INCOMPLETE NEUTRALIZATION AND TASK EFFECTS IN EXPERIMENTALLY-ELICITED SPEECH:

EVIDENCE FROM THE PRODUCTION AND PERCEPTION OF WORD-FINAL DEVOICING IN RUSSIAN.

Viktor Kharlamov

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the PhD degree in Linguistics

Department of Linguistics
Faculty of Arts
University of Ottawa

ABSTRACT

This dissertation investigates the role of grammatical versus methodological influences in the production and perception of final devoicing in experimentally-elicited speech from Russian. It addresses the question of how the partial preservation of the phonological voicing contrast in word-final obstruents is affected by (i) task-independent factors that reflect phonological and lexical properties of stimuli words (underlying voicing, word length, lexical competition) and (ii) task-dependent biases that arise due to the nature of the experimental task performed by the speaker (availability of orthographic inputs, presence of minimal pairs among the stimuli). Results of a series of acoustic production and perceptual identification tasks reveal that task-dependent factors account for the presence of robust and perceptually salient differences in the parameter of phonetic voicing. Several types of stimuli items also show limited but statistically significant differences in closure/frication duration and release duration that are independent of the presence of orthography or inclusion of full minimal pairs among test items. Taken together, these findings indicate that non-grammatical factors can play a prominent biasing role in both production and perception of the voicing contrast in experimentally-elicited speech, such that certain voicing-dependent cues are maintained only in the presence of task-dependent pressures. However, not all incompletely neutralized differences between phonologically voiced versus voiceless final obstruents can be attributed to the effects of orthography or inclusion of minimal pairs among the stimuli. In the theoretical domain, these results are argued to favour a less restrictive definition of neutralization and a model of phonology that views devoicing as a loss of the primary acoustic cue to the underlying voicing contrast rather than complete identity of the [voiced] feature.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis advisor, Marie-Hélène Côté. Jamais je n'aurais pu écrire cette thèse sans son conseil, son encouragement, sa générosité et surtout sa brillance et sa connaissance de la phonologie. Merci beaucoup, Marie-Hélène!

I am also most grateful to the members of my committee: Lev Blumenfeld, Marc Brunelle, Jeff Mielke, and Joe Salmons. With the help of their comments, this thesis has improved tremendously and I have a much better idea of where to take my research next. Спасибо, Лев! Merci, Marc! Thank you, Jeff and Joe!

I would also like to express my gratitude to the participants of my experiments, the faculty of the Linguistics Department at the University of Ottawa, and the many attendees of the conferences at which I presented the preliminary findings of the thesis or the earlier work on methodological effects in syllabification that inspired the current investigation. I also acknowledge the financial support I received from the Social Sciences and Humanities Research Council of Canada and the Ministry of Colleges and Universities of Ontario.

Finally, I thank my family and friends for the encouragement and help I received from them during the years of graduate studies. I dedicate this thesis to my parents, Tatiana and Eduard, for their unconditional support of my academic endeavors.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF APPENDICES

LIST OF ABBREVIATIONS

C	consonant
d	disyllabic
Inf	infinitive
Instr	instrumental case
L1	first language
L2	second language
m	monosyllabic
MP	minimal pair
Nom	nominative case
Orth	orthography
OT	Optimality Theory
Part	participle
pl	plural
RM	repeated measures
RT	reaction time
sg	singular
Subj	subject
V	vowel
voi	voiced
vls	voiceless

CHAPTER 1: INTRODUCTION

1.1. Research Questions and Preview of the Findings

This dissertation examines the influences of grammatical versus methodological factors in the production and perception of final devoicing in laboratory speech. In recent years, formal analyses of phonological phenomena and entire linguistic frameworks have often been grounded in experimentally-elicited data (e.g., Articulatory Phonology; Browman & Goldstein, 1992). Laboratory speech offers many important advantages over spontaneous recordings of the same utterances, including the ability to elicit multiple tokens of each stimulus item and to control for prosodic patterns (Xu, 2010). However, both the production and perception of experimentally-elicited data are known to be sensitive to the nature of the task performed by the speaker and the speaker's access to meta-linguistic information during or before testing (Ohala, 1995a; Niedzielski, 1999; Hay et al., 2009, 2010; Pitt, 2009; Drager et al., 2010; Hay & Drager, 2010; Côté & Kharlamov, 2011; for similar effects in the domain of psychology, see Duncan et al., 1969; Rosnow & Rosenthal, 1997). Thus, the contrasts seen in laboratory speech may not be representative of the grammatical knowledge that plays a role in everyday communication. Instead, the observed effects may be driven by task-dependent biases and may correspond to a subtype of linguistic performance that is relevant only in experimental environments.

The general goal of the current investigation is to advance our understanding of the biasing role of methodological influences in experimentally-elicited devoicing data, including the theoretical consequences of the presence of incompletely neutralized

contrasts. This goal is addressed in a series of acoustic production and perceptual identification tasks that investigate the voicing contrast in word-final obstruents in Russian, a language that has so far received limited attention in the experimental literature on laryngeal neutralization despite being a classical example of a devoicing language.

Crucially to the present discussion, the Russian language is known to neutralize the underlying voicing contrast in phonetic realizations of word-final obstruents, with words such as /kɔt/ 'cat' and /kɔd/ 'code' presumably sharing the same output form [kɔt] (Avanesov, 1956; Wade, 2010). This process is known as 'neutralization' or '(word-final) devoicing' and, in addition to Russian, it is thought to affect final obstruents in German, Dutch, Afrikaans, Catalan, Turkish, Polish, Czech, Slovak and many other languages and dialects (Moulton, 1962; Benni, 1964; Heim, 1976; Wierzchowska, 1980; Hualde, 1992; Donaldson, 1993; Rubach, 1993; Booij, 1995). However, a growing number of experimental studies done in the past 40 years have reported that speakers of neutralizing languages maintain significant durational differences between phonologically voiced versus voiceless obstruents and that listeners can use such cues to infer the intended voicing specification of final segments at above-chance level (among others, Charles-Luce, 1985; Port & O'Dell, 1985; Port & Crawford, 1989, Piroth & Janker, 2004; Warner et al., 2004; Dmitrieva et al., 2010, Matsui, 2011; Röttger et al., 2011). In other words, neutralization of the voicing contrast has been found to be incomplete in both production and perception, a finding that is sometimes claimed to be incompatible with those formal analyses of devoicing that operate on abstract underlying representations (Port, 1996;

Port & Leary, 2005; Ernestus & Baayen, 2007a, 2007b) or the traditional two-way laryngeal classification of obstruent phonemes (Ernestus, 2000). At the same time, other studies have argued that the apparent lack of neutralization in laboratory speech does not necessarily signal preservation of the underlying contrast but may instead be due to task-dependent factors, including speakers' exposure to orthographic forms during testing or inclusion of full minimal pairs among the stimuli (among others, Fourakis & Iverson, 1984; Mascaró, 1987; Jassem & Richter, 1989; Manaster Ramer 1996a, 1996b). Hence, incomplete neutralization of the voicing contrast in final obstruents may be driven by the nature of the experimental task performed by the speaker or the composition of the stimulus list rather than the grammar.

To evaluate the effects of grammatical versus methodological influences on the degree to which the voicing contrast is neutralized in the production and perception of laboratory speech, two types of factors are specifically manipulated in the current investigation: (i) task-independent/phonological and lexical factors (place and manner of articulation of the final consonant, word length in syllables, availability of lexical competition) and (ii) task-dependent/methodological influences (presence of orthographic representations during testing, inclusion of full minimal pairs among stimuli items). For speech production, results of the thesis reveal that significant voicing-dependent differences are present at the group level for the acoustic parameter of closure voicing and, for a subset of test items, closure/frication duration and release duration but not preceding vowel duration. Crucially, differences in closure/frication duration and release duration signal underlying voicing regardless of whether or not task-dependent factors

encourage preservation of the contrast. Differences in closure voicing, however, show a strong influence of task-dependent factors and are significant (or marginally significant) only when participants are exposed to orthographic representations of the stimuli or presented with minimally contrasting item pairs. At the individual level, differences in glottal pulsing are more robust and consistent than differences in consonantal duration. For speech perception, the current investigation shows that listeners are able to identify intended voicing with limited accuracy only, with the majority of experimental stimuli showing a voiceless bias and with response rates of around 50% not always signaling chance-level performance. Identification responses are also found to be highly sensitive to the presence of task-dependent acoustic cues.

The present findings confirm that methodological factors can have a strong biasing effect on the production and perception of word-final devoicing, such that many of the observed voicing-dependent effects can be attributed to a task-dependent manipulation of the acoustic cue of glottal pulsing. At the same time, the incompletely neutralized contrasts in consonantal duration appear to be maintained for task-independent reasons. These results are argued to favour a less restrictive definition of neutralization and a model of phonology that views devoicing as a loss of the primary acoustic cue to the underlying voicing contrast rather than complete identity of the [voiced] feature.

1.2. Overview of the Chapters

The dissertation is organized as follows. Chapter 2 introduces the topic of the present investigation by reviewing the background literature on final devoicing and pointing out the relevant grammatical and methodological influences. Chapter 3 is concerned with production aspects of the voicing contrast in word-final obstruents in Russian and it reports the results of an acoustic production experiment that investigated the effects of task-independent versus task-dependent factors on the retention of cues to underlying voicing. Perception-related issues are addressed in Chapter 4, which presents the findings of a series of perceptual identification tasks that examined Russian listeners' sensitivity to the voicing-dependent differences found in production. Chapter 5 discusses the experimental findings and their implications. Conclusions and possibilities for future research are presented in Chapter 6, followed by a list of references and appendices.

CHAPTER 2: BACKGROUND

2.1. Introduction

Chapter 2 begins with overviews of the cross-linguistic typology of laryngeal neutralization and the traditional (prescriptive and formal) accounts of final devoicing (Section 2.2.). This is followed by a summary of the previous experimental findings in the domains of production (Section 2.3.) and perception (Section 2.4.). Given the methodological focus of the present dissertation, descriptions of previous studies include information on the experimental procedures, stimuli items, participants, and other relevant methodological aspects (the reader who is not interested in such detailed reviews may proceed to Section 2.3.6., which offers a summary of the previous production results, and Section 2.4.6., which is an overview of the previous perception findings). The remainder of Chapter 2 presents the debate on the roles of grammatical versus methodological factors in experimentally-elicited devoicing data (Section 2.5.), followed by a brief summary of the background claims (Section 2.6.).

2.2. Final Devoicing: a Theoretical Overview

2.2.1. Typology

The majority of the world's languages that have both voiced and voiceless obstruents in their segmental inventories allow the two types of consonants to surface in word-initial and word-medial positions (Denes, 1963). However, many languages differ with respect to whether or not phonetically voiced obstruents may be found word-finally. In *neutralizing* languages and dialects, only voiceless stops and fricatives are allowed in

word-final positions, with phonologically voiced obstruents being devoiced in order to meet this restriction (e.g., articulating /d#/ as [t]). This process of word-final devoicing is often described as a phonetically natural, easy-to-acquire phonological regularity that can be attested even in the speakers of languages without word-final obstruents (e.g., during the production of L2 words; Stampe, 1969). Notable examples of neutralizing languages and dialects include German, Dutch, Afrikaans, Catalan, Turkish, Polish, Russian, Singapore English, Ojibwa, and Wolof (for general summaries, see Brockhaus, 1995; Blevins, 2006; Myers, 2012).

In *non-neutralizing* languages and dialects, the voicing contrast is maintained both underlyingly and on the surface, such that phonemically voiced word-final obstruents are pronounced as voiced (e.g., articulating /d#/ as [d]). At the phonetic level, phonological voicing is usually signalled by vocal fold vibration during the closure stage of the obstruent, shortening of the final consonant, lengthening of its preceding vowel, and absence of aspiration (for a detailed list of voicing cues, see Lisker, 1986). Examples of non-neutralizing languages and dialects include the majority of English varieties spoken in North America and the British Isles, Spanish, French, Ukrainian, and Sorbian.

2.2.2. Descriptive Accounts

Descriptive accounts of word-final devoicing have traditionally been based on impressionistic data, and they represent the view that no distinction is maintained in either production or perception between consonants that are phonologically voiceless and those that are underlyingly voiced but phonetically devoiced. For example, the

underlying representation of the Russian word /kɔd/ 'code' ends in the voiced coronal plosive /d/ (with underlying voicing seen in morphologically-related forms in which the critical segment in pre-vocalic and is pronounced as [d]; e.g., the plural form /kɔda/). However, the phonetic output form of /kɔd/ is assumed to be [kɔt], which is homophonous to that of /kɔt/ 'cat' that ends in the lexically voiceless coronal stop /t/ (as evidenced by the plural form /kɔta/ that is articulated with a word-medial [t]) (among others, Avanesov, 1956; Timberlake, 2004; Wade, 2010).

Hence, a word-final [t] in Russian may correspond to either an underlying /d/ or an underlying /t/, with the true identity of the consonant presumably not being recoverable from phonetics. Such loss of the underlying voicing contrast is also assumed in the majority of (largely descriptive) phonological grammars of German (e.g., Moulton, 1962), Dutch (e.g., Booij, 1995), Afrikaans (e.g., Donaldson, 1993), Catalan (e.g., Hualde, 1992), Polish (e.g., Benni, 1964; Wierzchowska, 1980), and other neutralizing languages. Notably, despite the general agreement on the completeness of neutralization, descriptive accounts often vary in how they formulate the environment in which neutralization applies. In many grammars, devoicing is presented as a syllable-final phenomenon (e.g., Hualde (1992) for Catalan; Booij (1995) for Dutch). In other studies, devoicing is argued to be word-final (e.g., Donaldson (1993) for Afrikaans; Wade (2010) for Russian) or even phrase-final in some of the same languages (e.g., Timberlake (2004) for Russian) (also, see Hock (1999) and Blevins (2006) for diachronic discussions of the domain of devoicing in languages such as German).

2.2.3. Formal Analyses

Formal analyses of devoicing have been developed largely on the basis of descriptive data and they vary depending on which particular phonological framework is adopted and which pressure is thought to mitigate against preservation of the voicing contrast in final obstruents. For example, many researchers have argued that the devoicing process is driven by such grammar-internal pressures as syllable structure and markedness (e.g., Rubach, 1990; Lombardi, 1994; Kiparsky, 2008). Others have proposed that final devoicing has phonetic or perceptual motivations and that it should be attributed to such factors as the voicelessness of the word-final boundary (Lass, 1971; Lightner, 1972; Ingram, 1989), phrase-final lengthening and the presence of laryngeal spreading and closing gestures (Blevins, 2004, 2006), the general aerodynamic constraints that inhibit the production of voicing in obstruents (Ohala, 1983; 1997), or the low perceptual salience of word-final segments (Steriade, 1999).

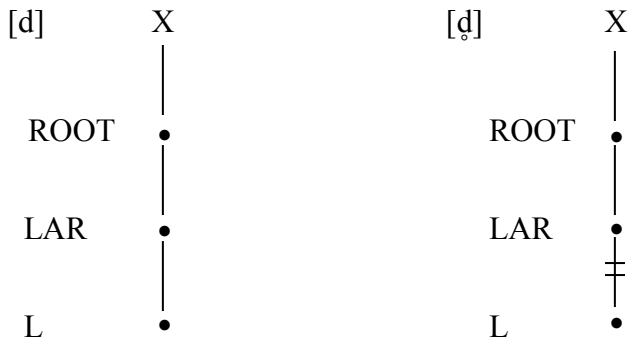
In models based on the SPE (Chomsky & Halle, 1968), both syllabic and non-syllabic accounts of final devoicing are traditionally represented as a linear rule which dictates that all obstruents must be realized as voiceless when word-final (syllable-final, adjacent to a prosodic boundary, occurring in a non-salient environment, etc.). A classic example of the devoicing rule is shown in (1) (based on Vennemann, 1968).[1]

(1) [-sonorant] -> [-voiced] / __ #

[1] The rule in (1) may also be represented with monovalent features, may refer to syllabic, morphemic, phrasal, etc. boundaries and/or may use [obstruent] instead of [sonorant] or [tense] instead of [voice]. For examples of various versions of the rule, see Kenstowicz & Kisseberth (1979), Lass (1984), and Hall (1992).

In frameworks such as Government Phonology (Kaye et al., 1985) and Autosegmental Phonology (Goldsmith, 1990), neutralization of the voicing contrast is attributed not to a feature-changing process but rather to the absence or presence of a particular laryngeal element for phonologically voiced segments, such as delinking of a laryngeal node/feature (Lombardi, 1994; Brockhaus, 1995; Ferme, 2006) or addition of [spread glottis] to underlyingly voiced (lenis) consonants (Iverson & Salmons, 1995, 1999, 2007; Vaux & Samuels, 2005). A sample autosegmental analysis of the difference between a fully voiced [d] and a devoiced [d̥] is provided in (2), with 'L' representing the laryngeal element that undergoes delinking (based on Brockhaus, 1995).

(2)



In Optimality Theory (Prince & Smolensky, 2004), neutralization of the voicing contrast in final obstruents is usually analyzed as an effect of a markedness constraint against voiced obstruents in syllable-final environments (e.g., '*VOICED-CODA: coda obstruents are voiceless', Kager, 1999) or, in non-syllabic analyses, to a constraint against voiced stops and fricatives that are not released into a sonorant sound (e.g.,

'*[voice]/V_[-son]', Steriade, 1999; 'FINALDEVOICING (FD): voiced obstruents are only allowed in a position preceding a tautosyllabic sonorant', van Oostendorp, 2008). When ranked above the relevant faithfulness constraint that prohibits changes to the voicing specification of speech sounds (e.g., 'IDENT-IO(VOICE)', Kager, 1999), the markedness constraint penalizes voiced realizations of word-final (coda-final, phrase-final, etc.) obstruents. Alternatively, the same prediction can also be achieved if using a context-free markedness constraint against voiced obstruents (e.g., 'VOICED OBSTRUENT PROHIBITION (VOP), *[+voi, -son], no obstruent must be voiced', Kager, 1999) or a general constraint against laryngeal specification (e.g., '*LAR: Do not have laryngeal specifications', Lombardi, 1999). Such constraints penalize voiced realizations of obstruents when ranked below domain-specific faithfulness constraints that require preservation of voicing in non-neutralizing environments (e.g., 'IDONSLAR: A pre-sonorant consonant in an onset should be faithful to underlying laryngeal specification', Lombardi, 1999). A sample OT analysis of final devoicing is provided in (3) below (based on Kager, 1999).[2]

(3)

Input: /d#/	*VOICED-CODA	IDENT-IO(VOICE)
[d#]	*!	
☞ [t#]		*

Notably, although the analyses outlined above use different phonological frameworks and attribute final devoicing to different factors, they all rely on the restricted

[2] For an example of the approach using a context-free markedness constraint, see Petrova et al. (2006).

definition of neutralization according to which the outcome of the neutralization process is a complete change in category membership (from voiced to voiceless; Kiparsky, 1976). In other words, formal analyses of word-final devoicing aim to ensure that underlyingly voiced and voiceless obstruents have the same featural organization at the phonetic level, with only fully voiced and fully voiceless candidates being considered as potential output forms of underlyingly voiced final obstruents. In addition, for languages such as German, the assumption that the underlying contrast is lost in production remains applicable even when the neutralization process is described as final fortition and the underlying contrast is analyzed as involving [tense] or [spread glottis] instead of [voice] (e.g., Iverson & Salmons, 1995, 2007; Jessen, 1998, 2001; Jessen & Ringen, 2002; Beckman et al., 2009).

2.3. Experimental Production Findings

Contrary to the traditional view on final devoicing, experimental production studies done in the past 40 years have repeatedly shown that speakers of neutralizing languages preserve acoustic cues to the underlying laryngeal specification of word-final obstruents, with robust and usually statistically significant voicing-dependent differences seen in such parameters as preceding vowel duration, consonantal closure and release duration, and closure voicing (among others, Chen, 1970; Dinnsen & Charles-Luce, 1984; Port & O'Dell, 1985; Slowiaczek & Dinnsen, 1985; Pye, 1986; Charles-Luce, 1993; van Rooy et al., 2003; Warner et al., 2004; Smith et al., 2009; Dmitrieva et al.,

2010; Röttger et al., 2011).[3] However, a few experimental studies have found no evidence for preservation of the voicing contrast in some of the same neutralizing languages (Fourakis & Iverson, 1984; Inozuka, 1991; Kopkalli, 1993), which has led some researchers to question the validity of incomplete neutralization findings (Fourakis & Iverson, 1984; Manaster Ramer, 1996a, 1996b). Acoustic measurements and experimental tasks that have been used in previous neutralization studies are described in Section 2.3.1. that follows. Methodologies and main findings of selected published works are summarized in Section 2.3.2. through Section 2.3.5. and are grouped by the language of investigation as well as the authors' conclusions with respect to the maintenance of the voicing contrast. A general summary of the previous studies is offered in Section 2.3.6.

2.3.1. Procedures and Acoustic Measurements

Experimental studies of the voicing contrast in word-final obstruents usually report the results of a reading task during which orthographic representations of stimuli words are presented in isolation or embedded in carrier sentences and participants are instructed to read each word or sentence out loud in their regular voice and intonation (e.g., Chen, 1970; Port & O'Dell, 1985; Kopkalli, 1993; van Rooy et al., 2003; Warner et al., 2004; Dmitrieva et al., 2010). Non-reading tasks have been used in very few previous studies and have included oral conjugation of verbs (Fourakis & Iverson, 1984), oral depluralization of nouns (Röttger et al., 2011), and word-guessing on the basis of short descriptive questions (Jassem & Richter, 1989). In such tasks, speakers are not exposed

[3] For additional unpublished evidence, see Tieszen & Read (1993) and Tieszen (1997) (Polish), Shrager (2006) (Russian), Abu Bakar et al. (2007) (Malay), Kuzla & Dubeda (2008) (Czech), Nicolae & Nevins (2009) (Turkish), and Iosad (2010) (Breton).

to orthographic representations of test items but rather recall wordforms from memory when asked to provide words that are morphologically and/or semantically related to specific trigger tokens. For both reading and non-reading tasks, stimuli items are usually randomized and mixed together with an equal number of filler words (in order to mask the goal of the study) and multiple repetitions of all test items are often produced by each speaker.

Findings of acoustic production studies are usually based on the analyses of 4 parameters for plosives (preceding vowel duration, consonantal closure duration, consonantal release duration, and closure voicing) and 3 parameters for fricatives and affricates (preceding vowel duration, closure/frication duration, and voicing into the consonant). Preceding vowel duration is traditionally calculated in milliseconds from the onset of visible F2 and F3 to their offset (Chen, 1970; Fourakis & Iverson, 1984; Port & O'Dell, 1985; Dinnsen & Charles-Luce, 1985; Port & Crawford, 1989; van Rooy et al., 2003). For words with a sonorant consonant adjacent to the final vowel (e.g., /Vl/, /Vn/), the measurement can include the entire vowel+consonant sequence (Port & O'Dell, 1985; Piroth & Janker, 2004).

Consonantal closure/frication duration is measured in milliseconds from the end of the preceding vocalic nucleus to the beginning of the release burst of the plosive or to the end of the high frequency noise associated with the final fricative or affricate (Fourakis & Iverson, 1984; Charles-Luce, 1985; Port & O'Dell, 1985; Pye, 1986; Dinnsen & Charles-Luce, 1985; Port & Crawford, 1989; Wissing & van Rooy, 1992; van

Rooy et al., 2003). For words containing nasal stops followed by oral plosives, the measurement can include the entire closure period (Port & Crawford, 1989).

Consonantal release duration is calculated only for final stops. The measurement is done in milliseconds, starting from the onset of the release burst and continuing up to the end of the burst or, if a consonant is aspirated, to the end of the aspiration noise (e.g., Port & O'Dell, 1985; van Rooy et al., 2003; Piroth & Janker, 2004). In some studies, burst and aspiration duration are calculated separately (e.g., Fourakis & Iverson, 1984).

Closure voicing (='voicing into the consonant', 'glottal pulsing') is determined on the basis of one or more of the following parameters: periodicity of the waveform, presence of a voice bar in the spectrogram, audible vocal fold activity in the acoustic output, and presence of striations during the closure period of the consonant (e.g., the first 3 parameters are used in Ernestus & Baayen, 2007a; only the last parameter is used in Pye, 1986). Considering that phonetic voicing is often absent during the release stage of final stops (e.g., as many as 98% of plosives in Piroth & Janker, 2004), only the closure part of the plosive is traditionally used to measure glottal pulsing (e.g., Dinnsen & Charles-Luce, 1984; Port & O'Dell, 1985; Piroth et al., 1991; Piroth & Janker, 2004). For fricatives and affricates, voicing is calculated on the basis of vocal fold vibration during the entire period of frication (e.g., Slowiaczek & Dinnsen, 1985). The measurement can be done in milliseconds (e.g., Port & O'Dell, 1985; Pye, 1986; Dmitrieva et al., 2010), as a ratio (Inozuka, 1991; see also Snoeren et al., 2006), or as a number of glottal pulses (e.g., Port & Crawford, 1989; van Rooy et al., 2003; also, the production study presented in Chapter 3). The latter measurement is often preferred because, as pointed out in

Fourakis & Iverson (1984), voicing is not always stable during consonantal closure and it often stops but then reappears again, which makes any measurements in milliseconds (including ratios) problematic.

Other acoustic parameters that have occasionally been used to investigate the voicing contrast in obstruents in both neutralizing and non-neutralizing languages and environments include duration of the voiceless-only part of the closure (Keating et al., 1983), duration of the entire consonant (Mitleb, 1981; Shrager, 2006), coda or rhyme duration (Piroth & Janker, 2004), word stem duration (Piroth et al., 1991), formant transitions and F0 values (Raphael, 1981; Wardrip-Fruin, 1982; van Rooy et al., 2003), and spectral characteristics of the release burst (Revoile et al., 1982; Röttger et al., 2011). However, such measurements do not generally show statistically reliable effects in the case of word-final obstruents in neutralizing languages (e.g., van Rooy et al., 2003) and these additional acoustic parameters are not commonly used in experimental studies of final devoicing.

2.3.2. Data from German

Up to this date, much of the experimental evidence showing preservation of voicing-dependent cues in production has come from German, with several acoustic studies finding statistically significant durational differences between the surface forms of underlyingly voiced versus voiceless final obstruents (Charles-Luce, 1985; Port & O'Dell, 1985; Port & Crawford, 1989; Piroth et al., 1991; Piroth & Janker, 2004; Smith et al., 2009; Röttger et al., 2011). Such findings contradict the traditional view of German

being a language in which all word-final obstruents are discretely voiceless (Moulton, 1962; Brockhaus, 1995) and also go against the null results reported in other experimental investigations of final devoicing in German (Fourakis & Iverson, 1984; Inozuka, 1991). Methodologies and main results of the German studies are described below. The studies are grouped by the authors' conclusion with respect to the completeness of neutralization, starting with the investigations that advocated for retention of voicing-dependent cues in production.

2.3.2.1. Port & O'Dell (1985)

One of the earliest published studies of the voicing contrast in word-final obstruents in German is presented in Port & O'Dell (1985) who examined the production of voicing cues in labial, coronal, and velar plosives (/p, b, t, d, k, g/).[4] Native speakers of German (n=10) performed a word-reading task with two repetitions. The list of stimuli words included a small set of test items (n=20, 10 minimal pairs; 4 ending in labials, 10 in coronals, 6 in velars) and a larger set of filler words (n=40). Preceding vowel duration, closure duration, release duration and closure voicing were measured and subjected to analyses of variance (by-subject), with underlying voicing used as the only predictor variable.

Results of the study revealed that, for underlyingly voiced stops, preceding vowels were longer (15 ms difference), consonantal closures were shorter (5 ms difference), release durations were shorter (15 ms difference), and there was more

[4] The findings presented in Port & O'Dell (1985) were based in part on earlier unpublished work by the same authors (e.g., Port et al., 1981; O`Dell & Port, 1983). The study also investigated perceptual aspects of the voicing contrast. For details on the perception experiment, see Section 2.4.2.1.

phonetic voicing for voiced segments compared to their voiceless counterparts (5 ms difference). All differences were statistically significant and went in the direction that could be expected on the basis of the effects observed in the studies of voicing cues in intervocalic plosives (e.g., Mitleb, 1981). Although these results were based on small samples of both speakers and experimental items and 6 (out of 10) stimuli pairs were problematic (due to the presence of borrowed lexical items, archaic words, regional forms, and wordforms without paradigmatic alternations; Piroth & Janker, 2004), Port & O'Dell interpreted their findings as evidence for the maintenance of the underlying voicing contrast in production and argued that the process of final devoicing in German was a phonetic implementation rule rather than a phonological phenomenon.

2.3.2.2. Charles-Luce (1985)

Charles-Luce (1985) investigated the voicing contrast in word-final coronal plosives and fricatives (/t, d, s, z/). In addition to determining whether voicing-dependent differences were present in production, the study aimed to establish whether the maintenance of such cues would be affected by phonetic environment and sentential position. Speakers of German (n=5) performed a reading task with 5 repetitions. Stimuli items (n=8, 4 minimal pairs; half ending in plosives) were embedded in 4 carrier sentences, such that each word was found in 4 different environments: (i) clause-final followed by a consonant-initial word, (ii) clause-final followed by a vowel-initial word, (iii) non-clause-final followed by a consonant-initial word, and (iv) non-clause-final followed by a vowel-initial word. Filler sentences (n=40) were also included, and

participants were familiarized with the full list of sentences prior to testing. Analyses of variance (by-subject) examined the effects of underlying voicing, sentential position and phonetic environment on the acoustic parameters of preceding vowel duration, closure/frication duration, and closure voicing.

For final plosives, production results revealed that closure voicing showed effects of sentential position, phonetic environment, and phonological voicing (5 ms longer period of vocal fold vibration for /d/ than /t/ when clause-final and followed by a consonant). The parameter of preceding vowel duration demonstrated effects of the phonetic environment for voiced stops only (12~13 ms longer voiced stops when followed by a vowel-initial word). For fricatives, preceding vowel duration showed a significant main effect of underlying voicing and a marginally-significant interaction of phonological voicing and sentential position, with longer vowels found before clause-final /z/ (10 ms difference). No other significant voicing-dependent effects were observed for either plosives or fricatives. Although these findings were based on very small samples of speakers and test items and very few significant voicing-dependent effects were actually observed, Charles-Luce concluded that the voicing contrast was preserved in German and that the degree of neutralization was affected by both the following phonetic environment and the sentential position. However, as pointed out in Piroth & Janker (2004), any interpretation of Charles-Luce's results is further complicated by the fact that several stimuli items and the VOS (verb, object, subject) type of one of the carrier sentences are not necessarily characteristic of everyday German speech and that

the '_#V' environment may also be analyzed as '_#$C_{Voiceless}$' due to German having a productive process of word-initial glottal stop epenthesis for words starting in vowels.

2.3.2.3. Port & Crawford (1989)

Port & Crawford (1989) examined the production of voicing cues in word-final coronal stops (/t, d/), with a secondary goal of determining whether the degree of neutralization could also be affected by pragmatic factors.[5] Speakers of German (n=5) performed 4 different tasks: (i) reading words embedded phrase-medially in non-contrasting sentences, (ii) reading words embedded phrase-medially in contrasting sentences (i.e., 'I said X, not Y'; where 'X' and 'Y' are members of a minimal pair), (iii) reading sentences for the purpose of dictation (i.e., for an assistant to write them down), and (iv) reading a wordlist. Three repetitions were elicited in Task 1 (reading 2 times, reciting from memory one time), one repetition was recorded for Task 2 and Task 3, and four repetitions were produced in Task 4. Stimuli items were existing words of German (n=6, 3 minimal pairs; including 2 forms without voicing alternations in their paradigms). Analyses of variance (by-subject, by-item) and discriminant analyses were performed on the measurements of preceding vowel duration, closure duration, burst duration, and closure voicing, with underlying voicing used as the predictor variable.

The study revealed that, when the production data were pooled across all experimental tasks, burst duration was the only parameter that was affected by underlying voicing at a statistically significant level in both by-subject and by-item tests.

[5] Port & Crawford (1989) also investigated perceptual aspects of the voicing contrast. Details of the perception experiment are presented in Section 2.4.2.2.

Discriminant analyses showed that all acoustic cues other than closure voicing could contribute to the maintenance of the underlying contrast, with the rates of correct classification of 55-56% for Task 1, 63% for Task 2, 78% for Task 3, and 62% for Task 4. Individual results revealed that 4 (out of 5) speakers showed differences in burst duration. One participant produced differences in glottal pulsing. In discriminant analyses, only burst duration was identified as a significant contributor for all 5 speakers (contribution of other factors varied across participants). Although only 3 minimal pairs and 5 speakers were tested, Port & Crawford concluded that the voicing contrast was preserved in all conditions and that speakers of German produced more prominent differences when dictating words, less prominent differences when reading sentences, and intermediate differences when pronouncing individual words. Furthermore, even though only two stimuli items did not alternate in their paradigms and burst duration was the sole parameter that showed a significant effect of underlying voicing in analyses of variance, Port & Crawford suggested that full neutralization did not occur regardless of the availability of paradigmatic evidence for phonological voicing and that speakers used several cues to signal the underlying contrast.

2.3.2.4. Piroth et al. (1991)

Piroth et al. (1991) studied the voicing contrast in word-final and morpheme-final labial, coronal, and velar stops (/p, b, t, d, k, g/). Participants (n=2) performed a reading task with 3 repetitions. Stimuli words (n=18, including one minimal pair; 6 words per place of articulation) were embedded in carrier phrases in 5 different prosodic positions:

(i) word-final and utterance-final, (ii) word-final but not utterance-final, (iii) morpheme-final (word-medial) followed by a voiced consonant, (iv) morpheme-final (word-medial) followed by a voiceless consonant, and (v) intervocalic. Preceding vowel duration, closure duration, release duration and duration of word stem were used as dependent variables in analyses of variance that examined the effects of underlying voicing and sentential position (using data pooled over subjects). Voicing during closure and the gap between the final plosive and the following consonant were also measured but not analyzed statistically.

Both preceding vowel duration and word stem duration showed a main effect of underlying voicing, with longer vowels seen before voiced plosives and longer stems observed in voiceless stops.[6] For these two parameters, underlying voicing did not interact with sentential position, suggesting that the observed effects were significant regardless of the immediate phonetic environment. In contrast, the voicing factor interacted with sentential position for both closure duration and release duration. Significantly longer closures for voiceless plosives were observed when stimuli items were word-final and utterance final, morpheme-final followed by a voiceless consonant, and intervocalic. In the remaining two environments, differences in closure duration were only marginally significant. Longer releases for voiceless stops were seen in 3 environments only: word-final and utterance-final, word-final but not utterance-final, and intervocalic. Piroth and colleagues also noted that phonetic voicing was absent word-finally for almost all test items and not present in the intervocalic environment for

[6] Piroth et al. (1991) report durational differences in a graph format and do not provide the exact numerical values.

approximately half of the tokens. Although these results were based on a sample of two speakers and 9 stimuli pairs, Piroth et al. interpreted their findings as supporting the view that the voicing contrast was maintained in German and also stated that it would be an oversimplification to analyze the near-absence of voicing cues in final obstruents as an instance of phonological devoicing.

2.3.2.5. Piroth & Janker (2004)

Piroth & Janker (2004) explored how the voicing contrast was produced across 3 different dialects of German in word-final and word-medial labial, coronal and velar plosives and labial and coronal fricatives (/p, b, t, d, k, g, f, v, s, z/). Speakers of West German (n=2), Central German (n=2) and South German (n=2) performed a reading task with 3 repetitions. Stimuli items (n=26, one minimal pair; 18 plosive-final, 8-fricative-final; distributed evenly across places of articulation) were embedded in carrier sentences in 5 different environments: (i) word-final and utterance-final, (ii) word-final sentence-medial followed by a nasal, (iii) word-medial prevocalic, (iv) word-medial (morpheme-final) followed by a voiceless consonant, and (v) word-medial (morpheme-final) followed by a voiced consonant. Each speaker was also presented with 142 filler sentences. Preceding vowel duration, closure/frication duration, release duration, coda duration, rhyme duration and closure voicing were used as dependent measures in analyses of variance (performed on non-averaged data). Underlying voicing and sentential environment were the only predictor variables. Voicing during burst was inspected visually but not analyzed.

For words with final plosives that were embedded in neutralizing environments, no voicing-dependent differences were observed in preceding vowel duration or closure voicing in any of the 6 speakers. For coda duration, no significant effects were found in the participants who spoke West German and Central German but differences were significant in the two South German speakers (31~34 ms longer codas for voiceless stops). For words with final fricatives placed in neutralizing contexts, no significant differences were produced by any speaker in any of the acoustic measurements. For all obstruents, significant differences in the expected direction were observed in the parameters of preceding vowel duration, closure voicing and coda duration in all non-neutralizing (control) environments (e.g., prevocalic). Even though the study tested 6 participants and only two speakers per dialect, Piroth & Janker concluded that the voicing contrast was not preserved in final fricatives in any of the 3 dialects and that neutralization in final plosives was dialect-specific, with South German speakers maintaining the underlying contrast. Piroth & Janker further noted that glottal pulsing was not a strong voicing cue in German even in non-neutralizing environments and also suggested that the voicing contrast was likely realized using several parameters.

2.3.2.6. Smith et al. (2009)

Smith et al. (2009) investigated how speakers of German (n=13) produced the voicing contrast in word-final coronal and velar plosives (/t, d, k, g/) in L1 German and

L2 English in phrase-medial and phrase-final positions.[7] Participants performed a reading task with 10 repetitions of each test word (5 times per environment). A total of 56 stimuli items were included in the study (26 from German, 30 from English; minimal and near-minimal pairs). However, the analysis concentrated only on 12 words (6 minimal pairs; 5 ending in coronals, one in velars) that were orthographically and/or phonologically similar across the two languages (e.g., the *Bad* 'bath' ~ *bat* 'asked' pair in German, which was similar to the *bad* ~ *bat* pair in English). The stimuli were embedded phrase-medially and phrase-finally in carrier frames, with members of the same minimal pair never occurring in the same sentence. Measurements of preceding vowel duration, closure duration, release duration and closure voicing were compared in a series of t-tests.

Results of the study demonstrated that, when speaking their native language, German participants neutralized the difference in closure duration but maintained significant differences in preceding vowel duration (13 ms longer when followed by voiced plosives), release duration (12 ms longer for voiceless stops), and closure voicing (4 ms longer for voiced plosives). Individual data further revealed that the effects of underlying voicing were observed in 10 (out of 13) speakers for the parameter of preceding vowel duration and 12 speakers for closure voicing. Differences in closure duration went in the expected direction in 6 participants only. When speaking English, their second language, the same speakers showed significant differences in all 4 acoustic parameters, with many of the observed effects also found at the individual level. Smith et

[7] The primary goal of the study in Smith et al. (2009) was to examine the role of inter-language transfer and L1-L2 intelligibility benefits, including the production and perception of voicing cues in English.

al. concluded that German participants maintained the voicing contrast in both L1 German and L2 English. However, Smith and colleagues acknowledged that their study did not control for usage frequency, morphological complexity, place and manner of articulation of final consonants and other similar grammatical factors that could have affected the results of their investigation. The observed results were also not necessarily representative of the German language in general, since the participants of the study were raised speaking different varieties of German and, at the time of testing, all of them were residing in the USA.

2.3.2.7. Röttger et al. (2011)

Röttger et al. (2011) studied the production of the voicing contrast in labial, coronal and velar stops (/p, b, t, d, k, ɡ/).[8] German speakers (n=16) performed an oral depluralization task. Stimuli items (n=48; 16 ending in labials, 14 in alveolars, 18 in velars) were minimally-contrasting non-words (e.g., *Drud ~ Drut*). Test tokens were embedded in a carrier sentence and presented auditorily in the plural form such that the critical consonant was in a non-neutralizing environment (e.g., *die Drud̲e*). Participants were asked to change the non-word to its singular form in which the critical consonant would be word-final (e.g., *ein Drut̲*) and to pronounce it phrase-medially in another carrier sentence. Each stimulus item was elicited only once. Test tokens and filler items (n=96) were presented in 4 experimental blocks, with members of each minimal pair always separated from each other. Preceding vowel duration, closure duration, release

[8] The study in Röttger et al. (2011) also examined the perception of voicing cues in German. Details of the latter experiment are provided in Section 2.4.2.3.

duration and burst intensity were used as dependent measures in a linear mixed effects model, with underlying voicing included as a fixed effect and subjects and items as random effects.

The study revealed the presence of statistically significant voicing-dependent differences in the parameters of preceding vowel duration (8 ms longer vowels before voiced stops) and burst intensity (1.24 dB louder bursts for voiceless plosives). Neither closure duration nor release duration showed a significant effect of underlying voicing. Röttger et al. concluded that the voicing contrast was preserved in German even in the case of an auditory task and even when using non-word items and described the observed incomplete neutralization of voicing cues as a "robust phenomenon whose phonological implications should be taken seriously (p. 1725)". However, the exact interpretation of Röttger et al.'s results is not clear considering that their stimuli were non-words that contrasted minimally, which could have encouraged the participants to produce voicing-dependent differences that are not typical of more natural speech and existing lexical items.

2.3.2.8. Fourakis & Iverson (1984)

Unlike the German investigations summarized above, Fourakis & Iverson (1984) advocated for complete loss of voicing-dependent cues in production. In their study, German participants (n=4) performed two experimental tasks: (i) oral conjugation of verbs containing stem-final coronal and velar stops and coronal fricatives (/t, d, k, g, s, z/) and (ii) reading a list of words ending in labial, coronal and velar stops (/p, b, t, d, k, g/).

In the conjugation task, speakers heard an infinitival form of a verb and were instructed to repeat the infinitive followed by two additional grammatical forms, including one form in which the stem-final obstruent was word-final (e.g., hearing *raten* 'advise (Inf.)' and producing *raten*, *riet* 'advise (Past)' and *geraten* 'advise (Past Part.)'). During testing, test verbs (n=11, no minimal pairs; 4 ending in coronal stops, 5 in velar plosives, 2 in coronal fricatives) and filler words (n=5) were each repeated twice. In the word-list reading task, speakers were presented with a small list of test items (n=10, 5 minimal pairs; 2 items ending in labials, 6 in coronals, 2 in velars) that were mixed together with an equal number of fillers (n=10). Every item was read 4 times. Separate t-tests were performed for each speaker and each item using the measurements of preceding vowel duration, closure/frication duration, burst duration and aspiration duration (for the conjugation task) or preceding vowel duration and closure duration (for the word-list reading task).

In the case of oral conjugation, no significant differences were observed in the pooled data. Individual results showed that one speaker produced shorter closures for voiced coronal plosives and for voiced velar stops that followed [u:] but not [a:]. Another participant maintained shorter closures for all voiced coronal plosives. The third subject showed longer [a:] before voiced velars. The fourth speaker demonstrated no consistent differences. In the word-list reading task, differences in preceding vowel duration were significant at the group level in one word pair only (25 ms longer vowels before voiced stops). No significant differences were observed for closure duration. Individual data revealed that one participant maintained differences in vowel duration for 4 word pairs and closure duration for two pairs. Another speaker produced differences in vowel

duration (in the unexpected direction) for one word pair. The remaining two speakers showed no reliable voicing-dependent effects. Despite the fact that these results were based on very few subjects and test items, Fourakis & Iverson argued that the underlying voicing contrast was not maintained in word-final environments in German and attributed the findings to the contrary (e.g., Port et al., 1981; O'Dell & Port, 1983) to the effects of hyperarticulation. However, Port & Crawford (1989) argued that a reanalysis of Fourakis & Iverson's data would show statistically significant differences at the group level if results were pooled across items rather than subjects. Charles-Luce (1993) further pointed out that speakers repeated a base form followed by two conjugated forms and, consequently, the presence of disambiguating information (i.e., morphologically related words) could have reduced the need to maintain the voicing contrast. More recently, Piroth & Janker (2004) questioned the inclusion of outdated verbal forms among the stimuli and criticized Fourakis & Iverson for failing to control for list effects in the conjugation task.

2.3.2.9. Inozuka (1991)

Complete neutralization of cues to underlying voicing in word-final obstruents in German is also claimed in Inozuka (1991) who conducted a small-scale acoustic study of velar stops (/k, g/). German speakers (n=6) performed a reading task with two repetitions. Stimuli items (n=18, no minimal pairs) were monosyllabic and disyllabic words that contained voiced and voiceless velar stops in word-final, word-initial and word-medial environments (6 words per environment). Test items were embedded phrase-medially in a

carrier sentence. Acoustic measurements included preceding vowel duration (for word-final and word-medial plosives only), closure duration, release duration, and vocalization ratio. Statistical analyses were limited to a series of t-tests.

Inozuka did not find any significant differences between underlyingly voiced versus voiceless velar plosives in the word-final environment, yet such differences were observed in all acoustic parameters in the case of non-neutralizing positions (i.e., word-initial, word-medial). Phonetic outputs of word-final /k, g/ also appeared distinct from those of an underlying /k/ in word-initial and word-medial positions. Although only 3 pairs of stimuli words were tested per environment and only 6 speakers participated in the study, Inozuka concluded that the voicing contrast was not preserved in word-final obstruents in German and also proposed that neutralized velar stops constituted a distinct phonetic category.

2.3.3. Data from Dutch and Afrikaans

Voicing-dependent differences between underlyingly voiced versus voiceless final obstruents have also been found in Dutch (Warner et al., 2004; Ernestus & Baayen, 2006, 2007a, 2007b) and Afrikaans (Wissing & van Rooy, 1992; van Rooy et al., 2003). As in the case of German, these findings contradicted the assumptions of the traditional analyses of final devoicing in both languages (e.g., Donaldson, 1993; Booij, 1995) and also went against the claims of complete neutralization made in the experimental literature on the voicing contrast in Dutch (Jongman et al., 1992; Baumann, 1995). However, despite finding evidence of incomplete neutralization of the voicing contrast,

Warner et al. (2004) and van Rooy et al. (2003) questioned the role of the grammar in the observed effects. Methodologies and main findings of the latter two studies are described below.[9]

2.3.3.1. Warner et al. (2004)

The Dutch investigation in Warner et al. (2004) examined the production of the voicing contrast in word-final coronal stops (/t, d/).[10] Participants (n=15) performed a word-reading task with two repetitions. Stimuli items were existing Dutch words that either ended in plosives (n=40, 20 minimal pairs) or contained stops in intervocalic positions (n=32, 16 minimal pairs). For plosive-final items, 37 (out of 40) words had morphologically related forms in which the critical consonant was prevocalic (i.e., the underlying voicing specification of the final consonant could be inferred from the word's paradigm). Approximately half of the pairs had phonemically long vowels. Members of the same minimal pair were always separated from each other in the experimental list, and filler items were added at the beginning and the end of the list to account for intonation effects. Participants were familiarized with the stimuli prior to testing. Preceding vowel duration, closure duration, release duration and closure voicing were used as dependent variables in analyses of variance (by-subject, by-item; with repeated

[9] Only representative studies are discussed here because production aspects of the voicing contrast were not the primary goal of investigation in Jongman et al. (1992), Baumann (1995) or Ernestus & Baayen (2006, 2007a, 2007b) and the study in Wissing & van Rooy (1992) was published in Afrikaans and only a summary of the findings is available in English in van Rooy et al. (2003).

