw & J. v. d. Weijer, eds, 'Optimality theory: phonology, syntax and acquisition', Oxford: Oxford University Press, pp. 88–120.

Hayes, B. & MacEachern, M. (1998), 'Quatrain form in English folk verse', *Language* **74**, 473–507.

Hayes, B. & Steriade, D. (2003), The phonetic bases of markedness, *in* B. Hayes, D. Steriade & R. Kirchner, eds, 'Phonetically based phonology', Cambridge: Cambridge University Press.

Hermans, B. (2001), Yer vocalisation in suffixed forms; the case of Russian, *in* M. van Oostendorp & E. Anagnostopoulou, eds, 'Progress in grammar: articles at the 20th anniversary of the comparison of grammatical models group in Tilbourg', http://www.roquade.nl/meertens/progressingrammar/hermans.pdf.

Hume, E. (1998), 'Metathesis in phonological theory: The case of Leti', *Lingua* **104**(3-4), 147–186.

Hume, E. (1999), The role of perceptibility in consonant/consonant metathesis, *in* K. Shahin, S. Blake & E.-S. Kim, eds, 'The Proceedings of the 17th West Coast Conference on Formal Linguistics', Stanford, CA: Center for the Study of Language and Information, pp. 293–307.

Hume, E. & Johnson, K. (2001*a*), A model of the interplay of speech perception and phonology, *in* 'Ohio State University Working Papers in Linguistics', pp. 1–22.

Hume, E. & Johnson, K. (2003), The impact of partial phonological contrast on speech perception, *in* M. J. Sole, D. Recasens & J. Romero, eds, 'Proceedings of the International Congress of Phonetic Sciences', Barcelona: UAB.

Hume, E. & Johnson, K., eds (2001*b*), *The role of perception in phonology*, Orlando, London: Academic Press.

Hume, E. & Odden, D. (1996), 'Reconsidering [consonantal]', *Phonology* **13**, 345–376.

Ito, J. & Mester, A. (1999), The structure of the phonological lexicon, *in* N. Tsujimura, ed., 'The handbook of Japanese linguistics', Oxford: Blackwell.

Ito, J. & Mester, A. (2001), 'Covert generalisation in optimality theory: the role of stratal faithfulness constraints', *Studies in Phonetics, Phonology and Morphology* **7**(2), 273–299.

Jacobs, H. & Gussenhoven, C. (2000), Loan phonology: perception, salience, the lexicon and OT, *in* J. Dekkers, F. v. d. Leeuw & J. v. d. Weijer, eds, 'Optimality theory: phonology, syntax and acquisition', Oxford: Oxford University Press, pp. 193–210.

Jakobson, R. (1929), *Remarques sur l'evolution phonologique du russe comparee a celle des autres langues slaves*, Prague: Jedotna Ceskoslovenskych Matematiku a Fysiku.

Jakobson, R. (1956), Die Verteilung der stimmhaften und stimmlosen Geräuschlaute im Russischen, *in* M. Woltner & H. Bruer, eds, 'Festschrift für Max Vasmer zum 70. Geburtstag am 28. Februar 1956', Wiesbaden: In Kommission bei Otto Harrassowitz.

Johnson, K. (2004), Cross-linguistic perceptual differences emerge from the lexicon, *in* A. Agwuele, W. Warren & S.-H. Park, eds, ‘Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception’, Sommerville, MA: Cascadilla Press, pp. 26–41.

Jones, D. & Ward, D. (1969), *The phonetics of Russian*, Cambridge: Cambridge University Press.

Jun, J. (1995), Perceptual and articulatory factors in place assimilation: an optimality-theoretic account, PhD thesis, UCLA.

Kager, R. (1999), *Optimality theory*, Cambridge: Cambridge University Press.

Kang, Y. (2003), ‘Perceptual similarity in loanword adaptation: adaptation of English postvocalic word-final stops in Korean’, *Phonology* **20**(2), 219–273.

Keating, P. (1989), The phonetics-phonology interface, *in* F. J. Newmeyer, ed., ‘Linguistics: The Cambridge Survey. Volume 1’, Cambridge: Cambridge University Press, pp. 281–302.

Kenstowicz, M. (2001), ‘The role of perception in loanword phonology’, *Linguistique africaine* **20**.

Kenstowicz, M. (2004), ‘Tone loans: the adaptation of English loanwords into Yoruba’. Paper presented at the 78th Annual Meeting of the Linguistic Society of America.