[10] Warner et al. (2004) also presented the results of a series of perceptual identification tasks. For details on the perception component of their study, see Section 2.4.3.1.

measures).[11] Separate analyses were conducted for word-final and intervocalic consonants, with underlying voicing and phonemic vowel length used as predictor variables.

For word-final stops, significant voicing-dependent differences were found in the group data in both by-subject and by-item analyses for the parameters of preceding vowel duration (3.5 ms longer vowels when followed by voiced plosives) and release duration (9 ms longer releases in voiceless stops after phonemically long vowels). No significant differences were observed in either closure duration or closure voicing. In comparison, intervocalic plosives showed significant differences in vowel duration (20 ms longer vowels when followed by /d/) and release duration (30 ms longer releases for /t/) but not closure duration. Warner et al. concluded that both vowel duration and release duration served as cues to the underlying voicing contrast in Dutch. Although the presence of orthography was not manipulated experimentally for final obstruents, Warner and colleagues also suggested that orthographic influences were likely contributing to the apparent preservation of voicing cues in experimentally-elicited tokens and further pointed out the smaller magnitude of the durational differences observed in Dutch compared to the values reported for other neutralizing languages, such as German and Polish (e.g., differences of 10 ms or more in Port and O'Dell, 1985; Slowiaczek and Dinnsen, 1985).

[11] Closure voicing was not investigated for intervocalic plosives.

2.3.3.2. Van Rooy et al. (2003)

The Afrikaans study in van Rooy et al. (2003) investigated the production of voicing cues in word-final coronal stops (/t, d/).[12] Participants (n=8) performed a reading task with two repetitions. Stimuli items (n=20) included a set of words from an earlier study in Wissing & van Rooy (1992) (n=10, all contrasting minimally) as well as several new items (n=10; a mix of existing and non-existing wordforms; 5 ending in a voiced grapheme, 5 in a voiceless grapheme). The stimuli were embedded phrase-medially in carrier sentences, with the critical consonant always followed by a vowel-initial preposition. The format of the carrier sentences was chosen such that it would encourage speakers to produce differences in voicing (e.g., 'I have not said **X** or so, I have said **Y** or so', where 'X' and 'Y' contrast in voicing and are presented in bold typeface). In addition to the traditional measurements of preceding vowel duration, closure duration, release duration and closure voicing, van Rooy et al. also calculated duration of final formant transitions and F0 at the midpoint and the offset of the preceding vowel. Acoustic measurements were subjected to analyses of variance (by-subject, by-item) with underlying voicing used as the sole independent predictor. Separate analyses were performed for existing words versus non-words. Production data were also analyzed in a series of (stepwise) regressions which aimed to predict underlying voicing on the basis of acoustic parameters.

For existing words, production results revealed significant voicing-dependent differences in preceding vowel duration for phonemically long vowels (15 ms longer

[12] The study in van Rooy et al. (2003) also investigated perceptibility of voicing cues. Details of the perception experiments are given in Section 2.4.3.3.

when followed by voiced plosives), closure duration (9 ms longer for voiceless stops), release duration (14 ms longer for voiceless sounds), and closure voicing (0.3 periods more for voiced obstruents). For non-words, significant differences were observed in preceding vowel duration for phonemically short vowels (14 ms longer when followed by voiced plosives), closure duration (11 ms longer for voiceless stops), release duration (12 ms longer for voiceless obstruents), closure voicing (0.9 periods more for voiced segments), and duration of final formant transitions (0.8 periods longer for voiced plosives). Regression analyses showed that both closure duration and release duration contributed significantly to the maintenance of the voicing contrast for existing words as well as non-words. Closure voicing played only a limited role for the two types of items. Preceding vowel duration was a significant contributor for existing words only. F0 was a non-significant contributor for nonsense stimuli. Other acoustic parameters did not enhance the statistical model. Individual results revealed that longer closures for voiceless plosives were maintained by 6 (out of 8) speakers, longer release durations for voiceless stops were observed in 5 participants, and more phonetic voicing for voiced segments was produced by two speakers. For preceding vowel duration, individual differences were small and inconsistent.

Van Rooy et al. concluded that F0 and formant transitions did not signal the underlying voicing contrast in Afrikaans and that phonetic voicing was only a marginal cue. In contrast, differences in release duration, closure duration and preceding vowel duration were all reliable indicators of the voicing specification in final plosives. A

hierarchy of voicing cues for Afrikaans was also proposed and is reproduced in (4) below.

(4) Release duration > Closure duration > Preceding V duration > (Closure voicing)

Notably, since phonetic voicing played only a limited role and release duration came out as the most robust acoustic cue even though final plosives are not usually aspirated in Afrikaans under normal speaking conditions, van Rooy et al. reinterpreted the underlying contrast as involving [tense] instead of [voice] and, without testing normal speaking conditions, further proposed that voicing cues would not be preserved in non-laboratory speech.

2.3.4. Data from Catalan and Turkish

In addition to Germanic languages, retention of cues to underlying voicing in word-final obstruents has also been investigated in Catalan (Dinnsen & Charles-Luce, 1984; Charles-Luce & Dinnsen, 1987; Charles-Luce, 1993) and Turkish (Kopkalli, 1993). Both Catalan and Turkish are known in the traditional literature to neutralize the voicing contrast in word-final stops and fricatives (Lees, 1961; Hualde, 1992). However, unlike the majority of neutralizing languages in which different graphemes are usually used for underlyingly voiced versus voiceless final consonants, Catalan and Turkish provide only limited orthographic cues to underlying voicing. In both Catalan and Turkish, word-final plosives and fricatives are typically represented with voiceless graphemes irrespective of their phonological specification (e.g., *cap* ‘head’ (Catalan) and *kap* ‘container’ (Turkish)

are written with the voiceless grapheme ‘p’ even though the two words end in /b/ at the phonological level). The Catalan studies by Dinnsen & Charles-Luce (1984), Charles-Luce & Dinnsen (1987) and Charles-Luce (1993) all argued for retention of voicing cues in production even in the absence of graphemic cues to voicing. The investigation by Kopkalli (1993) found complete loss of the underlying contrast in Turkish. Methodologies and main findings of the Catalan and Turkish studies are summarized below, starting with the work on Catalan.

2.3.4.1. Dinnsen & Charles-Luce (1984), Charles-Luce & Dinnsen (1987)

Dinnsen & Charles-Luce (1984) explored how speakers of Catalan (n=5) produced voicing cues in word-final labial, coronal, and velar stops (/p, b, t, d, k, g/). Subjects performed a reading task with 5 repetitions. Test words (n=10, all contrasting minimally; 4 ending in labials, 2 in coronals, 4 in velars) were embedded sentence-medially and were followed by either a vowel-initial word (‘_#V’) or a consonant-initial word (‘_#C’). Some test items were highly infrequent, and Spanish glosses were provided for each Catalan wordform to ensure that participants knew the meanings of all the stimuli. Each speaker also produced a total of 560 filler items. Preceding vowel duration, closure duration and closure voicing were tested as dependent variables in analyses of variance (by-subject) that examined the effects of underlying voicing and phonetic environment.

Results of the study showed that voicing-dependent differences were absent in the pooled data for all 3 acoustic measurements but present at the individual level for two

(out of 5) speakers. The first subject produced longer vowels before voiced stops in the '_#C' environment (5 ms difference), and the other participant articulated longer closures for voiced plosives (14 ms difference; the direction was not expected). None of the speakers maintained significant differences in closure voicing. Although Catalan subjects as a group showed no statistically significant effects and individual results of one participant went against the general expectation of shorter closures for voiced plosives, Dinnsen & Charles-Luce argued that their findings supported the view that voicing-dependent cues were maintained in Catalan. However, Mascaró (1987) pointed out several confounding factors that made any interpretation of Dinnsen & Charles-Luce's findings problematic. For example, since not all stimuli words had paradigmatic forms in which the critical consonant was prevocalic, the underlying voicing specification of some obstruents was not apparent. Phonetic realization of the final obstruent was also likely to have been affected by voicing assimilation across the word boundary in the '_#C' condition. In addition, some 'C#C' sequences were likely to undergo gemination and surface as [C:], which would make it difficult to determine the exact durations of consonantal closures. Finally, although the actual Spanish glosses were not provided in the description of the study, the Spanish words were most likely cognates of the Catalan wordforms in which the voicing contrast would be represented graphemically (e.g., the Spanish equivalent of the Catalan word *cap* 'head' (ending in /b/) is *ca<u>b</u>eza*, which contains a voiced grapheme). As such, Spanish orthography could have biased the way the stimuli from Catalan were pronounced.

In response to Mascaró's criticisms (prior to their publication), Charles-Luce & Dinnsen (1987) reanalyzed the data from Dinnsen & Charles-Luce (1984). They used the same statistical model as before and retained the data from all 5 original speakers but included only those stimuli items that showed paradigmatic alternations and were not expected to undergo gemination (n=4, 2 minimal pairs; 2 words ending in /p, b/, 2 ending in /t, d/). Reanalysis of the data revealed that no voicing-dependent effects were present for either preceding vowel duration or closure duration. Closure voicing showed a significant influence of the underlying contrast (1 ms longer for voiced stops). Individual results went in the expected direction but were not statistically significant. Charles-Luce & Dinnsen claimed that these updated findings confirmed that Catalan speakers maintained voicing-dependent cues in production. This conclusion, however, was once again challenged in Mascaró (1987) who expressed doubt that a difference of 1 ms could have any real consequences for either production or perception of speech and also mentioned an additional potential problem of not controlling the stimuli for differences in stress patterns.

2.3.4.2. Charles-Luce (1993)

Charles-Luce (1993) investigated the effects of underlying voicing and semantic predictability on the production of word-final coronal and velar stops (/t, d, k, ɡ/) that underwent voicing assimilation across word boundary. Bilingual Catalan/Spanish speakers (n=5) performed a reading task with 5 repetitions. Test words (n=10, 5 minimal pairs; 4 items with final coronals, 6 with final velars) and fillers (n=10) were embedded

sentence-medially in semantically biasing and semantically neutral contexts. Two phonetic environments were used for each context (followed by an [s]-initial word, followed by an [r]-initial word). The environments were chosen such that they would both trigger regressive voicing assimilation across the word edge. Preceding vowel duration, closure duration and closure voicing were entered as dependent variables into analyses of variance (by-subject, with repeated measures), with underlying voicing, semantic context and phonetic environment used as independent factors.

Statistical analyses revealed that differences in preceding vowel duration were significant in two (out of 5) stimuli pairs in the semantically biasing condition (20~22 ms longer vowels before voiced stops; significant in one phonetic environment for each pair) and also significant in 3 pairs in the semantically neutral condition (9~36 ms difference in the expected direction; significant in both phonetic environments for all 3 pairs). No effect of underlying voicing was present for closure duration or closure voicing. On the basis of these data from 5 speakers and 5 item pairs, Charles-Luce concluded that the voicing contrast was maintained more in the absence of semantic bias and that Catalan speakers used preceding vowel duration to signal the voicing contrast. However, in later work, Charles-Luce (1997) acknowledged that these conclusions were problematic because, for example, the observed durational effects could have been affected by prosodic lengthening. Burton & Robblee (1997) also pointed out that the two consonants used in the following phonetic environments ([s, r]) differed in more parameters than just underlying voicing, which could have also influenced the degree of neutralization.

2.3.4.3. Kopkalli (1993)

Kopkalli (1993) studied how speakers of Turkish (n=5) produced the voicing contrast in word-final labial, coronal and velar plosives (/p, b, t, d, k, g/) during a reading task.[13] Test items (n=60, one minimal pair; 18 words with final labials, 22 with final coronals, 20 with final velars) and fillers (n=60) were monosyllabic and disyllabic nouns and verbs (including several archaic words) that were pronounced in isolation as well as phrase-medially followed by a vowel-initial word. A total of 4 repetitions were recorded (two repetitions per environment), and all speakers were familiarized with the stimuli prior to testing. Acoustic measurements of preceding vowel duration, closure duration, release duration and closure voicing were analyzed using separate Mixed Models ANOVAs for each participant (with 'word' as a random effects factor). To determine whether the speakers knew the underlying forms of all final consonants, participants were also subjected to a post-test during which they were asked to add a vowel-initial suffix at the end of each test item and to pronounce the inflected form. Their productions were examined impressionistically.

The study revealed that, as a group, Turkish speakers maintained only small (1~3 ms) and non-significant differences in vowel duration, closure duration, release duration, and closure voicing. Individual data showed that one speaker produced longer releases for voiced plosives in words pronounced in isolation (9 ms difference; the direction was not expected). Another participant had longer vowels before voiced stops (6 ms difference) and longer closures in voiceless plosives in words produced in isolation (7 ms

[13] Kopkalli (1993) also investigated perceptual aspects of the voicing contrast in Turkish. Details of the perception study are provided in Section 2.4.4.1.

difference). The remaining 3 participants did not show any voicing-dependent differences. Crucially, the effects of underlying voicing diminished substantially when speakers' familiarity with the stimuli items (i.e., whether or not they knew the underlying representation of the critical consonant) was added to the statistical model as an additional predictor variable. Kopkalli concluded that Turkish speakers did not preserve the voicing contrast in final obstruents. However, since very few speakers took part in the study, it remained to be confirmed that these results were in fact representative of the Turkish language in general.

2.3.5. Data from Polish and Russian

Experimental investigations of the voicing contrast in word-final obstruents have also been published for Polish (Slowiaczek & Dinnsen, 1985; Jassem & Richter, 1989) and Russian (Chen, 1970; Pye, 1986; Dmitrieva et al., 2010). The two Slavic languages are both known in the traditional literature to prohibit voiced stops and fricatives in word-final positions (Polish: Benni, 1964; Wierzchowska, 1980; Russian: Avanesov, 1956; Wade, 2010). However, only the Polish study in Jassem & Richter (1989) claimed that voicing-dependent cues were not maintained in production. The rest of the experimental investigations all reported retention of cues to underlying voicing (Polish: Slowiaczek & Dinnsen, 1985; Russian: Chen, 1970; Pye, 1986; Dmitrieva et al., 2010). Descriptions of the Polish and Russian studies are presented below, starting with the research on Polish.

2.3.5.1. Slowiaczek & Dinnsen (1985)

Slowiaczek & Dinnsen (1985) studied how speakers of Polish (n=5) realized the voicing contrast in 14 different word-final plosives, fricatives, and affricates (/p, b, t, d, k, ɡ, f, v, ʃ, ʒ, ɕ, ʑ, tɕ, dʑ/). Participants performed a reading task with 4 repetitions. Test items (n=30, 15 minimal pairs; 18 ending in plosives, 6 in fricatives, 6 in affricates) and filler items (n=36) were a mix of monosyllabic nouns, verbs, prepositions and numerals, including some inflected forms. The items were embedded phrase-medially in two different carrier frames such that they were followed by either a consonant-initial word or a vowel-initial word. Measurements of preceding vowel duration, closure/frication duration and closure voicing were used as dependent variables in analyses of variance (by-subject) that examined the effects of underlying voicing and phonetic environment. Separate analyses were performed for each measurement and each place and manner of articulation.

For preceding vowel duration, voicing-dependent differences were found for all obstruents (10% longer vowels when followed by voiced sounds; seen in 4 (out of 5) speakers). Differences in closure voicing were observed only for labial plosives (13 ms longer in voiced stops; seen in 4 speakers). Closure duration differences were found only in the individual data of one participant (13 ms longer in voiceless obstruents). For some speakers, underlying voicing also interacted with phonetic context. On the basis of these data from 5 speakers, Slowiaczek & Dinnsen concluded that word-final devoicing in Polish was not neutralizing. However, Jassem & Richter (1989) argued that Slowiaczek & Dinnsen had misjudged the domain of final devoicing (according to the former,

devoicing should have been viewed as utterance-final rather than word-final) and that they also failed to take into account the potential interference from English (all 5 speakers were residing in the USA at the time of testing).

2.3.5.2. Jassem & Richter (1989)

Jassem & Richter (1989) tested the same set of 14 word-final stops, fricatives, and affricates as Slowiaczek & Dinnsen (1985) (i.e., /p, b, t, d, k, g, f, v, ʃ, ʒ, ɕ, ʑ, tɕ, dʑ/).[14] However, unlike the previous study, they recruited monolingual speakers from Poland (n=4) and did not expose participants to orthographic representations of test items. Instead, stimuli words (n=34, 17 minimal pairs; 2 ending in /p, b/, 8 in /t, d/, 4 in /k, g/, 4 in /f, v/, 4 in /s, z/, 6 in / ʃ, ʒ/, 2 in /ɕ, ʑ/, and 4 in /tɕ, dʑ/) were elicited as unprepared one-word utterances in response to a series of oral questions asked by an experimenter. Questions were based on general knowledge (e.g., asking which bird has the most beautifully colored tail; the expected answer would be 'peacock') as well as linguistic knowledge (e.g., asking to provide the diminutive form of a noun). Preceding vowel duration, closure/frication duration, and burst duration were entered as dependent variables into analyses of variance that were performed on non-averaged data from 16 (out of 17) word pairs. Underlying voicing, consonantal manner, type of question and speaker were used as independent factors.

Jassem & Richter reported that, overall, vowels were shorter before voiceless obstruents (4 ms difference). The expected pattern was found in 3 (out of 4) speakers but

[14] Jassem & Richter (1989) also investigated perceptual aspects of the voicing contrast. For details on the perception experiment, see Section 2.4.5.2.

reversed in one participant, who produced longer vowels before voiceless segments (2 ms difference). Preceding vowel duration was also affected by question type. Vowels were shorter before voiceless obstruents in the linguistic knowledge condition (6 ms difference) but longer before voiced obstruents in the general knowledge condition (2 ms difference). Voiceless obstruents were longer than their voiced counterparts regardless of the type of knowledge involved (4 ms difference). Although only 4 speakers were tested and some voicing-dependent differences were in fact found, Jassem & Richter stated that the observed effects were small and not consistent and concluded that there was no reliable evidence of the maintenance of the underlying voicing contrast in Polish.

2.3.5.3. Chen (1970)

For Russian, preservation of cues to underlying voicing is reported in Chen (1970) who conducted a cross-linguistic (English, French, Korean, Russian) acoustic study of the effects of obstruent voicing on the parameter of preceding vowel duration. Participants (one per language) performed a word-reading task. Monosyllabic and disyllabic members of minimal and near-minimal pairs that contained word-final and word-medial (intervocalic, pre-consonantal) stops and fricatives served as stimuli items. Labial, coronal and velar stops (/$p^{(j)}$, $b^{(j)}$, t, d, k, g/) were tested for Russian. Plosive-final items (n=10, 3 minimal pairs; 4 ending in labials, 4 in coronals, 2 in velars) and tokens containing stops in word-medial positions (n=12) were all repeated 6 times (3 times in isolation, 3 times in a minimal/near-minimal pair). The measurement of preceding vowel duration was calculated with data pooled across all places and manners of articulation,

word lengths, and consonantal positions. Only overall means were reported, and inferential statistics were not performed.

The study revealed that Russian vowels were longer before voiced obstruents (28 ms difference). A similar pattern was also observed in the remaining 3 languages (29~92 ms difference). Although these results were based on descriptive data from one speaker per language and very few test words (that were all averaged together), Chen proposed that preceding vowel duration was a major cue to obstruent voicing in Russian. Furthermore, Chen suggested that the observed effect was not a compensatory adjustment but rather the result of inherent differences in transitions from vowels to voiced versus voiceless consonants and that presence of longer vowels before voiced obstruents should be a language-universal phenomenon (with the exact amount of lengthening being language-specific). Finally, even though phonetic voicing was not measured experimentally, Chen also stated that glottal pulsing was not used in Russian to signal the voicing specification of final stops.

2.3.5.4. Pye (1986)

Pye (1986) investigated the production of labial, coronal and dorsal plosives and fricatives in Russian (/p, b, $t^{(j)}$, $d^{(j)}$, k, g, s, z, ʃ, ʒ/). Participants of the experiment (n=5) performed a reading task with two repetitions. Test items (n=36, 18 minimal pairs) and fillers (n=34) were a mix of nouns, verbs, and prepositions. Words with final stops (n=26; 4 ending in labials, 14 in coronals, 8 in velars) and items with final fricatives (n=10; 8 ending in coronals, 2 in palatals) were embedded in carrier sentences utterance-finally.

Measurements of preceding vowel duration, closure/frication duration and closure voicing were reported only as overall means and not subjected to inferential statistics.

Pye observed voicing-dependent differences in preceding vowel duration (5~20 ms longer vowels before voiced obstruents), closure/frication duration (6~30 ms shorter for voiced segments), and closure voicing (1~33 ms longer for voiced stops and fricatives). Differences in vowel duration were less robust for coronal obstruents as well as during the second repetition of test words. Differences in closure/frication duration were also smaller for coronals but, in contrast to vowel duration, they increased during the second reading of stimuli items. At the individual level, one speaker demonstrated a pattern that was consistent with the view that the voicing contrast was maintained in production. Three participants showed clear exaggeration of voicing cues, which Pye attributed to their awareness of the goal of the study and described as a case of spelling pronunciation. The remaining one participant showed no significant differences for any of the acoustic parameters during the first repetition of stimuli items. Even though Pye was aware of the limitations of the study (e.g., absence of inferential statistics, lack of a consistent pattern across all participants, undue emphasis on orthographic forms, speakers' awareness of the goal of the experiment), the conclusion of the investigation was that Russian speakers maintained the underlying voicing contrast in final obstruents.

2.3.5.5. Dmitrieva et al. (2010)

More recently, Dmitrieva et al. (2010) investigated the voicing contrast in 12 different word-final obstruents (/p, b, t, d, k, g, f, v, s, z, ʃ, ʒ/) in the speech of

monolingual speakers of Russian who resided in Russia (n=4) versus speakers of Russian who resided in the USA and spoke English as their L2 (n=7).[15] Participants performed a word-reading task with 3 repetitions. Test words (n=68, 32 minimal pairs, 2 near-minimal pairs; half ending in plosives) and filler items (n=58) were existing Russian words that belonged to various grammatical categories (nouns, verbs, pronouns, adverbs; including rare/archaic words, abbreviations, inflected forms). Test words were either monosyllabic (n=64) or disyllabic (n=4). For words with final stops, the critical consonant was labial (n=10), coronal (n=12), or velar (n=12). For words ending in fricatives, the final segment was labial (n=4), coronal (n=22), or palatal (n=8). Mean lexical frequency of items ending in voiced obstruents was matched to that of words ending in voiceless stops and fricatives. Stimuli items were arranged in a specific order that was meant to create an impression of a rhyming/semantic task. Members of the same minimal pair were always separated from each other. Acoustic measurements of preceding vowel duration, closure/frication duration (a single measurement averaged across plosives and fricatives), release duration and closure voicing were based on the recordings from two repetitions. The data were subjected to analyses of variance (by-subject; with repeated measures) as well as a series of t-tests.

Dmitrieva et al. found an asymmetry in the production of voicing cues between the speakers residing in Russia versus the participants living in the USA. For speakers from Russia, significant differences were found only for the parameters of closure duration and release duration (in both cases, 16 ms longer for voiceless obstruents). For

[15] Dmitrieva et al. (2010) also tested how the same contrast was produced by native speakers of English who were learning Russian. Results of this group are omitted from the present discussion.

Russian speakers from the USA, significant differences were observed in all 4 acoustic parameters, including longer vowels before voiced stops and fricatives (8 ms difference), longer closures for voiceless obstruents (15 ms difference), longer releases for voiceless consonants (17 ms difference), and longer closure voicing for phonologically voiced sounds (6 ms difference). Dmitrieva and colleagues concluded that both groups maintained the underlying voicing contrast in production and that exposure to English resulted in increased robustness of acoustic cues. However, Dmitrieva et al. noted that the group-dependent variability in the use of acoustic cues could be attributable to differences in experimental setup rather than differences in linguistic background (i.e., the group from the USA was recorded in a research lab, whereas the group from Russia was tested in a non-laboratory environment). Furthermore, the conclusions in Dmitrieva et al. were largely based on a series of t-tests that compared very small sets of data (e.g., two sets of 4 means, with only 3 degrees of freedom remaining in the model). The authors also did not discuss the possible effects of language attrition that are well known to affect the production of speech in populations with limited exposure to L1 (Polinsky & Kagan, 2007).

2.3.6. Production Data - Summary

As shown in the preceding sections, previous production studies of the voicing contrast in word-final obstruents in neutralizing languages have often reported inconsistent or contradictory results, with some investigations advocating for retention of voicing-dependent differences but others arguing for complete loss of acoustic cues to

underlying voicing in phonetic outputs. However, many of the previous studies tested very small samples of speakers and test items, did not control for exogenous variability and used very different experimental methodologies, which makes any interpretation of such findings or their direct comparison problematic.

A general summary of the group-level results of the previous investigations and some of the relevant methodological differences are provided in Table 1. In the table, the 'Task' column specifies the nature of the experimental task (reading versus oral) and the number of repetitions of each stimulus item. The 'MPs' column shows whether full minimal pairs were included ('Y') or excluded ('N') from the list of test items (when such pairs constituted only a small subset of test items, the exact number of minimally contrasting pairs is provided in parenthesis). The 'Subj' column lists the number of participants. The 'Stim' column shows the number of stimuli items ending in an obstruent. The 'Final Cs' column specifies the final consonants of test tokens. The 'Results' column presents the group-level findings for the acoustic parameters of preceding vowel duration ('Vdur'), closure/frication duration ('CLdur'), release duration ('Rdur'), and closure voicing ('Voi'). Differences between underlyingly voiced and voiceless obstruents are marked with 'Y' (if observed) or 'N' (if not observed). An asterisk ('*') is placed next to the results based on descriptive statistics only.

Table 1: Previous production findings – Summary

Language	Study	Task (repetitions)	MPs	Subj	Stim	Final Cs	Results			
							Vdur	CLdur	Rdur	Voi
German	*Port & O'Dell (1985)*	reading (x 2)	Y	10	20	p,b,t,d,k,g	Y	Y	Y	Y
	Charles-Luce (1985)	reading (x 5)	Y	5	8	t,d,s,z	Y	N	n/a	Y
	Port & Crawford (1989)	reading (x 1-4)	Y	5	6	t,d	N	N	Y	N
	Piroth et al. (1991)	reading (x 3)	Y (1)	2	18	p,b,t,d,k,g	Y	Y	Y	n/a
	Piroth & Janker (2004)	reading (x 3)	Y (1)	6	26	p,b,t,d,k,g, f,v,s,z	N	Y (C-dur)		N
	Smith et al. (2009)	reading (x 10)	Y	13	12	t,d,k,g	Y	N	Y	Y
	Röttger et al. (2011)	oral (x 1)	Y	16	48	p,b,t,d,k,g	Y	N	N	n/a
	Fourakis & Iverson (1984)	oral (x 2)	N	4	11	t,d,k,g,s,z	N	N	N	n/a
		reading (x 4)	Y		10	p,b,t,d,k,g	Y	N	n/a	n/a
	Inozuka (1991)	reading (x 2)	N	6	18	k,g	N	N	N	N
Dutch	*Warner et al. (2004)*	reading (x 2)	Y	15	40	t,d	Y	N	Y	N
Afrikaans	*Van Rooy et al. (2003)*	reading (x 2)	Y	8	20	t,d	Y	Y	Y	Y
Catalan	*Dinnsen & Charles-Luce (1984); Charles-Luce & Dinnsen (1987)*	reading (x 5)	Y	5	10	p,b,t,d,k,g	N	N	n/a	N
					4	p,b,k,g	N	N	n/a	Y
	Charles-Luce (1993)	reading (x 5)	Y	5	10	t,d,k,g	Y	N	n/a	N
Turkish	*Kopkalli (1993)*	reading (x 4)	Y (1)	5	60	p,b,t,d,k,g	N	N	N	N
Polish	*Slowiaczek & Dinnsen (1985)*	reading (x 4)	Y	5	30	p,b,t,d,k,g, f,v,ʃ,ʒ,ɕ,ʑ, tɕ,dʑ	Y	N	n/a	Y
	Jassem & Richter (1989)	oral (x 1)	Y	4	34	p,b,t,d,k,g, f,v,ʃ,ʒ,ɕ,ʑ, tɕ,dʑ	Y	Y (C-dur)		n/a
Russian	*Chen (1970)*	reading (x 6)	Y	1	10	$p^{(j)}$,$b^{(j)}$,t,d, k,g	Y*	n/a	n/a	n/a
	Pye (1986)	reading (x 2)	Y	5	36	p,b,$t^{(j)}$,$d^{(j)}$, k,g,s,z,ʃ,ʒ	Y*	Y*	n/a	Y*
	Dmitrieva et al. (2010)	reading (x 3)	Y	11	68	p,b,t,d,k,g, f,v,s,z,ʃ,ʒ	N/Y	Y/Y	Y/Y	N/Y
							(monoling./biling.)			

2.4. Experimental Perception Findings

Maintenance of the underlying voicing contrast in final obstruents has also been attested in perception, with several experimental studies claiming that voicing-dependent properties of test items can affect listeners' responses in perceptual tasks (Port & O'Dell, 1985; Port & Crawford, 1989; Slowiaczek & Szymanska, 1989; Ernestus & Baayen, 2007a; Matsui, 2011; Röttger et al., 2011). However, the magnitude of the observed effects has generally been modest, and some researchers have questioned the perceptual relevance of voicing cues even if above-chance-level performance was attested (Port & Crawford, 1989; Slowiaczek & Szymanska, 1989; Röttger et al., 2011). Furthermore, similarly to production results, perception findings have also been inconsistent, with a number of studies reporting chance-level performance or finding a strong bias for either voiceless or voiced responses (Jassem & Richter, 1989; Kahlene-Halstenbach, 1989; Kopkalli, 1993; van Rooy et al., 2003). Behavioral measurements and general procedures of perception studies are described in Section 2.4.1. that follows. Methodologies and findings of selected published experiments are summarized in Section 2.4.2. through Section 2.4.5., with studies grouped by language of investigation and authors' conclusions on whether or not the underlying contrast is preserved in perception. A general summary of the previous findings is given in Section 2.4.6.

2.4.1. Procedures and Behavioral Measurements

The majority of previous perception studies on the voicing contrast in word-final obstruents used a forced choice two-alternative identification task (e.g., Port & O'Dell,

1985; Port & Crawford, 1989; Warner et al., 2004; Röttger et al., 2011). In such a task, stimuli items are presented auditorily one word/sentence at a time, and listeners are asked to identify each stimulus by circling one of two choices on an answer sheet or pressing one of two designated keys on a computer keyboard or a button box. The two choices are usually shown side-by-side and include the actual stimulus item and its minimal pair counterpart. In addition to (or instead of) forced-choice identification tasks, some previous studies utilized open-choice identification (i.e., participants write down/type the stimulus item; van Rooy et al., 2003), a rating task (i.e., participants rate on a Likert scale how voiced/voiceless a given token sounds; Ernestus & Baayen, 2007a), or a discrimination task (i.e., participants listen to 2 or 3 stimuli items presented consequently with a short pause between them and indicate whether the two items represent the same category or, in the case of 3 items, whether the token in the middle is identical to the first item or the last item; Kopkalli, 1993; Matsui, 2011). For all types of experimental procedures, perception data are usually analyzed using parametric statistical tests (t-tests, analyses of variance, regression analyses; e.g., Port & O'Dell, 1985; Kopkalli, 1993; van Rooy et al., 2003; Warner et al., 2004; Matsui, 2011; Röttger et al., 2011). However, some previous studies relied on non-parametric analyses (Jassem & Richter, 1989) or did not perform any inferential statistics (Kahlene-Halstenbach, 1989).

With respect to behavioral measurements, all neutralization studies discussed in the present section registered and analyzed only response types (i.e., whether the final obstruent was heard as 'voiced' or 'voiceless' or, alternatively, how much voicing was heard). In some studies, participants' responses were converted into error rates, with data

pooled across both underlyingly voiced and voiceless tokens (e.g., Port & Crawford, 1989). Other investigations separated responses between the two voicing categories and took into consideration both the hit rate and the false positive rate by converting responses into d′ (d-prime) scores (e.g., Slowiaczek & Szymanska, 1989) or by performing statistics on the proportions of voiced (or voiceless) responses (e.g., Warner et al., 2004).

2.4.2. Data from German

Perceptual aspects of the voicing-dependent differences in word-final obstruents in German have so far been investigated in very few published works (Port & O'Dell, 1985; Port & Crawford, 1989; Kahlene-Halstenbach, 1989; Röttger et al., 2011). The studies by Port & O'Dell (1985), Port & Crawford (1989) and Röttger et al. (2011) all argued for listeners' ability to perceive acoustic cues to intended voicing at above-chance level. Kahlene-Halstenbach (1989) concluded that voicing-dependent differences were fully neutralized in perception. Methodologies and main findings of the German studies are summarized below, starting with the investigations supporting partial neutralization of voicing cues in perception.

2.4.2.1. Port & O'Dell (1985)

Port & O'Dell (1985) explored how speakers of German perceived voicing cues in word-final labial, coronal and velar stops (/p, b, t, d, k, g/). Participants (n=10, 4 also took part in the production experiment) performed an identification task (forced-choice two-alternative). Test tokens were based on the set of 20 test words (10 minimal pairs)

that were recorded during the production component of the study.[16] Participants' responses were converted into error rates and subjected to (i) a correlatory analysis that examined the relation between intended voicing and identification accuracy and (ii) a multiple regression analysis that tested the extent to which participants' responses could be modeled on the basis of the acoustic parameters measured in the production experiment (i.e., preceding vowel duration, closure duration, release duration, and closure voicing).

Results of the study demonstrated that speakers could correctly identify intended voicing in 59% of the stimuli, with error rates varying by up to 8% depending of consonantal place of articulation (60% correct for labials, 55% correct for coronals, and 63% correct for velars). Given that chance level was estimated at 50%, these findings were interpreted as above-chance performance. However, correlatory analysis revealed that underlying voicing could explain only 12% of variation in listeners' judgments, with results of the multiple regression analysis showing that preceding vowel duration, closure duration, release duration and closure voicing were all significant predictors of listeners' responses. Port & O'Dell concluded that the voicing contrast was preserved in German not only in production but also in perception.

2.4.2.2. Port & Crawford (1989)

Port & Crawford (1989) examined perceptual aspects of the voicing contrast in word-final coronal stops (/t, d/). Speakers of German (n=5, none took part in the

[16] Only a subset of the original production data was used in the perception experiment (50% of all tokens; usually taken from the first two repetitions). Details of the production experiment are presented in Section 2.3.2.1.

production study) performed an identification task (forced-choice two-alternative). Test tokens were based on 6 German words (3 minimal pairs, including both alternating and non-alternating words). Stimuli items were a subset of the acoustic data recorded during the production module of the study.[17] The tokens selected for the perception experiment were produced by two speakers during dictation and word-reading. Listeners' responses were converted into error rates and analyzed using discriminant analyses that tested whether the rate of correct identification could be predicted from the measurements of preceding vowel duration, closure duration, burst duration, and closure voicing.

Port & Crawford reported an overall accuracy rate of 69%, with slightly better performance observed for words produced during dictation. Discriminant analyses could correctly model listeners' responses in 71% of the cases on the basis of all 4 acoustic parameters (i.e., preceding vowel duration, closure duration, burst duration, and closure voicing). Port & Crawford interpreted these results as better-than-chance performance on underlying voicing and proposed that German speakers relied on a variety of cues when asked to disambiguate between underlyingly voiced versus voiceless final obstruents. However, although voicing-dependent differences were in fact significant in the experimental data, Port & Crawford stated that such cues were unlikely to play a role in the perception of non-laboratory speech.

2.4.2.3. Röttger et al. (2011)

Röttger et al. (2011) tested the perception of voicing cues in word-final labial, coronal and velar stops (/p, b, t, d, k, g/). German speakers (n=8, none participated in the

[17] For details on the production experiment, see Section 2.3.2.3.

production experiment) took part in an identification task (forced-choice two-alternative). Test tokens (n=192) were a subset of the 24 minimal pair pseudo-words produced by 16 different subjects during the production component of the same study.[18] Before the experiment, listeners were told that half of the stimuli ended in voiced plosives and the other half contained voiceless sounds. Participants' responses were analyzed as a dependent variable in a mixed logit model, with intended voicing included as an independent factor.

Röttger and colleagues found that although participants showed a general bias for voicelessness, their identification responses corresponded to underlying voicing in 54% of the cases, with individual accuracy rates ranging from 50% to 59%. These results were argued to be significantly above chance-level and were viewed as evidence of the maintenance of the voicing contrast in perception even when using non-words and when listening to many different voices during the experiment. At the same time, the relatively low accuracy rates were taken to be an indication of "the minor functional relevance of IN [incomplete neutralization] (p. 1725)".

2.4.2.4. Kahlene-Halstenbach (1989)

Kahlene-Halstenbach (1989) investigated how speakers of German (n=10) perceived word-final labial, coronal and velar plosives (/p, b, t, d, k, g/). Participants performed an identification task (forced-choice two-alternative). Test words (n=40, 20 minimal pairs; 2 ending in labials, 36 in coronals, 2 in velars) and fillers (n=40) were

[18] Details of the production experiment are given in Section 2.3.2.7.

nouns, verbs, adjectives, and adverbs (including some inflected forms). The stimuli were originally recorded in a word-list reading task and came from both male and female speakers (n=6). During the identification experiment, each participant heard only half of the tokens, with either the voiced or the voiceless member taken from each minimal pair. Listeners were told that the goal of the experiment was to investigate gender-related differences in speech. Participants' responses were converted into error rates. Only descriptive statistics were reported. No acoustic or statistical analyses were done on the production data that served as stimuli items.

The study revealed that 58% of the stimuli were attributed to the correct voicing category. Although inferential statistics were not performed, Kahlene-Halstenbach argued that the 8% difference was non-substantial and stated that German listeners did not perceive the intended voicing of final plosives even though voicing-dependent acoustic cues were most likely present in the stimuli. The study concluded that final devoicing was fully neutralizing in perception and that laryngeal neutralization had psychological reality in the minds of German speakers.

2.4.3. Data from Dutch and Afrikaans

Perception of the voicing contrast in word-final obstruents has also been examined in Dutch (Warner et al., 2004; Ernestus & Baayen, 2007a) and Afrikaans (van Rooy et al., 2003). The two Dutch studies found that voicing-dependent differences were perceptible. However, whereas Ernestus & Baayen (2007a) interpreted their findings as robust evidence of incomplete neutralization of voicing cues, Warner et al. (2004) argued

that the magnitude of the observed effects was small and that any cues to underlying voicing that were present in production played either little or no role in perception. The Afrikaans study in van Rooy et al. (2003) found a bias for voicelessness and argued for perceptual neutralization of the underlying contrast. Methodologies and main findings of these studies are described below, starting with the work on Dutch.

2.4.3.1. Warner et al. (2004)

Warner et al. (2004) studied how speakers of Dutch (n=62) perceived the voicing contrast in word-final coronal stops (/t, d/). Participants performed a series of identification tasks (forced-choice two-alternative) and were assigned to one of 3 experimental conditions: (i) listening to non-modified stimuli (30 subjects, 14 also took part in the production task), (ii) listening to stimuli items with adjusted vowel duration (16 subjects), and (iii) listening to tokens with adjusted closure duration (16 subjects). Perceptual tokens were created on the basis of 40 words (20 minimal pairs) produced by 4 speakers during the production component of the same investigation.[19] The speakers were selected such that the strength of voicing-dependent cues would vary across their productions. Namely, one speaker maintained robust differences in both preceding vowel duration and release duration. Another speaker showed robust differences in vowel duration but not release duration. The third speaker produced no differences in vowel duration and a small difference in the unexpected direction in release duration. The fourth speaker demonstrated differences in the unexpected direction in both vowel duration and release duration. Modifications in vowel duration were achieved by adding or removing

[19] Details of the production study are provided in Section 2.3.3.1.

one to 5 periods in the middle of the vowel. Closure duration was manipulated by lengthening or shortening the closure by 10 ms to 50 ms. Participants' responses (proportion of voiceless responses; also, d′ scores for a subset of the data) were analyzed using analyses of variance (by-subject, by-item; with repeated measures) that examined the effects of underlying voicing, speaker, phonemic vowel length and, for a subset of test items, listeners' participation in the production experiment.

For the non-modified stimuli items, perception results demonstrated that listeners' responses varied depending on which speaker the data came from. For the two speakers who maintained robust vowel duration differences (but differed with respect to the release duration cue), around 62% of the items ending in voiceless stops and 50% of the items ending in voiced plosives were classified as voiceless. For the speaker who did not maintain differences in vowel duration and produced small differences in the unexpected direction in burst duration, 54~55% of all items were identified as voiceless regardless of the underlying voicing specification. For the speaker who showed reversal of differences in both vowel duration and release duration, approximately 60% of the items ending in /t/ and 61% of the words with final /d/ were attributed to the voiceless category. These findings suggested that Dutch listeners were relying on the acoustic cue of preceding vowel duration. As expected, when vowel duration was manipulated experimentally, fewer voiceless responses were observed for tokens with longer vowels, confirming that longer nuclei were associated in perception with an underlying /d/. Manipulation of closure duration affected participants' responses only for those stimuli that contained phonemically long vowels, with longer closures attributed to voicelessness. Warner et al.

concluded that listeners performed slightly above chance level. At the same time, since the difference in vowel duration that was observed in the production experiment was small (i.e., 3.5 ms), they argued that vowel length may be a reliable cue in experimentally-elicited data but not necessarily so in more natural speech. The effect of closure duration (that did not show significant differences in production but was identified as an important cue in perception for a subset of test items) was interpreted as listeners' ability to rely on cues that were not produced consistently but could be associated with the same underlying contrast in other environments (e.g., in intervocalic positions).

2.4.3.2. Ernestus & Baayen (2007a)

Ernestus & Baayen (2007a) examined how speakers of Dutch (n=40) perceived voicing cues in word-final labial and coronal stops and fricatives (/p, b, t, d, f, v, s, z/). Participants performed a rating task in which they heard one token at a time and were asked to rate on a 5-point Likert scale how much voicing was present in the final obstruent. Half of the participants heard the entire wordform. The remaining listeners heard only the final rhyme that started with the steady-state of the preceding vowel. Stimuli items (n=94, no minimal pairs) were both existing and non-existing monosyllabic wordforms that were produced by a speaker of Dutch who was specifically instructed to maintain the voicing contrast. Test items ending in plosives (n=59) included 30 words with underlyingly voiced final stops and 29 words with underlyingly voiceless plosives (10 labial, 49 coronal). For fricative-final items (n=35), 17 ended in voiced segments and

18 in voiceless sounds (14 labial, 21 coronal). Listeners' response scores were averaged and subjected to (stepwise) ANOVAs that tested for the effects of underlying voicing, surface voicing, consonantal manner, and stimulus type (full words vs. final rhymes).

The study revealed that surface voicing (i.e., the presence of phonetic voicing in the tokens) was a better predictor of participants' responses than underlying voicing, which was largely due to the fact that effects of underlying voicing were robust in plosives but not fricatives. Furthermore, although both full and truncated tokens showed similar voicing-dependent differences, full word tokens showed a stronger effect of underlying voicing. Ernestus & Baayen attributed the observed asymmetry between plosives and fricatives to Dutch using voiced graphemes for final plosives but not final fricatives (e.g., morphologically related plosive-final words *mand* 'basket' and *manden* 'baskets' are both spelled with *d*; fricative-final *baas* 'boss' and *bazen* 'bosses' and spelled with *s* and *z*, respectively). The finding of an asymmetry between full words and their rhymes was explained as interference from lexical knowledge. Ernestus & Baayen concluded that Dutch listeners identified voicing at above-chance level and that they relied on not only acoustic cues but also lexical factors.

2.4.3.3. Van Rooy et al. (2003)

The Afrikaans study in van Rooy et al. (2003) examined the perceptual salience of acoustic cues to underlying voicing in word-final coronal stops (/t, d/). Two different identification tasks (open-choice, two-alternative forced-choice) were administered to a total of 39 speakers (19 performed the open-choice task, 20 participated in the forced-

choice task). Stimuli items (n=10, 5 minimal pairs) were created using the production data from Wissing & van Rooy (1992).[20] Test items were monosyllabic words produced by 4 speakers of Afrikaans who performed a reading task with 2 repetitions. Listeners' identification responses were converted into error rates as well as d′ scores and subjected to analyses of variance (using non-averaged data). Underlying voicing and the phonetic parameters of preceding vowel duration, closure duration, burst-only duration, release duration and closure voicing were used as fixed effects predictors. Stimulus pair was used as the random effects factor.

Results of the open-choice task revealed better identification accuracy for /t/ (70% correct) than /d/ (41% correct), with preceding vowel duration being the only significant predictor of participants' responses. The pattern was observed in 3 (out of 5) word pairs, and some listeners scored 100% correct for voiceless stops and 0% correct for voiced plosives. The effect was reversed and not significant in the remaining two pairs, which was attributed to the influences of usage frequency (both pairs included low frequency words with final /t/ and high frequency items with final /d/). In the forced-choice task, better-than-chance performance was observed for /t/ but not /d/ (d′ scores of 1.22 and 0.38, respectively), and both preceding vowel duration and release duration were shown to be significant predictors. Thus, the exact selection of cues on which listeners relied when identifying intended voicing varied depending on the nature of the task (i.e., open-choice versus forced-choice) and listeners' identification performance showed an overall bias for voicelessness as well as potential effects of usage frequency. Based on these

[20] For details on the recording procedure, see Wissing & van Rooy (1992) (published in Afrikaans) or an English summary in van Rooy et al. (2003).

findings, van Rooy et al. argued that the voicing contrast was neutralized in perception in Afrikaans.

2.4.4. Data from Turkish

For Turkish, perceptual aspects of the voicing contrast in word-final were investigated in Kopkalli (1993). As noted earlier, the Turkish language provides only limited graphemic cues to the underlying voicing specification of word-final obstruents, and Kopkalli reported that Turkish speakers could not correctly identify the intended voicing of final segments. A more detailed description of Kopkalli's study is provided below.

2.4.4.1. Kopkalli (1993)

Kopkalli (1993) tested how speakers of Turkish (n=7, 4 also took part in the production study) perceived voicing cues in word-final labial, coronal, and velar plosives (/p, b, t, d, k, g/). Participants performed both an identification task (forced-choice two-alternative) and a discrimination task (AX). Test tokens were created on the basis of 30 disyllabic words that were recorded during the production experiment.[21] All tokens came from one speaker who produced the most robust voicing-dependent differences. In the identification task, perceptual tokens were either final rhymes or final rhymes followed by a vowel-initial word. In the discrimination task, test tokens were final rhymes only. Participants' responses were converted into error rates. Only descriptive statistics were provided.

[21] For details on the production component of the study in Kopkalli (1993), see Section 2.3.4.3.

Results of the identification task indicated that listeners could identify intended voicing in 49% of the tokens when exposed to final rhymes and 48% of the items when listening to final rhymes followed by a vowel-initial word. The rate of correct identification was around 57~58% for words with voiced final plosives and 40~42% for items with voiceless final stops. Thus, identification performance did not improve when final plosives were released into a vowel, and participants showed an overall bias for voiced responses. Results of the discrimination task revealed an overall accuracy rate of 51%, which varied between 46% to 55% across subjects and 32% to 66% across items. Although these results were not subjected to inferential statistics, Kopkalli attributed the observed differences in error rates to the higher lexical frequency of stimuli items with voiced final plosives and concluded that Turkish listeners could neither identify nor discriminate the acoustic cues to underling voicing in word-final stops.

2.4.5. Data from Polish and Russian

Neutralization of the voicing contrast in perception has also been investigated for Polish (Slowiaczek & Szymanska, 1989; Jassem & Richter, 1989) and Russian (Matsui, 2011). The Polish study in Slowiaczek & Szymanska (1989) and the Russian study in Matsui (2001) both argued for above-chance-level performance on intended voicing. The Polish investigation in Jassem & Richter (1989) observed a strong bias for voicelessness. Notably, although the two Polish investigations reported contrasting results, the studies agreed in their general conclusion that voicing-dependent cues were unlikely to play a

significant role in the perception of everyday speech. Descriptions of the Polish and Russian studies are provided below, starting with the work on Polish.

2.4.5.1. Slowiaczek & Szymanska (1989)

Slowiaczek & Szymanska (1989) explored the perception of voicing cues in 12 different plosives, fricatives, and affricates (/p, b, t, d, k, g, f, v, ʃ, ʒ, tɕ, dʑ/). Speakers of Polish (n=21) and a control group of English speakers (n=18) performed an identification task (forced-choice two-alternative). Stimuli items (n=26, 13 minimal pairs) were created on the basis of the production data from Slowiaczek & Dinnsen (1985).[22] Speech tokens from both original phonetic contexts were used, with stronger cues to underlying voicing present in the perception stimuli than in the full set of the production data (e.g., the average difference in preceding vowel duration was 10% in the production study; the same difference was equal to 55% for the tokens tested in the perception experiment). Participants' responses were converted into error rates and analyzed using signal detection analysis and analyses of variance (by-subject), with underlying voicing and phonetic context used as independent variables.

Results of the experiment indicated that Polish listeners could correctly identify underlying voicing in 61% of the tokens. In comparison, English listeners were correct 59% of the time. Both groups also showed a similar bias for voicelessness, and the small difference in accuracy rates between the Polish and the English speakers was not statistically significant. Hence, although above-chance level performance was attested, neither direct experience with the devoicing process in one's native language nor lexical

[22] A general summary of the production study in Slowiaczek & Dinnsen (1985) is given in Section 2.3.5.1.

knowledge of stimuli words appeared to have affected identification accuracy. The observed effects of underlying voicing were also not as robust as expected given the strong cues to intended voicing that were present in the stimuli. Slowiaczek & Szymanska concluded that any acoustic differences that may have been observed in production were not reliable cues to underlying voicing in perception and suggested that voicing-dependent effects were unlikely to be found in identification of non-laboratory speech.

2.4.5.2. Jassem & Richter (1989)

Jassem & Richter (1989) investigated the perception of voicing cues in 14 different word-final obstruents (/p, b, t, d, k, ɡ, f, v, ʃ, ʒ, ɕ, ʑ, tɕ, dʑ/). Participants (n=10; 3 also took part in the production experiment) performed an identification task (forced-choice two-alternative). Stimuli items (n=34, 17 minimal pairs) were based on the set of Polish words that were elicited during the production experiment as unprepared one-word utterances in response to a series of oral questions.[23] Listeners' responses were transformed into error rates and used as a dependent variable in a non-parametric Chi-square statistic that compared the data across the two voicing categories.

For words with underlyingly voiceless final obstruents, Jassem & Richter reported that intended voicing was identified correctly in 67% of the cases. For words ending in phonologically voiced segments, the rate of correct identification was only 37% (i.e., 63% of all voiced plosives were attributed to the voiceless category). This finding revealed a strong overall bias for voiceless responses that was independent of the

[23] For details on the production experiment, see Section 2.3.5.2.

underlying voicing specification of the final obstruents. Given that only small and marginally-significant differences were observed in the production study, Jassem & Richter concluded that the voicing contrast was not preserved in either production or perception and expressed their support for the traditional view that final devoicing was fully neutralizing in Polish.

2.4.5.3. Matsui (2011)

For Russian, Matsui (2011) conducted a small-scale perception study of the voicing contrast in word-final velar stops (/k, g/). Participants (n=6) performed both an identification task (forced-choice two-alternative) and a discrimination task (AXB). Stimuli items (n=4, 2 minimal pairs) contained durational differences in preceding vowel duration (6~28 ms longer when followed by a voiced stop), closure duration (11~20 ms longer for voiceless plosives), release duration (15~22 ms longer for voiceless sounds), and closure voicing (4~6 ms more phonetic voicing for voiced stops). All test tokens were monosyllabic and were produced by the same speaker during an oral task. Participants' responses were converted into percentages and subjected to analyses of variance (by-subject; with repeated measures), with underlying voicing and word pair used as predictor variables. To compare listeners' performance across identification and discrimination tasks, d′ scores were also calculated and analyzed in an RM ANOVA that tested for the effects of experimental task and word pair.

In the case of the identification task, the effect of underlying voicing was only marginally significant, with individual rates of correct responses ranging from around

30% to almost 100%.[24] For the discrimination task, the effect of underlying voicing was significant and individual accuracy scores ranged from around 40% to 100%. Analysis of the d′ scores further confirmed that identification and discrimination results were significantly different from each other. Although these findings were based on very limited data from 6 participants and 4 test items and only the discrimination task showed a statistically significant effect of voicing, Matsui concluded that voicing cues were perceptible and that final devoicing was incomplete in Russian in the domain of perception.

2.4.6. Perception Data: Summary

As in the case of production, perception findings on the voicing contrast in word-final obstruents have been inconsistent, with some studies reporting above-chance level performance on intended voicing but other investigations arguing for complete neutralization of voicing-dependent cues in perception. Furthermore, also similarly to the work on production, many perception studies relied on small samples of data and tested the stimuli that were elicited using very different experimental methodologies. A general summary of the group-level results and relevant methodological differences are shown in Table 2. In the table, the 'Task' column specifies the experimental task (identification, discrimination, rating). The 'Subj' column lists the number of participants. The 'Stimuli' column shows the relevant properties of the stimuli, including the elicitation procedure (reading, oral), inclusion ('Y') or exclusion ('N') of full minimal pairs in the stimuli lists,

[24] Matsui (2011) does not report overall means. Individual results are estimated from graphs.

final consonants, and presence of voicing-dependent differences in production. Results are summarized in the final column, with 'Y' indicating that the voicing contrast was judged to be perceived at above-chance level and 'N' marking either chance-level performance or a voiced/voiceless bias (for the rating task, 'Y' shows that ratings were affected by underlying voicing). When applicable, percentages of correct responses and/or presence of a voiced/voiceless bias are shown inside brackets. An asterisk ('*') marks the results based on descriptive statistics only. A question mark ('?') indicates marginally significant findings, ambiguous results or incomplete descriptions of the data.

Table 2: Previous perception findings – Summary

Language	Study	Task	Subj	Stimuli					Contrast perceived? (%-correct; voi/vls bias)
				Items	Elicit. proced.	MPs	Final Cs	Voi. cues	
German	*Port & O'Dell (1985)*	identification	10	20	reading	Y	p,b,t,d, k,g	Y	Y (59%)
	Port &Crawford (1989)	identification	5	6	reading	Y	t,d	Y	Y (69%)
	Kahlene-Halstenbach (1989)	identification	10	40	reading	Y	p,b,t,d, k,g	Y?	N?* (58%)
	Röttger et al. (2011)	identification	8	24	oral	Y	p,b,t,d, k,g	Y	Y (54%; vls bias)
Dutch	*Warner et al. (2004)*	identification	62	40	reading	Y	t,d	Y	Y (variable %; vls bias for many tokens)
	Ernestus & Baayen (2007a)	rating	40	94	reading	N	p,b,t,d, f,v,s,z	Y	Y
Afrikaans	*Van Rooy et al. (2003)*	identification: -open-choice -forced-choice	19 20	10	reading	Y	t,d	Y?	N (41%-70%; vls bias) N (chance for voi)
Turkish	*Kopkalli (1993)*	identification discrimination	7	30	reading	Y (1)	p,b,t,d, k,g	Y	N* (40%-58%; voi bias) N* (51%)
Polish	*Slowiaczek & Szymanska (1989)*	identification	21	26	reading	Y	p,b,t,d, k,g,f,v, ʃ,ʒ,tɕ,dʑ	Y	Y (61%)
	Jassem & Richter (1989)	identification	10	34	oral	Y	p,b,t,d, k,g,f,v, ʃ,ʒ,ɕ,ʑ, tɕ,dʑ	Y	N (37%-67%; vls bias)
Russian	*Matsui (2011)*	identification discrimination	6	4	oral	Y	k,g	Y	Y? (30-100%) Y (40-100%)

2.5. Interpretation of Experimental Findings

Presence of partial neutralization in experimentally-elicited data as well as the observed discrepancy in experimental findings both across and within languages have been attributed in the literature to two opposing causes: (i) the influence of grammatical factors (underlying voicing specification, paradigm uniformity, homophony avoidance, etc.) versus (ii) the biasing effects of experimental methodology (the orthographic nature of the reading task, presence of full minimal pairs among the stimuli, reliance on small sample sizes, inadequate statistical treatment, etc.). These alternative explanations and the specific task-independent and task-dependent factors that are most relevant to the present discussion are presented in the sections that follow, starting with the view that the voicing contrast is maintained because of the influences coming from the grammar.

2.5.1. Grammatical Pressures

2.5.1.1. Underlying Voicing and Paradigmatic Influences

The majority of previous neutralization studies that observed production and perception differences between the surface realizations of underlyingly voiced versus voiceless final obstruents attributed the apparent presence of voicing-dependent effects to pressures from the underlying voicing specification of the final obstruent and the effects from within the word's paradigm (e.g., Port & Crawford, 1989; Port, 1996; Ernestus & Baayen, 2007a, 2007b). In other words, any final obstruent that is fully voiced in morphologically related forms is expected to retain some of the cues to its [+voiced] specification, which presumably ensures that all segments surface in accordance with

their underlying form or, if one's view of phonology does not rely on abstract representations (e.g., Ernestus & Baayen, 2007a, 2007b), in accordance with the specification that is attested throughout the word's inflectional and derivational paradigms. For example, in the case of Russian, production of a partially devoiced [d̥] at the end of /kɔd/ 'code' (instead of a discretely voiceless [t]) would allow preservation of the underlying voicing setting of the /d/ and would lead to a greater uniformity throughout the different forms of the same lexeme.

2.5.1.2. Homophony Avoidance

Speakers' attempt to preserve the voicing contrast in production has also been attributed to grammatical influences from outside of the word's paradigm, such as the pressure to avoid producing potentially homophonous wordforms (Port & Crawford, 1989; also, Snoeren et al., 2006). This means that acoustic cues to underlying voicing may be specifically maintained for those words that have minimally-contrasting lexical competitors in order to disambiguate the two competing members of the same minimal pair. For example, if neutralization were complete, the phonetic form [kɔt] would represent two distinct nouns in Russian - /kɔt/ 'cat' and /kɔd/ 'code'. If voicing-dependent cues were neutralized only partially and the word were pronounced as [kɔd̥] (with [d̥] taken to be phonetically distinct from [t]), the underlying representation /kɔt/ would no longer need to be considered and [kɔd̥] would be mapped directly onto /kɔd/.

2.5.1.3. Consonantal Place of Articulation

In addition to motivating retention of voicing cues, grammatical factors may also affect the specific degree to which the underlying voicing contrast is maintained. For example, the acoustic parameter of closure voicing is known to be sensitive to such task-independent factors as consonantal place of articulation (Gamkrelidze, 1975; Ohala, 1983, 1997). Given that vocal fold vibration requires a continuous airflow through the larynx, glottal pulsing can be maintained longer for those segments that are articulated with a greater volume of space between the glottis and the oral constriction or with a shorter period of oral closure. As such, phonetic voicing is inherently easier to sustain in consonants with more anterior constriction locations or in apical sounds for which the oral closure is formed with the tip of the tongue (which is a faster articulator than the tongue blade or tongue body). Therefore, due to the task-independent effects of the consonantal place of articulation, phonetic devoicing can be expected to be more complete in Russian words such as /lug/ 'meadow' that end in a dorsal plosive compared to such nouns as /kɔd/ 'code' or /krab/ 'crab' that end in coronal or labial stops.

2.5.1.4. Theoretical Implications

Considering the grammatical nature of such influences as underlying voicing, paradigm uniformity, lexical competition and consonantal place of articulation, many studies have argued that the apparent lack of discrete neutralization requires integration into the existing models of final devoicing and that the phonological nature of devoicing, the binarity of the voicing feature and the underlying representations of final obstruents

could all be questioned on the basis of incomplete neutralization data (e.g., Dinnsen, 1983, 1985; Port & Crawford, 1989; Port, 1996; Ernestus, 2000; Port & Leary, 2005; Ernestus & Baayen, 2007a, 2007b).

More specifically, traditional phonological theory is often thought to be more appropriate for dealing with forms that can be unambiguously attributed to specific categories, such as 'voiced' versus 'voiceless', but not with the more gradient phenomena, such as 'partially voiced', that are relegated to other components of the grammar (e.g., phonetic implementation rules; among others, Kiparsky, 1976). Considering that incomplete neutralization does provide evidence of non-categorical behaviour, Dinnsen (1983, 1985) argued that the type of discrete neutralization that is assumed in the majority of phonological analyses is not likely to be attested in any of the world's languages and that the presence of residual but sufficiently robust cues to underlying voicing confirms that final devoicing is not a phonological process. Similarly, Port (1996) and Port & Leary (2005) claimed that the observed lack of complete neutralization in laboratory speech provides robust evidence against the need for formal phonological analyses of devoicing. Port (1996) also stated that incomplete neutralization is "a worrisome chink in the dam that supports all of current symbol-based phonological theory (p. 492)" and that, in the absence of categorical neutralization, theoretical phonology "would appear to have no place to begin work (p. 492)". Port (1996) further questioned the validity of the impressionistic data upon which much of the generative phonology has been based by saying that linguistic transcriptions are inherently biased

and that many phonological phenomena could have been "completely missed due to reliance on the traditional symbolic transcriptions of speech (p. 502)".

Traditional descriptions of word-final devoicing have also been challenged on a language-specific basis (Charles-Luce, 1985; Port & O'Dell, 1985; Port & Crawford, 1989; Ernestus & Baayen, 2003, 2004, 2007a, 2007b). Charles-Luce (1985), for example, stated that final devoicing in German must be viewed as a phonetic process that does not affect any underlying properties of wordforms. Port & Crawford (1989) wrote that the "German syllable-final devoicing rule cannot be stated in terms of the same [-voice] feature that is employed in the lexical specification of words" (p. 257). More recently, Ernestus & Baayen (2007a, 2007b) suggested that the apparent preservation of the voicing contrast in the production and perception data from Dutch corroborates the view that final devoicing in Dutch is best accounted for with such non-categorical models as Exemplar theory (Bybee, 2001; Pierrehumbert, 2001, 2002) rather than an abstract phonological analysis.[25]

Maintenance of the underlying voicing contrast in final obstruents has also been claimed to have important consequences for phonetic and phonological classifications of speech sounds (Inozuka, 1991; Steriade, 1999; Ernestus, 2000). For example, Inozuka (1991) argued that devoiced segments are phonetically distinct from both voiced and voiceless realizations of the same speech sounds and, therefore, a 3-way contrast between voiced, voiceless and devoiced obstruents is needed at the phonetic level. Similarly,

[25] In earlier work, Ernestus & Baayen (2003) proposed that incomplete neutralization effects favoured using Stochastic Optimality Theory (Boersma, 1998). However, this view appears to have been abandoned in the more recent work (e.g., Ernestus & Baayen, 2007a, 2007b).

Steriade (1999) proposed that phonetically voiced and voiceless consonants have specific articulatory targets (that are achieved by means of active voicing or devoicing) but that laryngeally neutralized obstruents lack such targets at the phonetic level (i.e., they are voiced or devoiced only passively). In other words, an underlying /d/ would be pronounced as [d] in a pre-vocalic position but as [D] when it is word-final, with the latter realization sounding largely voiceless but nevertheless articulatory distinct from [t]. In formal phonology, this view has been modeled with such mechanisms as non-binary evaluation of the (surface) voicing feature (e.g., by including partially devoiced candidates in OT tableaux; van Oostendorp, 2008).

A more radical view on the nature of the voicing feature and the underlying representation of the neutralizing segment can be found in Ernestus (2000) who claimed that the 3-way distinction between voiced, voiceless and devoiced obstruents is present even at the phonological level. According to Ernestus, onset consonants are marked in the phonology as either voiced or voiceless and they undergo active voicing (or devoicing) at the phonetic level. Unlike segments in onsets, coda/word-final obstruents are not marked for voicing in the phonology and their surface forms are the result of passive voicing, meaning that they are voiced if followed by a voiced sound (e.g., if a vowel-initial suffix is added onto the stem) and devoiced if followed by a voiceless sound or a prosodic boundary (e.g., when word-final). To account for this type of behaviour in formal phonology, underlying representations of word-final obstruents would need to be not only [+/-voiced] but also unmarked for [voiced].

2.5.2. Methodological Influences

In contrast to the view that incomplete neutralization is the result of grammatical pressures, a number of researchers have proposed that the presence of partial neutralization in experimental data could be driven by methodological influences rather than grammatical factors (among others, Fourakis & Iverson, 1984; Mascaró, 1987; Manaster Ramer, 1996a, 1996b; van Rooy et al., 2003). In other words, incomplete neutralization would be an unintended consequence of the experimental manipulation and partial devoicing would not need to be modeled in formal analyses. Fourakis & Iverson (1984), for example, stated that word-final devoicing in German is "a bona fide instance of phonological neutralization (p. 149)" and expressed full support for the traditional accounts of laryngeal neutralization (e.g., Moulton, 1962). Manaster Ramer (1996a, 1996b) argued that devoicing is a categorical phonological process in Catalan, German, Polish and Russian and proposed that preservation of voicing cues is "nothing more than an experimental or statistical mirage (1996b, p. 517)". Manaster Ramer also expressed disbelief that linguistically naïve speakers but not trained phoneticians can perceive the underlying contrast. More recently, van Rooy et al. (2003) pointed out that underlying voicing is often signaled by a single acoustic parameter (e.g., burst duration in Port & Crawford, 1989) and proposed that manipulation of one cue is a conscious and task-driven adjustment of articulation.

The list of methodological factors that have been implicated in the appearance of partial neutralization in experimentally-elicited speech includes the biasing effects of orthography, presence of full minimal pairs and/or non-alternating forms among the

stimuli, failure to control test words for segmental or usage-based differences, small sample sizes, lack of proper statistical treatment, elicitation of multiple repetitions and other similar influences that are discussed in more detail in the sections that follow.

2.5.2.1. Orthographic Influences

Considering that the majority of neutralizing languages maintain the underlying voicing contrast at the orthographic level, with different graphemes used for underlyingly voiced versus voiceless obstruents (e.g., 'П' /p/ versus 'Б' /b/ in Russian), and that many experimental studies administered a reading task, several researchers have proposed that preservation of voicing cues in phonetic outputs may be driven (at least in part) by orthographic influences (Fourakis & Iverson, 1984; Jassem & Richter, 1989; Kopkalli, 1993; Manaster Ramer, 1996a, 1996b; Piroth & Janker, 2004; Warner et al., 2004, 2006; Beckman et al., 2009). This proposal is corroborated by the fact that incomplete neutralization of voicing-dependent differences has usually been observed in word-reading experiments (e.g., Charles-Luce, 1985; Port & O'Dell, 1985; Piroth et al., 1991; Dmitrieva et al., 2010) but not in the case of the oral tasks in Fourakis & Iverson (1984) and Jassem & Richter (1989).

However, the exact biasing role of orthography in the preservation of the voicing contrast in word-final obstruents has not been clear. Kopkalli (1993), for example, proposed that the absence of orthographic representations during experimental tasks can only reduce the magnitude of voicing-dependent differences but cannot account for the appearance of such effects in the first place. Van Rooy et al. (2003) pointed out that the

effects observed in their study of final plosives in Afrikaans could not have been caused by orthography exclusively, since voicing cues were enhanced only slightly for non-existing words (the pronunciation of which can be expected to be strongly affected by orthography). Piroth & Janker (2004) argued that orthography alone could not explain why, in their study, South German speakers maintained the underlying contrast in plosives but speakers of West German and Central German did not (even though the 3 dialects use the same orthographic representations) and further proposed that orthographic influences may be limited to a subset of lexical items, such as infrequent words. Ernestus & Baayen (2007a) pointed out that orthographic differences cannot explain the interparadigmatic effects found in the perception of non-word rhymes in Dutch. Finally, Röttger et al. (2011) observed significant voicing-dependent differences in German even though their participants performed oral depluralization of non-words.

2.5.2.2. Presence of Minimal Pairs

Preservation of voicing-dependent differences in production has also been attributed to the common practice of including full minimal pairs among the stimuli (Jassem & Richter, 1989; Piroth & Janker, 2004). Jassem & Richter (1989), for example, suggested that the presence of minimal pair counterparts in the lists of stimuli items results in hypercorrection of one's pronunciation. Piroth & Janker (2004) argued that exposing speakers to both members of a minimal pair increases the amount of attention they pay to orthographic representations. No significant differences were also observed in the case of the oral conjugation task in Fourakis & Iverson (1984) and the word-reading

tasks in Inozuka (1991) and Kopkalli (1993), which are some of the studies that did not use any minimal pairs among the stimuli or included very few test words that contrasted minimally (e.g., 2 (out of 60) words in Kopkalli, 1993).

At the same time, voicing-dependent differences have in fact been observed when test items included very few minimal pairs (e.g., one pair (out of 9) in Piroth et al., 1991; one pair (out of 13) in Piroth & Janker, 2004). Effects of underlying voicing have also been found when using nonword forms that do not have existing minimal pair counterparts in the lexicon (e.g., Ernestus & Baayen, 2006). Hence, direct exposure to lexical competition may not be required for the voicing contrast to be preserved in experimentally-elicited speech. Moreover, the absence of significant differences in Fourakis & Iverson (1984) could have been motivated by the non-orthographic nature of the oral conjugation task rather than exclusion of minimal pair counterparts from the list of stimuli items, and the null results in Inozuka (1991) and Kopkalli (1993) could have been the consequence of using small samples of speakers and very few test items.

2.5.2.3. Segmental, Prosodic and Lexical Differences

Many previous experimental investigations reported production and perception data that were averaged across different places and manners of articulation, word lengths, segmental environments and/or prosodic positions (e.g., Fourakis & Iverson, 1984; Inozuka, 1991; Dmitrieva et al., 2010). However, much of the phonetic research done in a variety of neutralizing and non-neutralizing languages has shown that many of the parameters associated with underlying voicing (e.g., VOT, closure duration, vowel

duration) can also be affected by such factors as consonantal place and manner of articulation, vowel height, stress placement, word length, complexity of the phonological environment, and adjacency to a prosodic boundary (among others, Lindblom, 1967; Lehiste, 1970; Lisker, 1974; Gamkrelidze, 1975; Westbury & Keating, 1980; Ohala, 1983; Slis, 1986; Nearey & Rochet, 1994; Maddieson, 1997; Cho & Ladefoged, 1999; Byrd et al., 2005). As such, both the presence and the absence of significant effects in previous neutralization studies may be attributable to the lack of control for segmental and prosodic factors.

Moreover, many earlier findings could have been biased by the practice of using prepositions, adverbs and other stimuli items that do not have any paradigmatic forms in which the critical consonant is prevocalic. Since such words cannot be modified by adding a vowel-initial suffix, the underlying identity of the word-final obstruent is sometimes thought to be ambiguous (Fourakis & Iverson, 1984; Mascaró, 1987; Manaster Ramer, 1996a). In other words, those consonants that can never be found in a non-neutralizing position may be voiced, voiceless or underspecified for voicing at the phonemic level. Although such forms can potentially inform us on the role of orthography in final devoicing, the absence of paradigmatic evidence for underlying voicing in a subset of test items complicates any interpretation of the findings that are based on an uneven number of alternating and non-alternating wordforms (e.g., Port & O'Dell, 1985; Dmitrieva et al., 2010).

Voicing-dependent differences may also be masked by the effects of lexical frequency. Given that more frequent wordforms are known to be articulated faster

(Whalen, 1992), if lexical frequencies of items ending in voiced versus voiceless obstruents are not controlled for in the experimental design, exogenous effects of usage frequency may lead to either appearance or disappearance of durational differences between the two voicing groups and could be erroneously analyzed as influences of underlying voicing. However, very few previous experimental studies matched usage frequencies across experimental conditions or tested for frequency effects in statistical models (notable exceptions include Ernestus & Baayen, 2007a; Dmitrieva et al., 2010). Furthermore, many previous studies used a mix of inflected and uninflected wordforms and items belonging to different grammatical categories (e.g., a mix of Nominative singular and Genitive plural nous, infinitival and perfective verbal forms, adverbs and abbreviations is used in Dmitrieva et al., 2010). Although the exact effects of introducing this additional variability are yet to be explored, different inflectional forms and grammatical categories may be expected to have different likelihoods of occurring in the utterance-final position, which is the only environment where final devoicing is unlikely to be blocked (or masked) by other phonological processes (e.g., voicing assimilation). Hence, the strength of voicing cues may vary across test items depending on whether a particular form is encountered more often in phrase-final versus phrase-medial environments, especially if a given wordform tends to co-occur with another word that, for example, triggers regressive voicing assimilation across the word boundary.

2.5.2.4. Participant Selection, Sample Size and Statistical Treatment

Many previous findings may also be questioned on the basis of subject recruitment, sample size and/or statistical treatment of the data. For example, the Russian findings in Chen (1970) and Pye (1986) were based on recordings from one and 5 subjects, respectively. Piroth & Janker (2004) collected production data from only two speakers per dialect. Dmitrieva et al. (2010) tested only 4 monolingual Russian speakers. Given that the existence of inter-speaker variability is well known in the phonological literature (Burton & Robblee, 1997), those voicing-dependent effects that are found in small datasets are not always representative of larger populations. The practice of making generalizations on the basis of very few speakers (as well as very few experimental items) becomes even more problematic considering that a subset of test tokens is often not analyzed due to the presence of unreleased plosives, creaky voicing, and other similar factors that make the data unusable (e.g., 36% of the tokens in the '_#C' condition in Charles-Luce & Dinnsen, 1987).

Several previous neutralization studies also recruited participants who had significant exposure to (or knowledge of) a non-neutralizing language. For example, the German speakers in Charles-Luce (1985), the Polish speakers in Slowiaczek & Dinnsen (1985) and Slowiaczek & Szymanska (1989) and a subset of the Russian speakers in Pye (1986) were all residing in the USA or the UK at the time of testing (for up to 37 years in the case of Slowiaczek & Szymanska, 1989). Testing participants who are exposed to non-neutralizing languages is not problematic if research goals include investigating inter-language differences and effects of bilingualism (as in the case of Smith et al., 2009;

Dmitrieva et al., 2010). However, the vast majority of previous experimental studies were not interested in L2 issues and generalized their findings to all speakers of a particular language even if predominantly bilingual participants were tested (e.g., Charles-Luce, 1985; Pye, 1986; Slowiaczek & Szymanska, 1989). Notably, the German investigation in Smith et al. (2009) and the Russian study in Dmitrieva et al. (2010) both showed that linguistic experience can bias the way the voicing contrast is produced in not only L2 but also L1 (also, see Simon (2010) for similar Dutch-to-English interference effects). Hence, many of the previously reported voicing-dependent effects may not be characteristic of the differences maintained in monolingual groups or the language in general.

Furthermore, some earlier investigations reported descriptive statistics only (Chen, 1970; Pye, 1986; Kahlene-Halstenbach, 1989; also, the perception data in Kopkalli, 1993), which makes it impossible to evaluate the generalizability of the observed effects. Other previous findings were based exclusively on the results of analyses of variance without repeated measures (e.g., Charles-Luce & Dinnsen, 1987; Slowiaczek & Szymanska, 1989). However, since in devoicing studies the same participant usually produces or evaluates both voiced and voiceless tokens, repeated measures (RMs) are required in order to account for the variability that is due to the fact that a given subject provides responses under different experimental conditions (Vonesh & Chinchilli, 1997).[26] Moreover, the majority of previous devoicing studies limited their analyses to by-subject tests (e.g., Charles-Luce & Dinnsen, 1987; Slowiaczek & Szymanska, 1989; Matsui, 2011). In such tests, whenever a factor comes out as

[26] Examples of the few studies that used repeated measures in their statistical analyses include Charles-Luce (1993), Warner et al. (2004), Dmitrieva et al. (2010), and Matsui (2011).

significant, it can be assumed that the same variable would also be significant if a larger sample of participants from the same population were tested on the same set of stimuli items. Neutralization studies, however, are not interested in finding out how speakers of a given language produce or perceive the underlying voicing contrast in a small set of wordforms (e.g., 6 words in the production experiment in Inozuka, 1991; 4 words in the perception study in Matsui, 2011). Instead, the aim of experimental research is to predict speakers' performance for all similar words that may exist in the lexicon.

One way of ensuring generalizability of results across both subjects *and* items, when item variability has been constrained experimentally, is to conduct by-subject (F_1) as well as by-item (F_2) analyses and to consider an effect significant only if it is significant in both F_1 and F_2 tests (i.e., when it meets the F_1 x F_2 Criterion; Clark, 1973; Raaijmakers et al., 1999).[27] In other words, participants' responses are averaged across items (F_1) and across subjects (F_2). Both sets of data are then tested in separate analyses of variance, with F_1 analyses revealing whether larger samples of subjects would show similar effects when tested on the same items and F_2 analyses showing how the same participants would perform on a new set of similar words. If both F_1 and F_2 tests reveal significant effects of the same predictor variable, the finding can be assumed to generalize to other speakers as well as other lexical items of the same language. However, with the exception Warner et al. (2004), the F_1 x F_2 Criterion has not been

[27] An alternative way to account for variability that is due to both subjects and items is to use mixed effects modeling with 'subjects' and 'items' as random effects factors (e.g., Röttger et al., 2011). Although mixed effects modeling may have advantages over traditional analyses of variance (Locker et al., 2007; Baayen, 2008), conducting by-subject and/or by-item ANOVAs has been the most common statistical procedure in previous neutralization studies and, to ensure greater comparability of the present findings with those of earlier investigations, the current study also relies on F_1 and F_2 ANOVAs and leaves mixed effects modeling of the Russian data for future work.

applied in previous neutralization studies even when both by-subject and by-item analyses were conducted (e.g., Port & Crawford, 1989; van Rooy et al., 2003).

2.5.2.5. Multiple Repetitions and Other Procedural Effects

Researchers working on the voicing contrast in final obstruents have usually attempted to mask the aim of their investigation by adding filler items and creating an impression of a different goal (e.g., 'gender-related differences in speech' in Kahlene-Halstenbach, 1989; 'a semantic task' in Dmitrieva et al., 2010). However, in the majority of previous neutralization studies, participants were asked to repeat the same experimental list several times and/or were presented with test items prior to the recording session so that they could familiarize themselves with the experimental tokens (e.g., Charles-Luce, 1985; Charles Luce, 1993; Warner et al., 2004; Dmitrieva et al., 2010). Asking participants to produce the same stimuli items more than once as well as exposing them to test items before the experimental session can be expected to increase the likelihood of the results of the study being affected by speaker's awareness of the goal of the experiment and the resultant hyper-articulation of the phonological contrast under investigation. Pye (1986), for example, noted that 3 (out of 5) speakers included in her study seemed to have discovered the goal of the investigation and produced exaggerated voicing cues in final obstruents.

For perception findings, part of the observed variability in identification data may also be attributable to general differences in how responses are registered (Burton & Robblee, 1997; also, see Côté & Kharlamov (2011) for task-dependent asymmetries in

syllabification judgments). More specifically, experimental studies ask participants to perform slightly different actions (e.g., circling an answer on a sheet of paper versus pressing a computer key) and, depending on the exact nature of the task, participants' responses may be expected to be more or less sensitive to the same factors. For example, since computer-registered answers are faster than written responses provided on a sheet of paper, answers recorded electronically may be more affected by acoustic-phonetic properties of stimuli items and may also be less sensitive to such factors as graphemic frequency than the answers that are written down on a sheet of paper.

2.6. General Summary

Contrary to the claims made in the traditional accounts of final devoicing, experimental research has often shown that speakers of neutralizing languages produce and perceive cues to the underlying voicing specification of word-final obstruents. Such findings have been attributed to a variety of grammatical factors and have been argued to warrant re-consideration of traditional analyses of final devoicing and re-evaluation of the underlying voicing properties of final consonants. However, other experimental studies done in many of the same languages have failed to observe statistically significant differences in the production and perception of voiced versus voiceless obstruents in neutralizing environments and have attributed the presence of incompletely neutralized cues to the effects of task-dependent factors and various methodological issues. Notably, many of the claims with respect to neutralization of voicing cues have been based on limited, impressionistic or otherwise problematic data and none of the previous studies

have attempted to specifically manipulate the presence of not only orthographic representations but also full minimal pairs in order to determine the exact biasing role of task-dependent influences. This lack of empirical evidence is what motivated the current dissertation, which specifically investigates the effects of grammatical versus methodological factors in the production (Chapter 3) and perception (Chapter 4) of voicing cues in word-final obstruents in Russian and discusses the implications of the experimental findings for the phenomenon of final devoicing and phonology in general (Chapter 5).

CHAPTER 3: PRODUCTION STUDY

3.1. Introduction

Chapter 3 describes the acoustic production study of the voicing contrast in word-final obstruents in Russian that was conducted as part of the present doctoral research. The chapter starts with stating the goals, research questions and hypotheses of the production experiment (Section 3.2.), followed by a detailed description of the experimental methodology (Section 3.3.) and results of the study (Section 3.4.).

3.2. Goals, Research Questions and Hypotheses

3.2.1. Aims of the Study

The production component of the current dissertation represents a large-scale and systematic study dedicated to exploring the effects of (i) grammatical (phonological, lexical) influences versus (ii) methodological factors in the realization of the voicing contrast in word-final obstruents in Russian. The main goal of the production study is to investigate the sources of the incomplete neutralization effects that are seen in laboratory speech. The scope of the investigation is limited to word-final/utterance-final stops and fricatives that are not adjacent to any other obstruent (i.e., all test items end in a single obstruent and are produced as isolated one-word utterances that are followed by a pause; cases of word-medial (coda) devoicing or cases involving obstruent+obstruent clusters are not examined).

3.2.2. Research Questions

The following research questions are addressed in the production study:

R1: *Do speakers of Russian, as a group, maintain statistically significant differences in preceding vowel duration, closure/frication duration, release duration and/or closure voicing in word-final obstruents during experimentally-elicited speech?*

R2: *If attested, can voicing-dependent differences be attributed primarily to speakers' exposure to orthographic inputs and/or presence of full minimal pairs among the stimuli?*

R3: *Are voicing-dependent differences found irrespective of phonological and lexical properties of stimuli words (consonantal place and manner of articulation, word length in syllables, membership in a minimal pair)?*

R4: *Do the magnitude and direction of different voicing-dependent effects vary across individual speakers and items or do all acoustic parameters show the same degree of individual-level variability?*

3.2.3. Hypotheses

In general, it is hypothesized that statistically significant voicing-dependent durational differences will be found at the group level, with effects of underlying voicing expected to be more robust when orthography or lexical competition encourage preservation of the voicing contrast. Task-independent effects are also anticipated, such

as stronger cues to underlying voicing for items with potentially homophonous lexical competitors, yet they are expected to be less robust than the effects of task-dependent factors. At the individual level, substantial between-speaker and between-item variability is predicted for those acoustic parameters that are non-primary cues to underlying voicing. If the current predictions are met, the findings will show that grammatical influences are present in experimentally-elicited devoicing data, yet experimental methodology exerts a prominent bias on speakers' performance in linguistic tasks, with task-dependent factors accounting for a substantial part of the observed results.

3.3. Method

3.3.1. Participants and Location

Native speakers of Russian (n=78; 23 male, 55 female; 18 to 28 years old; mean age of 20.1) were recruited and tested at a university campus in Perm, Russia. All participants were undergraduate, graduate or postdoctoral members of the university community. None majored in linguistics, modern languages or any other field which involves explicit training in phonetics or phonology. All participants spoke the same local variety of Russian. None self-reported any language-related problems. The majority of speakers (n=67) identified themselves as monolingual. The remaining participants (n=11) were raised in bilingual households (Russian and Komi, Tatar or Udmurt); however, at the time of testing, they used only the Russian language actively in their daily lives. All participants indicated having low to average proficiency in at least one foreign language (English, French, German, Spanish).

3.3.2. Stimuli

3.3.2.1. The Consonantal System of Russian

The phonemic system of Russian includes the following types of consonants: plosives (bilabial, alveolar, velar), fricatives (labiodental, alveolar, post-alveolar, velar), affricates (alveolar, alveopalatal), nasals (bilabial, alveolar), trills (alveolar), and approximants (alveolar, palatal). With the exception of the velar fricative and the two affricates, all obstruents can be either voiced or voiceless. The majority of segments also have both plain and palatalized versions that contrast phonemically. A sample consonantal chart of Russian is presented in Figure 1 (based on Bogdanova, 2001; the obstruents tested in the present study are listed inside shaded areas).

	Bilabial	**Labiodent.**	**Alveolar**	**Post-alv. & palatal**	**Velar**
Plosive	p b		t d		k g
	p^j b^j		t^j d^j		k^j g^j
Fricative		f v	s z	ʃ ʒ	x
		f^j v^j	s^j z^j	$ɕ^j$	x^j
Affricate			t͡s	t͡ʃ	
Nasal	m		n		
	m^j		n^j		
Trill			r		
			r^j		
Approx.			l		
			l^j	j	

Figure 1: Consonantal phonemes of Russian

As can be seen in Figure 1, the segments of interest to the current investigation are limited to (non-palatalized) labial, coronal and dorsal plosives and fricatives that are paired in Russian with respect to the voicing feature.

3.3.2.2. Stimulus Selection Criteria

Production stimuli (n=150; 75 ending in voiced obstruents, 75 in voiceless; 90 with final plosives, 60 with final fricatives; distributed evenly across places of articulation) were singular masculine nouns in the Nominative case. All items were existing words of Russian, had ultimate stress, and could be inflected by adding a vowel-initial suffix. The stimuli belonged to one of the 5 word types shown in (4) below.

(4) *Production stimuli*

(i) plosive-final monosyllabic minimal pairs
(n=30; e.g., /kɔt/ 'cat', /kɔd/ 'code')

(ii) plosive-final monosyllabic non-minimal pairs
(n=30; e.g., /zlak/ 'grass', /flag/ 'flag')

(iii) plosive-final disyllabic non-minimal pairs
(n=30; e.g., /pirat/ 'pirate', /parad/ 'parade')

(iv) fricative-final monosyllabic non-minimal pairs
(n=30; e.g., /trus/ 'coward', /gruz/ 'load')

(v) fricative-final disyllabic non-minimal pairs
(n=30; e.g., /tarif/ 'tariff', /zaliv/ 'bay')

Disyllabic plosive-final and fricative-final minimal pairs were not included due to the low number of relevant lexical items in Russian that belong to the same part of speech, represent the same inflectional category, and are not archaic.

To mitigate the potential effects of exogenous variability, monosyllabic and disyllabic non-minimal pairs consisted predominantly of near-minimal pairs and words with matching final onsets and/or nuclei. Test items were also controlled for length in graphemes (M_{voiced}=4.4, SD=1.1; $M_{voiceless}$=4.3, SD=1.0; $|t|(74)<1$; $p>0.1$) and lexical frequency (M_{voiced}=27.4 ipm, SD=90.6; $M_{voiceless}$=9.95 ipm, SD=18.3; $|t|(74)<2$; $p>0.1$).[28] In the absence of detailed information on the positional frequency of Russian words, it was not possible to also control for the effects of positional frequency. However, since the stimuli represented the same grammatical category and the same inflectional form, any residual effects of positional frequency were not expected to affect the overall results.

Given the nature of the production study, stimuli words were chosen such that each item could be represented as a short fill-in-the-blank description (for hard-to-image objects and concepts; e.g., /kɔd/ 'code') or a vector image (for easy-to-image objects; e.g., /arbuz/ 'watermelon'). Fill-in-the-blank descriptions (n=100) were created on the basis of dictionary definitions in Ozhegov & Shwedova (2000). Vector images (n=50) came from several standardized sets, including the database of 260 colored pictures from

[28] Frequency values were calculated on the basis of the Russian National Corpus (Lyashevskaya & Sharoff, 2009; <http://ruscorpora.ru>). The sub-corpus used for the purposes of the present study included over 113 million individual word tokens and was based on approximately 181,000 newspaper articles published between 2000 and 2010. Notably, although an overall difference in mean values of instances per million (ipm) was observed between words ending in voiced versus voiceless obstruents, the difference was largely due to a small number of frequent items and was not statistically significant.

Rossion & Pourtois (2004), 400 black-and-white images from Cycowicz et al. (1997), and 590 black-and-white pictures from Severens et al. (2005).[29] Following the suggestion in Rossion & Pourtois (2004) that addition of color facilitates recognition of depicted objects, the present study used only colored pictures.[30]

To ensure that vector images and fill-in-the-blank descriptions would result in a high rate of expected responses, 168 undergraduate and graduate students (from the same university; none participated in the production study) filled out anonymous written questionnaires. The questionnaires asked to provide semantic associations to a given word, to denote names of objects appearing in vector images, and/or to complete fill-in-the-blank descriptions of various objects and concepts. Descriptions and the selection of vector images were repeatedly modified on the basis of the answers provided in the questionnaires until a set of 150 tokens was formed, with over 90% of the respondents producing the expected wordform for each non-orthographic representation.

In addition to test items, participants were exposed to an equal number of filler words (n=150) that were taken from a pool of Nominative singular masculine and feminine nouns ending in vowels and sonorant consonants. Similarly to test items, both orthographic and non-orthographic representations were used for filler tokens. For those participants who were exposed only to one member of each minimal pair, a supplementary set of monosyllabic filler words (n=15) was added to the stimulus list to ensure that all participants were exposed to the same overall number of experimental

[29] Due to licensing considerations, the full set of images used in the present study cannot be reproduced in the Appendix and should be accessed via the original sources cited herewith.

[30] When colored versions were unavailable, color was added manually using the Paint.NET image manipulation application (<www.getpaint.net>).

items (i.e., 300). The additional words did not have minimal pair counterparts in the lexicon and ended in voiced or voiceless plosives. They were distributed evenly across places of articulation and shared the final onset and/or nucleus with the one member of each minimal pair which was retained in the stimulus list (e.g., if the word /krap/ 'specks' was not included among test items, /krab/ 'crab' was paired up with /trap/ 'ramp'; */trab/ is not an existing word of Russian).

Table 3 provides a summary of the stimuli items used in the production study.[31]

[31] The full list of test items is given in Appendix 1. Fill-in-the-blank descriptions and a list of vector images are provided in Appendix 2 and Appendix 3, respectively.

Table 3: Stimuli items (production) - Summary

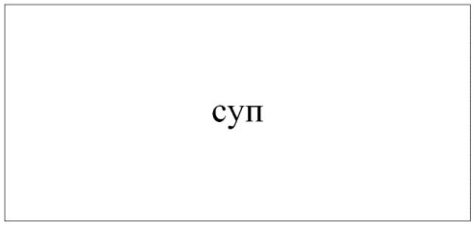
Obstruent type	Place	Syllabicity	Minimal Pair	Voicing	Example
Final plosives (n=90)	Labial (n=30)	Monosyllabic (n=20)	yes (n=10)	Voi (n=5)	/snɔb/
				Vls (n=5)	/snɔp/
			no (n=10)	Voi (n=5)	/zub/
				Vls (n=5)	/sup/
		Disyllabic (n=10)	no (n=10)	Voi (n=5)	/saskɔb/
				Vls (n=5)	/padkɔp/
	Coronal (n=30)	Monosyllabic (n=20)	yes (n=10)	Voi (n=5)	/kɔd/
				Vls (n=5)	/kɔt/
			no (n=10)	Voi (n=5)	/spad/
				Vls (n=5)	/skat/
		Disyllabic (n=10)	no (n=10)	Voi (n=5)	/parad/
				Vls (n=5)	/pirat/
	Velar (n=30)	Monosyllabic (n=20)	yes (n=10)	Voi (n=5)	/lug/
				Vls (n=5)	/luk/
			no (n=10)	Voi (n=5)	/flag/
				Vls (n=5)	/zlak/
		Disyllabic (n=10)	no (n=10)	Voi (n=5)	/avrag/
				Vls (n=5)	/barak/
Final fricatives (n=60)	Labial (n=20)	Monosyllabic (n=10)	no (n=10)	Voi (n=5)	/xlɛv/
				Vls (n=5)	/blɛf/
		Disyllabic (n=10)	no (n=10)	Voi (n=5)	/zaliv/
				Vls (n=5)	/tarif/
	Coronal (n=20)	Monosyllabic (n=10)	no (n=10)	Voi (n=5)	/gruz/
				Vls (n=5)	/trus/
		Disyllabic (n=10)	no (n=10)	Voi (n=5)	/arbuz/
				Vls (n=5)	/indus/
	Palatal (n=20)	Monosyllabic (n=10)	no (n=10)	Voi (n=5)	/muʒ/
				Vls (n=5)	/duʃ/
		Disyllabic (n=10)	no (n=10)	Voi (n=5)	/mulʲaʒ/
				Vls (n=5)	/gulʲaʃ/

3.3.3. Experimental Groups and Procedures

Experimental sessions were conducted in a sound-treated university lab in Perm, Russia. Participation in the production study consisted of attending a single session which lasted up to one hour. Prior to testing, all participants provided informed consent and filled out a background questionnaire (information about age, gender, linguistic experience, places of residence, and speech-related physiological difficulties). During the recording session, speakers were instructed to name the experimental items that were appearing on the computer screen. Participants' responses (acoustic outputs) were recorded for analyses. At the end of the session, all speakers completed a post-test questionnaire which asked about the perceived goal of the study.