Kenstowicz, M. (2005), The phonetics and phonology of Korean loanword adaptation, *in* S.-J. Rhee, ed., ‘Proceedings of the First European Conference on Korean Linguistics’.

Kenstowicz, M. (in press), ‘Salience and similarity in loanword adaptation: a case study from Fijian’, *Language Sciences* .

Kenstowicz, M. & Suchato, A. (2006), ‘Issues in loanword adaptation: a case study from Thai’, *Lingua* **116**(7), 921–949.

Khumalo, J. (1984), ‘A preliminary survey of Zulu adoptives’, *African Studies* **43**, 205–216.

Kiparsky, P. (1993), ‘Variable rules’. Talk presented at Rutgers Optimality Workshop.

Kochetov, A. (2006), ‘Testing licensing by cue: A case of Russian palatalized coronals’, *Phonetica* **63**, 113–148.

Kochetov, A. & Goldstein, L. (2001), ‘Competing recoverability factors and intergestural phasing in Russian stop clusters’. Paper presented at the LSA Annual Meeting.

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-language research’, Timonium, MD: York Press.

Kučera, H. & Monroe, G. K. (1968), *A comparative quantitative phonology of Russian, Czech, and German*, New York: American Elsevier.

Ladefoged, P. (2003), *A course in phonetics (3rd ed.)*, Orlando: Harcourt Brace.

Lehiste, I. (1964), *Acoustical Characteristics*, The Hague: Mouton.

Lightner, T. M. (1965), Segmental phonology of contemporary standard Russian, PhD thesis, MIT.

Lightner, T. M. (1972), *Problems in the theory of phonology, vol. 1*, Edmonton: Linguistic Research.

Liljencrants, J. & Lindblom, B. (1972), 'Numerical simulation of vowel quality systems: the role of perceptual contrast', *Language* **48**, 839–862.

Lindblom, B. (1975), 'Experiments in sound structure', *Revue de Phonétique Appliquée* **51**, 155–189.

Lodge, M. (1981), *Magnitude scaling: quantitative measurement opinions*, Beverly Hills, London: Sage.

Lombardi, L. (2000), 'Second language data and constraints on manner: explaining substitutions for the English interdentals'.

Lukatela, G. & Turvey, M. (1990), 'Phonemic similarity effects and prelexical phonology', *Memory and Cognition* **18**(2), 128–152.

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

Massaro, D. W. & Cohen, M. M. (1983), 'Phonological context in speech perception', *Perception and Psychophysics* **34**(4), 338–348.

Matthews, J. & Brown, C. (1998), Qualitative and quantitative differences in the discrimination of second language speech sounds, *in* 'Proceedings of the Annual Boston University Conference on Language Development', pp. 499–510.

McMahon, A. & McMahon, R. (2003), 'Finding families: Quantitative methods in language classification', *Transactions of the Philological Society* **101**, 7–55.

Mennen, I. (to appear), Phonological and phonetic influences in non-native intonation, *in* J. Trouvain & U. Gut, eds, 'Non-native prosody: phonetic description and teaching practice', Berlin: Mouton de Gruyter.

Miao, R. (2005), Loanword adaptation in Mandarin Chinese: Perceptual, phonological and sociolinguistic factors, PhD thesis, Stony Brook University.

Miller, G. & Nicely, P. (1955), 'An analysis of perceptual confusions among some English consonants', *Journal of the Acoustical Society of America* **27**, 338–352.

Morais, J. & Kolinsky, R. (1994), 'Perception and awareness in phonological processing: the case of the phoneme', *Cognition* **50**, 287–297.

Moreton, E. (2002), 'Structural constraints in the perception of English stop-sonorant clusters', *Cognition* **84**(1), 55–71.

Mueller, S. T., Seymour, T. L., Kieras, D. E. & Meyer, D. E. (2003), 'Theoretical implications of articulatory duration, phonological similarity and phonological complexity in verbal working memory', *Journal of Experimental Psychology: Learning, Memory and Cognition* **29**(6), 1353–1380.

Olive, J. P., Greenwood, A. & Coleman, J., eds (1993), *Acoustics of American English speech*, New York: Springer.

Padgett, J. & Zygis, M. (2003), The evolution of sibilants in Polish and Russian, *in* T. A. Hall & S. Hamann, eds, 'Papers in Phonology and Phonetics. ZAS Papers in Linguistics', Vol. 23, Berlin: ZAS, pp. 155–174.

Paradis, C. (1995), Derivational constraints in phonology: Evidence from loanwords and implications, *in* 'Papers from the Regional Meetings, Chicago Linguistic Society', pp. 360–374.