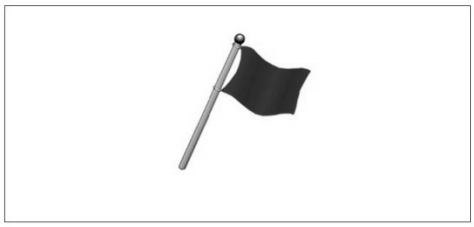
Participants were assigned on a random basis to one of 4 experimental groups. The groups differed with respect to two parameters: (i) presence of orthographic representations of test items and (ii) presence of full minimal pairs among the stimuli. The first parameter was manipulated by asking participants to perform either a word-reading task (henceforth, Orth+) or a picture-naming/fill-in-the-blank task (henceforth, Orth-). During the word-reading task, speakers saw orthographically-represented stimuli items that stayed on the computer screen for 1 second and were instructed to pronounce the words as soon as they appeared. The picture-naming/fill-in-the-blank description task involved seeing vector images or short semantically-based descriptions of test words, and speakers were asked to name the non-orthographically represented items as soon as they recognized them. Vector images were displayed for 1 second. Fill-in-the-blank descriptions stayed on the screen until a computer key was pressed (with a 4 second time-

out).[32] Sample screenshots demonstrating the different experimental tasks are provided in Figure 2 (word-reading task), Figure 3 (picture-naming task), and Figure 4 (fill-in-the-blank description task).



(/sup/ 'soup')

Figure 2: Word-reading task - Sample screenshot



(/flag/ 'flag')

Figure 3: Picture-naming task - Sample screenshot

[32] For fill-in-the-blank descriptions, participants were instructed to pronounce only the missing word, which was sentence-final in 91 (out of 100) carrier sentences and sentence-medial in the remaining 9 cases. Inclusion of carrier phrases with non-final target items helped achieve a high rate of expected responses, and reanalyses of the production data showed that exclusion of the 9 test items did not affect the overall results.

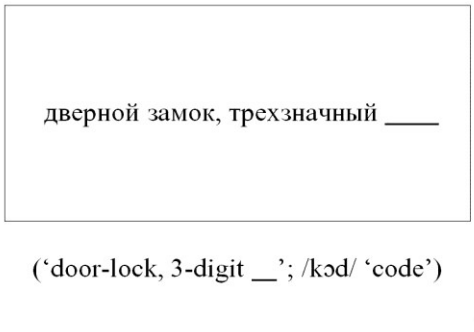


Figure 4: Fill-in-the-blank description task - Sample screenshot

The second parameter, presence of full minimal pairs among the stimuli, was manipulated by exposing participants from Orth+ and Orth- groups to either (i) both members of each minimal pair (henceforth, MP+) or (ii) only one member of each minimal pair (henceforth, MP-). For example, participants in the MP+ groups encountered both /kɔt/ 'cat' and /kɔd/ 'code' during the recording session. Speakers in the MP- groups were exposed to /kɔt/ 'cat' or /kɔd/ 'code' only. A general summary of the 4 experimental groups is given in Table 4.

Table 4: Experimental groups (production) - Summary

Experimental group	Task	Minimal pair items
1. **ORTH- MP-**	picture-naming/ fill-in-the-blank	one item (e.g., /kɔt/)
2. **ORTH- MP+**	picture-naming/ fill-in-the-blank	full pair (e.g., /kɔt/~/kɔd/)
3. **ORTH+ MP-**	word-reading	one item (e.g., /kɔd/)
4. **ORTH+ MP+**	word-reading	full pair (e.g., /kɔt/~/kɔd/)

Every experimental session started with trial items (n=10), followed by 6 experimental blocks with 50 stimuli items per block. Experimental tokens were pseudo-randomized, such that the first 5 items in each block were fillers and the rest of the items occurred in a random order (randomization never resulted in members of the same minimal pair being immediately adjacent to each other). Participants were encouraged to take a short break after each block.

3.3.4. Instruments

Stimuli items were presented in the center of a portable PC computer screen. Randomization and stimulus delivery were done with Presentation® software (Version 14, www.neurobs.com). A unidirectional Shure SM10A microphone connected to a portable Marantz recorder (PMD670) was used to capture speakers' acoustic outputs. Recording (mono) was done continuously with a sampling rate of 44.1 kHz and a depth of 16 bit.

3.3.5. Data Treatment and Analyses

3.3.5.1. Segmentation and Acoustic Measurements

Acoustic tokens were extracted from continuous recordings and were labeled using the PRAAT software (Boersma & Weenink, 2009). An annotated TextGrid was created for each token. For items ending in plosives, TextGrids included boundary information for the entire wordform, nucleus of the final vowel, consonantal closure, and consonantal release (including both burst and aspiration, if applicable). For words ending in fricatives, TextGrids contained information about word boundaries, edges of the final

nucleus, and duration of the frication period. Word-initial and word-final edges were marked automatically using a pause-detection script. Word-internal boundaries were inserted manually on the basis of spectrograms and oscillograms. All boundaries were re-verified visually before calculating measurement values.

Location of the nucleus-initial boundary depended on the type of the preceding segment. When preceded by a plosive, the boundary was placed at the end of the release noise of the stop and the beginning of F2 and F3 associated with the vowel. For nuclei preceded by fricatives or affricates, the initial boundary was located after the offset of frication noise. For vowels preceded by sonorant consonants and glides, the boundary was placed at the point of abrupt changes in the waveform and the spectral pattern. The end of the nucleus was marked at the point of interruption in F2 and F3. The closure period of plosives was marked from the end of the preceding nucleus to the beginning of the release burst of the stop, which was defined as the point of significant increase in waveform amplitude and concurrent appearance of a noise pattern in the spectrogram. The release stage of plosives was marked from the end of the closure period to the end of burst or, if a segment was aspirated, the end of aspiration. For fricatives, frication duration was marked from the end of the preceding nucleus to the end of the high-frequency noise associated with the consonant. Sample waveforms and spectrograms showing the placement of segmental boundaries are provided in Figure 5 (final plosives) and Figure 6 (final fricatives).

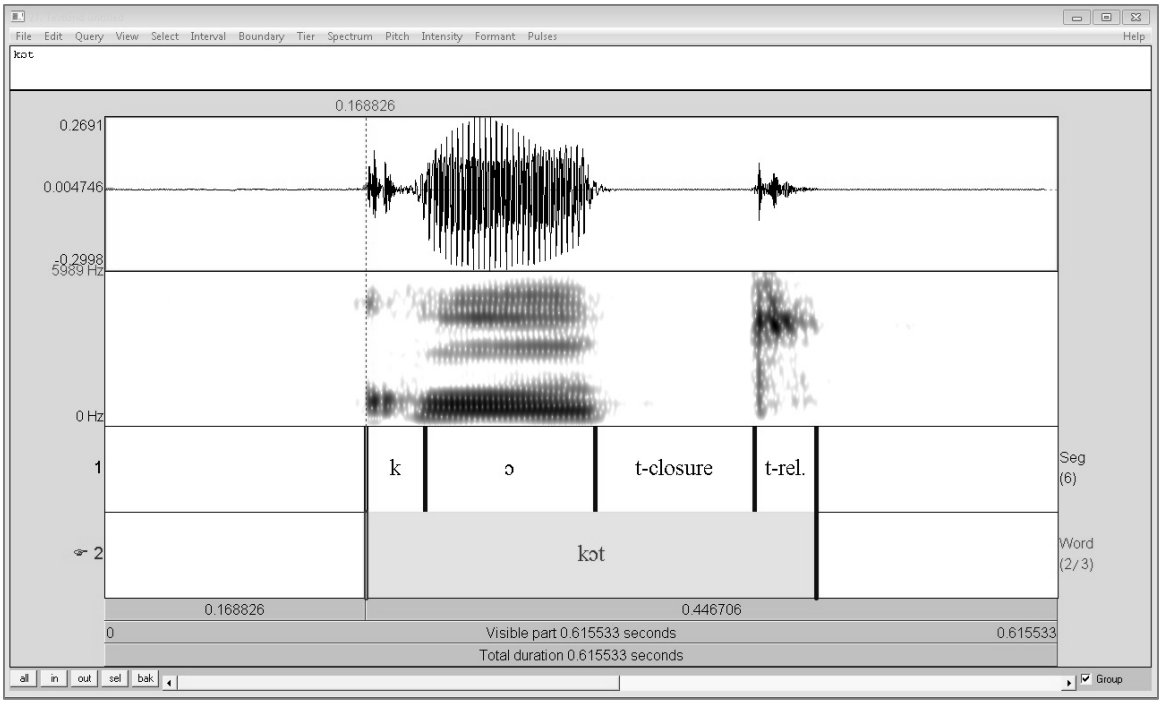


Figure 5: Segmentation of plosive-final tokens (/kɔt/ 'cat')

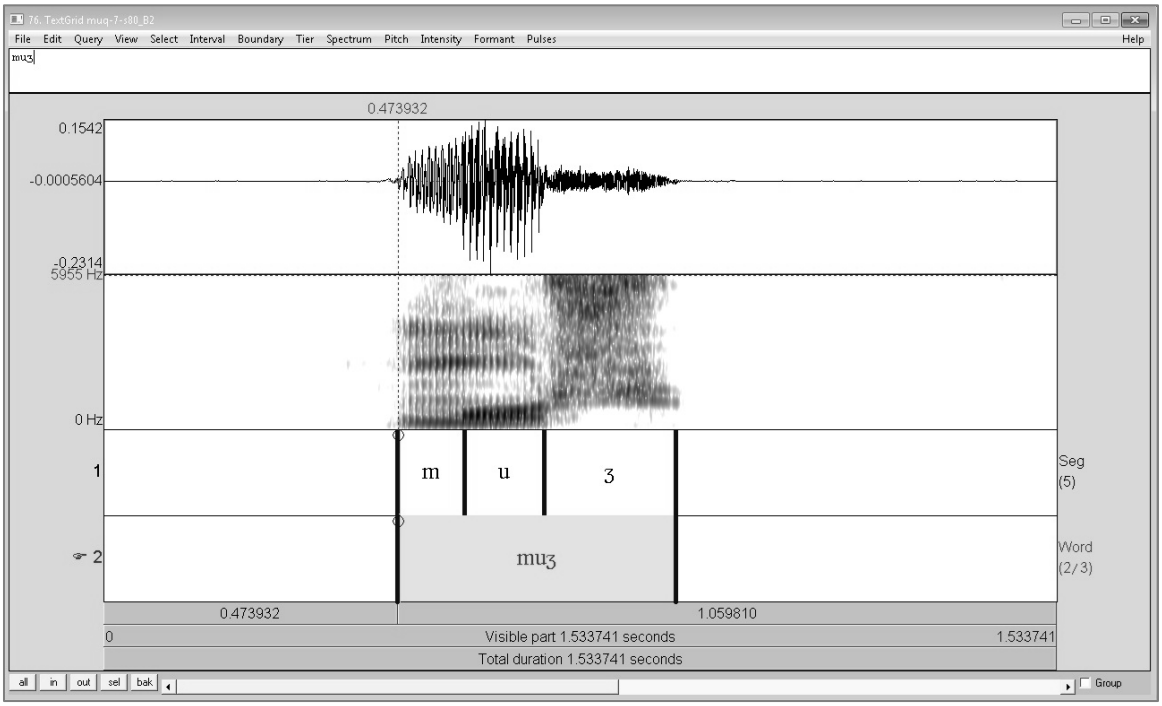


Figure 6: Segmentation of fricative-final tokens (/muʒ/ 'husband')

TextGrids and corresponding sound files were used to calculate 4 acoustic parameters for final plosives (listed in (5) below) and 3 measurements for final fricatives (listed in (6) below).

(5) *Acoustic measurements (plosives)*

(i) preceding vowel duration;

(ii) closure duration;

(iii) release duration;

(iv) closure voicing.

(6) *Acoustic measurements (fricatives)*

(i) preceding vowel duration;

(ii) frication duration;

(iii) closure voicing (=voicing into frication).

Preceding vowel duration, closure/frication duration and release duration were measured in milliseconds (rounded to the nearest full millisecond). Closure voicing was measured as a number of glottal pulses. As noted in Chapter 2, these acoustic parameters have been used in the majority of previous neutralization studies that examined the voicing contrast in word-final obstruents.

3.3.5.2. Statistical Analyses

Acoustic measurements of preceding vowel duration, closure/frication duration, release duration and closure voicing were subjected to a series of analyses of variance

with repeated measures (RM ANOVAs). The 3 factors listed in (7) were tested as predictor variables (factor levels are specified inside brackets).

(7) *Predictor variables*

(i) UR-voice (=underlying voicing; voiced, voiceless);

(ii) C-place (=place of articulation; labial, coronal, dorsal);

(iii) Group (=experimental group; Orth-MP-, Orth-MP+, Orth+MP-, Orth+MP+).

The IBM SPSS Statistics software package (Version 19; www.ibm.com) was used to run the tests. Separate analyses were performed for each dependent variable and each word type. The 2 (UR-voice) x 3 (C-place) x 4 (Group) RM ANOVAs were run on pooled data using both by-subject (F_1) and by-item (F_2) factorial designs. Only those differences that met the F_1 x F_2 Criterion (i.e., were significant at the α=.05 level in both F_1 and F_2 analyses) were judged to be statistically significant. In F_1 analyses, UR-voice and C-place were used as within-subjects independent variables. Group was a between-subjects predictor. In F_2 analyses, Group was a within-items variable. UR-voice and C-place were between-items predictors. Analyses of variance examined all main effects and interactions. When applicable, pairwise comparisons (t-tests, 2-tailed) were used to resolve significant effects and interactions and Greenhouse-Geisser or Bonferroni corrections were utilized to account for the presence of multiple comparisons (with corrected values reported alongside the original degrees of freedom).

The relationship between acoustic measurements and the results of the post-test questionnaire was examined using regression analyses that were conducted on the by-

subject difference scores (subtracting values for voiceless obstruents from values for voiced obstruents; for MP- groups, difference scores were calculated by averaging the data for each voicing category across the 3 places of articulation). Separate analyses were performed for each of the 5 stimuli types. Subjects' awareness of the contrast under investigation (aware, unaware) was used as the predictor variable.

A total of 10,810 acoustic tokens were retained for analyses, which represents 92.4% of the maximum possible number of tokens (n=11,700; 150 words X 78 speakers). The difference of 7.6% (i.e., 890 tokens) was due to speakers not producing the expected wordform in non-reading tasks (n=805) or the reading task (n=12), background noise (n=15), pause-insertion (n=24), absence of release (n=10), and other miscellaneous problems that made production data unusable (n=24). Table 5 provides a summary of the distribution of tokens with respect to word type and consonantal place and manner of articulation.

Table 5: Acoustic tokens retained for analyses - Summary

Word type	Final Cs	Num. of tokens
monosyllabic MPs (e.g., /kɔt/~/kɔd/)	/p, b/	575
	/t, d/	573
	/k, g/	570
monosyllabic & disyllabic non-MPs (e.g., /sup/, /arbuz/)	/p, b/	1,512
	/t, d/	1,520
	/k, g/	1,516
	/f, v/	1,504
	/s, z/	1,526
	/ʃ, ʒ/	1,514
	Total:	10,810

3.4. Results

Production results are presented below in the following order: (i) plosive-final monosyllabic minimal pairs (Section 3.4.1.), (ii) plosive-final monosyllabic and disyllabic non-minimal pair items (Section 3.4.2.), and (iii) fricative-final monosyllabic and disyllabic non-minimal pairs (Section 3.4.3.). For each stimulus type, group results are followed by individual-level findings and a brief summary of the observed effects. Acoustic findings are followed by the results of the post-test questionnaire (Section 3.4.4.) and a general summary of the effects found in production (Section 3.4.5.). Discussion of the results is offered in Chapter 5.

3.4.1. Plosive-Final Minimal Pairs

3.4.1.1. Group Results

Table 6 shows group averages and 95% confidence intervals for the parameters of preceding vowel duration, closure duration, release duration and closure voicing in plosive-final monosyllabic minimal pairs (averaged across all 4 experimental groups). As reflected in the table, differences in closure duration, release duration and closure voicing but not preceding vowel duration are statistically significant.

Table 6: Preceding V duration, closure duration, release duration & closure voicing in plosive-final monosyllabic MPs - Group means & 95% confidence intervals.

Measure	Voiced	Voiceless	diff.
Vowel duration (ms)	130 (125-134)	126 (122-130)	4
Closure duration (ms)	99 (95-103)	106 (102-110)	-7**
Release duration (ms)	65 (62-69)	71 (68-75)	-6**
Closure voicing (cycles)	4.19 (3.63-4.75)	2.07 (1.80-2.33)	2.12***

* - significant @ .05 level
** - significant @ .01 level
*** - significant @ .001 level

Significant ANOVA results are presented in Table 7, which shows main effects and interactions involving UR-voice.[33]

[33] Detailed results are provided in Appendix 4.

Table 7: Production of plosive-final monosyllabic MPs - Summary of the significant ANOVA results

Factor	Measure	By-Subject		By-Item	
		F_1	p	F_2	p
UR-voice	Vowel duration	18.53	.000	-	-
	Closure duration	40.07	.000	11.59	.002
	Release duration	33.21	.000	9.75	.005
	Closure voicing	73.94	.000	207.03	.000
C-place	Vowel duration	64.57	.000	-	-
	Closure duration	22.18	.000	7.98	.002
	Release duration	26.95	.000	17.56	.000
	Closure voicing	11.26	.000	8.68	.001
Group	Vowel duration	-	-	8.78	.000
	Closure duration	-	-	-	-
	Release duration	-	-	-	-
	Closure voicing	5.00	.003	52.86	.000
UR-voice x Group	Vowel duration	-	-	-	-
	Closure duration	-	-	-	-
	Release duration	-	-	-	-
	Closure voicing	9.43	.000	44.13	.000
UR-voice x C-place	Vowel duration	-	-	-	-
	Closure duration	-	-	-	-
	Release duration	5.34	.006	-	-
	Closure voicing	6.22	.003	-	-

As depicted in Table 7, a significant main effect of UR-voice was observed for the parameters of closure duration ($F_1(1,74)=40.07$, $p<.001$; $F_2(1,24)=11.59$, $p=.002$), release duration ($F_1(1,74)=33.21$, $p<.001$; $F_2(1,24)=9.75$, $p=.005$), and closure voicing ($F_1(1,74)=73.94$, $p<.001$; $F_2(1,24)=207.03$, $p<.001$). For preceding vowel duration, a main effect of UR-voice did not meet the criterion of statistical significance as it was

observed in the by-subject analysis only (F_1>18, p<.001; F_2<1, p>.1). For closure duration and release duration, neither a main effect of Group nor a significant UR-voice x Group interaction was found in any analysis (all Fs<2.6, all ps>.05), which shows that that the effect of underlying voicing was independent of experimental group membership. For preceding vowel duration, a significant main effect of Group was present in the by-item analysis only (F_1<1, p>.1; F_2>8, p<.001) and Group did not interact with UR-voice (all Fs<1, all ps>.1),

For closure voicing, analyses of variance indicated a significant main effect of Group (F_1(1,74)=5.0, p=.003; F_2(1,24)=52.86, p<.001). Pairwise comparisons revealed that Orth-MP- and Orth-MP+ groups were both significantly different from the Orth+MP+ group (all ps<.05). A significant UR-voice x Group interaction was also found (F_1(1,74)=9.43, p<.001; F_2(1,24)=44.13, p<.001), and pairwise tests showed that differences in closure voicing were significant in the Orth-MP+, Orth+MP- and Orth+MP+ groups (all ps<.05) but not in the Orth-MP- group (p>.1). The two Orth+ groups also produced significantly more voicing for phonologically voiced stops than the two Orth- groups (all ps<.001). Hence, Orth-MP- was the only experimental group that did not maintain a significant difference in closure voicing between underlyingly voiced versus voiceless final stops, and more phonetic voicing was observed for underlyingly voiced plosives during word-reading. Mean values of closure voicing across the 4 experimental groups are shown in Figure 7, with significance of pairwise comparisons marked as '*' (p<.05), '**' (p<.01), '***' (p<.001), and 'n.s.' (p>.1).

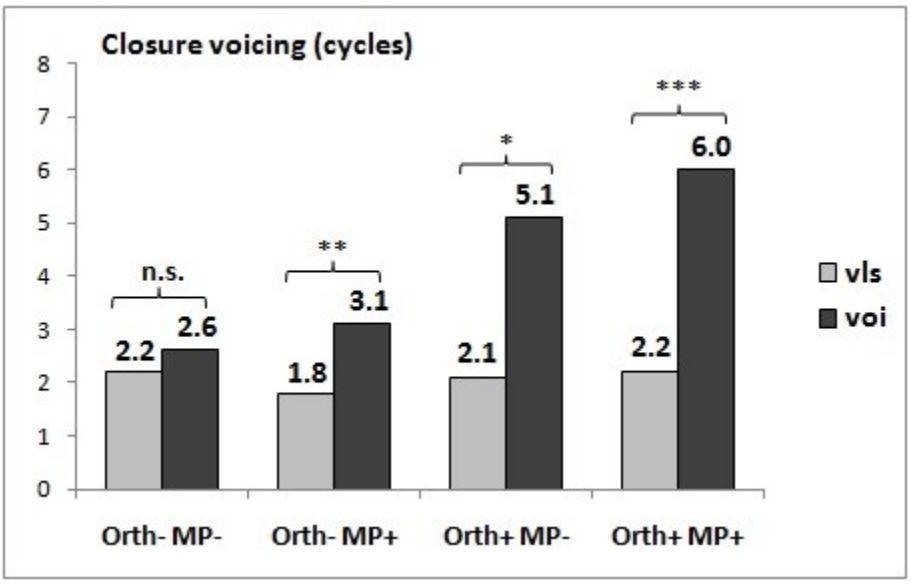


Figure 7: Closure voicing in plosive-final monosyllabic MPs - By-group differences.

A significant main effect of C-place was also found in both F_1 and F_2 analyses for closure duration, release duration, and closure voicing (as well as in the by-subject analysis for preceding vowel duration) (all *F*s>7.0, all *p*s<.01). However, the UR-voice x C-place interaction was significant for release duration and closure voicing only in the by-subject analyses (all *F*s>5, all *p*s<.01) and not significant in the rest of the tests (all *F*s<2, all *p*s >.1).

3.4.1.2. Individual Results

Individual results for plosive-final monosyllabic minimal pairs are provided in a text format in Table 8 and in a graph format in Figure 8 through Figure 15. The first line in Table 8 presents the by-subject results for the Orth-MP- group and indicates that (i) longer vowels before voiced plosives were found in 14 participants (67% of subjects;

1~16 ms diff.), (ii) the pattern was reversed in 5 speakers (24%; 4~11 ms diff.), and (iii) the observed differences were less than 0.5 ms in the remaining 2 participants (9%). The second line shows the by-item results for the same Orth-MP- group and reveals that (i) vowels were longer before voiced stops in 8 item pairs (53%; 1~19 ms diff.), (ii) the pattern was reversed in 4 word pairs (27%; 4~9 ms diff.), and (iii) differences were around zero in the remaining 3 pairs (20%). All other lines follow the same general format.

Table 8: Production of plosive-final monosyllabic MPs - Summary of individual results.

Measure	Group		Expected pattern		Opposite pattern		No diff.
			num. (%)	range	num. (%)	range	num. (%)
Vowel duration (ms)	Orth-MP-	Subj.	14 (67%)	1~16	5 (24%)	1~8	2 (9%)
		Item	8 (53%)	1~19	4 (27%)	4~9	3 (20%)
	Orth-MP+	Subj.	17 (81%)	1~19	4 (19%)	2~12	-
		Item	11 (73%)	4~8	3 (20%)	4~13	1 (7%)
	Orth+MP-	Subj.	12 (67%)	1~19	5 (28%)	1~8	1 (5%)
		Item	11 (73%)	1~16	2 (13%)	3~8	2 (14%)
	Orth+MP+	Subj.	13 (72%)	1~10	5 (28%)	2~4	-
		Item	12 (80%)	1~9	2 (13%)	3~8	1 (7%)
Closure duration (ms)	Orth-MP-	Subj.	16 (76%)	2~27	4 (19%)	1~14	1 (5%)
		Item	11 (73%)	1~20	2 (13%)	1~8	2 (14%)
	Orth-MP+	Subj.	12 (57%)	1~22	9 (43%)	2~6	-
		Item	9 (60%)	3~14	5 (33%)	1~9	1 (7%)
	Orth+MP-	Subj.	11 (61%)	1~27	7 (39%)	1~14	-
		Item	12 (80%)	3~19	3 (20%)	1~5	-
	Orth+MP+	Subj.	14 (78%)	3~26	3 (17%)	3~5	1 (5%)
		Item	14 (93%)	1~21	1 (7%)	3	-
Release duration (ms)	Orth-MP-	Subj.	15 (71%)	1~19	6 (29%)	4~17	-
		Item	9 (60%)	2~21	6 (40%)	2~13	-
	Orth-MP+	Subj.	13 (62%)	1~16	8 (38%)	1~10	-
		Item	8 (53%)	2~18	6 (40%)	1~5	1 (7%)
	Orth+MP-	Subj.	15 (83%)	1~20	3 (17%)	3~6	-
		Item	10 (67%)	4~28	4 (27%)	3~11	1 (7%)
	Orth+MP+	Subj.	15 (83%)	3~23	2 (11%)	1~6	1 (5%)
		Item	14 (93%)	1~20	-	-	1 (7%)
Closure voicing (cycles)	Orth-MP-	Subj.	9 (43%)	0.8~2.9	2 (9%)	0.9~2.0	10 (48%)
		Item	6 (40%)	0.6~1.9	2 (13%)	0.6~0.7	7 (47%)
	Orth-MP+	Subj.	14 (67%)	0.6~4.9	1 (5%)	0.6	6 (29%)
		Item	13 (87%)	0.8~2.9	-	-	2 (13%)
	Orth+MP-	Subj.	15 (83%)	0.9~10.5	-	-	3 (17%)
		Item	15 (100%)	1.0~4.1	-	-	-
	Orth+MP+	Subj.	18 (100%)	0.6~9.8	-	-	-
		Item	15 (100%)	1.9~5.9	-	-	-

Figure 8 through Figure 15 below provide graphical representations of the results shown in Table 8, starting with the individual findings for the acoustic parameter of preceding vowel duration (Figure 8, by-subject; Figure 9, by-item). In all plots, results are arranged by the experimental group and the difference scores of individual subjects and item pairs (subtracting voiceless from voiced). For preceding vowel duration, closure duration and release duration, the y-axis represents seconds (e.g., the value of 0.02 corresponds to 1/50 of a second or 20 ms). For closure voicing, the y-axis shows the number of glottal cycles. Those differences that are either in the unexpected direction or round down to zero are shown inside rectangular boxes. For preceding vowel duration, values above zero (i.e., outside of the boxes) indicate longer vowels before voiced stops (which is the expected pattern).

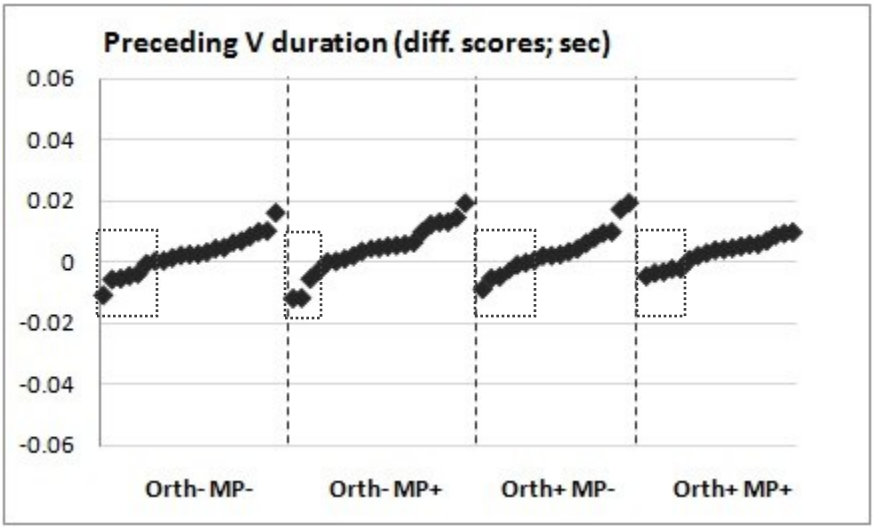


Figure 8: Preceding V duration in plosive-final monosyllabic MPs - By-subject means.

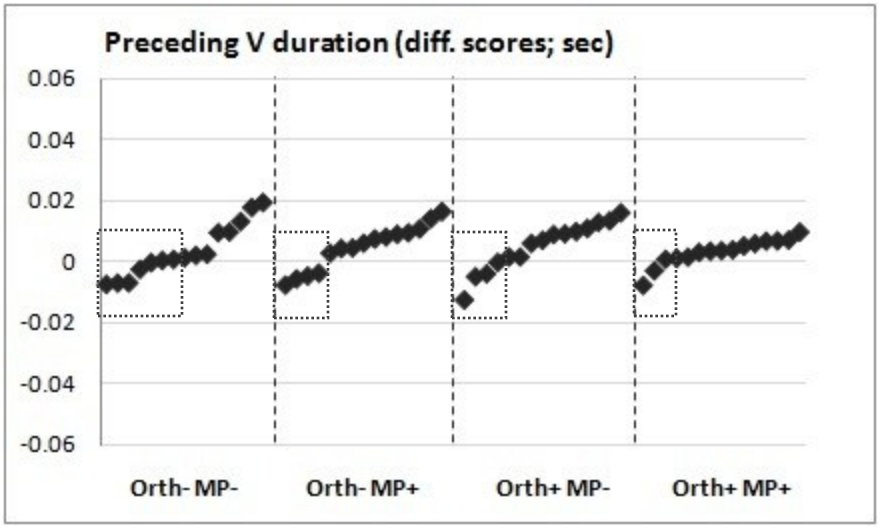


Figure 9: Preceding V duration in plosive-final monosyllabic MPs - By-item means.

Individual results for the acoustic cue of closure duration are shown in Figure 10 (by-subject) and Figure 11 (by-item). In the two figures, scores below zero (outside of the boxes) indicate longer voiceless closures (which is expected).

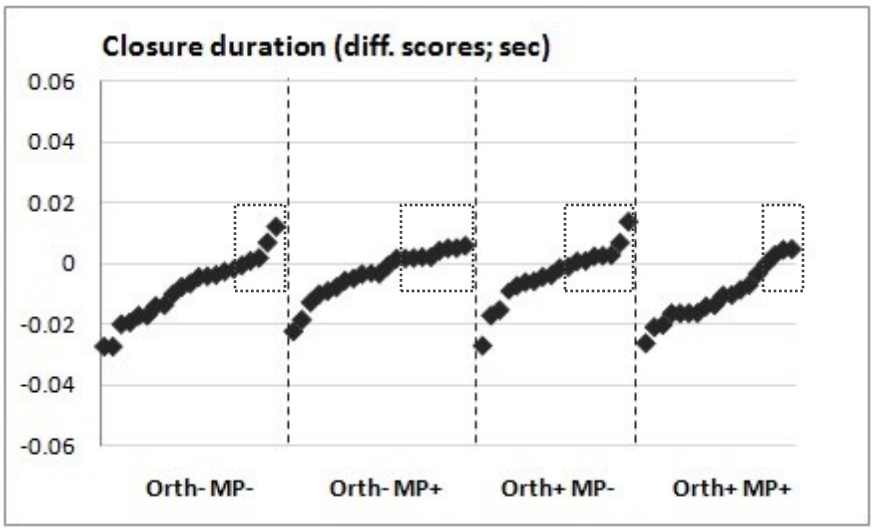


Figure 10: Closure duration in plosive-final monosyllabic MPs - By-subject means.

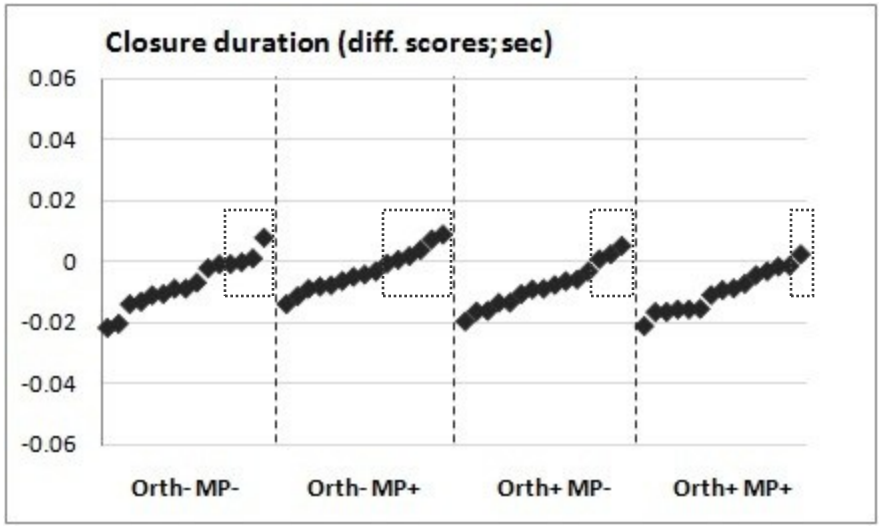


Figure 11: Closure duration in plosive-final monosyllabic MPs - By-item means.

Individual results for the acoustic parameter of release duration are shown in Figure 12 (by-subject) and Figure 13 (by-item). Values below zero (outside of the boxes) denote longer voiceless releases (which is expected).

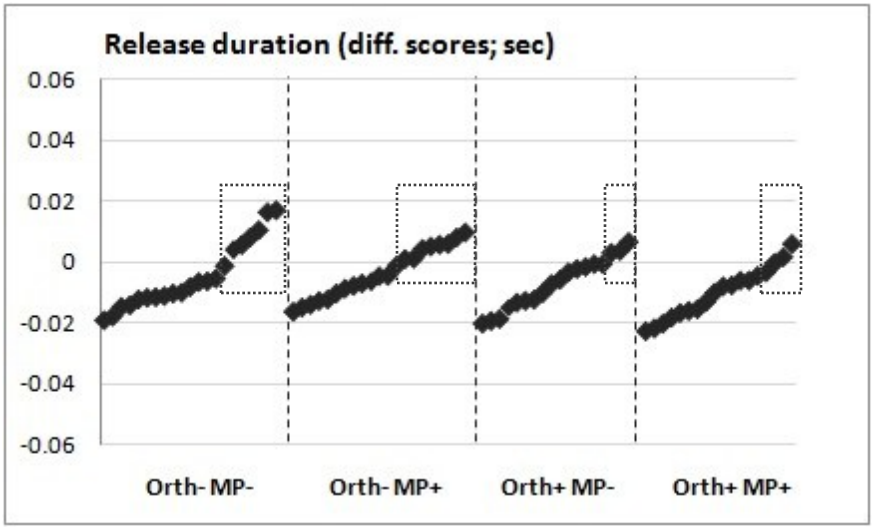


Figure 12: Release duration in plosive-final monosyllabic MPs - By-subject means.

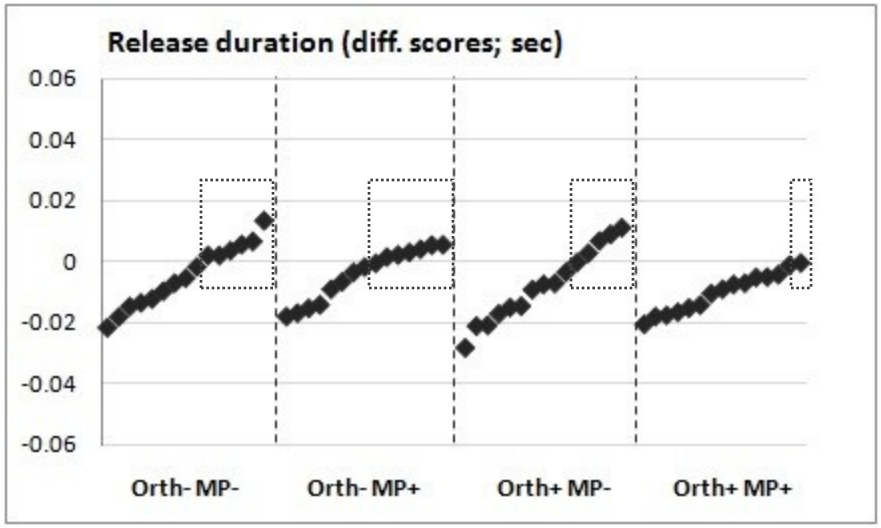


Figure 13: Release duration in plosive-final monosyllabic MPs - By-item means.

Individual results for closure voicing are shown in Figure 14 (by-subject) and Figure 15 (by-item). Scores above zero (outside of the boxes) indicate more phonetic voicing for underlyingly voiced obstruents (which is expected).

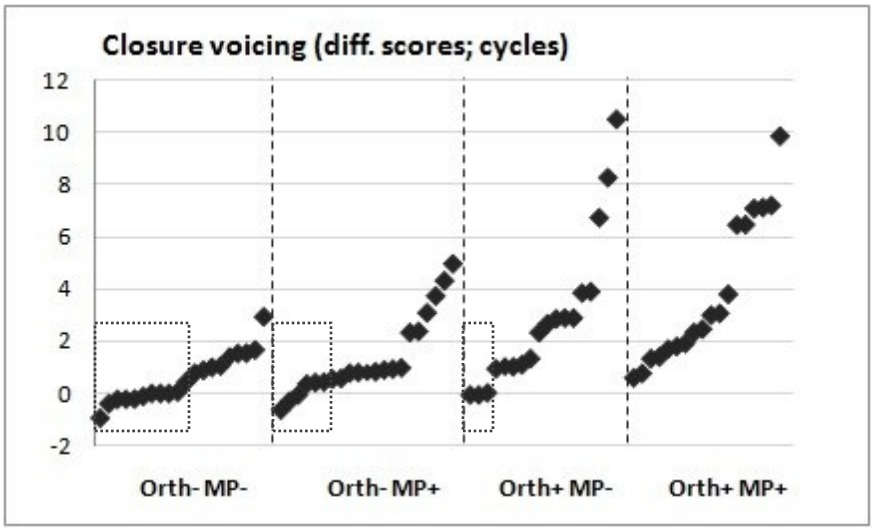


Figure 14: Closure voicing in plosive-final monosyllabic MPs - By-subject means.

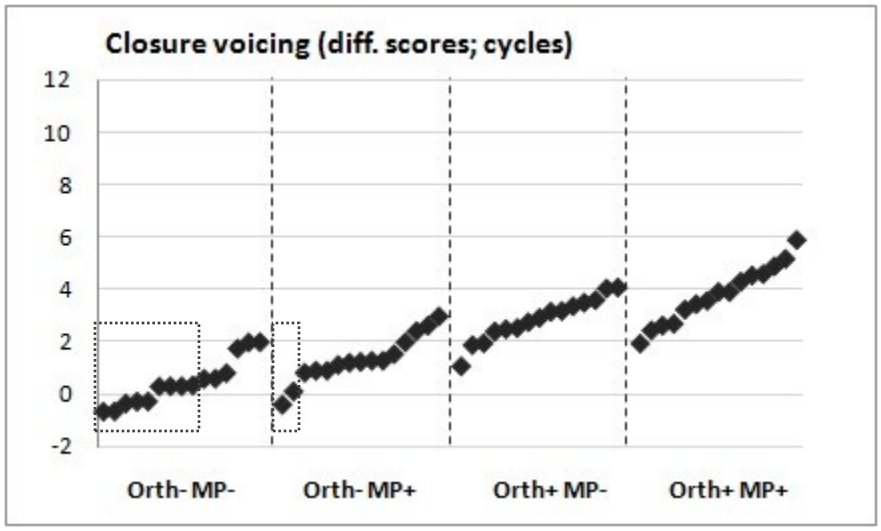


Figure 15: Closure voicing in plosive-final monosyllabic MPs - By-item means.

3.4.1.3. Summary

For plosive-final monosyllabic words with minimal pair counterparts in the lexicon (e.g., /kɔt/ 'cat' ~ /kɔd/ 'code'), group data reveal small but statistically significant differences in closure duration (7 ms longer for underlyingly voiceless stops), release duration (6 ms longer for underlyingly voiceless plosives), and closure voicing (2.12 cycles more for underlyingly voiced stops). For closure duration and release duration, the observed differences are independent of experimental group membership. For closure voicing, significantly more glottal pulsing is seen for voiced final stops only in the data of those participants who were exposed to orthographic representations of stimuli words and/or both members of each minimal pair. Differences in phonetic voicing are also more robust in the case of word-reading than picture-naming/word-guessing. Differences in preceding vowel duration do not meet the criterion of statistical

significance, and the interaction between underlying voicing and consonantal place of articulation is also not significant for any of the acoustic parameters measured.

Individual results show robust between-speaker and between-item variability in the extent and direction of the durational differences in preceding vowel duration, closure duration, and release duration. For example, up to 27 ms longer closures for underlyingly voiceless stops are seen in 68% of speakers and, depending on the experimental group, 60% to 93% of test items. At the same time, up to 14 ms longer closures for phonologically voiced plosives (which is unexpected) are present in 29% of participants and up to 33% of test items. Similar results that often go in the unexpected direction across all 4 experimental groups are also obtained for preceding vowel duration and release duration, with many of the observed differences being on the lower part of the range (e.g., around 1 ms). For closure voicing, individual results show more phonetic voicing for underlyingly voiced stops in 67% to 100% of speakers and 87% to 100% of minimal pairs in the Orth-MP+, Orth+ MP- and Orth+MP+ groups (with some of the differences being in excess of 6 glottal cycles) but only 43% of speakers and 40% of item pairs in the Orth-MP- group (with none of the differences exceeding 3 cycles). Thus, individual-level differences in closure voicing are generally robust and go in the expected direction in the data elicited during word-reading and/or with full minimal pairs included among the stimuli. The same differences are highly limited in the data recorded in non-reading tasks and without minimally contrasting wordforms present among test items. As shown in Section 3.4.2. that follows, similar results were also obtained for plosive-final monosyllabic and disyllabic non-minimal pairs.

3.4.2. Plosive-Final Non-Minimal Pairs

3.4.2.1. Group Results

Group means and 95% confidence intervals for the parameters of preceding vowel duration, closure duration, release duration and closure voicing in plosive-final monosyllabic and disyllabic non-minimal pairs are shown in Table 9. As noted in the table, only the differences in release duration and closure voicing are statistically significant for monosyllables and only the difference in closure voicing is significant for disyllables.

Table 9: Preceding V duration, closure duration, release duration & closure voicing in plosive-final monosyllabic & disyllabic non-MPs - Group means & 95% confidence intervals.

Measure	Monosyllables			Disyllables		
	Voiced	Voiceless	diff.	Voiced	Voiceless	diff.
Vowel duration (ms)	129 (124-133)	127 (122-131)	2	118 (114-122)	115 (111-119)	3
Closure duration (ms)	101 (97-104)	103 (99-106)	-2	93 (89-97)	94 (90-98)	-1
Release duration (ms)	67 (60-70)	72 (67-75)	-5**	47 (43-50)	48 (44-51)	-1
Closure voicing (cycles)	3.15 (2.79-3.51)	2.18 (1.92-2.45)	0.97**	2.93 (2.65-3.22)	1.99 (1.76-2.22)	0.94***

* - significant @ .05 level
** - significant @ .01 level
*** - significant @ .001 level

Table 10 summarizes the significant ANOVA results for plosive-final monosyllabic and disyllabic non-minimal pairs.[34]

[34] The table shows significant main effects and interactions involving UR-voice. Detailed results are provided in Appendix 5 (monosyllables) and Appendix 6 (disyllables).

Table 10: Production of plosive-final monosyllabic & disyllabic non-MPs - Summary of the significant ANOVA results

Factor	Measure	Monosyllables				Disyllables			
		By-Subject		By-Item		By-Subject		By-Item	
		F_1	p	F_2	p	F_1	p	F_2	p
UR-voice	Vowel dur.	-	-	-	-	5.28	.024	-	-
	Closure dur.	-	-	-	-	-	-	-	-
	Release dur.	13.26	.000	16.48	.000	-	-	-	-
	Closure voi.	50.22	.000	327.87	.000	72.54	.000	124.10	.000
C-place	Vowel dur.	53.02	.000	17.68	.000	65.24	.000	17.25	.000
	Closure dur.	14.95	.000	28.72	.000	13.17	.000	18.39	.000
	Release dur.	26.15	.000	69.09	.000	23.51	.000	24.51	.000
	Closure voi.	17.63	.000	48.19	.000	8.73	.000	8.97	.000
Group	Vowel dur.	-	-	3.60	.029	-	-	5.40	.002
	Closure dur.	-	-	-	-	-	-	-	-
	Release dur.	-	-	5.21	.003	-	-	-	-
	Closure voi.	3.43	.021	82.77	.000	6.28	.001	106.93	.000
UR-voice x Group	Vowel dur.	-	-	-	-	-	-	-	-
	Closure dur.	-	-	-	-	-	-	-	-
	Release dur.	-	-	-	-	-	-	-	-
	Closure voi.	9.54	.000	56.67	.000	14.60	.000	55.30	.000

As reflected in Table 10, monosyllabic wordforms showed a significant main effect of UR-voice for release duration ($F_1(1,74)=13.26$, $p<.001$; $F_2(1,24)=16.48$, $p<.001$) and closure voicing ($F_1(1,74)=50.22$, $p<.001$; $F_2(1,24)=327.87$, $p<.001$). For preceding vowel duration and closure duration, the main effect of voicing was marginally significant in by-subject analyses but not significant in by-item tests (all F_1s<3.5, all ps>.05; all F_2s<2.2, all ps>.1). In disyllabic words, only closure voicing showed a significant main effect of UR-voice in both F_1 and F_2 analyses ($F_1(1,74)=72.54$, $p<.001$; $F_2(1,24)=124.10$, $p<.001$). For preceding vowel duration, UR-voice was significant in the by-subject

ANOVA only (F_1>5, p<.05; F_2<1, p>.1). Closure duration and release duration did not show a main effect of UR-voice in either F_1 or F_2 analyses (all Fs<1.0, all ps>.1).

Main effect of Group was found for closure voicing in monosyllables as well as disyllables (monosyllables: $F_1(1,74)=3.43$, p=.021; $F_2(1,24)=82.77$, p<.001; disyllables: $F_1(1,74)=6.28$, p<.001; $F_2(1,24)=106.93$, p<.001). For both types of words, pairwise comparisons revealed that the two Orth- groups were significantly different from the Orth+MP+ group (all ps<.05). For preceding vowel duration, the Group factor was significant for monosyllables and disyllables in by-item analyses only (all F_1s>3, all ps<.05; all F_2s<1, all ps >.1). For release duration, a main effect of Group was present in monosyllabic items only in the F_2 analysis (F_1<1, p>.1; F_2>5, p<.01). Thus, as in the case of plosive-final monosyllabic minimal pairs, the main effect of Group met the criterion of statistical significance only for closure voicing.

A significant UR-voice x Group interaction was observed for closure voicing both in monosyllables and disyllables (monosyllables: $F_1(1,74)=9.54$, p<.001; $F_2(1,24)=56.67$, p<.001; disyllables: $F_1(1,74)=14.60.54$, p<.001; $F_2(1,24)=55.30$, p<.001). Pairwise comparisons showed that differences in closure voicing were always significant in the Orth-MP+ and Orth+MP+ groups (all ps<.01) and always non-significant in the Orth-MP- group (p>.1). For the Orth-MP+ group, the difference was marginally significant in monosyllabic words (p<.1) but significant in disyllables (p<.001). Furthermore, for both monosyllabic and disyllabic items, the two Orth+ groups produced significantly more phonetic voicing for phonologically voiced plosives than the two Orth- groups (all ps<.001). Hence, only the speakers in the Orth-MP- group never maintained a significant

difference in closure voicing between underlyingly voiced versus voiceless final stops, and more phonetic voicing for voiced stops was observed during the reading task. Mean values of closure voicing across the 4 experimental groups are shown in Figure 16 (monosyllables) and Figure 17 (disyllables).

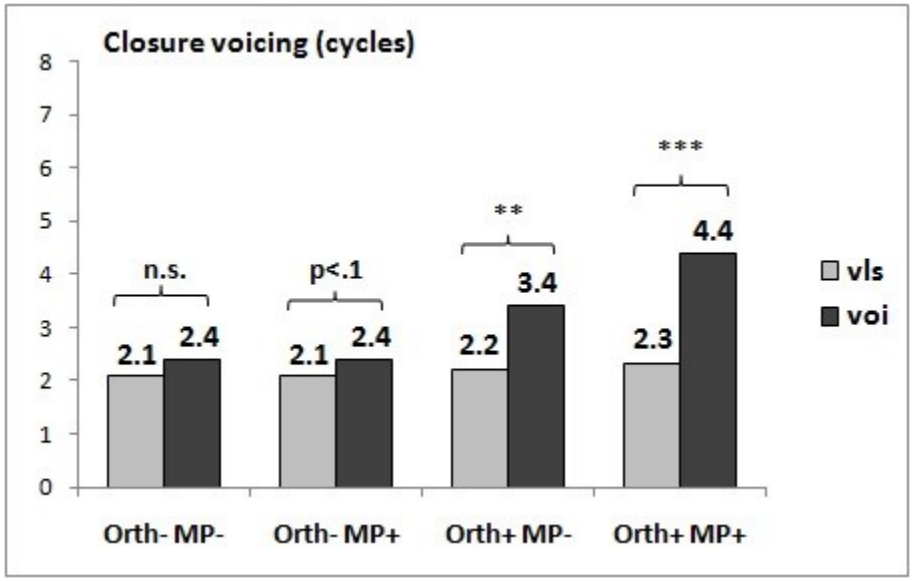


Figure 16: Closure voicing in plosive-final monosyllabic non-MPs - By-group differences.

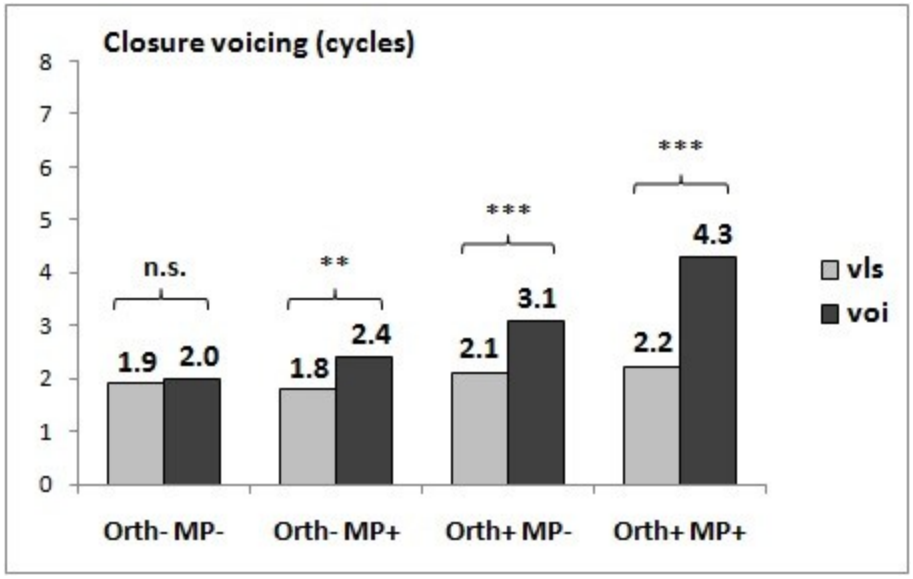


Figure 17: Closure voicing in plosive-final disyllabic non-MPs - By-group differences.

For all acoustic measurements, a significant main effect of C-place was found in monosyllabic as well as disyllabic items (all Fs>8.0, all ps<.001). However, UR-voice did not interact with C-place for any of the dependent variables, and none of the remaining interactions were significant in both by-subject and by-item analyses (all F_2s<3, all ps >.05).

3.4.2.2. Individual Results

Individual results for plosive-final non-minimal pairs are presented in Table 11 (monosyllables) and Table 12 (disyllables) and are also plotted in Figure 18 through Figure 33. The tables and the graphs follow the same format as Table 8 and Figures 8 through 15 (Section 3.4.1.2.).

Table 11: Production of plosive-final monosyllabic non-MPs - Summary of individual results.

Measure	Group		Expected pattern		Opposite pattern		diff<.5
			num. (%)	range	num. (%)	range	num. (%)
Vowel duration (ms)	Orth-MP-	Subj.	12 (57%)	1~21	8 (38%)	2~20	1 (5%)
		Item	8 (53%)	1~23	6 (40%)	2~18	1 (7%)
	Orth-MP+	Subj.	11 (52%)	1~15	9 (43%)	1~16	1 (5%)
		Item	8 (53%)	3~17	6 (40%)	2~18	1 (7%)
	Orth+MP-	Subj.	10 (55%)	1~16	8 (44%)	1~18	-
		Item	10 (67%)	3~10	5 (33%)	2~13	-
	Orth+MP+	Subj.	12 (67%)	2~19	6 (33%)	1~12	-
		Item	11 (73%)	1~19	4 (27%)	3~7	-
Closure duration (ms)	Orth-MP-	Subj.	15 (71%)	1~12	6 (29%)	1~8	-
		Item	11 (73%)	2~10	3 (20%)	4~10	1 (7%)
	Orth-MP+	Subj.	10 (48%)	1~18	10 (48%)	1~26	1 (5%)
		Item	8 (53%)	1~8	6 (40%)	1~16	1 (7%)
	Orth+MP-	Subj.	10 (55%)	4~17	7 (39%)	9~17	1 (5%)
		Item	7 (47%)	2~13	7 (47%)	1~13	1 (7%)
	Orth+MP+	Subj.	14 (78%)	4~13	4 (22%)	4~15	-
		Item	12 (80%)	1~14	3 (20%)	2~4	-
Release duration (ms)	Orth-MP-	Subj.	15 (71%)	1~29	5 (24%)	8~23	1 (5%)
		Item	8 (53%)	2~14	6 (40%)	1~14	1 (7%)
	Orth-MP+	Subj.	13 (62%)	1~19	7 (33%)	2~14	1 (5%)
		Item	10 (67%)	2~18	4 (27%)	4~8	1 (7%)
	Orth+MP-	Subj.	15 (83%)	1~28	3 (17%)	7~29	-
		Item	8 (53%)	1~15	5 (33%)	1~8	2 (13%)
	Orth+MP+	Subj.	12 (67%)	3~40	6 (33%)	1~20	-
		Item	15 (100%)	1~18	-	-	-
Closure voicing (cycles)	Orth-MP-	Subj.	10 (48%)	0.7~1.7	1 (5%)	0.6	10 (48%)
		Item	3 (20%)	0.6~1.1	1 (7%)	0.9	11 (73%)
	Orth-MP+	Subj.	5 (24%)	0.6~1.9	2 (9%)	0.6~0.7	14 (67%)
		Item	4 (27%)	0.6~1.6	1 (7%)	0.9	10 (67%)
	Orth+MP-	Subj.	11 (61%)	0.7~4.1	-	-	7 (39%)
		Item	13 (87%)	0.7~1.8	-	-	2 (13%)
	Orth+MP+	Subj.	15 (83%)	0.9~5.7	-	-	3 (17%)
		Item	15 (100%)	1.1~2.8	-	-	-

Table 12: Production of plosive-final disyllabic non-MPs - Summary of individual results.

Measure	Group		Expected pattern		Opposite pattern		diff<.5
			num. (%)	range	num. (%)	range	num. (%)
Vowel duration (ms)	Orth-MP-	Subj.	15 (71%)	1~20	6 (29%)	4~27	-
		Item	8 (53%)	2~22	7 (47%)	1~12	-
	Orth-MP+	Subj.	10 (48%)	2~27	9 (43%)	1~16	2 (9%)
		Item	9 (60%)	3~18	5 (33%)	1~16	1 (7%)
	Orth+MP-	Subj.	9 (50%)	2~21	8 (44%)	1~16	1 (5%)
		Item	7 (47%)	8~21	8 (53%)	1~15	-
	Orth+MP+	Subj.	11 (61%)	5~20	7 (39%)	1~13	-
		Item	8 (53%)	4~25	7 (47%)	1~14	-
Closure duration (ms)	Orth-MP-	Subj.	12 (57%)	3~21	8 (38%)	1~14	1 (5%)
		Item	9 (60%)	2~14	4 (27%)	6~8	2 (13%)
	Orth-MP+	Subj.	9 (43%)	2~15	12 (57%)	3~20	
		Item	5 (33%)	2~9	10 (67%)	2~14	-
	Orth+MP-	Subj.	9 (50%)	3~21	9 (50%)	1~21	-
		Item	7 (47%)	1~10	6 (40%)	2~14	2 (13%)
	Orth+MP+	Subj.	11 (61%)	1~24	6 (33%)	3~21	1 (5%)
		Item	9 (60%)	1~16	6 (40%)	1~4	-
Release duration (ms)	Orth-MP-	Subj.	10 (48%)	4~20	11 (52%)	1~20	-
		Item	8 (53%)	2~18	7 (47%)	1~14	-
	Orth-MP+	Subj.	9 (43%)	2~21	11 (52%)	1~17	1 (5%)
		Item	8 (53%)	2~16	6 (40%)	2~10	1 (7%)
	Orth+MP-	Subj.	7 (39%)	1~16	10 (56%)	2~21	1 (5%)
		Item	7 (47%)	1~9	7 (47%)	2~10	1 (7%)
	Orth+MP+	Subj.	11 (61%)	1~22	5 (28%)	3~9	2 (11%)
		Item	12 (80%)	1~12	3 (20%)	3~7	-
Closure voicing (cycles)	Orth-MP-	Subj.	6 (29%)	0.6~1.4	3 (14%)	0.8~1.8	12 (57%)
		Item	3 (20%)	0.6~0.7	1 (7%)	0.6	11 (73%)
	Orth-MP+	Subj.	10 (48%)	0.6~2.5	1 (5%)	0.6	10 (48%)
		Item	8 (53%)	0.6~0.9	-	-	7 (47%)
	Orth+MP-	Subj.	12 (67%)	0.6~3.7	-	-	6 (33%)
		Item	13 (87%)	0.6~1.9	-	-	2 (13%)
	Orth+MP+	Subj.	16 (89%)	0.7~4.5	-	-	2 (11%)
		Item	15 (100%)	1.1~2.9	-	-	-

Individual-level results for the parameter of preceding vowel duration are shown in Figure 18 (monosyllabic, by-subject), Figure 19 (monosyllabic, by-item), Figure 20 (disyllabic, by-subject), and Figure 21 (disyllabic, by-item). As before, scores above zero (outside of the boxes) represent longer vowels before voiced stops (which is expected).

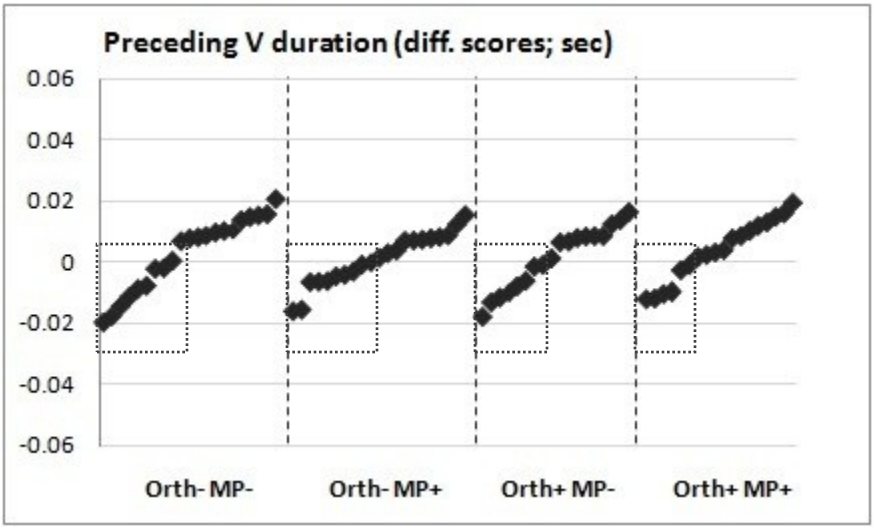


Figure 18: Preceding V duration in plosive-final monosyllabic non-MPs - By-subject means.

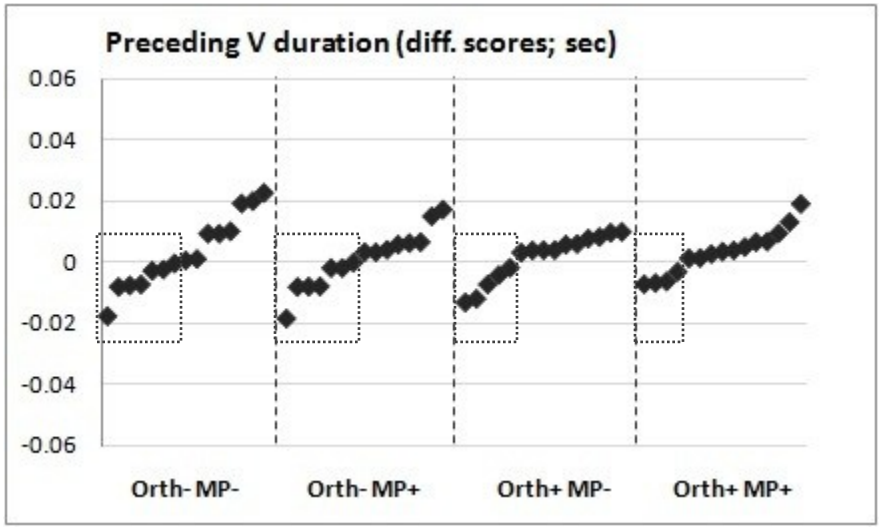


Figure 19: Preceding V duration in plosive-final monosyllabic non-MPs - By-item means.

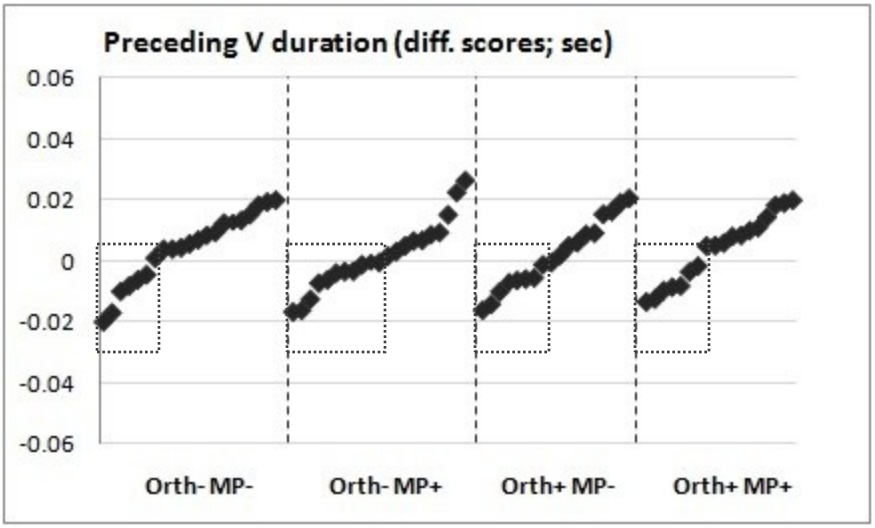


Figure 20: Preceding V duration in plosive-final disyllabic non-MPs - By-subject means.

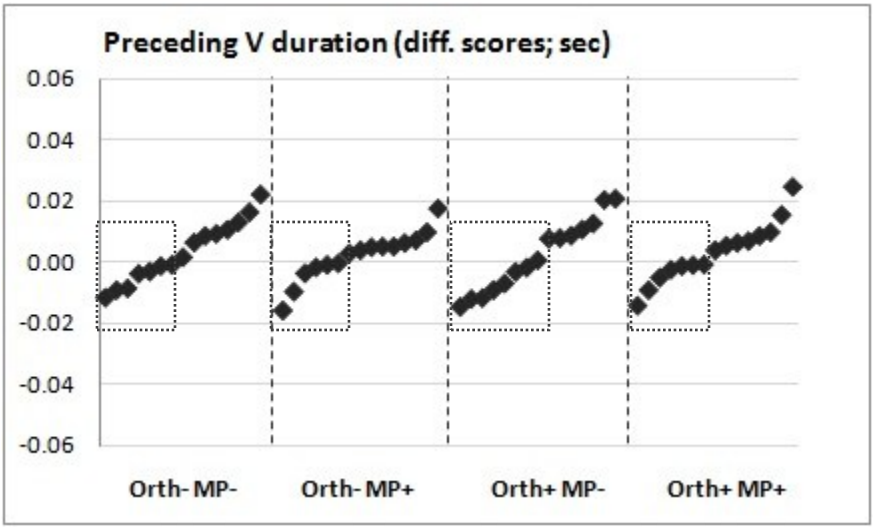


Figure 21: Preceding V duration in plosive-final disyllabic non-MPs - By-item means.

Results for the parameter of closure duration are shown in Figure 22 (monosyllabic, by-subject), Figure 23 (monosyllabic, by-item), Figure 24 (disyllabic, by-subject), and Figure 25 (disyllabic, by-item). Scores below zero (outside of the boxes) indicate longer voiceless closures (which is expected).

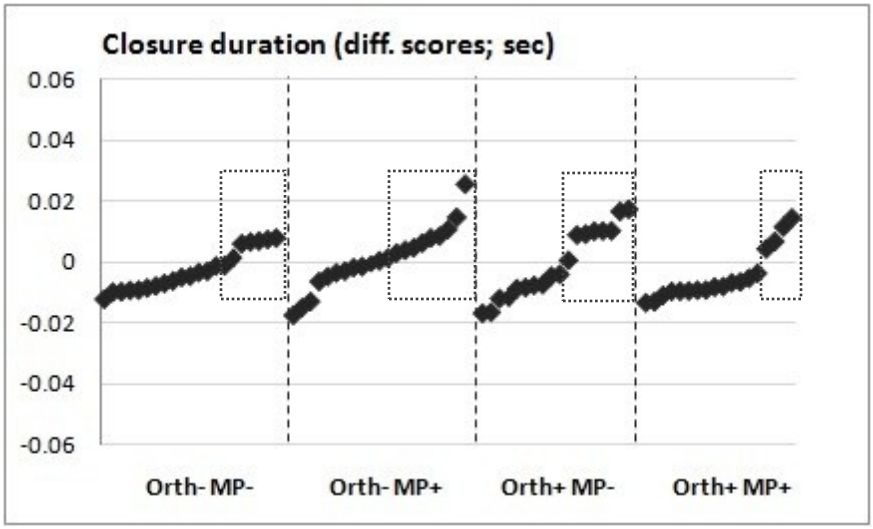


Figure 22: Closure duration in plosive-final monosyllabic non-MPs - By-subject means.

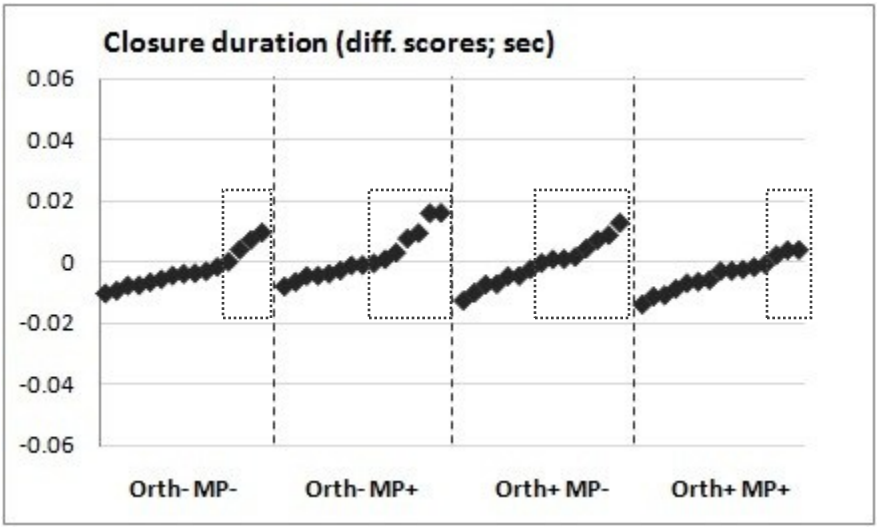


Figure 23: Closure duration in plosive-final monosyllabic non-MPs - By-item means.

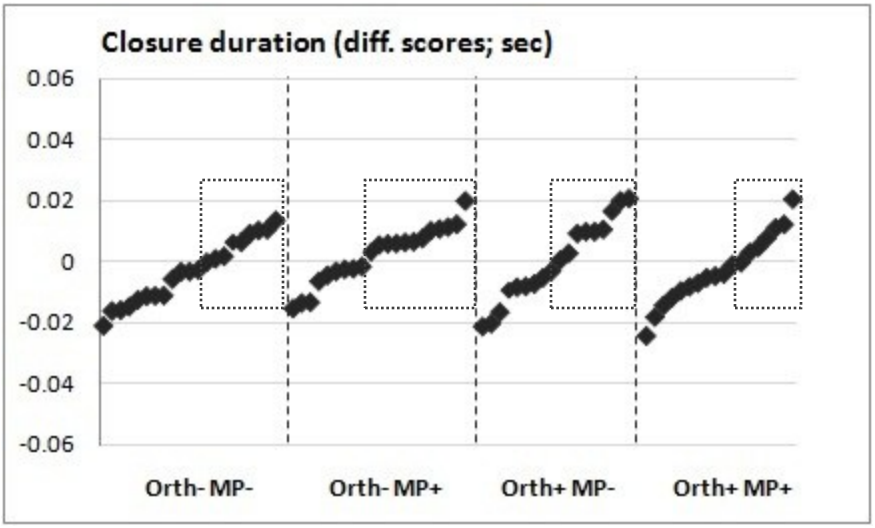


Figure 24: Closure duration in plosive-final disyllabic non-MPs - By-subject means.

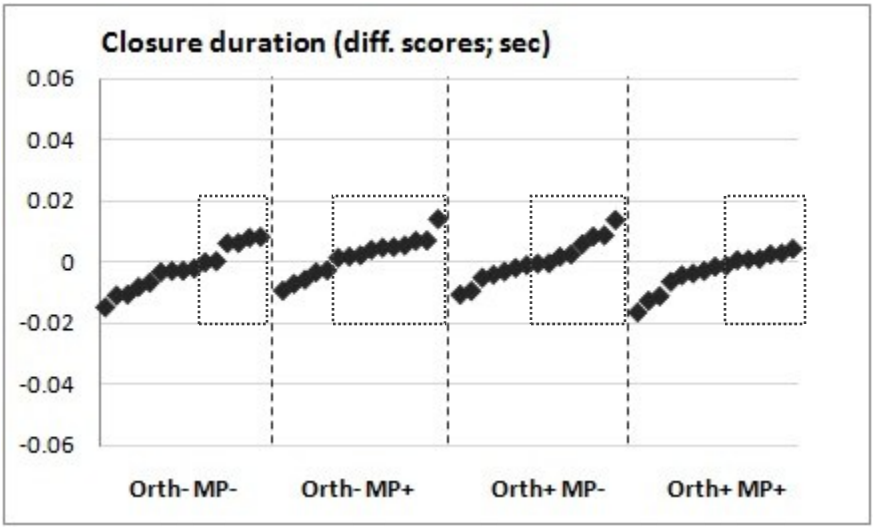


Figure 25: Closure duration in plosive-final disyllabic non-MPs - By-item means.

Individual results for the parameter of release duration are shown in Figure 26 (monosyllabic, by-subject), Figure 27 (monosyllabic, by-item), Figure 28 (disyllabic, by-subject), and Figure 29 (disyllabic, by-item). Values below zero (outside of the boxes) denote longer voiceless releases (which is expected).

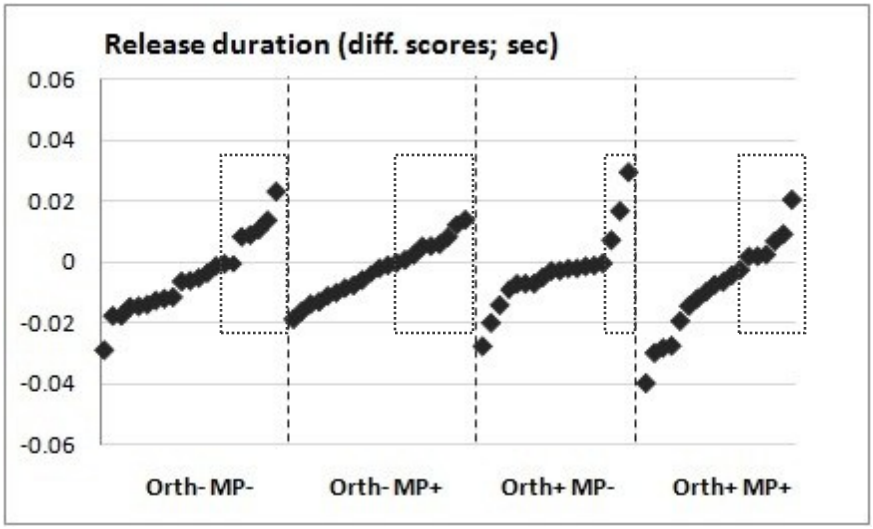


Figure 26: Release duration in plosive-final monosyllabic non-MPs - By-subject means.

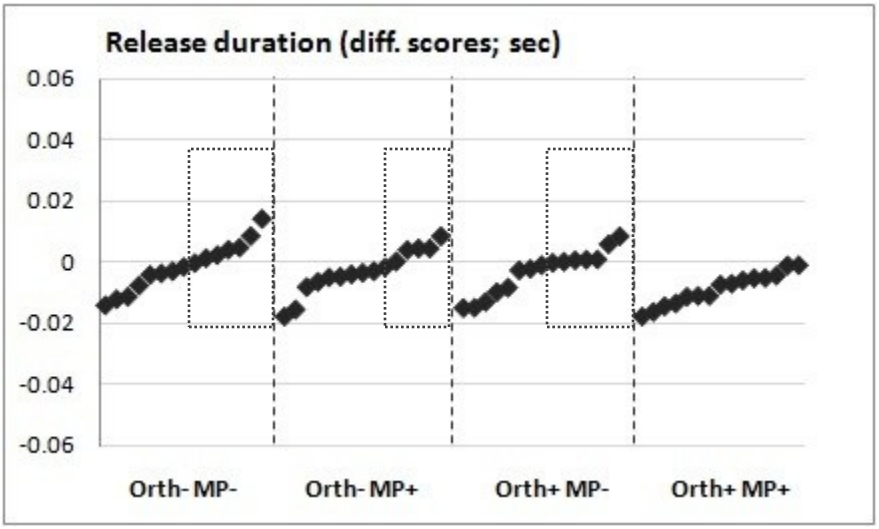


Figure 27: Release duration in plosive-final monosyllabic non-MPs - By-item means.

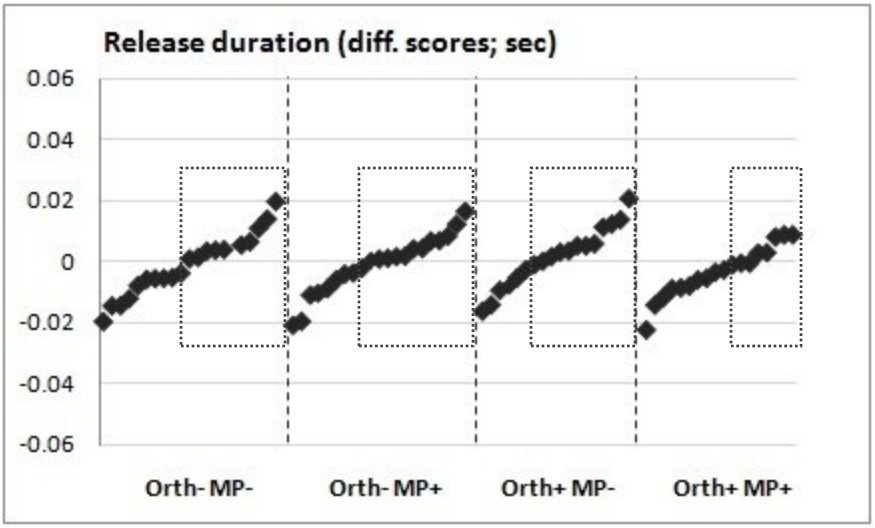


Figure 28: Release duration in plosive-final disyllabic non-MPs - By-subject means.

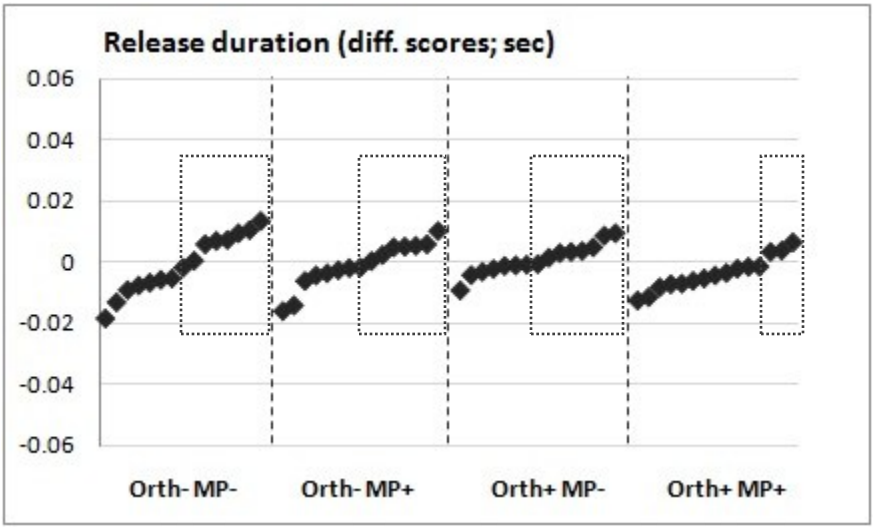


Figure 29: Release duration in plosive-final disyllabic non-MPs - By-item means.

Results for closure voicing are shown in Figure 30 (monosyllabic, by-subject), Figure 31 (monosyllabic, by-item), Figure 32 (disyllabic, by-subject), and Figure 33 (disyllabic, by-item). Scores above zero (outside of the boxes) signal more phonetic voicing for underlyingly voiced segments (which is expected).

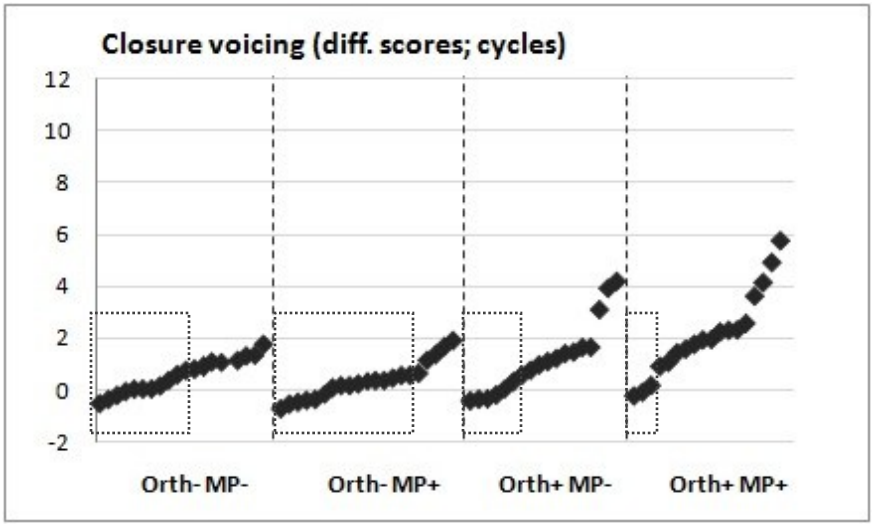


Figure 30: Closure voicing in plosive-final monosyllabic non-MPs - By-subject means.

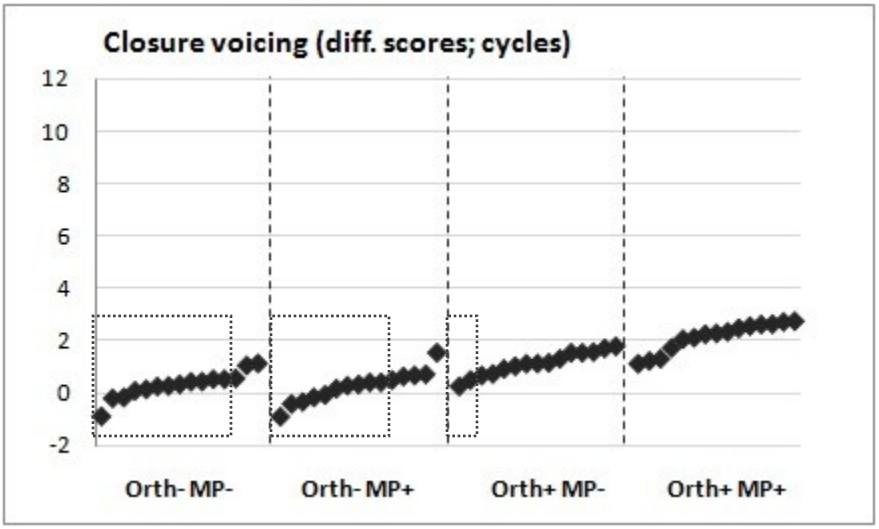


Figure 31: Closure voicing in plosive-final monosyllabic non-MPs - By-item means.

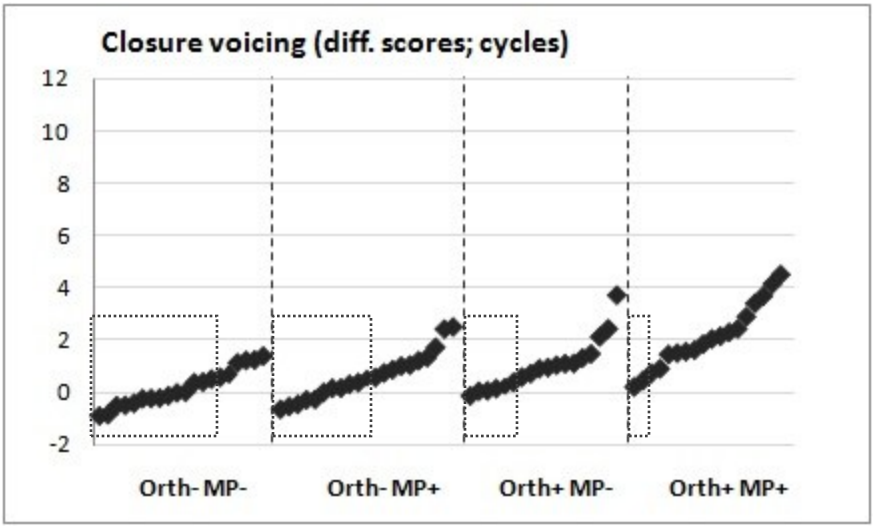


Figure 32: Closure voicing in plosive-final disyllabic non-MPs - By-subject means.

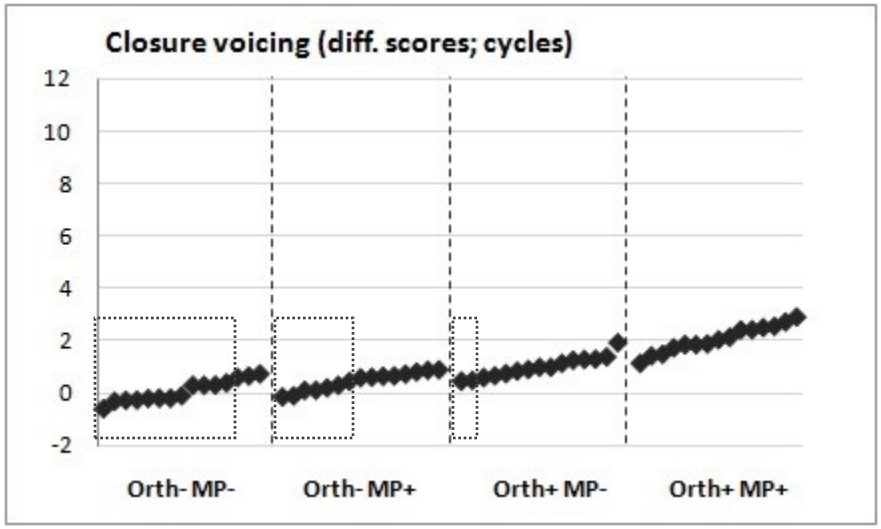


Figure 33: Closure voicing in plosive-final disyllabic non-MPs - By-item means.

3.4.2.3. Summary

For monosyllabic plosive-final words without potentially homophonous competitors in the lexicon (e.g., /flag/ 'flag'; */flak/ is not an existing word of Russian), group results reveal the presence of longer releases for underlyingly voiceless stops (5 ms difference) and more closure voicing for phonologically voiced plosives (0.97 cycles difference). In disyllabic items (e.g., /parad/ 'parade'; */parat/ is not an existing word of Russian), only the difference in the amount of glottal pulsing is significant (0.94 cycles more for underlyingly voiced plosives). However, whereas differences in release duration are found in monosyllabic items independently of experimental group, differences in closure voicing are significant (or marginally significant) only in the Orth-MP+, Orth+MP- and Orth+MP+ groups but not the Orth-MP- group. The two Orth+ groups also show more phonetic voicing for phonologically voiced plosives. Thus, closure voicing signals the underlying contrast only during word-reading or, to a lesser extent, during picture-naming/word-guessing if full minimal pairs are present among the stimuli. The remaining acoustic parameters show durational differences that are in the expected direction but do not meet the criterion of statistical significance, and the observed effects are also found to be consistent across different places of articulation.

Individual results show robust between-speaker and between-item variability in the extent and direction of voicing-dependent differences in release duration (as well as preceding vowel duration and closure duration). For example, for monosyllables, releases are up to 40 ms longer in underlyingly voiceless stops in 71% of speakers and 53% to 100% of item pairs, yet the same difference in release duration are up to 29 ms in the

opposite direction in 27% of subjects and 27% to 40% of stimuli pairs. Many means are also around 1 ms, and similar degrees of variability are seen for all experimental groups. Unlike release duration, closure voicing shows a clear asymmetry for both monosyllables and disyllables between the two Orth- groups and the two Orth+ groups. More glottal pulsing for underlyingly voiced stops is found in over 60% of speakers and up to 100% of item pairs in the two Orth+ groups, with mean differences reaching as high as 5.7 cycles. The same pattern is present in no more than 48% of subjects and 53% of items in the two Orth- groups, with averaged values never exceeding 2.5 cycles. Section 3.4.3. below shows that comparable results were also obtained for items with final fricatives.

3.4.3. Fricative-Final Non-Minimal Pairs

3.4.3.1. Group Results

Group means and 95% confidence intervals for the parameters of preceding vowel duration, frication duration and closure voicing in fricative-final monosyllabic and disyllabic non-minimal pair items are shown in Table 13. As noted in the table, only the differences in frication duration and closure voicing are statistically significant.

Table 13: Preceding V duration, frication duration & closure voicing in fricative-final monosyllabic & disyllabic non-MPs - Group means & 95% confidence intervals.

Measure	Monosyllables			Disyllables		
	Voiced	Voiceless	diff.	Voiced	Voiceless	diff.
Vowel duration (ms)	126 (121-130)	123 (118-127)	3	120 (116-124)	117 (112-121)	3
Frication duration (ms)	162 (157-167)	175 (169-181)	-13***	155 (150-161)	163 (158-169)	-8***
Closure voicing (cycles)	2.37 (1.81-2.92)	0.85 (0.75-0.96)	1.52***	2.14 (1.67-2.61)	0.82 (0.71-0.93)	1.32***

* - significant @ .05 level
** - significant @ .01 level
*** - significant @ .001 level

A summary of the significant ANOVA results for fricative-final items without minimal pair counterparts in the lexicon is provided in Table 14.[35]

[35] The table lists significant main effects and interactions involving UR-voice. Detailed results are given in Appendix 7 (monosyllables) and Appendix 8 (disyllables).

Table 14: Production of fricative-final monosyllabic & disyllabic non-MPs - Summary of the significant ANOVA results

Factor	Measure	Monosyllables				Disyllables			
		By-Subject		By-Item		By-Subject		By-Item	
		F_1	p	F_2	p	F_1	p	F_2	p
UR-voice	V dur.	12.62	.001	-	-	15.19	.000	-	-
	Frication dur.	79.32	.000	17.82	.000	23.78	.000	14.94	.001
	Closure voi.	31.32	.000	75.52	.000	34.14	.000	345.98	.000
C-place	V dur.	7.07	.001	-	-	3.96	.002	-	-
	Closure dur.	87.59	.000	29.65	.000	79.07	.000	53.82	.000
	Closure voi.	-	-	-	-	-	-	-	-
Group	V dur.	-	-	9.33	.000	-	-	7.88	.000
	Frication dur.	-	-	6.60	.001	-	-	6.01	.002
	Closure voi.	5.10	.003	51.15	.000	5.27	.002	156.24	.000
UR-voice x Group	V dur.	3.27	.026	-	-	-	-	-	-
	Frication dur.	-	-	4.62	.006	-	-	3.58	.025
	Closure voi.	5.68	.001	47.02	.000	5.62	.002	130.63	.000
UR-voice x C-place	V dur.	13.75	.000	-	-	4.98	.008	-	-
	Frication dur.	4.59	.013	-	-	3.75	.028	-	-
	Closure voi.	-	-	-	-	-	-	-	-
UR-voice x C-place x Group	V dur.	-	-	-	-	-	-	-	-
	Frication dur.	-	-	-	-	-	-	-	-
	Closure voi.	-	-	-	-	-	-	3.95	.004

As can be seen in the table above, a significant main effect of UR-voice on frication duration was found for both monosyllables and disyllables (monosyllables: $F_1(1,74)=79.32$, $p<.001$; $F_2(1,24)=17.82$, $p<.001$; disyllables: $F_1(1,74)=23.78$, $p<.001$; $F_2(1,24)=14.94$, $p<.001$). A significant main effect of UR-voice on closure voicing was also observed for both stimuli types (monosyllables: $F_1(1,74)=31.32$, $p<.001$; $F_2(1,24)=75.52$, $p<.001$; disyllables: $F_1(1,74)=34.14$, $p<.001$; $F_2(1,24)=345.98$, $p<.001$).

For preceding vowel duration, a main effect of UR-voice was significant only in the by-subject ANOVAs (all F_1s >12, all ps<.001; all F_2s<1, all ps>.1).

For preceding vowel duration and frication duration, a main effect of Group was found in monosyllables and disyllables only in by-item analyses (all F_1s<1, all ps>.1; all F_2s>6, all ps<.01) and UR-voice interacted with Group in either by-subject or by-item ANOVAs (all Fs>3, all ps<.05) but not in both types of tests at the same time (all remaining Fs<2.5, all ps>.05). Closure voicing showed a main effect of Group regardless of word length in syllables (monosyllables: $F_1(1,74)=5.10$, $p=.003$; $F_2(1,24)=51.15$, $p<.001$; disyllables: $F_1(1,74)=5.27$, $p=.002$; $F_2(1,24)=156.24$, $p<.001$), and pairwise comparisons revealed that the two Orth- groups were always significantly different from the Orth+MP+ group (all ps<.05).

For both monosyllables and disyllables, closure voicing also showed a significant UR-voice x Group interaction (monosyllables: $F_1(1,74)=5.68$, $p=.001$; $F_2(1,24)=47.02$, $p<.001$; disyllables: $F_1(1,74)=5.62$, $p=.002$; $F_2(1,24)=130.63$, $p<.001$). Pairwise comparisons indicated that the difference in the amount of phonetic voicing between underlyingly voiced versus voiceless final fricatives was never significant in the Orth-MP- group (all ps>.1) but always significant in the Orth-MP+, Orth+MP- and Orth+MP+ groups (all ps<.01). For monosyllabic words, differences in the amount of voicing produced for phonologically voiced fricatives were significant across all experimental groups (all ps<.05). For disyllables, the two Orth- groups produced significantly less voicing for underlyingly voiced segments than the two Orth+ groups (all ps<.01). Mean

values of closure voicing across the 4 experimental groups are shown in Figure 34 (monosyllables) and Figure 35 (disyllables).

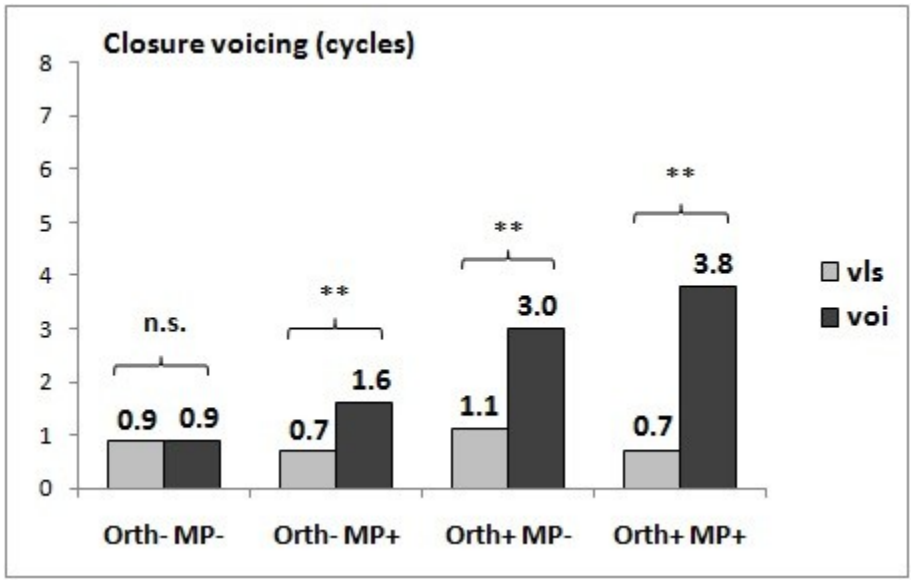


Figure 34: Closure voicing in fricative-final monosyllabic non-MPs - By-group differences.