Paradis, C. (2005), 'The unnatural /Cju/ (<foreign /Cy/) sequence in Russian loanwords: a problem for the perceptual view', *Lingua* **116**(7), 976–995.

Paradis, C. & LaCharité, D. (1997), 'Preservation and minimality in loanword adaptation', *Journal of Linguistics* **33**, 379–430.

Peperkamp, S. (2003), 'Towards a new theory of loanword adaptation'. Ms., University of Paris 8.

Peperkamp, S. & Dupoux, E. (2001), 'Loanword adaptations: three problems for phonology (and a psycholinguistic solution)'.

Peperkamp, S., Dupoux, E., Nakamura, K. & Vendelin, I. (2003), 'Reinterpreting loanword adaptations: the role of perception'. Proceedings of the International Congress of Phonetic Sciences.

Peperkamp, S., Dupoux, E., Nakamura, K. & Vendelin, I. (2004), 'A psycholinguistic theory of loanword adaptations'. Talk presented at the 30th Annual Meeting of the Berkeley Linguistics Society.

Pierrehumbert, J. B. (1990), 'Phonological and phonetic representation', *Journal of Phonetics* **18**, 375–394.

Pierrehumbert, J. B. (2000), 'The phonetic grounding of phonology', *Bulletin de la Communication Parlee* **5**, 7–23.

Pierrehumbert, J. B. (2003), 'Phonetic diversity, statistical learning, and acquisition of phonology', *Language and Speech* **46**(2-3), 115–154.

Pike, K. L. & Fries, C. C. (1949), 'Coexistent phonemic systems', *Language* **25**, 25–50.

Pitt, M. A. (1998), 'Phonological processes and the perception of phonotactically illegal consonant clusters', *Perception and Psychophysics* **60**(6), 941–951.

Prince, A. & Smolensky, P. (1993), Optimality theory: constraint interaction in generative grammar, *in* 'Technical Report, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick NJ, and Computer Science Department, University of Colorado, Boulder'.

Repp, B. (1984), Categorical perception: issues, methods and findings, *in* N. J. Lass, ed., 'Speech and Language, Vol. 10, Advances in Basic Research and Practice', Orlando, London: Academic Press.

Rochet, B. L. (1995), Perception and production of second-language speech sounds by adults, *in* W. Strange, ed., 'Speech perception and linguistic experience: issues in cross-language research', Timonium, MD: York Press.

Rose, Y. (1999), 'A structural account of root node deletion in loanword phonology', *Canadian Journal of Linguistics* **44**(4), 389–404.

Rubach, J. (1986), 'Abstract vowels in three-dimensional phonology: the yers', *The Linguistic Review* **5**, 247–280.

Rubach, J. (2000), 'Backness switch in Russian', *Phonology* **17**, 39–64.

Selkirk, E. O. (1984), On the major class features and syllable theory, *in* M. Aronoff & R. Oerhle, eds, 'Language sound structure', Cambridge: MIT Press.

Shinohara, S. (2004), Emergence of universal grammar in foreign word adaptations, *in* R. Kager, J. Pater & W. Zonneveld, eds, 'Fixing prorities: constraints in phonological acquisition', Cambridge: Cambridge University Press.

Sievers, E. (1881), *Grundzüge der Phonetik*, Leipzig: Breitkopf und Härtel.

Silverman, D. (1992), 'Multiple scansions in loanword phonology: evidence from Cantonese', *Phonology* **9**, 289–328.

Singh, R. (1995), 'Rethinking the relationship between phonology and loan phonology', *Folia Linguistica* **29**(3-4), 395–405.

Skalozub, L. G. (1963), *Palatogrammy i rentgenogrammy soglasnyx fonem Russkogo literaturnogo jazyka*, Kiev: Izdatelstvo Kievskogo Universiteta.

Sorace, A. & Keller, F. (2005), 'Gradience in linguistic data', *Lingua* .

Stampe, D. (1979), *A dissertation on natural phonology*, New York: Garland.

Stemberger, J. P. & Bernhardt, B. H. (1999), The emergence of faithfulness, *in* B. MacWhinney, ed., 'The emergence of language', Mahweh: Erlbaum, pp. 417–446.

Steriade, D. (1982), Greek prosodies and the nature of syllabification, PhD thesis, MIT.

Steriade, D. (1994*a*), 'Licensing by cue'. Ms., UCLA.

Steriade, D. (1994*b*), 'Positional neutralization and the expression of contrast'. Ms., UCLA.