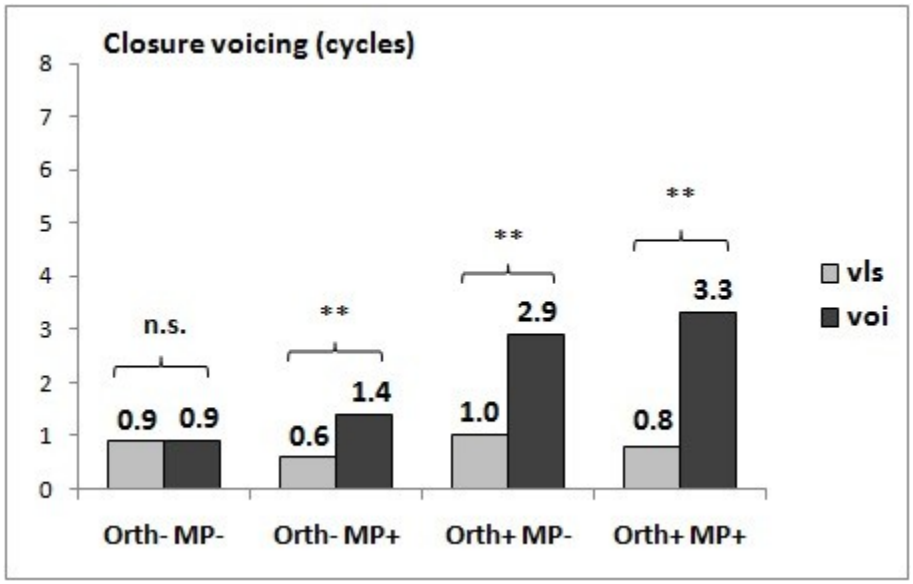


Figure 35: Closure voicing in fricative-final disyllabic non-MPs - By-group differences.

Main effect of C-place was significant for the parameter of frication duration in all analyses of variance (all Fs>29, all ps<.001) and for preceding vowel duration in by-item analyses only (all F_1s>3, all ps<.01; all F_2s<1, all ps >.1). However, none of the interactions involving C-place met the criterion of statistical significance. Namely, a significant UR-voice x C-place interaction was found for preceding vowel duration and frication duration in by-subject ANOVAs only (all F_1s>3, all ps<.05; F_2s<1, all ps>.1), and a 3-way interaction between UR-voice, C-place and Group was seen for closure voicing in disyllabic words only in the by-item analysis (F_2 <2, p>.1; F_2>3, p<.01).

3.4.3.2. Individual Results

Individual results for fricative-final non-minimal pairs are shown in Table 15 (monosyllables) and Table 16 (disyllables) and are plotted in Figure 36 through Figure 47, with all tables and graphs following the same format as in the case of plosive-final stimuli (Section 3.4.1.2. and Section 3.4.2.2.).

Table 15: Production of fricative-final monosyllabic non-MPs - Summary of individual results.

Measure	Group		Expected pattern		Opposite pattern		diff<.5
			num. (%)	range	num. (%)	range	num. (%)
Vowel duration (ms)	Orth-MP-	Subj.	9 (43%)	1~12	9 (43%)	2~9	3 (14%)
		Item	9 (60%)	2~13	6 (40%)	1~15	-
	Orth-MP+	Subj.	10 (48%)	1~11	7 (33%)	1~7	4 (19%)
		Item	6 (40%)	2~17	9 (60%)	1~8	-
	Orth+MP-	Subj.	11 (61%)	1~19	5 (28%)	7~19	2 (11%)
		Item	7 (47%)	1~22	7 (47%)	1~14	1 (7%)
	Orth+MP+	Subj.	16 (89%)	1~20	2 (11%)	1~10	-
		Item	10 (67%)	2~24	5 (33%)	4~12	-
Frication duration (ms)	Orth-MP-	Subj.	20 (95%)	1~24	1 (5%)	9	-
		Item	14 (93%)	3~30	1 (7%)	2	-
	Orth-MP+	Subj.	17 (81%)	1~30	4 (19%)	3~7	-
		Item	12 (80%)	2~23	3 (20%)	2~5	-
	Orth+MP-	Subj.	14 (78%)	1~40	2 (11%)	9~10	2 (11%)
		Item	12 (80%)	1~32	3 (20%)	4~8	-
	Orth+MP+	Subj.	16 (89%)	3~63	2 (11%)	1~6	-
		Item	15 (100%)	4~43	-	-	-
Closure voicing (cycles)	Orth-MP-	Subj.	1 (5%)	0.6	-	-	20 (95%)
		Item	2 (13%)	0.6~0.8	1 (7%)	0.9	12 (80%)
	Orth-MP+	Subj.	13 (62%)	0.6~4.7	-	-	8 (38%)
		Item	10 (67%)	0.6~2.5	-	-	5 (33%)
	Orth+MP-	Subj.	11 (61%)	0.8~10.7	-	-	7 (39%)
		Item	14 (93%)	0.8~3.2	-	-	1 (7%)
	Orth+MP+	Subj.	17 (95%)	0.8~13.9	-	-	1 (5%)
		Item	15 (100%)	1.8~4.9	-	-	-

Table 16: Production of fricative-final disyllabic non-MPs - Summary of individual results.

Measure	Group		Expected pattern		Opposite pattern		diff<.5
			num. (%)	range	num. (%)	range	num. (%)
Vowel duration (ms)	Orth-MP-	Subj.	11 (52%)	1~16	8 (38%)	2~7	2 (9%)
		Item	10 (67%)	1~16	5 (33%)	5~15	-
	Orth-MP+	Subj.	8 (38%)	5~10	9 (43%)	1~14	4 (19%)
		Item	5 (33%)	4~20	8 (53%)	1~11	2 (13%)
	Orth+MP-	Subj.	13 (72%)	2~20	5 (28%)	3~16	-
		Item	10 (56%)	1~16	4 (27%)	2~5	1 (7%)
	Orth+MP+	Subj.	15 (83%)	1~20	3 (17%)	3~8	-
		Item	11 (73%)	1~25	4 (27%)	2~10	-
Frication duration (ms)	Orth-MP-	Subj.	17 (81%)	2~26	4 (19%)	1~15	-
		Item	11 (73%)	2~26	2 (13%)	1~7	2 (13%)
	Orth-MP+	Subj.	14 (67%)	1~30	6 (29%)	2~15	1 (5%)
		Item	9 (60%)	2~26	6 (40%)	1~10	-
	Orth+MP-	Subj.	10 (56%)	1~41	7 (39%)	1~27	1 (5%)
		Item	9 (60%)	6~27	5 (33%)	1~10	1 (7%)
	Orth+MP+	Subj.	14 (78%)	3~57	4 (22%)	1~16	-
		Item	14 (93%)	4~27	1 (7%)	3	-
Closure voicing (cycles)	Orth-MP-	Subj.	1 (5%)	0.6	-	-	20 (95%)
		Item	1 (7%)	0.7	-	-	14 (93%)
	Orth-MP+	Subj.	10 (48%)	0.7~3.4	-	-	11 (52%)
		Item	9 (60%)	0.7~1.4	-	-	6 (40%)
	Orth+MP-	Subj.	12 (67%)	0.8~9.2	-	-	6 (33%)
		Item	15 (100%)	1.1~2.8	-	-	-
	Orth+MP+	Subj.	12 (67%)	0.7~9.8	-	-	6 (33%)
		Item	15 (100%)	1.6~3.2	-	-	-

Results for the parameter of preceding vowel duration in individual subjects and items are shown in Figure 36 (monosyllabic, by-subject), Figure 37 (monosyllabic, by-item), Figure 38 (disyllabic, by-subject), and Figure 39 (disyllabic, by-item). As before, scores above zero (outside of the boxes) indicate longer vowels before voiced consonants (which is the expected direction).

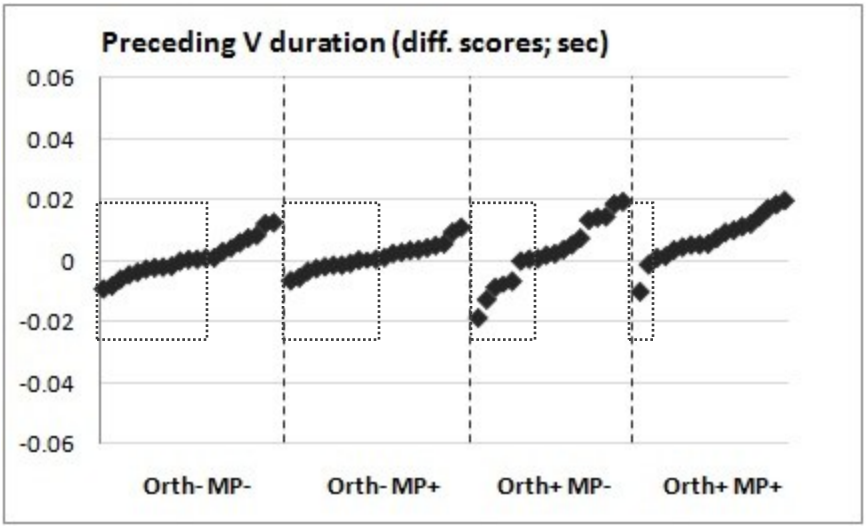


Figure 36: Preceding V duration in fricative-final monosyllabic non-MPs - By-subject means.

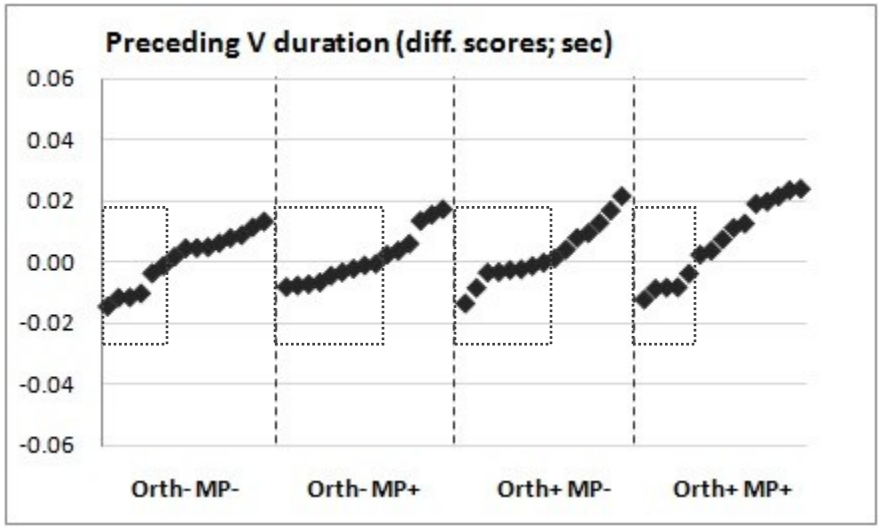


Figure 37: Preceding V duration in fricative-final monosyllabic non-MPs - By-item means.

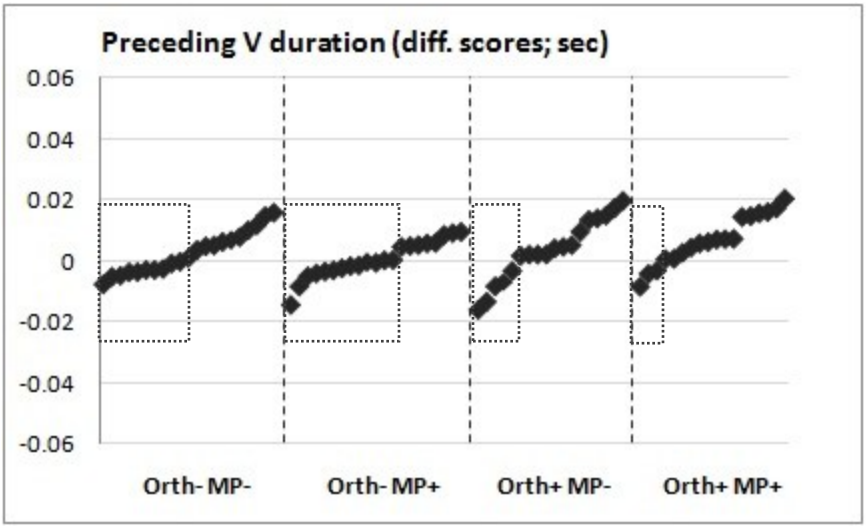


Figure 38: Preceding V duration in fricative-final disyllabic non-MPs - By-subject means

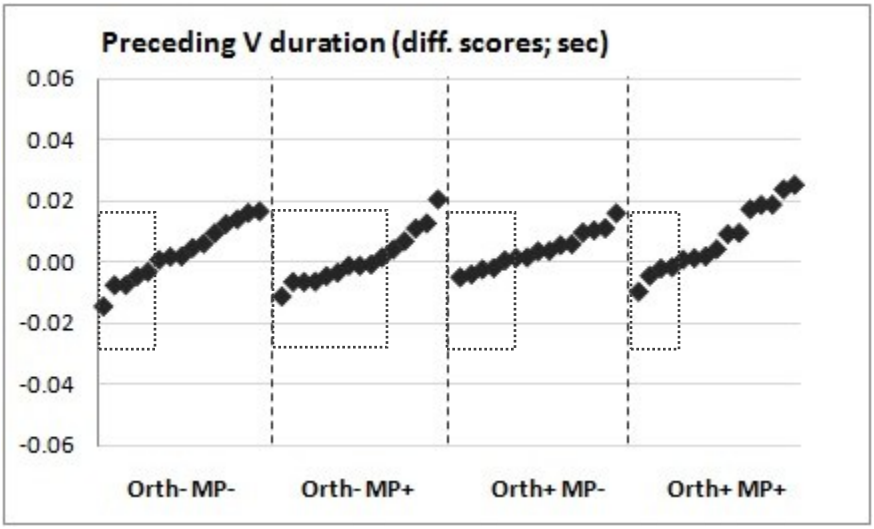


Figure 39: Preceding V duration in fricative-final disyllabic non-MPs - By-item means.

Individual results for the parameter of frication duration are shown in Figure 40 (monosyllabic, by-subject), Figure 41 (monosyllabic, by-item), Figure 42 (disyllabic, by-subject), and Figure 43 (disyllabic, by-item). Values below zero (outside of the boxes) indicate longer frication for voiceless sounds (which is expected).

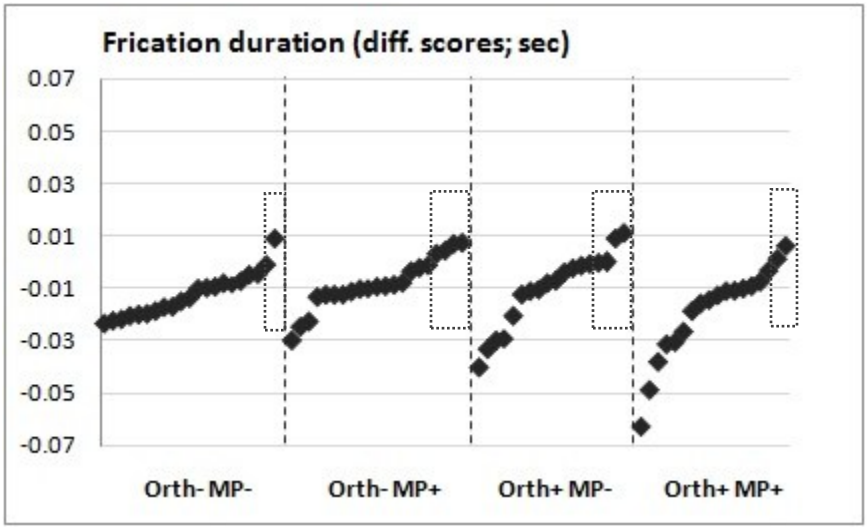


Figure 40: Frication duration in fricative-final monosyllabic non-MPs - By-subject means.

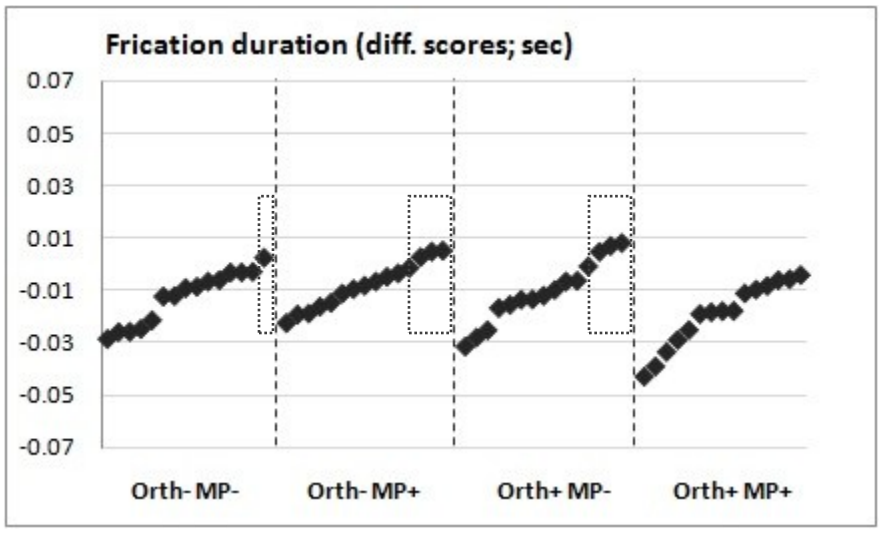


Figure 41: Frication duration in fricative-final monosyllabic non-MPs - By-item means.

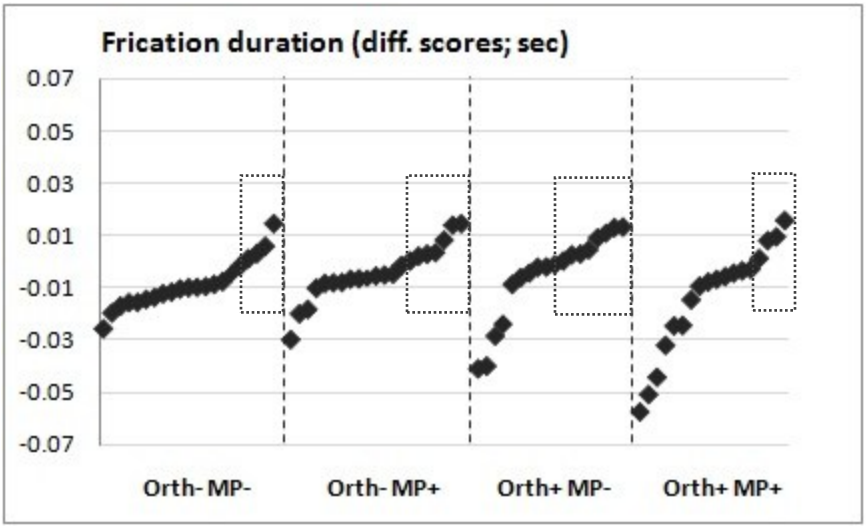


Figure 42: Frication duration in fricative-final disyllabic non-MPs - By-subject means.

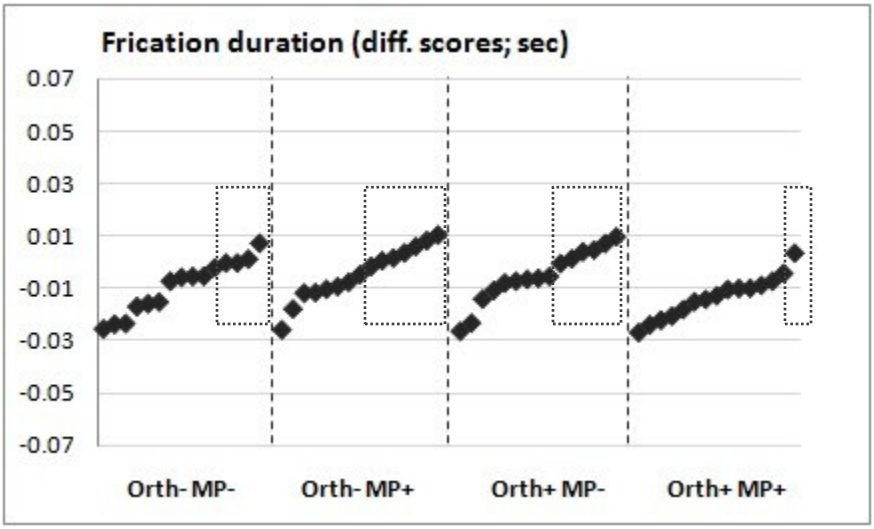


Figure 43: Frication duration in fricative-final disyllabic non-MPs - By-item means.

Results for the parameter of closure voicing are shown in Figure 44 (monosyllabic, by-subject), Figure 45 (monosyllabic, by-item), Figure 46 (disyllabic, by-subject) and Figure 47 (disyllabic, by-item). Scores above zero (outside of the boxes) denote more phonetic voicing for underlyingly voiced segments (which is expected).

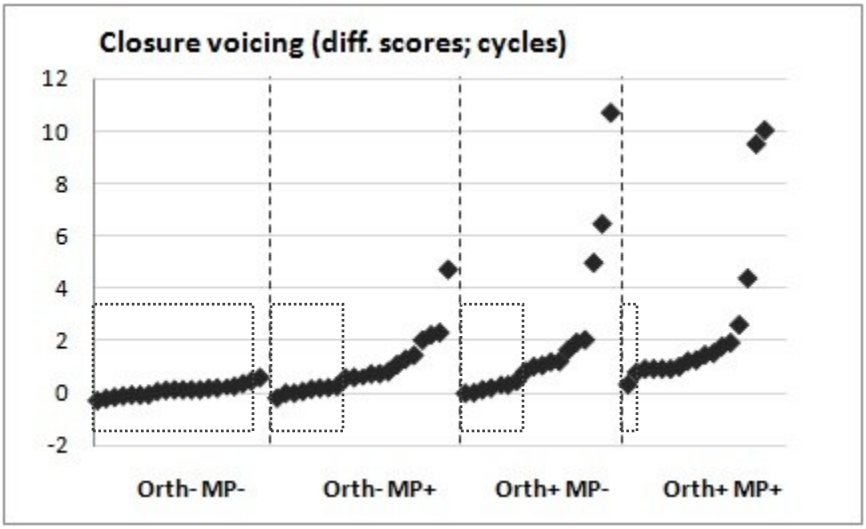


Figure 44: Closure voicing in fricative-final monosyllabic non-MPs - By-subject means.

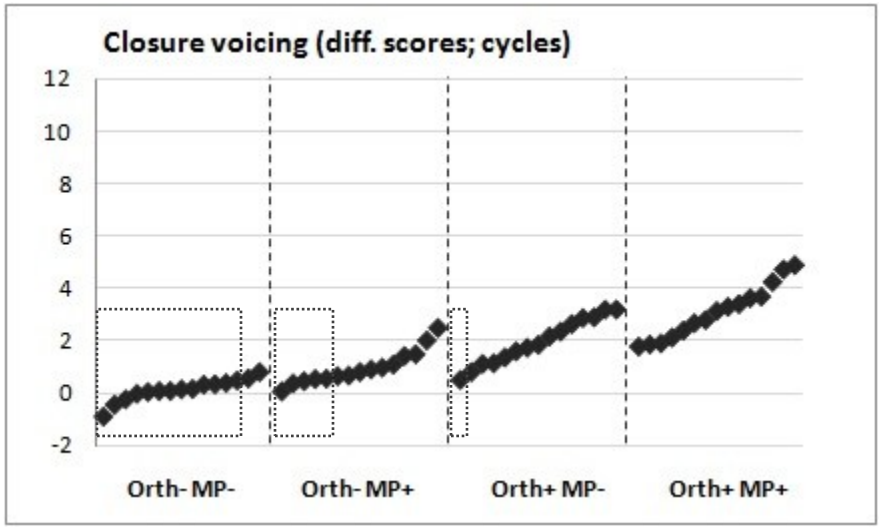


Figure 45: Closure voicing in fricative-final monosyllabic non-MPs - By-item means.

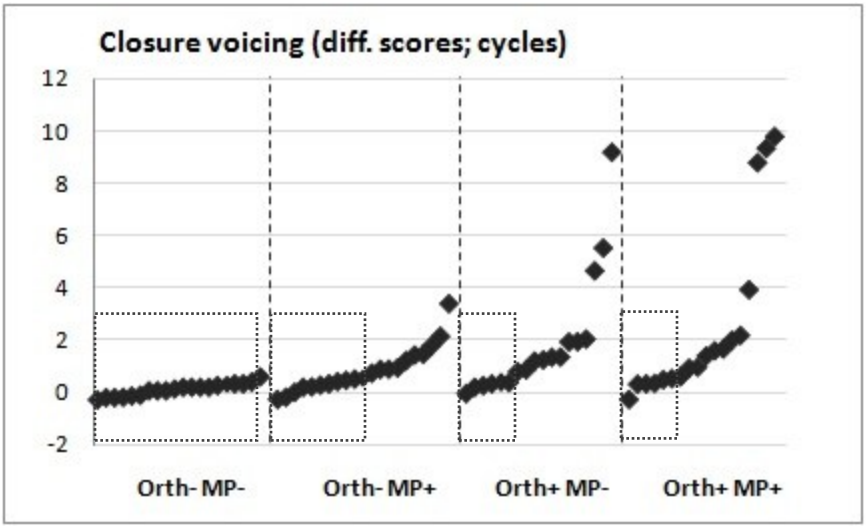


Figure 46: Closure voicing in fricative-final disyllabic non-MPs - By-subject means.

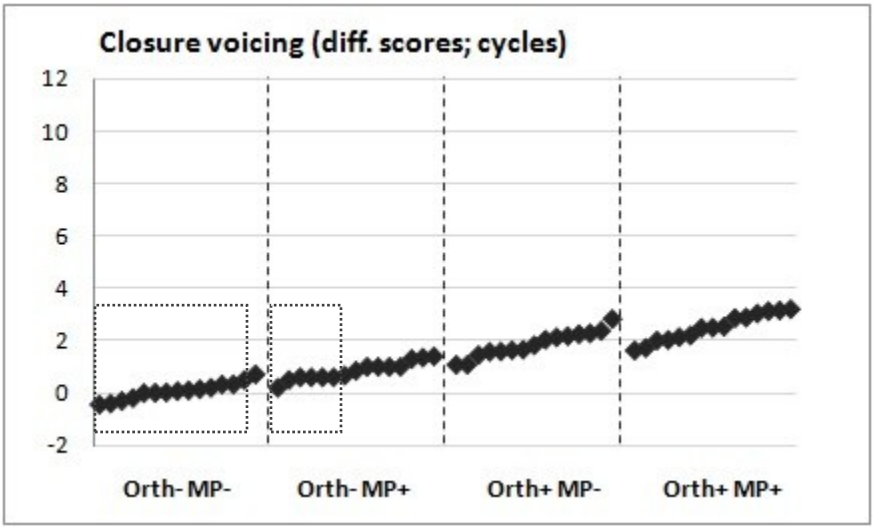


Figure 47: Closure voicing in fricative-final disyllabic non-MPs - By-item means.

3.4.3.3. Summary

For fricative-final monosyllabic and disyllabic words without minimal pair competitors in the lexicon (e.g., /nɔs/ 'nose', /arbuz/ 'watermelon'; */nɔz/ and */arbus/ are not existing words of Russian), group results show statistically significant differences in the parameters of frication duration and closure voicing. For frication duration, the period of high frequency noise is up to 13 ms longer for underlyingly voiceless final segments in both monosyllables and disyllables and the observed differences are independent of experimental group membership. For closure voicing, up to 1.52 cycles more glottal pulsing is seen for phonologically voiced fricatives in both monosyllabic and disyllabic words. However, closure voicing also shows an effect of task-dependent factors, such that significantly more glottal pulsing for phonologically voiced fricatives is found only when experimental methodology specifically favours preservation of the underlying contrast (i.e., in the Orth-MP+, Orth+MP- and Orth+MP+ groups but not the Orth-MP- group). For both monosyllabic and disyllabic tokens, preceding vowel duration shows only small and non-significant differences, and effects of consonantal place of articulation are also not significant.

Individual data reveal robust between-speaker and between-item variability in frication duration. Depending on word length in syllables, voiceless fricatives are up to 67 ms longer in as many as 86% of speakers and, depending on the experimental group, in 60% to 100% of item pairs. The opposite pattern of up to 27 ms longer frication noise for underlyingly voiced fricatives is seen in up to 27% of subjects and 40% of word pairs. Notably, variability in frication duration is found across all 4 experimental groups, with

many mean values also being low (e.g., 1 ms). Unlike frication duration, closure voicing shows a clear group-dependent asymmetry. More closure voicing for underlyingly voiced fricatives is observed in 48% to 95% of participants and 60% to 100% of item pairs in the Orth-MP+, Orth+MP- and Orth+MP+ groups. However, the same pattern is present in only 5% of speakers and no more than 13% of stimuli pairs in the Orth-MP- group. Section 3.4.4. below presents the results of the post-test questionnaire which examined the influence of study goal awareness on the acoustic differences found in production.

3.4.4. Post-Test Questionnaire

3.4.4.1. Results

The post-test questionnaire showed that the general goal of the study (i.e., investigating the voicing contrast in word-final obstruents) became known to one subject in the Orth-MP- group (5%), 4 subjects in the Orth-MP+ group (29%), 4 subjects in the Orth+MP- group (22%), and 14 speakers in the Orth+MP+ group (78%). When asked to illustrate the pattern that was being investigated, these participants stated that the study examined the pronunciation of word-final obstruents and/or provided examples of minimally contrasting item pairs from the stimulus list (e.g., /kɔt/ 'cat' ~ /kɔd/ 'code'). The remaining subjects indicated that the experiment aimed to investigate general knowledge of the Russian lexicon (14 speakers in the Orth-MP- group, 67%; 12 speakers in the Orth-MP+ group, 57%; one speaker in the Orth+MP+ group, 5%), dialectal variation in the pronunciation of words in general (11 speakers in the Orth+MP- group, 61%), pronunciation of unstressed vowels (one speaker in the Orth-MP- group, 5%),

reading skills (3 speakers in the Orth+MP- group, 17%; 3 speakers in the Orth+MP+ group, 17%) or did not identify any goals or patterns (5 participants in the Orth-MP- group, 24%; 3 speakers in the Orth-MP+ group, 14%).

Table 17 presents the findings of the regression analyses that aimed to predict the production differences in preceding vowel duration, closure/frication duration, release duration and closure voicing on the basis of speakers' awareness of the goal of the study. In the table, the values in brackets show the amount of variance in voicing-dependent differences that can be accounted for by study goal awareness.

Table 17: Effect of experimental goal awareness on acoustic difference scores - Results of regression analyses

Acoustic Measure	Plosive-final			Fricative-final	
	Monosyll. MPs	**Monosyll. non-MPs**	**Disyll. non-MPs**	**Monosyll. non-MPs**	**Disyll. non-MPs**
Vowel duration	-	-	-	-	-
Closure/frication duration	-	-	-	-	-
Release duration	* (6.4%)	-	-	-	-
Closure voicing	** (11.8%)	** (10.5%)	*** (15.9%)	** (8.5%)	* (6.3%)

* - significant @ .05 level
** - significant @ .01 level
*** - significant @ .001 level

As depicted in the table, speakers' awareness of the phonological contrast under investigation could not predict the difference scores in either closure/frication duration or

preceding vowel duration (all βs<.15, all $|t|$s<1.5, all R^2s<.02, all Fs<1, all ps>.1) but it could predict the scores for the parameters of closure voicing and release duration.

For release duration, study goal awareness predicted difference scores in plosive-final minimal pairs only (M_{aware}=10 ms, $M_{unaware}$=5 ms) and could account for 6.4% of the observed variance (minimal pairs: β=-.252, $t(76)$=-2.72, R^2=.064, $F(1, 76)$=5.16, p<.05; non-minimal pairs: all βs<.15, all $|t|$s<1.5, all R^2s<.02, all Fs<1, all ps>.1). For closure voicing, speakers' awareness of the goal of the study predicted between 6.3% and 15.9% of variance for all 5 stimuli types (plosive-final minimal pairs: β=.343, $t(76)$=3.18, R^2=.118, $F(1, 76)$=10.13, p<.01; plosive-final monosyllabic non-minimal pairs: β=.324, $t(76)$=2.99, R^2=.105, $F(1, 76)$=8.94, p<.01; plosive-final disyllabic non-minimal pairs: β=.398, $t(76)$=3.78, R^2=.159, $F(1, 76)$=14.33, p<.001; fricative-final monosyllabic non-minimal pairs: β=.292, $t(76)$=2.66, R^2=.085, $F(1, 76)$=7.06, p=.01; fricative-final disyllabic minimal pairs: β=.250, $t(76)$=3.19, R^2=.063, $F(1, 76)$=5.07, p<.05). Mean closure voicing difference scores across the two awareness conditions (aware, unaware) are shown in Figure 48 below. In the figure, 'm' and 'd' stand for 'monosyllabic' and 'disyllabic', respectively.

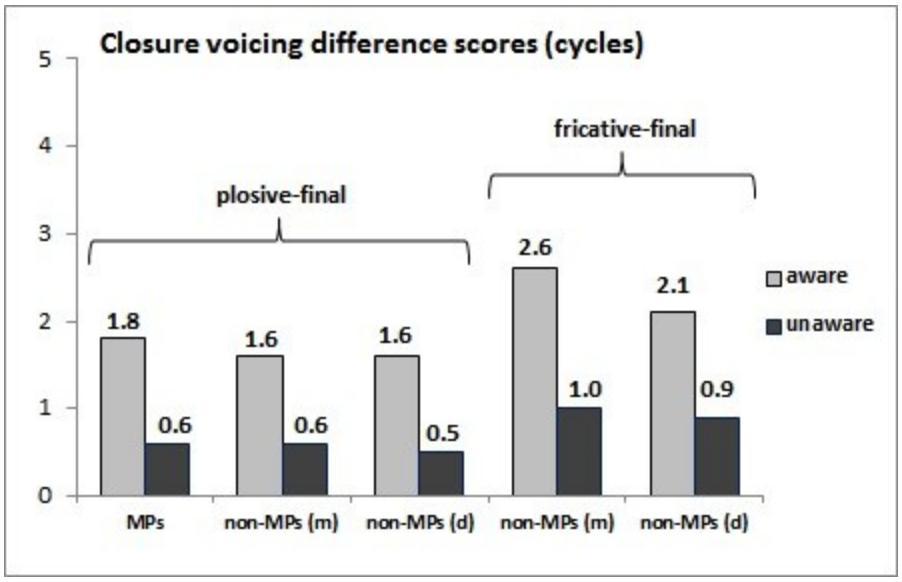


Figure 48: Effect of study goal awareness on closure voicing difference scores

3.4.4.2. Summary

Results of the post-test questionnaire show that the exact goal of the study became apparent to the majority of those participants who were exposed to orthographic forms as well as full minimal pairs during the recording session (i.e., 78% of the speakers in the Orth+MP+ group). In comparison, no more than 29% of speakers in the remaining 3 groups could correctly identify the topic of investigation, with the lowest number (5%, i.e. one participant) observed in the case of the Orth-MP- group that performed a non-reading task and did not encounter any minimal pairs among the stimuli. Furthermore, regression analyses reveal that speaker's awareness of final devoicing being the phenomenon of interest can significantly predict the overall magnitude of the production differences in the amount of glottal pulsing and, to a lesser extent, the voicing-dependent differences in release duration, although the observed R^2 scores are generally low.

3.4.5. General Summary

Overall, production results reveal the presence of statistically significant effects of underlying voicing on the acoustic parameter of glottal pulsing and, for a subset of stimuli, on closure/frication duration and release duration but not for the cue of preceding vowel duration. Differences in closure/frication duration and release duration are observed independently of the orthographic nature of the experimental task and regardless of whether or not full minimal pairs are included among the stimuli. Both parameters also show substantial individual-level variability in the degree and direction of the observed effects, with many speakers and experimental items demonstrating only small differences or unexpected patterns. Unlike the results for consonantal duration, differences in closure voicing are found at the group level only when experimental methodology specifically encourages preservation of the underlying contrast, and these differences are also robust and consistent at the individual level. Perceptual aspects of the voicing-dependent cues found in production are examined in Chapter 4 that follows. A detailed discussion of the experimental findings is given in Chapter 5.

CHAPTER 4: PERCEPTION STUDY

4.1. Introduction

Chapter 4 presents the perception module of the current doctoral research. The chapter describes the goals, research questions and hypotheses of the perception experiment (Section 4.2.), experimental design (Section 4.3.), and results of the investigation (Section 4.4.).

4.2. Goals, Research Questions and Hypotheses

4.2.1. Aims of the Study

Perceptual aspects of the voicing contrast in word-final obstruents in Russian are examined in the current thesis in a series of forced-choice two-alternative identification tasks. The general goal of the perception study is to determine how identification of intended voicing (rates of voiced responses and reaction times) varies across the stimuli produced with and without methodological factors favouring preservation of the underlying contrast.

4.2.2. Research Questions

The specific research questions that are addressed in the perception study include:

R1: *When identifying the intended voicing setting of word-final obstruents, do the rates of voiced responses and/or reaction times show statistically significant effects of underlying voicing?*

R2: *If effects of underlying voicing are attested, is there a significant asymmetry in the rates of voiced responses and/or reaction times between the tokens produced by the Orth-MP-, Orth-MP+, Orth+MP-, and Orth+MP+ groups?*

R3: *Are the effects seen in perception consistent across all stimuli types (i.e., found irrespective of the consonantal place and manner of articulation, word length in syllables, and availability of potentially homophonous lexical competitors)?*

R4: *Can the observed rates of voiced responses be predicted on the basis of the acoustic differences seen in production for the parameters of preceding vowel duration, closure/frication duration, release duration and/or closure voicing?*

4.2.3. Hypotheses

Perception findings are expected to show that Russian listeners show limited identification accuracy (a voiceless bias) and that their performance differs across the tokens elicited from the 4 production groups. Higher rates of voiced responses to underlyingly voiced final obstruents are anticipated for the items recorded during word-reading and/or with full minimal pairs present among the stimuli (i.e., the tokens with robust closure voicing cues). The same rates of voiced responses to phonologically voiced segments are hypothesized to be lower for the items elicited during picture-naming/word-guessing and without full minimal pairs included among the stimuli (i.e., the tokens with non-significant or limited phonetic voicing cues). When modeling the

observed rates of voiced responses, glottal pulsing is expected to be identified as the most prominent predictor variable across all stimuli types.

If perception findings are as expected, they will confirm that Russian listeners' identification responses are based primarily on the parameter of glottal pulsing, which is a task-dependent cue in production, and that identification of intended voicing is less consistent with phonological specifications in the absence of robust differences in phonetic voicing. These results will corroborate the claim that factors related to experimental methodology can affect not only the production of voicing cues in word-final obstruents but also the subsequent perception of the contrast.

4.3. Method

4.3.1. Participants and Location

A total of 216 native speakers of Russian took part in the perception study (84 male, 132 female; 18 to 37 years old; mean age of 22; none took part in the production study). All participants were undergraduate and graduate university students or postdoctoral researchers in Perm, Russia. They were recruited and tested at a university campus. None majored in linguistics or a related field, and all spoke the same local variety of Russian. The majority of participants (n=190) identified themselves as monolingual. The remaining speakers (n=26) were raised in bilingual households (Russian and Komi, Tatar or Udmurt) but reported using only the Russian language actively in their daily lives. All participants indicated having low to average proficiency

in at least one foreign language (English, French, German, Spanish). None of the subjects self-reported any hearing, visual, or other language-related problems.

4.3.2. Stimuli and Experimental Groups

Perceptual test items were created on the basis of the 150 stimuli words elicited in the production study. Acoustic data from all 78 speakers were included. For plosive-final monosyllabic minimal pairs, perceptual tokens corresponded to entire wordforms and final rhymes only (e.g., both /kɔt/ and /ɔt/ were used for /kɔt/ 'cat'). For all other items, test tokens were based on final rhymes (i.e., using /at/ from /brat/ 'brother', /uz/ from /arbuz/ 'watermelon'). Rhyme-only tokens always started with the steady-state part of the vowel. All stimuli were normalized for volume using an automated PRAAT script, and each item was coded with respect to which production group it originated from (i.e., Orth-MP-, Orth-MP+, Orth+MP-, Orth+MP+).

Considering the large number of tokens in the perception study (n=12,528), participants were divided into 4 experimental groups and, within each group, into 3 sub-groups (on the basis of consonantal place of articulation). Listeners assigned to Group 1 (n=54; 18 per place of articulation) were exposed to full wordforms representing plosives-final monosyllabic minimal pairs (e.g., listening to /kɔt/ 'cat', /kɔd/ 'code'). Subjects from Group 2 (n=54; 18 per place of articulation) listened to the same wordforms as Group 1 but heard final rhymes only (e.g., listening to /ɔt/ and /ɔd/ excised from /kɔt/ 'cat' and /kɔd/ 'code'). Participants in Group 3 (n=54; 18 per place of articulation) heard final rhymes of monosyllabic and disyllabic plosive-final non-minimal

pairs (e.g., listening to /at/ from /brat/ ‘brother’, /ad/ from /parad/ ‘parade’). Group 4 (n=54; 18 per place of articulation) heard final rhymes of fricative-final monosyllabic and disyllabic items (e.g., /uʒ/ from /muʒ/ ‘husband’, /ɔs/ from /kakɔs/ ‘coconut’). For the latter two groups, each participant heard half of all rhymes from monosyllabic and disyllabic wordforms, with the selection of tokens counter-balanced across subjects.

A general summary of experimental groups and test items is provided in Table 18.

Table 18: Experimental groups & stimuli items (perception) - Summary.

Group	Token type	Length	Final Cs	Num. of tokens	Num. of subjects
1	plosive-final MPs (entire words) (e.g., /kɔt/~/kɔd/)	monosyllabic	/p, b/	575	18
			/t, d/	573	18
			/k, g/	570	18
2	plosive-final MPs (final rhymes) (e.g., /ɔd/ from /kɔd/)	monosyllabic	/p, b/	575	18
			/t, d/	573	18
			/k, g/	570	18
3	plosive-final non-MPs (final rhymes) (e.g., /at/ from /brat/)	monosyllabic & disyllabic	/p, b/	1,512 (756 per subj.)	18
			/t, d/	1,520 (760 per subj.)	18
			/k, g/	1,516 (758 per subj.)	18
4	fricative-final non-MPs (final rhymes) (e.g., /uʒ/ from /muʒ/)	monosyllabic & disyllabic	/f, v/	1,504 (752 per subj.)	18
			/s, z/	1,526 (763 per subj.)	18
			/ʃ, ʒ/	1,514 (757 per subj.)	18
			Total:	12,528	216

4.3.3. Procedures

Participants performed a forced-choice two-alternative identification task. Stimuli items were delivered auditorily via headphones. One token was presented at a time, with two alternative orthographic representations of each stimulus displayed concurrently on a computer screen in front of the listener. The two choices always differed minimally with respect to the final consonant. Upon presentation of each token, subjects were instructed to indicate which of the two choices corresponded to the auditory stimulus by pressing one of two designated buttons on a computer keyboard. Response times and types of responses were recorded for analyses. The next token was presented 1 sec after a response was registered or after a 3 sec time-out if no response was given. Perceptual tokens were presented in blocks (n=5) and were grouped by the minimal pair or the rhyme. To mitigate the potential effects of presentation sequencing, the order of blocks was counter-balanced across participants and the order of stimuli items was randomized within each block. Subjects took short breaks throughout the experimental session which lasted approximately one hour. Prior to the start of the experiment, all participants provided informed consent and filled out a linguistic background questionnaire.

4.3.4. Instruments

The DMDX software package (Forster & Forster, 2003) was used to present the auditory stimuli and to record participants' responses (type of response, reaction time). Stimuli items were delivered binaurally via over-ear Sennheiser headphones. The two

arrow keys on a computer keyboard were used to register response type and latency. Assignment of keys was counter-balanced across subjects.

4.3.5. Data Treatment and Analyses

Only those responses that were given after the offset of the acoustic signal and were within 3 standard deviations of each participant's mean were retained for analyses. Mean rates of voiced responses (%-voiced) and response times were calculated and entered as dependent variables into a series of RM ANOVAs. Independent predictors are listed in (8) below, with factor levels specified inside brackets.

(8) *Predictor variables*

(i) UR-voice (=underlying voicing; voiced, voiceless);

(ii) C-place (=place of articulation; labial, coronal, dorsal);

(iii) P-group (=production group; Orth-MP-, Orth-MP+, Orth+MP-, Orth+MP+).

Analyses of variance were performed on pooled data using both by-subject and by-item designs. Only those differences that met the F_1 x F_2 Criterion were taken to be statistically significant. In F_1 analyses, UR-voice and P-group were within-subjects predictors. C-place was a between-subjects factor. In F_2 analyses, P-group was a within-items predictor. UR-voice and C-place were between-items independent factors. Separate 2 (UR-voice) x 3 (C-place) x 4 (P-group) analyses were conducted for each dependent variable and each token type using the IBM SPSS Statistics software package (Version 19; www.ibm.com). The ANOVAs tested for all main effects and interactions. When

applicable, pairwise comparisons (t-tests, 2-tailed) were used to resolve significant main effects and interactions and Greenhouse-Geisser or Bonferroni corrections were applied to account for multiple comparisons.

The relationship between participants' perceptual judgments and the acoustic measurements obtained in the production study was examined in a series of (stepwise) regressions. Regression analyses aimed to predict the mean rates of voiced responses on the basis of voicing-dependent differences in the acoustic parameters of preceding vowel duration, closure/frication duration, release duration, and closure voicing. Separate analyses were performed for each type of perceptual stimuli.

4.4. Results

Perception results are presented in the following order: (i) plosive-final minimal pairs (Section 4.4.1.), (ii) plosive-final non-minimal pair items (Section 4.4.2.), and (iii) fricative-final non-minimal pairs (Section 4.4.3.). Each section contains the findings for the rate of voiced responses, reaction time data, and a summary of the observed effects. Identification findings are followed by the results of regression modeling (Section 4.4.4.) and a general summary of the observed effects (Section 4.4.5.). Discussion of the findings is provided in Chapter 5.

4.4.1. Plosive-Final Minimal Pairs

Table 19 provides group averages and 95% confidence intervals for the rates of voiced responses and reaction times to plosive-final monosyllabic minimal pairs (listening to full wordforms and final rhymes). As reflected in the table, only the

differences in the rates of voiced responses (but not reaction times) are statistically significant.

Table 19: Rates of voiced responses and reaction times to full word tokens and final rhymes of plosive-final monosyllabic MPs - Group means & 95% confidence intervals.

Measure	Full wordforms			Final rhymes		
	Voiced	Voiceless	diff.	Voiced	Voiceless	diff.
%-voiced	44.0 (41.1-46.9)	27.7 (24.4-31.1)	16.3***	43.0 (39.6-46.4)	27.1 (23.2-31.1)	15.9***
RT (ms)	479 (97-104)	481 (99-106)	-2	495 (460-530)	486 (452-520)	9

* - significant @ .05 level
** - significant @ .01 level
*** - significant @ .001 level

4.4.1.1. Rate of Voiced Responses

Table 20 provides a summary of the significant ANOVA results for the rates of voiced responses to plosive-final monosyllabic minimal pairs.[36]

[36] The table shows significant main effects and interactions involving UR-voice. Detailed results are provided in Appendix 9.

Table 20: Rates of voiced responses to full word tokens and final rhymes of plosive-final monosyllabic MPs - Summary of the significant ANOVA results.

Factor	Full wordforms				Final rhymes			
	By-Subject		By-Item		By-Subject		By-Item	
	F_1	p	F_2	p	F_1	p	F_2	p
UR-voice	660.09	.000	173.68	.000	465.23	.000	197.63	.000
C-place	-	-	4.26	.026	-	-	7.36	.003
P-group	15.25	.000	3.85	.016	17.12	.000	5.65	.002
UR-voice x C-place	45.16	.000	12.12	.000	37.82	.000	16.87	.000
UR-voice x P-group	66.13	.000	13.23	.000	43.40	.000	13.98	.000
UR-voice x C-place x P-group	7.81	.000	-	-	-	-	-	-

As noted in Table 20, a significant main effect of UR-voice was observed for both full-word and rhyme-only stimuli items (full words: $F_1(1,51)=660.09$, $p<.001$; $F_2(1,24)=173.68$, $p<.001$; rhymes: $F_1(1,51)=465.23$, $p<.001$; $F_2(1,24)=197.63$, $p<.001$). A main effect of P-group was also found for the two token types (full words: $F_1(1,51)=15.25$, $p<.001$; $F_2(1,24)=3.85$, $p=.016$; rhymes: ($F_1(1,51)=17.12$, $p<.001$; $F_2(1,24)=5.65$, $p=.002$), and pairwise comparisons revealed that responses to the tokens produced by the Orth+MP+ group were significantly different from responses to the items elicited from all other production groups (all ps$<.05$).

Furthermore, a significant UR-voice x P-group interaction was also found for both stimuli types (full words: $F_1(1,51)=66.13$, $p<.001$; $F_2(1,24)=13.23$, $p<.001$; rhymes: $F_1(1,51)=43.40$, $p<.001$; $F_2(1,24)=13.98$, $p<.001$). Pairwise comparisons showed that all within-group differences in the rates of voiced responses to voiced versus voiceless

obstruents were significant (all *p*s<.001). For full words, significantly fewer voiced responses to items with phonologically voiceless final plosives were also found in the case of the tokens produced by the Orth+ groups compared to the Orth- groups (all *p*s<.01). For rhyme-only stimuli, the rate of voiced responses for voiceless final stops was significantly higher for the items elicited from the Orth-MP- group compared to all other groups (all *p*s<.01). For both types of stimuli items, between-group differences in the rates of voiced responses to tokens with underlyingly voiced final stops were significantly different (all *p*s<.05) with the exception of the Orth-MP- and Orth-MP+ pair (*p*>.1). By-group means of voiced responses are shown in Figure 49 (full words) and Figure 50 (final rhymes).

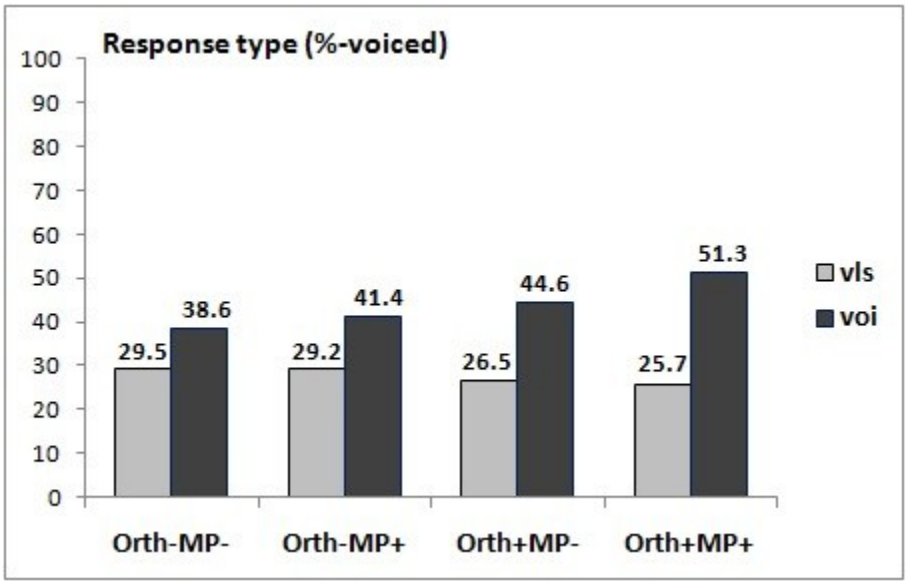


Figure 49: Rates of voiced responses to full word tokens of plosive-final MPs - By-group (production) means.

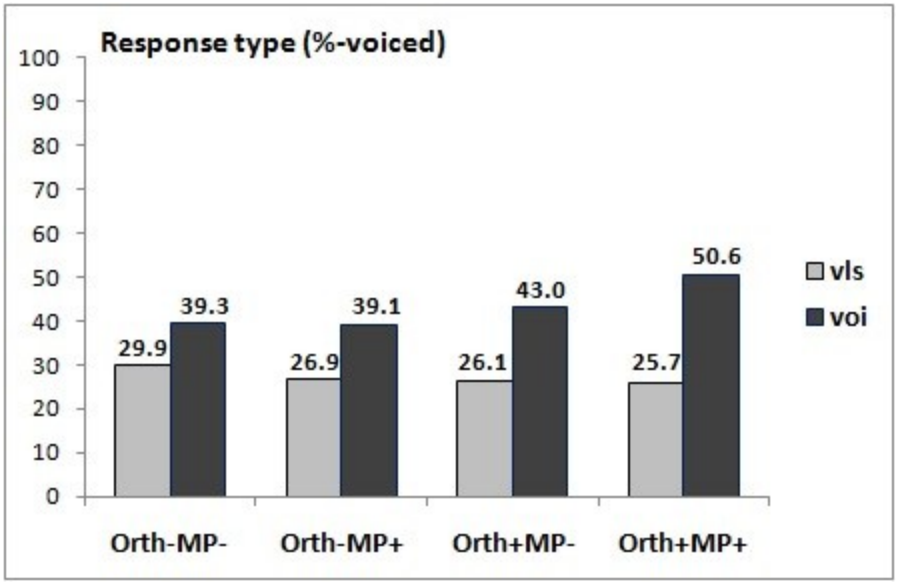


Figure 50: Rates of voiced responses to final rhymes of plosive-final MPs - By-group (production) means.

For either full or truncated perceptual tokens, a main effect of C-place was observed only in by-item analyses (all F_1s<1, all *p*s >.1; all F_2s>4, all *p*s<.05). A significant UR-voice x C-Place x P-group interaction was also found for full word tokens only in the by-subject ANOVA (F_1 >7, *p*<.001; F_2<2, *p*>.1). However, a significant interaction of UR-voice and C-place was observed in both F_1 and F_2 analyses for full words as well as final rhymes (full words: $F_1(2,51)=45.16$, *p*<.001; $F_2(2,24)=12.12$, *p*=.005; rhymes: $F_1(2,51)=37.82$, *p*<.001; $F_2(2,24)=13.98$, *p*<.001). For both stimuli types, pairwise comparisons indicated that differences in responses to underlyingly voiced versus voiceless final obstruents were significant within each place of articulation (all *p*s<.001). At the same time, perceptual items with voiceless velar plosives received significantly fewer voiced responses than tokens ending in labial or coronal stops (all *p*s<.001). More tokens with underlyingly voiced final stops were also attributed to the

voiced category in the case of velars than labials (all *p*s<.01). Mean values of voiced responses across the 3 places of articulation are shown in Figure 51 (full words) and Figure 52 (final rhymes).

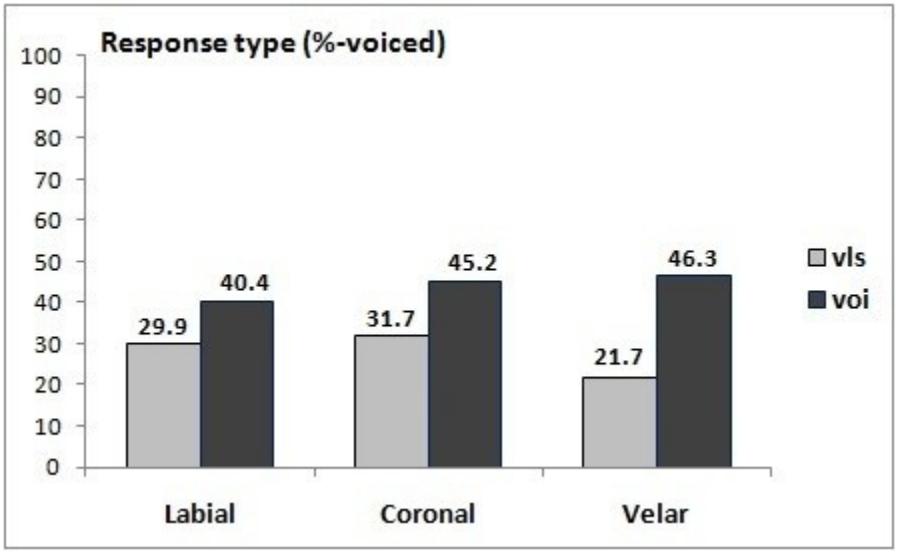


Figure 51: Rates of voiced responses to full word tokens of plosive-final MPs - By-place means.

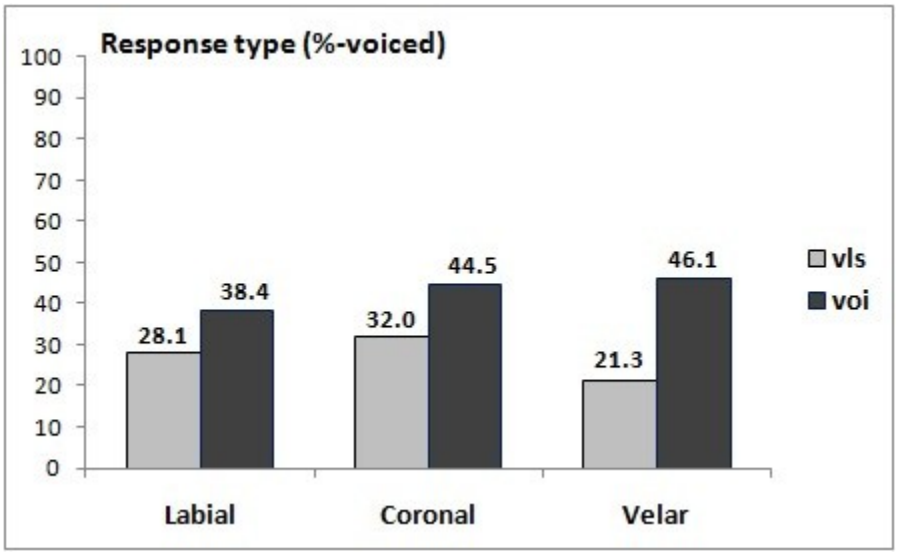


Figure 52: Rates of voiced responses to final rhymes of plosive-final MPs - By-place means.

The remaining interactions were not significant according to the F_1 x F_2 Criterion (all F_2s<3, all ps >.05).

4.4.1.2. Reaction Times

Significant ANOVA results for reaction times to plosive-final monosyllabic minimal pairs are listed in Table 21.[37]

Table 21: Reaction times to final rhymes of plosive-final monosyllabic MPs - Summary of the significant ANOVA results.

Factor	Full wordforms				Final rhymes			
	By-Subject		By-Item		By-Subject		By-Item	
	F_1	p	F_2	p	F_1	p	F_2	p
UR-voice	-	-	-	-	8.54	.005	-	-
C-place	-	-	23.57	.000	-	-	13.58	.000
P-Group	7.30	.000	-	-	12.40	.000	5.31	.005
UR-voice x C-place	3.89	.027	-	-	-	-	-	-

As can be seen in the table above, a main effect of UR-voice was not significant for full word items (all Fs<1, all ps>.1) and was observed for final rhymes only in the by-subject data (F_1>8, p<.01; F_2<2, p>.1). A main effect of P-group was significant for final rhymes only (final rhymes: $F_1(1,51)=12.40$, p<.001; $F_2(1,24)=5.31$, p=.005; full words: F_1>7, p<.001; F_2<3, p>.05). Pairwise comparisons revealed that responses to the rhyme-only tokens produced by the two Orth- groups were slower than responses to the items from

[37] The table lists significant main effects and interactions involving UR-voice. Detailed results are provided in Appendix 9.

the Orth+MP+ group (all *p*s<.05). Mean reaction times across the 4 production groups are shown in Table 22 (listening to final rhymes).

Table 22: Reaction times to final rhymes of plosive-final monosyllabic MPs - By-group means & 95% confidence intervals.

Measure	Production group			
	Orth-MP-	Orth-MP+	Orth+MP-	Orth+MP+
RT (ms)	499 (464-535)	494 (460-529)	486 (451-521)	482 (448-516)

Main effect of C-place was observed for reaction times only in by-item analyses for either full or truncated tokens (all F_1s<3, all *p*s >.05; all F_2s>13, all *p*s<.001). C-place interacted with UR-voice for full word items only in the by-subject ANOVA (F_1>3, p<.05; F_2<1, p>.1). None of the remaining interactions met the F_1 x F_2 Criterion (all F_2s<3, all *p*s >.05).

4.4.1.3. Summary

Overall, when listening to perceptual tokens created on the basis of plosive-final monosyllabic minimal pairs (e.g., /kɔt/ 'cat', /kɔd/ 'code'), speakers of Russian attributed 27~28% of all items ending in voiceless stops and 43~44% of tokens with voiced final plosives to the voiced category. However, full-word and rhyme-only tokens produced by the two Orth+ groups received 3~4% fewer voiced responses to underlyingly voiceless final segments and 11~13% more voiced responses to phonologically voiced final stops compared to the items produced by the two Orth- groups. For both full and truncated

stimuli, the highest rate of voiced responses for items with voiced final stops (51%) is found when the tokens come from the Orth+MP+ production group. In addition, items ending in velar plosives show higher rates of voiced responses to underlyingly voiced stops and lower rates of voiced responses to phonologically voiceless plosives, and up to 18 ms longer reaction times are observed for both voiced and voiceless tokens for the rhyme-only items produced by the two Orth- groups compared to the stimuli from the Orth+MP+ group. Section 4.4.2. below shows that similar results were obtained for perceptual tokens excised from plosive-final monosyllabic and disyllabic non-minimal pair items.

4.4.2. Plosive-Final Non-Minimal Pairs

Group means and 95% confidence intervals for the rates of voiced responses and reaction times to plosive-final monosyllabic and disyllabic non-minimal pairs are shown in Table 23. As noted in the table, differences in response types but not response latencies are statistically significant.

Table 23: Rates of voiced responses & reaction times to final rhymes of plosive-final non-MPs - Group means & 95% confidence intervals.

Measure	Monosyllabic			Disyllabic		
	Voiced	Voiceless	diff.	Voiced	Voiceless	diff.
%-voiced	43.4 (40.0-46.8)	26.4 (24.4-31.1)	17***	42.5 (39.0-46.0)	27.7 (23.4-31.9)	14.8***
RT (ms)	491 (457-526)	486 (451-521)	5	490 (455-526)	486 (453-519)	4

* - significant @ .05 level
** - significant @ .01 level
*** - significant @ .001 level

4.4.2.1. Rate of Voiced Responses

Table 24 summarizes the significant ANOVA results for the rates of voiced responses to final rhymes of plosive-final non-minimal pairs.[38]

Table 24: Rates of voiced responses to final rhymes of plosive-final non-MPs - Summary of the significant ANOVA results.

Factor	Monosyllabic				Disyllabic			
	By-Subject		By-Item		By-Subject		By-Item	
	F_1	p	F_2	p	F_1	p	F_2	p
UR-voice	393.60	.000	174.08	.000	253.18	.000	94.95	.000
C-place	-	-	5.14	.014	-	-	6.03	.008
P-Group	7.37	.000	3.61	.024	12.14	.000	5.76	.002
UR-voice x C-place	27.61	.000	11.95	.000	23.15	.000	8.28	.002
UR-voice x P-group	18.28	.000	9.57	.000	25.40	.000	12.12	.000

As can be seen above, analyses of variance revealed a significant main effect of UR-voice for the tokens excised from both monosyllables and disyllables (monosyllabic: $F_1(1,51)=393.60$, $p<.001$; $F_2(1,24)=174.08$, $p<.001$; disyllabic: $F_1(1,51)=253.18$, $p<.001$; $F_2(1,24)=94.95$, $p<.001$). A main effect of P-group was also found for both types of stimuli items (monosyllabic: $F_1(1,51)=7.37$, $p<.001$; $F_2(1,24)=3.61$, $p=.024$; disyllabic: $F_1(1,51)=12.14$, $p<.001$; $F_2(1,24)=5.76$, $p=.002$), and pairwise comparisons showed that responses to the Orth+MP+ tokens were significantly different from responses to test items from all other production groups (all ps$<.05$).

[38] The table shows significant main effects and interactions involving UR-voice. Detailed results are provided in Appendix 10 (monosyllables) and Appendix 11 (disyllables).

In addition, a significant UR-voice x P-group interaction was present for both types of stimuli (monosyllabic: $F_1(1,51)=18.28$, $p<.001$; $F_2(1,24)=9.57$, $p<.001$; disyllabic: $F_1(1,51)=25.40$, $p<.001$; $F_2(1,24)=12.12$, $p<.001$). Pairwise comparisons revealed that all within-group differences in the rates of voiced responses to voiced versus voiceless categories were significant (all *p*s<.001). At the same time, significantly more voiced responses to underlyingly voiceless final plosives were found for monosyllable-based tokens from the Orth-MP- production group compared to all other groups (all *p*s<.01) and for disyllable-based items from the Orth-MP- group compared to the Orth+MP+ group ($p<.01$). For items ending in phonologically voiced stops, more voiced responses were observed in the case of monosyllable-based rhymes produced by the Orth+MP+ group compared to any other group (all *p*s<.01). The same difference was significant for the final rhymes of disyllables in all pairwise comparisons (all *p*s<.05) except the Orth-MP- and Orth-MP+ pair ($p>.1$). Mean rates of voiced responses across the 4 production groups are shown in Figure 53 (monosyllables) and Figure 54 (disyllables).

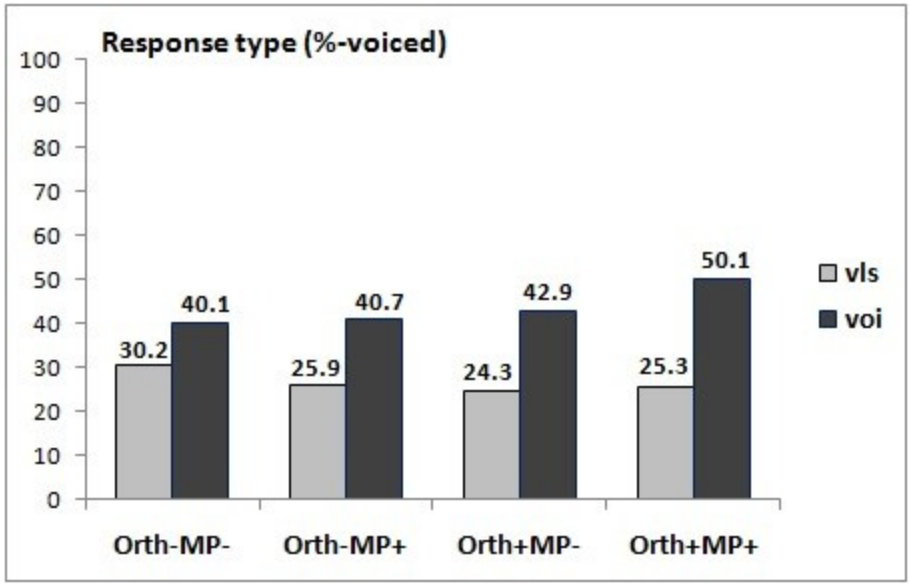


Figure 53: Rates of voiced responses to final rhymes of plosive-final monosyllabic non-MPs - By-group (production) means.

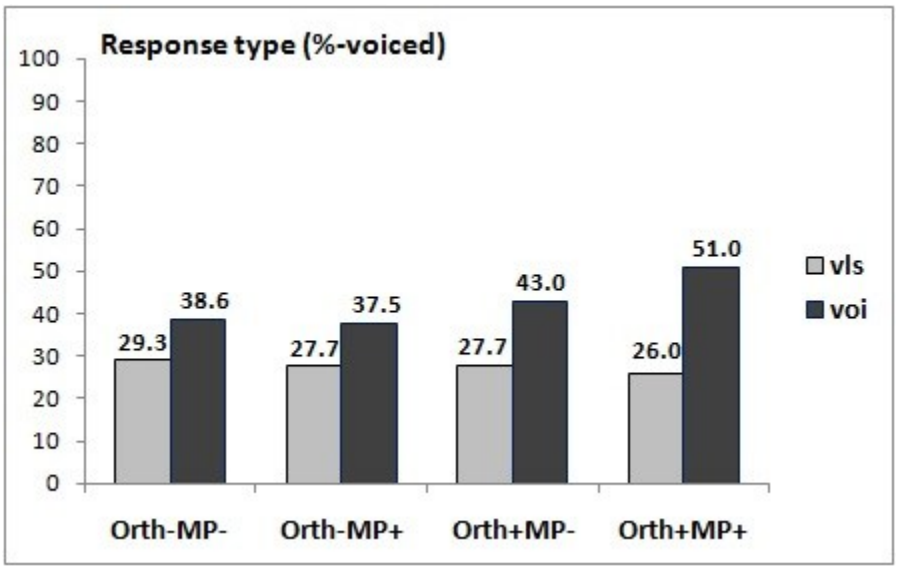


Figure 54: Rates of voiced responses to final rhymes of plosive-final disyllabic non-MPs - By-group (production) means.

Main effect of C-place was significant for response types only in by-item analyses regardless of whether the tokens were created from monosyllables or disyllables (all F_1s<1, all ps >.1; all F_2s>5, all ps<.05). However, the UR-voice x C-place interaction was significant for both types of stimuli items (monosyllables: $F_1(2,51)=27.61$, $p<.001$; $F_2(2,24)=11.95$, $p=.005$; disyllables: $F_1(2,51)=23.15$, $p<.001$; $F_2(2,24)=8.28$, $p=.002$). Pairwise comparisons showed that the differences in responses to underlyingly voiced versus voiceless final plosives were significant within each place of articulation (all ps<.001), yet voiceless velar stops received significantly fewer voiced responses than voiceless labials or voiceless coronals (all ps<.01). More underlyingly voiced velar plosives were also classified as voiced compared to labial stops (all ps<.01). Mean values of voiced responses across the 3 places of articulation are provided in Figure 55 (monosyllables) and Figure 56 (disyllables).

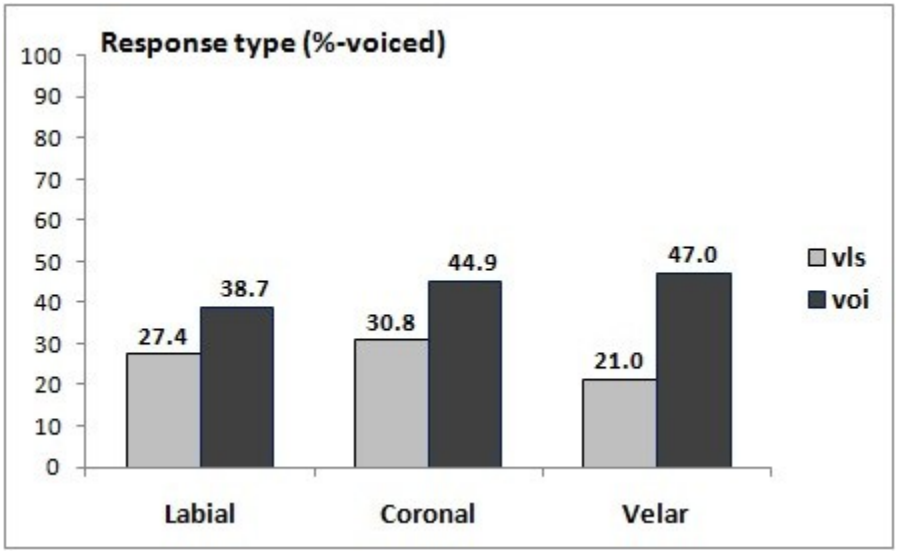


Figure 55: Rates of voiced responses to final rhymes of plosive-final monosyllabic non-MPs - By-place means.

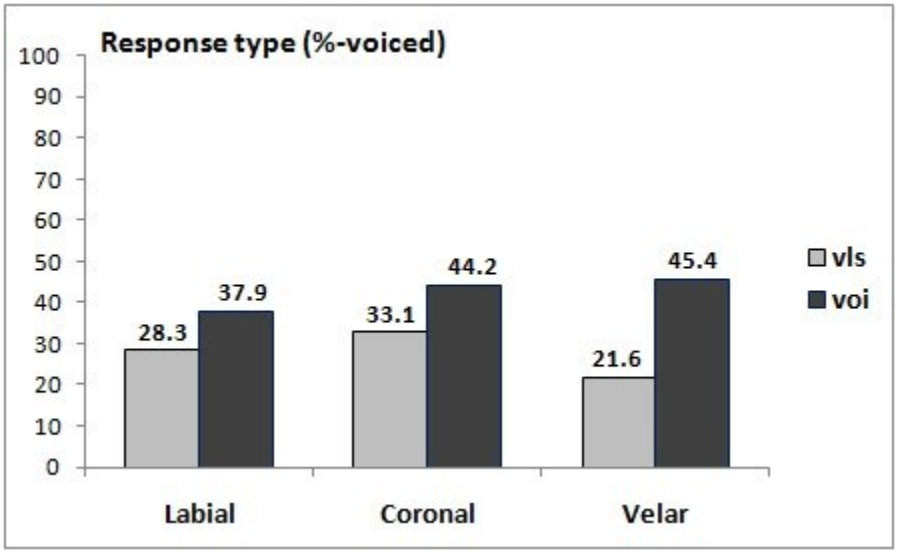


Figure 56: Rates of voiced responses to final rhymes of plosive-final disyllabic non-MPs - By-place means.

The rest of the interactions did not meet the criterion of statistical significance (all F_2s<3, all *p*s >.05).

4.4.2.2. Reaction Times

For the dependent variable of reaction times, results of the analyses of variance are summarized in Table 25.[39]

Table 25: Reaction times to final rhymes of plosive-final non-MPs - Summary of the significant ANOVA results.

Factor	Monosyllabic				Disyllabic			
	By-Subject		By-Item		By-Subject		By-Item	
	F_1	*p*	F_2	*p*	F_1	*p*	F_2	*p*
C-place	-	-	11.46	.000	-	-	20.94	.000
P-Group	7.20	.000	5.35	.005	-	-	-	-

[39] Table 25 shows significant main effects and interactions with UR-voice. Detailed results are provided in Appendix 10 (monosyllables) and Appendix 11 (disyllables).

As reflected in the table, a main effect of UR-voice was not observed for either monosyllable-based or disyllable-based rhymes (all Fs<2, all ps>.1). A main effect of P-group was found for final rhymes of monosyllables only ($F_1(2,51)=7.20$, $p<.001$; $F_2(2,24)=5.35$, $p=.005$). Pairwise comparisons showed that responses to the tokens produced by the two Orth- production groups were significantly slower than responses to the items elicited from the Orth+MP+ group (all ps<.01). Mean reaction times across the 4 production groups are shown in Table 26 (listening to tokens excised from monosyllabic words).

Table 26: Reaction times to final rhymes of plosive-final monosyllabic non-MPs - By-group means & 95% confidence intervals.

Measure	Production group			
	Orth- MP-	**Orth- MP+**	**Orth+ MP-**	**Orth+ MP+**
RT (ms)	501 (488-514)	491 (481-502)	486 (471-500)	477 (465-488)

Main effect of C-place was observed for reaction times only in F_1 analyses for the tokens created from either monosyllables or disyllables (all F_1s<1, all ps >.1; all F_2s>5, all ps<.05). None of the remaining interactions were significant in both by-subject and by-item data (all F_2s<3, all ps >.05).

4.4.2.3. Summary

For rhyme-only tokens excised from plosive-final monosyllabic and disyllabic non-minimal pairs (e.g., /at/ from /brat/ 'brother', /ad/ from /parad/ 'parade'),

identification results show that the overall rates of voiced responses are around 26~28% for tokens ending in phonologically voiceless stops and 43~44% for items with underlyingly voiced final plosives. Crucially, up to 5% more underlyingly voiceless stops are attributed to the voiced category in the data from the Orth-MP- group compared to the items produced by all other groups. For phonologically voiced final plosives, 2~5% fewer tokens are identified as voiced in the stimuli from the two Orth- production groups compared to the two Orth+ groups. The highest rates of voiced responses to tokens with underlyingly voiced final stops (50~51%) are seen for the stimuli elicited from the Orth+MP+ group. Additionally, items ending in velar plosives show higher rates of voiced responses to voiced stops and lower rates of voiced responses to voiceless plosives. Independently of underlying voicing, up to 24 ms longer listening times are seen for the stimuli produced by the two Orth- groups compared to the items elicited from the Orth+MP+ group. As shown in Section 4.4.3. that follows, comparable findings were also obtained for fricative-final stimuli items.

4.4.3. Fricative-Final Non-Minimal Pairs

Table 27 provides mean values and 95% confidence intervals for the rates of voiced responses and reaction times to final rhymes of fricative-final monosyllabic and disyllabic non-minimal pairs. As reflected in the table, only the differences in the rates of voiced responses are statistically significant.

Table 27: Rates of voiced responses & reaction times to final rhymes of fricative-final non-MPs - Group means & 95% confidence intervals.

Measure	Monosyllabic			Disyllabic		
	Voiced	Voiceless	diff.	Voiced	Voiceless	diff.
%-voiced	42.7 (40.2-45.2)	30.3 (27.3-33.2)	12.4***	42.4 (39.7-45.2)	29.8 (27.2-32.4)	12.6***
RT (ms)	477 (446-509)	474 (440-507)	3	478 (446-511)	472 (438-506)	6

* - significant @ .05 level
** - significant @ .01 level
*** - significant @ .001 level

4.4.3.1. Rate of Voiced Responses

Significant ANOVA results for the rate of voiced responses to monosyllabic and disyllabic fricative-final tokens are shown in Table 28.[40]

Table 28: Rates of voiced responses to final rhymes of fricative-final non-MPs - Summary of the significant ANOVA results.

Factor	Monosyllabic				Disyllabic			
	By-Subject		By-Item		By-Subject		By-Item	
	F_1	p	F_2	p	F_1	p	F_2	p
UR-voice	241.23	.000	38.60	.000	232.28	.000	32.51	.000
P-Group	2.82	.044	-	-	-	-	-	-
UR-voice x C-place	18.23	.000	-	-	20.145	.000	-	-
UR-voice x P-group	29.53	.000	10.45	.000	24.64	.000	10.30	.000
UR-voice x C-place x P-group	4.04	.001	-	-	7.20	.000	3.52	.007

[40] Table 28 lists significant main effects and interactions involving UR-voice. Detailed results are provided in Appendix 12 (monosyllables) and Appendix 13 (disyllables).

As can be seen in the table, a significant main effect of UR-voice was found for the tokens excised from monosyllabic as well as disyllabic wordforms (monosyllables: $F_1(1,51)=241.23$, $p<.001$; $F_2(1,24)=38.60$, $p<.001$; disyllables: $F_1(1,51)=232.28$, $p<.001$; $F_2(1,24)=32.51$, $p<.001$). A main effect of P-group was observed for monosyllable-based rhymes only in the by-item analysis ($F_1>2.8$, $p<.05$; $F_2<2$, $p>.1$) and was never significant for disyllable-based tokens ($F_1<2$, $p>.1$; $F_2<1$, $p>.1$).

Moreover, a significant UR-voice x P-group interaction was present for final rhymes of monosyllables as well as disyllables (monosyllables: $F_1(1,51)=29.53$, $p<.001$; $F_2(1,24)=10.45$, $p<.001$; disyllables: $F_1(1,51)=24.64$, $p<.001$; $F_2(1,24)=10.30$, $p<.001$). Pairwise comparisons demonstrated that all between-group differences in the rates of voiced responses to phonologically voiced versus voiceless fricatives were significant (all *p*s<.001). At the same time, significantly fewer voiced responses to underlyingly voiceless final fricatives were found for monosyllable-based tokens produced by the Orth+MP+ group compared to all other groups (all *p*s<.05). Furthermore, monosyllable-based rhymes showed significantly different rates of voiced responses to phonologically voiced final segments across all production groups (all *p*s<.05) except the Orth+MP- and Orth+MP+ pair ($p>.1$). For disyllable-based items, the observed rates of voiced responses to phonologically voiceless final stops were significantly different from each other in all pairwise comparisons (all *p*s<.05) with the exception of the Orth-MP- and Orth+MP- pair ($p>.1$). For the same tokens, differences in the rates of voiced responses to phonologically voiced final fricatives were also significant in all pairwise comparisons (all *p*s<.05) other

than the Orth-MP- and Orth-MP+ pair (p>.1). Mean values of voiced responses across the 4 production groups are shown in Figure 57 (monosyllables) and Figure 58 (disyllables).

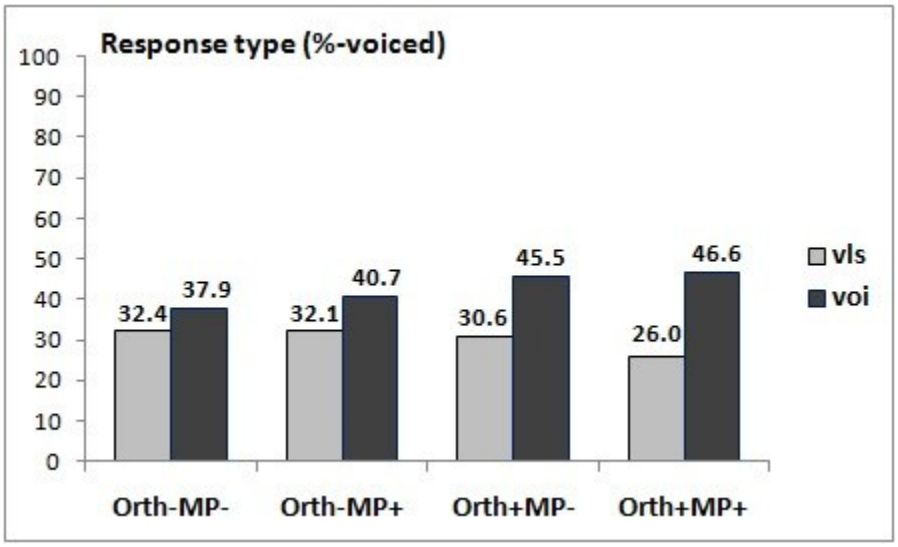


Figure 57: Rates of voiced responses to final rhymes of fricative-final monosyllabic non-MPs - By-group (production) means.

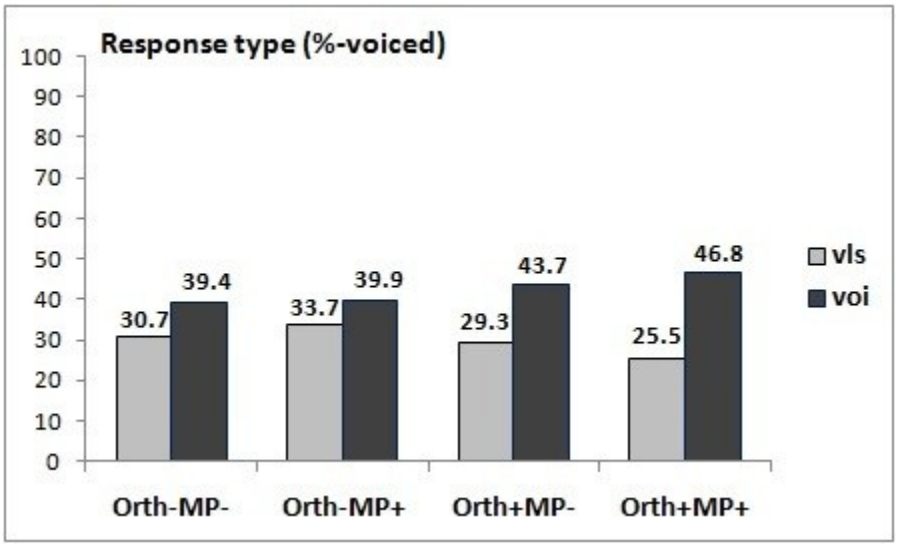


Figure 58: Rates of voiced responses to final rhymes of fricative-final disyllabic non-MPs - By-group (production) means.

For either monosyllable-based or disyllable-based items, a main effect of C-place was not found in any of the analyses (all *F*s<1, all *p*s>.1). A significant C-place x P-group interaction was observed only in the case of perceptual items excised from disyllabic words ($F_1(2,51)=7.20$, $p<.001$; $F_2(2,24)=3.52$, $p=.007$).[41] The remaining main effects and interactions were not significant according to the F_1 x F_2 Criterion (all F_2s<3, all *p*s >.05).

4.4.3.2. Reaction Times

Significant ANOVA results for the dependent variable of reaction times to fricative-final tokens are shown in Table 29.[42]

Table 29: Reaction times to final rhymes of fricative-final non-MPs - Summary of the significant ANOVA results.

Factor	Monosyllabic				Disyllabic			
	By-Subject		By-Item		By-Subject		By-Item	
	F_1	*p*	F_2	*p*	F_1	*p*	F_2	*p*
C-place	-	-	3.91	.034	-	-	-	-
P-Group	-	-	-	-	2.89	.042	-	-
UR-voice x P-group	-	-	-	-	3.62	.016	-	-

As depicted in the table, a main effect of UR-voice was not observed regardless of whether the tokens were excised from monosyllabic or disyllabic items (all *F*s<2.5, all *p*s>.1). A main effect of P-group and a significant UR-voice x P-group interaction were

[41] The significance of the interaction was due to the fact that coronal-final tokens produced by the Orth-groups received fewer voiced responses regardless of underlying voicing (all *p*s<.01). Since this asymmetry is not dependent on underlying voicing, it is outside of the scope of the present investigation.
[42] The table lists significant main effects and interactions involving UR-voice. Detailed results are provided in Appendix 12 (monosyllables) and Appendix 13 (disyllables).

found for disyllable-based tokens only in the by-subject ANOVAs (all F_1s>2.8, all ps<.05; all F_2s<2.6, all ps>.05). The C-place factor showed a significant main effect for final rhymes of monosyllables only in the by-item test (F_1<1, p>.1; F_2>3, p<.05). None of the remaining main effects and interactions were significant in both F_1 and F_2 analyses (all F_1s<3, all ps>.1).

4.4.3.3. Summary

For perceptual tokens excised from fricative-final monosyllabic and disyllabic non-minimal pairs (e.g., /af/ from /graf/ 'count', /uz/ from /arbuz/ 'watermelon'), the overall rates of voiced responses are around 30% for phonologically voiceless fricatives and 42~43% for underlyingly voiced consonants. At the same time, similarly to plosive-final items, 4~9% fewer tokens with phonologically voiceless segments are attributed to the voiced category in the data from the Orth+MP+ group, and up to 9% fewer tokens containing phonologically voiced final fricatives are identified as voiced for the tokens produced by the two Orth- groups. The highest rate of voiced responses to tokens with underlyingly voiced final fricatives (47%) and the lowest rate of voiced responses to rhymes ending in phonologically voiceless consonants (25~26%) are observed in the case of final rhymes produced by the Orth+MP+ group. Unlike plosive-final stimuli, items ending in fricatives do not show any place-dependent asymmetries, and no significant effects are observed for reaction time latencies. Section 4.4.4. below presents the findings of a series of regression analyses which investigated whether the observed rates of voiced responses could be predicted on the basis of the acoustic differences found in production.

4.4.4. Regression Modeling

4.4.4.1. Results

Results of regression modeling are presented in Table 30. In the table, perception findings are given alongside the production results for the same acoustic parameters.

Table 30: Effect of acoustic cues on mean rates of voiced responses - Results of regression analyses

Acoustic Measure		Plosive-final				Fricative-final	
		Monosyll. MPs (full words)	**Monosyll. MPs (final rhymes)**	**Monosyll. non-MPs (final rhymes)**	**Disyll. non-MPs (final rhymes)**	**Monosyll. non-MPs (final rhymes)**	**Disyll. non-MPs (final rhymes)**
Vowel duration	Perception:	* (1.7%)	-	-	-	-	-
	Production:	- (4 ms)	- (4 ms)	- (2 ms)	- (3 ms)	- (3 ms)	- (3 ms)
Closure/ frication duration	Perception:	-	-	-	-	-	-
	Production:	** (-7 ms)	** (-7 ms)	- (-2 ms)	- (-1 ms)	*** (13 ms)	*** (8 ms)
Release duration	Perception:	*** (7.6%)	*** (6.8%)	*** (5.8)	-	n/a	n/a
	Production:	** (-6 ms)	** (-6 ms)	- (-5 ms)	- (-1 ms)		
Closure voicing	Perception:	*** (47.2%)	*** (47.8%)	*** (37.7%)	*** (36.3%)	*** (13.8%)	*** (14.2%)
	Production:	*** (2.12)	*** (2.12)	** (0.97)	*** (0.94)	*** (1.52)	*** (1.32)
Regression model - total		*** (56.5%)	*** (54.6%)	*** (43.5%)	*** (36.3%)	*** (13.8%)	*** (14.2%)

* - significant @ .05 level
** - significant @ .01 level
*** - significant @ .001 level

As shown in Table 30, up to 56.5% of the variance in the mean rates of voiced responses to full-word tokens created on the basis of plosive-final monosyllabic minimal pairs could be predicted on the basis of 3 acoustic parameters: closure voicing, release duration, and preceding vowel duration. More specifically, production differences in the amount of phonetic voicing could explain 47.2% of the observed variance (β=.687, t(118)=10.28, p<.001; R^2=.472, F(1,118)=105.65, p<.001). The addition of the acoustic parameter of release duration increased the predictive ability of the regression model by 7.6% to 54.8% (β_1=.628, t_1(117)=9.87, p<.001, β_2=-.281, t_2(117)=-4.42, p<.001; R^2=.548, F(2,117)=70.86, p<.001). With preceding vowel duration introduced as the third predictor variable, the model accounted for 56.5% of the observed variance, which represents an additional increase of 1.7% (β_1=.631, t_1(116)=10.08, p<.001, β_2=-.282, t_2(116)=-4.5, p<.001, β_3=.133, t_3(116)=2.17, p<.05; R^2=.565, F(3,116)=50.31, p<.001). Closure duration was not a significant contributor (|t|<1, p>.1).

In the case of rhyme-only stimuli items excised from plosive-final monosyllabic minimal pairs, up to 54.6% of the variance in identification responses could be predicted using the acoustic measurements of closure voicing and release duration. Closure voicing alone accounted for 47.8% of the observed variance (β=.692, t(118)=10.40, p<.001; R^2=.478, F(1,118)=108.19, p<.001). With release duration added as the second predictor, the model improved by 6.8% and accounted for 54.6% of the variance (β_1=.635, t_1(117)=9.96, p<.001, β_2=-.266, t_2(117)=-4.17, p<.001; R^2=.546, F(2,117)=70.27, p<.001). Neither preceding vowel duration nor closure duration contributed to the model at a significant level (all |t|s<1.5, all ps>.1).

For tokens created on the basis of plosive-final monosyllabic non-minimal pairs, regression modeling could account for up to 43.5% of the variance in identification responses on the basis of closure voicing and release duration. Phonetic voicing explained 37.7% of all variance (β=.614, t(118)=8.44, p<.001; R^2=.377, F(1,118)=71.26, p<.001). When release duration was tested as the second independent variable, the model accounted for 43.5% of the variance, which is an enhancement of 5.8% (β_1=.588, t_1(117)=8.42, p<.001, β_2=-.243, t_2(117)=-3.48, p=.001; R^2=.435, F(2,117)=45.01, p<.001). The parameters of preceding vowel duration and closure duration were not identified as significant contributors (all |t|s<1.5, all ps>.1). For the tokens excised from plosive-final disyllables, closure voicing significantly predicted 36.3% of the variance in the mean rates of identification responses (β=.603, t(118)=8.20, p<.001; R^2=.363, F(1,118)=67.29, p<.001). None of the remaining acoustic cues enhanced the regression model for disyllable-based items (all |t|s<1.5, all ps>.1).

For perceptual tokens ending in fricatives, phonetic voicing was identified as the only significant predictor variable. For items excised from monosyllabic wordforms, closure voicing explained 13.8% of the variance in the rates of voiced responses (β=.372, t(118)=4.35, p<.001; R^2=.138, F(1,118)=18.95, p<.001). For disyllable-based items, glottal pulsing accounted for 14.2% of the observed variance (β=.377, t(118)=4.43, p<.001; R^2=.142, F(1,118)=19.60, p<.001). Neither preceding vowel duration nor frication duration contributed to the regression model for either monosyllable-based or disyllable-based rhymes (all |t|s<1.5, all ps>.1).

4.4.4.2. Summary

For the mean rates of voiced responses to final plosives, results of the regression analyses show that the acoustic parameters of closure voicing, release duration and preceding vowel duration can explain between 36.3% and 56.5% of the variance, with closure voicing alone accounting for up to 47.8% of the observed effects. For fricative-final items, closure voicing comes out as the only significant predictor and it explains around 14% of the variance. Hence, phonetic voicing comes out as the single most prominent predictor variable for all types of perceptual tokens. Preceding vowel duration, closure/frication and release duration either fail to contribute to the regression model at a statistically significant level (for plosive-final tokens excised from disyllables and for all fricative-final items) or they enhance the statistical model by 1.7% to 7.6% only.

4.4.5. General Summary

Results of the perception study reveal limited identification accuracy, a voiceless bias, and a statistically significant asymmetry in participants' identification responses. More underlyingly voiced segments are classified as voiced and more phonologically voiceless segments are judged to be voiceless when perceptual tokens are elicited during word-reading or with potentially homophonous items included in the stimuli lists. In contrast, listeners' responses are less consistent with underlying voicing when test tokens are recorded during non-orthographic tasks or when full minimal pairs are not included among the stimuli. In regression analyses, the acoustic parameter of closure voicing is

identified as the only predictor variable that is significant for all stimuli types. A general discussion of the experimental results is provided in Chapter 5.