Steriade, D. (2001), 'The phonology of perceptibility effects: the P-map and its consequences for constraint organisation'.

Steriade, D. (2003), Knowledge of perceptual similarity and its phonological uses: evidence from half-rhymes, *in* M. J. Sole, D. Recasens & J. Romero, eds, 'Proceedings of the International Congress of Phonetic Sciences', Barcelona: UAB.

Steriade, D. (2005), 'Perception grammar and loan adaptation'. Talk presented at DGfS Workshop *Speech perception in or outside phonology?*

Stevens, K. N. (1989), 'On the quantal nature of speech', *Journal of Phonetics* **17**(1-2), 47–143;145–157.

Stevens, K. N. (1998), *Acoustic phonetics*, Cambridge: MIT Press.

Stevens, S. S. (1956), 'The direct estimation of sensory magnitudes - loudness', *American Journal of Psychology* **69**, 1–25.

Strange, W., ed. (1995), *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press.

The Oxford English Dictionary (2007), 'OED Online', Oxford: Oxford University Press.

Trask, R. (1996), *A dictionary of phonetics and phonology*, London, New York: Routledge.

Treiman, R., Broderick, V., Tincoff, R. & Rodriguez, K. (1998), 'Children's phonological awareness: confusions between phonemes that differ only in voicing', *Journal of Experimental Child Psychology* **68**(1), 3–21.

Vendelin, I. & Peperkamp, S. (2006), 'The influence of orthography on loanword adaptations', *Lingua* **116**(7), 996–1007.

Vitevitch, M. S. & Luce, P. A. (1998), 'When words compete: Levels of processing in perception of spoken', *Psychological Science* **9**(4), 325–329.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B. & Auer, E. (1999), 'Phonotactics, neighborhood activation, and lexical access for spoken words', *Brain and Language* **68**(1-2), 306–311.

Vitz, P. & Winkler, B. S. (1972), 'Predicting the judged similarity of sound of English words', *Journal of Verbal Learning and Verbal Behavior* **12**, 373–388.

Walden, B. & Montgomery, A. (1975), 'Dimensions of consonant perception in normal and hearing-impaired listeners', *Journal of Speech and Hearing Research* **18**, 445–455.

Weinberger, S. (1990), Minimal segments in second language phonology, *in* J. Leather & A. James, eds, 'New Sounds 90: Proceedings of the 1990 Amsterdam Symposium on the acquisition of second language speech', University of Amsterdam.

Werker, J. F. & Tees, R. C. (1984), 'Cross-language speech perception: evidence for perceptual reorganisation during the first year of life', *Infant Behaviour and Development* **7**, 49–63.

Whitney, W. D. (1865), 'On the relation of vowels and consonants', *Journal of the American Oriental Society* **8**, 347–373.

Wingstedt, M. & Schulman, R. (1988), 'Listeners judgments of simplifications of consonant clusters', *Linguistics* **26**, 105–123.

Yearley, J. (1995), Jer vowels in Russian, *in* J. Beckman, L. Walsh Dickey & S. Urbanczyk, eds, 'Papers in Optimality Theory', Amherst: GLSA.

Yip, M. (1993), 'Cantonese loanword phonology and optimality theory', *Journal of East Asian Linguistics* **2**(3), 261–291.

Yip, M. (2002), 'Perceptual influences in Cantonese loanword phonology', *Journal of the Phonetic Society of Japan* **6**(1), 4–21.

Yip, M. (2006), 'The symbiosis between perception and grammar in loanword phonology', *Lingua* **116**(7), 950–975.

Zalizniak, A. A. (1977), *Grammaticheskij slovar' russkogo jazyka: slovoizmenenie*, Moscow: Russkij jazyk.

Zinder, L. R., Bondarko, L. V. & Berbitskaja, L. A. (1964), Akusticheskaja xarakteristika razlichija tverkyx i mjagkix soglasnyx v russkom jazyke, *in* M. I. Matusevich, ed., 'Voprosy Fonetiki', Leningrad: Leningrad State University.

Zsiga, E. C. (2000), 'Phonetic alignment constraints: consonant overlap and palatalization in English and Russian', *Journal of Phonetics* **28**, 69–102.

Zsiga, E. C. (2003), 'Articulatory timing in a second language: evidence from Russian and English', *Studies in Second Language Acquisition* **25**, 399–432.

Zwicky, A. (1972), Note on a phonological hierarchy in english, *in* R. P. Stockwell & R. K. Macauley, eds, 'Linguistic Change and Generative Theory', Bloomington: Indiana University Press.