CHAPTER 5: DISCUSSION

5.1. Introduction

Chapter 5 discusses the experimental findings of the present doctoral research. For production results (Section 5.2.), the chapter starts with a general overview and discussion of the data (Section 5.2.1.). This is followed by focused discussions of the results for the parameters of closure voicing (Section 5.2.2.), consonantal duration (Section 5.2.3.), and preceding vowel duration (Section 5.2.4.). For perception data (Section 5.3.), the chapter discusses the findings for response types (Section 5.3.1.) and response latencies (Section 5.3.2.). The remainder of Chapter 5 points out the main theoretical implications of the experimental results (Section 5.4.), followed by a general summary of the claims (Section 5.5.).

5.2. Production Findings

5.2.1. General Summary & Discussion

The acoustic production study presented in Chapter 3 examined the voicing contrast in word-final stops and fricatives in Russian. The goal of the investigation was to understand the nature of the voicing-dependent effects that are found in experimentally-elicited devoicing data and, more specifically, to determine whether the observed patterns could be attributed primarily to phonological and lexical properties of test words or, alternatively, to methodological influences. Production results reveal that the use of phonetic voicing is restricted to the reading task and the stimuli lists that contain full minimal pairs and, when observed, differences in closure voicing are robust and

consistent at the individual level. Unlike glottal pulsing, differences in closure/frication duration and release duration are seen regardless of the task-dependent availability of orthography or full minimal pairs, yet they also show robust individual-level variability in the magnitude and direction of the observed effects and are significant only in a subset of stimuli items. Namely, differences in closure/frication duration are found only in plosive-final monosyllabic minimal pairs and in fricative-final items, and differences in release duration are significant only in plosive-final monosyllables. The final parameter, preceding vowel duration, is not found to be a reliable cue to the voicing specification of word-final obstruents in Russian, and all the observed effects are independent of consonantal place of articulation.

Thus, the present results confirm that experimental methodology can strongly bias the production of underlying voicing in laboratory speech, with robust effects of task-dependent factors seen in the case of glottal pulsing. In other words, differences in closure voicing, the only acoustic parameter which shows significant effects across all 5 stimuli types, are largely due to non-grammatical factors and such differences are neutralized when the experimental paradigm does not specifically encourage preservation of the underlying contrast. However, unlike glottal pulsing, differences in closure/frication duration or release duration are retained independently of methodological bias (other than the experimental situation itself), with complete neutralization of voicing-dependent cues observed at the group level only for plosive-final disyllabic words without minimal pair counterparts in the lexicon. Thus, the current findings do not fully support the view that preservation of voicing-dependent cues can be

attributed *exclusively* to such task-dependent factors as presence of orthographic representations of test items or inclusion of full minimal pairs among the stimuli. At the same time, differences in consonantal duration are not always consistent at the individual level and they do not appear to be enhanced when the experimental task favours maintenance of the voicing contrast in surface outputs. Therefore, even though such differences are produced, Russian speakers appear to associate the voicing contrast in final obstruents with the phonetic voicing cue specifically.

In addition, the current investigation reveals the presence of substantial variability in individual data, with some participants and test items showing robust differences in the expected direction but others demonstrating unexpected patterns or differences that round to zero. Furthermore, although certain differences go in the expected direction in many subjects and test items (e.g., in the case of preceding vowel duration), they are not always statistically significant in the pooled data. Thus, the present study confirms that not all individual-level findings are necessarily representative of larger populations of speakers and lexical items of the same language, which in turn warrants re-evaluation of the conclusions of many previous experimental investigations that relied on small sets of production data or conducted either by-subject or by-item analyses only (e.g., Chen, 1970; Fourakis & Iverson, 1984; Charles-Luce, 1985; Pye, 1986; Inozuka, 1991; Piroth et al., 1991; Kopkalli, 1991; Matsui, 2011).

Production findings also show that the degree to which the voicing contrast is maintained in Russian is place-independent. This result is not consistent with the general expectation of less robust voicing cues in the case of, for example, dorsal plosives

(among others, Ohala, 1983). However, previous research has shown that the effects of consonantal place of articulation on the production of various voicing-dependent differences are not always robust (e.g., VOT values in word-initial plosives can vary across different places of articulation by as little as 1 ms; Lisker & Abramson, 1964). As such, a larger dataset may be required for any place-based asymmetries to become statistically significant, especially considering that some effects of consonantal place are found in the current study in the domain of perception. Hence, the present findings should not be taken to mean that consonantal place of articulation cannot affect the surface realization of the voicing-contrast in neutralizing environments, with the potential relevance of consonantal place further supported by the composition of segmental inventories in languages such as Dutch and Russian (that lack the voiced velar stop /ɡ/ and the voiced velar fricative /ɣ/, respectively) as well as the place-sensitive application of the devoicing process in languages such as Frisian and Tonkawa (that devoice dorsal obstruents only; Blevins, 2006).

5.2.2. Closure Voicing

As stated above, effects of methodological biases are seen in the present study only for the acoustic parameter of closure voicing, with differences in the amount of glottal pulsing being significant and robust only during word-reading or when minimal pairs are present among the stimuli. This confirms that the orthographic nature of the reading task and the inclusion of minimally contrasting item pairs in the stimulus list can have a strong biasing effect on the production of glottal pulsing in word-final obstruents.

This finding is especially important considering that the majority of previous studies administered a reading task, with minimal pairs present among the stimuli, and that neutralizing languages usually represent the underlying voicing specification of word-final consonants in orthographic forms. Even Catalan and Turkish, the two languages that are often cited in the devoicing literature as examples of orthographic systems that do not maintain the voicing distinction at the graphemic level, can in fact provide orthographic evidence for phonemic voicing. Catalan, for example, has a number of high frequency words that end in voiced graphemes (e.g., *verd* 'green', *llard* 'butter'). Certain Catalan graphemes or their combinations can also represent voiceless segments only, whereas others always represent voiced sounds (e.g., the grapheme *ç* is used exclusively for voiceless consonants; the digram *ig* is used for voiced sounds; Manaster Ramer, 1996a, 1996b). Similarly, the Turkish lexicon contains several monomorphemic items that end in voiced graphemes (e.g., *ad* 'name'; Comrie, 1997) and such words are even known to be pronounced with phonetically voiced final obstruents (Kallestinova, 2004).

The observed absence (or weaker magnitude) of voicing-dependent differences in glottal pulsing during non-reading tasks is in line with some of the previous reports of limited and inconsistent effects of underlying voicing in oral experiments (e.g., the oral conjugation task in Fourakis & Iverson, 1984) and it also corroborates the proposal in Warner et al. (2004, 2006) who argued that orthography can play an important biasing role in the preservation of an underlying contrast in experimentally-elicited speech. Namely, Warner et al. (2004) reported that, when pronouncing word pairs such as *baten* 'to avail' ~ *baatten* 'availed' and *laden* 'to load' ~ *laadden* 'loaded', speakers of Dutch

produced statistically significant differences in consonantal duration and preceding vowel duration between singleton and geminate segments even though gemination of the medial stop is largely an orthographic convention. Warner et al. (2006) further demonstrated that no durational differences were found for singleton versus geminate consonants in Dutch when gemination was underlying but not represented orthographically (e.g., 'ik heet' that ends in /t/ versus 'hij heet' that ends in /tt/ but is represented in orthography with a single word-final stop; the second coronal plosive is a third person singular morpheme that is not spelled out).

The attested sensitivity of closure voicing to the presence of graphemic cues to the underlying contrast is also consistent with a number of earlier studies on the role of orthography in the production (and perception) of various linguistic phenomena (production: Buben, 1935; Lupker, 1982; Treiman & Danis, 1988; Laks, 2005; perception: Seidenberg & Tanenhaus, 1979; Donnenwerth-Nolan et al., 1981; Taft & Hambly, 1985; Dijkstra et al., 1995; Ziegler & Ferrand, 1998; Hallé et al., 2000). Buben (1935), for example, famously pointed out that the orthographic system of French can explain why some speakers pronounce the so-called 'silent' final consonants (e.g., the 't' in *but* 'goal') or geminate singleton phonemes that are spelt with two identical letters (e.g., the /m/ in *grammaire* 'grammar'). However, such graphemic effects have been found even in the absence of orthographic representations of test items (among others, Treiman & Danis, 1988; Hallé et al., 2000). Treiman & Danis (1988), for example, observed orthographic influences in syllabification even though their participants performed an oral task and were not exposed to written stimuli. As such, orthographic

knowledge may be activated even during non-reading and, for highly literate populations (such as the university students who took part in the current study), orthographic awareness may even be part of phonological grammar (e.g., Laks, 2005). Thus, although the current results show that preservation of voicing cues in experimentally-elicited speech can be motivated by the task-dependent availability of graphemic representations, the present findings do not preclude the possibility that orthography may influence the production of test items even without immediate exposure to written forms, which may explain the presence of significant voicing-dependent differences in the word-guessing task in Jassem & Richter (1989) and the oral depluralization task in Röttger et al. (2011) (note, however, that both of these studies included full minimal pairs among the stimuli and that the findings in Jassem & Richter (1989) are based on the data from 4 speakers only).

This finding that the presence of full minimal pairs among the stimuli can also bias the production of phonetic voicing in word-final obstruents in Russian (with differences in the amount of glottal pulsing being either absent or less prominent in the absence of potentially homophonous item pairs in the experimental lists) is in line with the results of those previous investigations that either did not use minimally contrasting test items or used very few such tokens and did not find consistent differences between underlyingly voiced versus voiceless word-final obstruents (Fourakis & Iverson, 1984; Inozuka, 1991; Kopkalli, 1993). At the same time, some earlier studies have in fact observed voicing-dependent effects even when the majority of stimuli words did not contrast minimally (Piroth et al., 1991; Piroth & Janker, 2004). In the present

investigation, the Orth+MP- group (i.e., the group that performed a reading task but did not see full minimal pairs) also shows more robust differences in the amount of phonetic voicing than the Orth-MP+ group (i.e., the group for which the two parameters were reversed). Therefore, although inclusion of full minimal pairs among the stimuli is found to enhance voicing-dependent differences, effects of underlying voicing can be observed during word-reading even when very few or no minimal pairs are present in the list of test items. This suggests that the biasing role of graphemic representations is more prominent than the influence of direct exposure to lexical competition. However, presence of a greater bias in the case of orthography is not surprising for at least two reasons. First, orthographic forms are available during word-reading for all test tokens but potential homophony affects only a subset of stimuli items. Second, graphemic representations are present for all stimuli from the start of the reading task but it takes time to encounter both members of a given minimal pair. As such, whereas orthography can play a role from the very beginning of the experimental procedure, the task-dependent availability of direct lexical competition is not expected to have an effect on at least the earlier productions when participants do not yet know that minimally-contrasting item pairs are included among the stimuli (as long as speakers are not exposed to experimental tokens prior to testing).

Another important finding of the present research concerns the results of the post-test questionnaire. When full minimal pairs are present among test items during word-reading, the majority of participants become aware of final devoicing being the topic of investigation. Speakers' awareness of the goal of the study also significantly predicts part

of the observed variance in acoustic cues (i.e., closure voicing for all stimuli types; release duration for plosive-final monosyllabic minimal pairs). This suggests that the common practice of including minimally-contrasting test items among the stimuli strongly encourages conscious production of voicing-dependent differences, with Russian speakers using phonetic voicing (and, to a limited extent, consonantal release duration) to ensure that their outputs meet the perceived goal of the study. However, the observed R^2 scores are generally low, which is due to the fact that some of the subjects from the Orth-MP+ group and the Orth+MP- group did not successfully identify the goal of the study in the post-test questionnaire, yet they produced prominent voicing-dependent differences. Hence, retention of cues to underlying voicing in the production of laboratory speech cannot be attributed exclusively to speaker's knowledge of the topic of the investigation, and factors other than study goal awareness need to be identified in order to offer a full account of the observed effects.

Nevertheless, it is notable that the availability of full minimal pairs among the stimuli helps discover the true aim of the experiment even when minimally contrasting items constitute 10% of the stimuli (including fillers), only one repetition of each test token is elicited, and speakers are not exposed to the experimental tokens prior to testing. In comparison, previous neutralization studies often used very few stimuli items, with 50% of all test tokens being minimal pairs (e.g., Charles-Luce, 1993). Many of the earlier investigations also recorded multiple repetitions by the same speaker and allowed familiarization with the stimuli prior to testing (e.g., 10 repetitions of each word in Smith et al., 2009). As such, participants of the majority of previous devoicing studies were

undoubtedly aware of the contrast under investigation (even if they did not explicitly acknowledge that) and their phonetic outputs were most likely affected by hypercorrection, spelling pronunciation and other similar phenomena that can be expected to interfere with the influences of phonological and lexical factors.

With respect to the primary role of closure voicing, the observed reliance of Russian speakers on the parameter of glottal pulsing for signaling voicing-dependent differences is consistent with some of the earlier proposals of the voicing contrast being associated with a single primary parameter (Slis & Cohen, 1969; Kohler, 1984) and is likely due to the fact that Russian belongs to the so-called 'prevoicing' languages. In such languages, the underlying voicing specification is signalled in non-neutralizing positions with a negative voice onset time (VOT) for voiced plosives and a short-lag positive VOT for voiceless stops (Lisker & Abramson, 1964). In other words, phonetic voicing begins prior to the release stage of voiced stops but shortly after the release burst of their voiceless counterparts. In comparison, neutralizing languages such as German maintain the voicing contrast by producing voiced stops with a short-lag positive VOT and voiceless plosives with a long-lag positive VOT, which is usually achieved by aspirating underlyingly voiceless but not voiced stops (Lisker & Abramson, 1964; Iverson & Salmons, 1995, 1999; Jessen, 1998, 2001; Jessen & Ringen, 2002). Thus, whereas the acoustic parameter of release duration is known to be a major cue to underlying voicing in non-neutralizing positions for aspirating languages such as German (Moulton, 1962; Umeda, 1975; Lisker, 1986), the same contrast is thought to be signaled predominantly by glottal pulsing in pre-voicing languages such as Russian (Avanesov, 1956; Bondarko,

1977; Timberlake, 2004; Wade, 2010). The phonetic study by Bondarko (1977), for example, argued that the voicing distinction in word-initial and word-medial consonants in Russian is manifested primarily by the presence of vocal fold vibration, with the only secondary cue to underlying voicing that is mentioned explicitly being the weaker intensity of aperiodic noise during the release stage of voiced sounds (which Bondarko also attributed to the presence of glottal pulsing). As such, when the experimental paradigm specifically favours preservation of voicing-dependent differences, Russian speakers appear to make use of the single most prominent cue to underlying voicing that they are familiar with from non-neutralizing environments.

It is also worth noting that differences in closure voicing are significant in the present study for all test items produced by the Orth+MP+ group, yet the same differences were not significant in the monolingual group of 4 Russian speakers described in Dmitrieva et al. (2010). Such discrepancy in experimental findings may seem unexpected considering that both groups of participants performed a word-reading task and encountered full minimal pairs among the stimuli. However, the current investigation and the study by Dmitrieva and colleagues have important methodological differences. For example, stimuli items were randomized in the present study, but Dmitrieva et al. arranged their test tokens in a non-random order that was meant to create an impression of a rhyming/semantic task. Hence, in the latter study, the biasing effects of orthography (as well as the influences of the presence of full minimal pairs among the stimuli) might have been overridden by pragmatic factors. Dmitrieva and colleagues also recorded their speakers in a non-formal setting (i.e., at home), which is an environment

that favours hypoarticulated speech. In contrast, participants of the current study were tested in a university laboratory, which likely encouraged production of more hyperarticulated outputs. Dmitrieva et al. also tested 4 monolingual Russian speakers only and, given the robust individual-level variability that is seen in the present investigation, such small samples are not necessarily representative of the language in general.

Finally, on a general note, the findings for closure voicing corroborate the growing body of evidence on the biasing effects of experimental methodology in the production of phonological phenomena in experimentally-elicited speech (Hay et al., 2009, 2010; Pitt, 2009; Drager et al., 2010; Côté & Kharlamov, 2011). Côté & Kharlamov (2011), for example, showed that Russian speakers' syllabification judgments could be influenced by the nature of the experimental task, with positively correlated results obtained for less than half of the possible combinations of 5 distinct syllabification procedures, including 3 oral tasks (pause-insertion, 1st syllable repetition, 2nd syllable repetition) and two written procedures (slash-insertion, evaluation questionnaire). For example, when presented with a set of $C_1VC_2C_3VC_4$ stimuli items, Russian speakers tended to produce C_1VC_2 sequences when instructed to repeat the initial syllable (which suggests $C_1VC_2.C_3VC_4$ syllabification) but VC_4 sequences when asked to repeat the second syllable (which suggests $C_1VC_2C_3.VC_4$ syllabification). Côté & Kharlamov attributed the finding that the syllabic affiliation of C_3 was not consistent across different experimental tasks to the influence of such factors as parsimony, which is a non-

grammatical pressure that is similar to the task-dependent effects of orthography and lexical competition that are observed in the present investigation.

5.2.3. Consonantal Duration

Unlike closure voicing which shows task-dependent effects across all stimuli types (as long as the experimental task encourages maintenance of the underlying contrast), differences in closure/frication duration and release duration are task-independent and they vary depending on consonantal manner of articulation, word length in syllables, and lexical status of test items. This is shown in (8) below for final plosives and in (9) for final fricatives.

(8) *Task-independent cues in final plosives*

Monosyllabic MPs	>>	**Monosyllabic non-MPs**	>>	**Disyllabic non-MPs**
closure duration release duration		release duration		N/A

(9) *Task-independent cues in final fricatives*

Monosyllabic non-MPs	<< >>	**Disyllabic non-MPs**
frication duration		frication duration

As can be seen in (8), underlying voicing affects two acoustic parameters in plosive-final monosyllabic minimal pairs, one cue in plosive-final monosyllabic non-minimal pairs, and none of the parameters in plosive-final disyllabic non-minimal pair items. As

demonstrated in (9), the same type of asymmetry is not seen in words with final fricatives, with frication duration being a significant cue in both monosyllabic and disyllabic wordforms.

The observed retention of differences in consonantal duration is consistent with some of the previous experimental findings on final devoicing in Russian (e.g., Dmitrieva et al., 2010), with the general direction of the attested effects being in line with the results of many previous phonetic studies of the voicing contrast in non-neutralizing languages and environments (i.e., presence of shorter closures, shorter releases and/or shorter periods of frication noise in underlyingly voiced obstruents; Klatt, 1976; Kohler, 1977, 1979; Raphael, 1981; Ladefoged & Maddieson, 1996). The current findings also suggest a hierarchical structure for voicing cues in plosive-final wordforms, whereby more voicing-dependent parameters show significant differences in shorter (monosyllabic) words than longer (disyllabic) items and in words with existing minimal pair counterparts than words without potentially homophonous competitors.[43] This asymmetry in the distribution of voicing cues corroborates the view that preservation of a phonological contrast should be most favoured in shorter words with potentially homophonous competitors (e.g., monosyllabic minimal pairs), favoured to a limited extent in short items that do not have minimal pair counterparts but have greater neighborhood density (e.g., monosyllabic non-minimal pair words), and least favoured in longer wordforms with no lexical competition and low neighbourhood density (e.g., disyllabic non-minimal pair

[43] Only voicing-based homophony is considered here.

tokens) (among others, Hockett, 1967; Crosswhite, 1999; Blevins & Wedel, 2009; Silverman, 2010).

Since the Russian lexicon does not have a sufficient number of disyllabic plosive-final or fricative-final minimal pairs that can be controlled for exogenous variability (i.e., items that belong to the same grammatical category, have matching stress patterns, are not archaic, etc.), it remains to be determined whether the voicing contrast is signaled by additional acoustic parameters in other minimal pairs that may exist in the lexicon (which could be predicted on the basis of the present results) or whether, for example, the shorter overall duration of word-final rhymes in plosive-final disyllables would inhibit the production of any secondary differences in consonantal duration. Nevertheless, any finding of voicing-dependent asymmetries that are sensitive to such grammatical factors as word length and lexical competition highlights the importance of using well-controlled stimuli items and also shows that experimental results are not fully comparable between the studies that test minimal pairs only (e.g., Port & O'Dell, 1985; Charles-Luce, 1993; Dmitrieva et al., 2010) and those investigations that use mostly non-contrasting stimuli items (e.g., Fourakis & Iverson, 1984; Inozuka, 1991; Kopkalli, 1993).

Although the current study did not specifically investigate the sources of the observed differences in consonantal duration, such effects are likely coming from morphologically-related wordforms in which the stem-final consonant is not word-final and does not undergo devoicing. More specifically, many of the grammatical inflections that are used in Russian start with a vowel, which puts the stem-final obstruent in a non-neutralizing environment (e.g., the /d/ is word-final in the Nominative singular form /kɔd/

'code' but the same segment is prevocalic in the Nominative plural /kɔda/, Dative singular /kɔdu/, and Instrumental singular /kɔdɔm/). Considering that complete closure of the vocal tract (for plosives) and frication noise (for fricatives) are both known to inhibit production of phonetic voicing (Ohala, 1983, 1997), articulation of a fully voiced obstruent is facilitated when the consonant is short. Therefore, in non-neutralizing positions, underlyingly voiceless plosives and fricatives can be expected to have longer durations than their voiced counterparts, which has been the exact finding of those experimental studies that examined the voicing contrast in both neutralizing and non-neutralizing environments (e.g., Inozuka, 1991; Piroth et al., 1991; Piroth & Janker, 2004). Crucially, in the case of the Russian, bare-stem members of the inflectional paradigm in which the final obstruent is word-final and is subject to devoicing appear to be less frequent than the forms that end in grammatical inflections in which the critical consonant is not word-final and does not devoice. For example, in the Russian National Corpus (Lyashevskaya & Sharoff, 2009), the wordform /gɔd/ 'year (Nom. sg.)' occurs over 62,000 times, yet the total count for the corresponding lexeme is over 631,000 times. This suggests that over 90% of the time that Russian speakers encounter the lexeme, the stem-final consonant is not word-final and is likely to show robust voicing-dependent differences in such parameters as consonantal duration.[44] Hence, the high usage frequency of the members of the paradigm in which neutralization does not occur can offer a plausible explanation for the appearance of differences in closure/frication

[44] According to the same corpus, bare stem forms are also less frequent than their inflected counterparts in the case of such Russian words as /rɔd/ 'gender' (lexeme count: 46,532; bare-stem form count: 5,529), /kɔd/ 'code' (lexeme: 2,360; bare-stem form: 943), /lug/ 'meadow' (lexeme: 6,641; bare-stem form: 1,233), /stɔlb/ 'pole' (lexeme: 8,768; bare-stem form: 2,136), /plɔd/ 'fruit' (lexeme: 11,401; bare-stem form: 2,472), and other stimuli items used in the present study.

duration and release duration and may even explain why final devoicing in Russian does not get generalized from a word-level process to a stem-level phenomenon.

5.2.4. Preceding Vowel Duration

For the acoustic parameter of preceding vowel duration, the general direction of the observed effects is consistent with much of the acoustic research done on languages such as English (i.e., presence of longer vowels before voiced obstruents; Peterson & Lehiste, 1960; Lisker & Abramson, 1964; Halle & Stevens, 1967; Slis & Cohen, 1969; Klatt, 1976; Lisker, 1974; Maddieson, 1997). However, the attested differences are not statistically significant even when the experimental paradigm specifically encourages retention of the voicing contrast or even in those words that have potentially homophonous lexical competitors. This finding corroborates the results of Dmitrieva et al. (2010) for monolingual Russian subjects and is also in line with some of the experimental research showing that consonantal voicing and preceding vowel duration do not have a universal relationship even in languages such as English (e.g., Purnell et al., 2005a, 2005b). At the same time, the current results contradict the earlier claims made in Chen (1970) and Pye (1986) who both argued that Russian vowels are longer before voiced obstruents and, in the case of Chen's study, that vowel lengthening before voiced consonants is a language-universal phenomenon. However, the latter studies based their claims on very small samples of speakers and lexical items, which is clearly problematic considering the robust individual-level variability in the use of vowel duration that is seen in the present production data, and neither Chen (1970) nor Pye (1986) controlled for the

various grammatical factors that are known to interfere with vowel duration (e.g., stress placement, adjacency to a prosodic boundary, word length; among others, Harris & Umeda, 1974; Umeda, 1975; Klatt, 1976; Crystal & House, 1990; Eefting, 1991).

The finding that preceding vowel duration is not a reliable cue to underlying voicing in Russian is also consistent with the proposal that vowel length is not relevant for the segmental phonology of Russian (among others, Cubberley, 2002). More specifically, neutralizing languages such as German contrast inherently and/or phonemically long and short vowels and often restrict the occurrence of the two vowel types in pre-consonantal environments (e.g., in German, inherently short (lax) vowels are rare before voiced coda consonants; Kleber et al., 2010). Languages such as Russian, however, use vowel length as a correlate of lexical stress rather than an indicator of phonemic differences and do not provide any phonotactic evidence showing that the distribution of long versus short vowels is dependent on the voicing setting of the following obstruent.[45] The Russian language also allows both underlyingly voiced and voiceless obstruents to be preceded by stressed (i.e., longer) vowels in the paradigms of the wordforms that end in plosives and fricatives. For example, in the case of the /kɔt/ 'cat' ~ /kɔd/ 'code' pair, the vowel /ɔ/ is unstressed (shorter) when followed by the voiceless plosive /t/ in the plural form /kɔˈta/ but the same vowel is stressed (longer) in

[45] Shcherba (1912) proposed in his Master's thesis that vowel length in Russian could depend on the quality of the adjacent consonant, including the presence of longer vowels before voiced obstruents. This proposal is occasionally reiterated in descriptive grammars of Russian as evidence of vowel length being a cue to underlying voicing of the following consonant (e.g., Timberlake, 2004). However, Shcherba's claims were based on very limited and potentially problematic acoustic data (i.e., recordings from one speaker (the author himself) who performed a word-reading task with no filler items added to the stimuli). Shcherba also examined the pronunciation of vowels in non-final stressed syllables of disyllabic and trisyllabic words, which is not the environment that is relevant for word-final devoicing.

the plural form /ˈkɔda/. However, the vowel /ɔ/ is always unstressed (shorter) in such pairs as /plɔˈta/ ~ /plɔˈda/ (which are the plural forms of /plɔt/ 'raft' and /plɔd/ 'fruit') and it is always stressed (longer) in word pairs such as /ˈblɔka/ ~ /ˈblɔga/ (which are the plurals of /blɔk/ 'block' and /blɔg/ 'blog'). Hence, the underlying voicing setting of stem-final consonants in Russian cannot be inferred from (or is masked by) the differences in vowel length that are related to stress placement. In comparison, neutralizing languages such as German, Dutch, Afrikaans and Polish show more consistent stress patterns, with lexical stress usually staying on the initial syllable of the stem in German, the penultimate syllable of disyllables in Dutch and Polish, and the initial syllable in Afrikaans. As such, any voicing-dependent differences in preceding vowel duration that may exist in the latter type of languages are unlikely to be masked by the evidence coming from paradigmatically-related forms.

At the same time, considering that in the present production study the effects of underlying voicing on the parameter of vowel duration are not significant at the group level but voicing-dependent differences are in fact observed at the individual level in a subset of speakers and test items, the vowel duration cue may potentially become significant in pooled data with a greater number of participants or stimuli words. Preceding vowel duration is also a limited but statistically significant contributor in perception (for monosyllabic plosive-final minimal pairs). Therefore, although differences in vowel length are not found to be a primary cue to consonantal voicing in word-final obstruents in Russian, the relevance of this parameter for signaling the voicing contrast in neutralizing environments needs to be further investigated.

5.3. Perception Findings

5.3.1. Response Types

In the domain of perception, the findings for the dependent variable of response types (i.e., rates of voiced responses) indicate that underlying voicing is perceptible with limited success only for both final plosives and final fricatives and that Russian speakers show an overall bias for voiceless responses. Namely, the observed rates of voiced responses to voiced versus voiceless obstruents differ by no more than 10% for the items produced by the Orth-MP- group and, on average, no more than 44% of test tokens containing phonologically voiced obstruents are attributed to the voiced category (i.e., overall, underlying voicing is not identified correctly in 56% of voiced obstruents). These results suggest that Russian speakers perceive the majority of underlyingly voiced stops and fricatives as voiceless, which parallels the findings of the Polish study in Jassem & Richter (1989) and the Afrikaans investigation in van Rooy et al. (2003) who also found high proportions of voiceless responses to all final obstruents regardless of their phonological specification. However, even in the absence of significant differences in closure voicing (i.e., the items produced by the Orth-MP- group), the rates of voiced responses to phonologically voiced final obstruents are always at least 6% higher than the rates of voiced responses to underlyingly voiceless final segments. Therefore, Russian listeners appear capable of taking perceptual advantage of differences in parameters other than glottal pulsing.

Crucially, identification responses are also found to be more in line with underlying voicing for the stimuli recorded during word-reading but less consistent with

phonological specifications for the tokens elicited in non-orthographic tasks and, especially, without minimal pairs present among the stimuli. This finding confirms that perceptual identification of intended voicing can be critically affected by those acoustic cues that are present in experimentally-elicited speech for task-dependent reasons. Given that a similar asymmetry is present in production for the acoustic parameter of closure voicing, these results further suggest that identification responses are influenced primarily by the acoustic differences in the (task-dependent) parameter of glottal pulsing and that the contribution of other (task-independent) parameters is at best limited.

The primary role of glottal pulsing in perceptual identification of underlying voicing is also corroborated by the findings of regression modeling, with regression results showing that closure voicing can explain larger proportions of variance in mean rates of voiced responses than any other acoustic cue for plosive-final monosyllabic items and that it is the only significant predictor for plosive-final disyllables and all fricative-final tokens, although R^2 scores are generally low (i.e., around 14%) for the latter type of tokens.[46] However, regression findings also indicate that identification of underlying voicing can sometimes be predicted on the basis of those cues that are not statistically significant in production (i.e., preceding vowel duration for monosyllabic minimal pairs, release duration for monosyllabic non-minimal pairs) and, at the same time, rates of voiced responses appear to be unaffected by some of the significant voicing-dependent differences (i.e., durational differences in closure/frication for plosive-final monosyllabic

[46] The low R^2 scores observed in the case of fricative-final items are not unexpected considering that fewer cycles of glottal pulsing are generally seen in production for fricatives than plosives and that frication noise is known to inhibit phonetic voicing (among others, Ohala, 1983).

minimal pairs and all fricative-final items). Hence, for a subset of perception stimuli, rates of voiced responses are sensitive to those acoustic patterns that are produced inconsistently but insensitive to some of the statistically significant differences. This shows that speakers and listeners do not always rely on the same (task-independent) acoustic parameters when disambiguating the voicing contrast in word-final obstruents, which is in line with some of the previous experimental research indicating that listeners can take advantage of non-significant cues and/or can disregard some of the robust acoustic contrasts that exist in the stimuli (e.g., voicing cues in Dutch, Warner et al., 2004; also, the 'card '~'cod' opposition in English, Costa & Mattingly, 1981; the tense-lax contrast in Swedish vowels, Janson & Schulman, 1983).

The task-dependent nature of asymmetries in identification performance that is seen in the current study provides a plausible explanation for the discrepancies in the findings of many previous investigations into perceptual salience of voicing-dependent cues. As stated in Chapter 2, several neutralization studies used minimally-contrasting tokens that were created in reading tasks and found above-chance level performance for underlyingly voiced obstruents (e.g., Port & O'Dell, 1985; Port & Crawford, 1989; Slowiaczek & Szymanska, 1989; Warner et al., 2004; Ernestus & Baayen, 2007a). Other studies tested the stimuli that were elicited in non-orthographic tasks or created on the basis of non-minimal pairs and observed an overall bias for voiceless (or voiced) responses or response rates of around 50% (e.g., Jassem & Richter, 1989; Kopkalli, 1993). Thus, previous reports of above-chance level identification were often based on the stimuli that were recorded under very different experimental conditions than those

tokens that resulted in below-chance or chance-level performance, with the former type of test items likely containing robust task-dependent cues.

Furthermore, the current findings also provide empirical grounds for questioning the common assumption that response rates of around 50% necessarily indicate chance-level identification. In the present investigation, approximately half of all voiced final plosives and fricatives are attributed to the voiced category when test tokens originate from the Orth+MP+ group, yet no more than 41% of all phonologically voiced consonants are classified as voiced when the data comes from the two Orth- groups. Therefore, when compared across production groups, response rates of around 50% reveal an increase in the number of voiced responses to phonologically voiced final obstruents (i.e., higher identification accuracy for voiced sounds) rather than chance-level guessing. Considering that the majority of earlier devoicing studies administered a reading task and included minimal pairs among the stimuli (which is equivalent to the task performed by the Orth+MP+ group in the present investigation), this result warrants re-evaluation of the claims (or assumptions) of chance-level performance on underlying voicing that were made in several previous studies (e.g., Kahlene-Halstenbach, 1989; Kopkalli, 1993; van Rooy et al., 2003).

More generally, the present results expand the body of literature on the role of non-grammatical factors in speech perception in laboratory environments (Niedzielski, 1999; Pitt, 2009; Hay & Drager, 2010). Niedzielski (1999), for example, showed that, when asked to match natural speech to synthesized tokens, American English speakers were more likely to select tokens with vowel raising if the word ‘Canadian’ was printed

on the answer sheet. The psycholinguistic experiment by Pitt (2009) revealed that 'overhearing' a dialog between two experimenters resulted in speakers of English treating two phonetically distinct forms (*seny*, *senty*) as phonologically equivalent if the equivalence of the two words was mentioned explicitly during the conversation. Hay & Drager (2010) found that perceptual identification of vowels by speakers of New Zealand English varied depending on whether, during testing, they were exposed to a stuffed toy associated with Australia or a toy associated with New Zealand (e.g., a kangaroo versus a kiwi; seeing a kangaroo biased the participants towards Australian English). Hence, speech perception is highly sensitive to the availability of metalinguistic linguistic information and, as confirmed by the findings of the current investigation, can also be affected by the presence of task-dependent acoustic cues in the stimuli.

Finally, for a subset of test items, perception results demonstrate place-based differences in the rates of voiced responses, with identification of velar stops showing greater consistency with underlying voicing even though place-based effects are not significant in production. The observed asymmetry also does not appear to be attributable to the usage frequency of specific stimuli items. Namely, whereas the overall frequencies of test items ending in voiced versus voiceless final stops were matched in the original production study, minimally contrasting test words ending in voiced velar plosives were in fact less frequent than the items ending in voiceless velar stops (M_{Voiced}=2.1 ipm, $M_{Voiceless}$=16.0 ipm). As such, a stronger bias for voiceless responses could have been expected for voiced final velars due to the higher frequency of wordforms ending in voiceless velar stops, yet identification responses show an opposite pattern. Furthermore,

a similar advantage for final velars was observed in the previous perception study in Port & O'Dell (1985) who also found higher rates of correct identification in the case of velar obstruents. This suggests that the effect of consonantal place of articulation is likely to be systematic rather than accidental.

One possible explanation of the observed results is related to the fact that phonetic voicing is known to be more difficult to maintain in dorsal sounds than in labial or coronal segments (Gamkrelidze, 1975; Ohala, 1983, 1997) and, therefore, glottal pulsing may be less expected and more perceptually salient in the case of final velars even though velar stops do not show more robust voicing-dependent production differences than the rest of the plosives. However, the perception experiment that is presented in Chapter 4 was not designed to investigate those place-dependent asymmetries that were not affected by methodological factors, and the finding of better identification performance for velar consonants may be due to listeners' reliance on acoustic cues that were not analyzed in the production study or may be driven by various exogenous influences (e.g., the fact that a subset of rhyme-only tokens corresponded to existing lexical items). Hence, the exact role of consonantal place of articulation in the perception of voicing in word-final obstruents remains to be investigated independently.

5.3.2. Response Latencies

In addition to examining rates of voiced responses, the perception study also investigated the effects of task-independent and task-dependent factors on reaction time latencies. Notably, previous experimental work of final devoicing concentrated on response types and did not report reaction times even though the latter measure is well

known to be an important indicator of processing difficulty in perceptual tasks (among others, Kemps et al., 2005; Ernestus & Baayen, 2007b). In the current study, effects of underlying voicing are never significant for reaction times for any experimental group, which indicates that Russian listeners do not take longer to identify tokens from either the voiced or the voiceless category. At the same time, participants take more time to evaluate the final rhymes of plosive-final monosyllabic items elicited in the absence of orthographic representations.

Considering that, at the acoustic level, tokens produced during non-reading tasks contain statistically significant differences in closure duration and/or release duration but show either limited or non-significant differences in phonetic voicing, longer reaction times to the stimuli produced by the two Orth- groups likely reflect an increase in the processing load associated with less salient and less consistent cues to underlying voicing. At the same time, the additional processing time does not translate into higher rates of voiced responses to underlyingly voiced final stops, suggesting that cues other than phonetic voicing have limited effect on identification accuracy for the voicing contrast in word-final obstruents in Russian even when longer processing times are involved. Moreover, the same group-based asymmetry in reaction times is not seen for final plosives of disyllables, which is likely due to the fact that, unlike monosyllables, plosive-final disyllables do not show any significant group-level differences in either closure duration or release duration. The effect is also absent for final fricatives, suggesting that the presence of differences in frication duration also cannot mitigate the absence of robust closure voicing cues.

5.4. Theoretical Implications

As stated in Chapter 2, preservation of voicing-dependent cues in the production and perception of word-final obstruents in experimentally-elicited speech has often been attributed to linguistic factors and has been interpreted as evidence in favour of non-phonological (phonetic or exemplar-based) accounts of devoicing or 3-way laryngeal classifications of speech sounds (e.g., Dinnsen, 1983, 1985; Charles-Luce, 1985; Port & O'Dell, 1985; Pye, 1986; Port & Crawford, 1989; Port, 1996; Ernestus, 2000; Ernestus & Baayen, 2003, 2004, 2007a, 2007b; Port & Leary, 2005). However, other studies have argued that the observed effects are driven primarily by task-dependent biases and that the presence of incompletely neutralized contrasts does not warrant any changes to the traditional accounts of final devoicing that derive the surface forms of underlyingly voiced segments by, for example, manipulating the setting of the [voiced] feature or delinking the laryngeal node (among others, Fourakis & Iverson, 1984; Mascaró, 1987; Jassem & Richeter, 1989; Jongman et al., 1992; Baumann, 1995; Manaster Ramer, 1996a, 1996b).

In the present investigation, the experimental design takes into account the task-dependent effects of orthography and lexical competition as well as various potentially confounding factors (sample size, statistical treatment, number of repetitions, segmental and prosodic differences, usage frequency, etc.). Results of the current study reveal that the process of final devoicing shows two types of behaviour: (i) complete neutralization of differences in the acoustic parameter of phonetic voicing and (ii) incomplete neutralization of closure/frication duration and/or release duration cues. Therefore, the

present findings are partially compatible with the traditional analyses according to which neutralization is categorical. However, contrary to the traditional view, complete identity of surface forms does not appear to be required and the devoicing process is found to target phonetic voicing specifically rather than the [voice] feature in general. In other words, any analysis which assumes that devoiced segments must be fully voiceless at the phonetic level and simply changes the setting of the [voiced] feature (or delinks the laryngeal node/element) would result in the correct prediction only for a subset of the production data (i.e., plosive-final disyllabic non-minimal pair items). The findings for all other word types favour a model of devoicing that can account for the presence of non-categorical outcomes. As such, the current results are also compatible with the alternative analyses according to which partial preservation of voicing cues is part of grammar. However, as noted above, devoicing does appear to have a categorical component, which is the neutralization of differences in the amount of glottal pulsing.

Therefore, the current experimental results do not support the view that no residual cues to a given underlying contrast may be preserved in those phonetic outputs that have been affected by neutralization (Kiparsky, 1976; also, Charles-Luce, 1985; Port & O'Dell, 1985; Port & Crawford, 1989; Port, 1996; Port & Leary, 2005). However, the present findings are fully in line with the proposal that neutralization results in an approximation of the phonetic characteristics of one of the members of a phonological opposition rather than an obligatory change in category membership (Trubetzkoy, 1939). In other words, devoicing does not render the final obstruent identical to the voiceless member of the voiced-voiceless opposition. Instead, the neutralized segment becomes

devoiced when it is no longer characterized by the presence of robust differences in a subset of voicing-dependent cues or the one cue that is most important for the laryngeal phonology of a particular language.

For Russian, this means that neutralization is best defined as the absence of glottal pulsing (which is found to be the single most prominent disambiguating parameter not only in production but also in perception) and, following the line of thought in Steriade (2000), word-final devoicing is best represented as a suppression of the primary phonetic cue that signals the underlying contrast in the grammar. In other words, the absence of distinctive voicing in word-final obstruents in Russian entails the loss of the main cue that is used to maintain the voicing distinction in non-final positions (or for context-dependent reasons, such as during word-reading, during dictation, in hyperarticulated speech, etc.). Other voicing-dependent differences are allowed to remain in phonetic outputs. Such non-primary cues may be more relevant in non-neutralizing environments and may therefore be attested throughout the word's paradigm, yet they do not support the voicing distinction at the grammatical level and may be expected to be used inconsistently.

The fact that Russian speakers rely on differences in vocal fold vibration whenever voicing distinctions become directly relevant further suggests that glottal pulsing is part of the phonological competence of Russian speakers, yet it is used only in specific situations (e.g., when the experimental paradigm creates a contextual need to disambiguate the members of a minimal pair or, possibly, when dealing with potential homophony in non-experimentally elicited speech). This type of selective application of a

phonetic cue favours those formal models that allow differentiation between core phonology that deals with language use in general and peripheral phonology that accounts for language use in particular situations, such as in laboratory speech, during pronunciation of newly encountered or borrowed lexical items (e.g., the 'co-phonology' analysis of vowel reduction in native versus non-nativized words in Crosswhite, 2001)

When final devoicing is viewed as a process that affects the phonetic voicing cue specifically rather than the entire set of the acoustic correlates of the [voiced] feature, the present findings also show that formal modeling of laryngeal neutralization does not crucially depend on adopting a non-dichotomous phonological classification of speech sounds in which word-final obstruents are neither voiced nor voiceless at the underlying level (e.g., Ernestus, 2000) or on treating neutralization as the result of a phonetic implementation process (among others, Port & O'Dell, 1985; Port, 1996; Port & Leary, 2005). For example, the supposed advantage of the phonetic models of devoicing is that they allow phonological analyses of other linguistic phenomena to remain discrete and symbolic (Port, 1996; Port & Leary, 2005). In the case of Russian, such an argument would work for plosive-final monosyllables and all fricative-final words (that show incomplete neutralization of voicing cues irrespective of task-dependent bias) but not for plosive-final disyllables (that show full neutralization of voicing-dependent differences unless speakers perform a reading task or encounter full minimal pairs among the stimuli). Hence, final devoicing would be viewed as a phonological phenomenon for plosive-final disyllables for which a categorical phonological rule can correctly predict phonetic outputs. However, the same process would be analyzed using a phonetic

implementation rule for plosive-final monosyllables as well as fricative-final wordforms that all surface with partially devoiced obstruents. Therefore, relegation of word-final devoicing to phonetics would require a phonological analysis for some lexical items but a series of phonetic implementation rules for all other words of the same language. Arguably, this is less advantageous than treating laryngeal neutralization as a process that occurs within the same component of grammar but targets only the major cue to the underlying contrast.

The apparent preservation of voicing-dependent cues in the production and perception of the voicing contrast in word-final obstruents has also been claimed to support exemplar-based approaches to devoicing (e.g., Ernestus & Baayen, 2007a, 2007b). In traditional Exemplar theory, acoustic outputs are presumed to be remembered instances of previously encountered speech events ('exemplars') that do not have an abstract level of representation and are reproduced by averaging over a set of memory traces of the same lexical item (Johnson, 1997; Bybee, 2001). As such, surface outputs of exemplars are able to contain robust acoustic-phonetic detail and can account for gradient linguistic phenomena, such as partial neutralization of voicing cues. However, as shown in the current dissertation, word-final devoicing in Russian has both non-categorical and categorical components. Therefore, although the present production findings can be successfully modeled as the outcome of an exemplar selection process, the Russian data clearly favours a framework that can also allow categorical behaviour (e.g., the more recent version of Exemplar theory that permits categorical phonemic representations to co-exist with phonetically rich surface forms; Pierrehumbert, 2002).

More generally, results of the present investigation corroborate the view that final devoicing is an interactive process that is sensitive to a variety of both grammatical and non-grammatical influences and that a comprehensive model of neutralization needs to account for the effects of not only underlying voicing but also lexical, paradigmatic, orthographic, and other factors that may be identified in devoicing data (Charles-Luce, 1997; Warner et al., 2004). Thus, although the question of which particular phonological framework is best suited to model the effects observed in the present investigation remains to be addressed in future research, the current findings can be taken to favour those approaches to phonological theory that do not require a strict separation of phonology from other components of grammar, including phonetics and perception (among many others, e.g., Ohala, 1990, 1995b, 1997; Flemming, 1995, 2004; Kirchner, 1998, 2004; Boersma, 1998; Steriade, 1999; Côté, 2000, 2004a, 2004b; Blevins, 2004; Blevins & Garrett, 2004).

5.5. General Summary

Overall, production findings indicate that the acoustic cue of closure voicing shows categorical neutralization of the underlying contrast unless experimental methodology encourages preservation of voicing-dependent cues, that differences in closure duration and release duration are found irrespective of the task-dependent influences of orthography or lexical competition, that preceding vowel duration is not a primary voicing cue in Russian, and that robust individual-level variability is found for all task-independent cues to underlying voicing. Results of the perception study further

show that identification of intended voicing is done with limited accuracy, that listeners' responses show a voiceless bias, that identification responses are highly sensitive to the presence of task-dependent acoustic cues, and that response rates of approximately 50% do not always signal chance-level performance. The present results confirm that both production and perception of experimentally-elicited speech can be strongly affected by non-grammatical factors, yet the current study shows that not all of the observed voicing-dependent effects can be attributed to the biasing role of orthography or inclusion of minimal pairs among the stimuli. These findings are interpreted as favouring a definition of neutralization that requires partial rather than complete identity and an analysis of final devoicing that specifically targets the primary acoustic cue to the underlying contrast.

CHAPTER 6: CONCLUSION

The present dissertation examined the roles of grammatical and methodological factors in the production and perception of experimentally-elicited speech. Multiple procedures and conditions were tested with the goal of determining the impact of task-independent versus task-dependent factors on the apparent maintenance of the underlying voicing contrast in word-final obstruents in Russian. Acoustic production data reveal that both grammatical and methodological influences can play a role in preservation of underlying voicing. Effects of experimental bias are seen for the parameter of closure voicing, which shows consistent and statistically significant differences whenever task-dependent factors favour maintenance of the underlying contrast. Task-independent effects are found for the acoustic cues of closure/frication duration and release duration, with the observed differences being limited to a subset of test items and showing substantial variability at the individual level. Task-dependent effects are also seen in perception. Identification responses reveal limited accuracy and a strong voiceless bias, with listeners' judgments being more in line with underlying voicing for the perceptual tokens recorded during word-reading and, especially, with minimally contrasting items included among the stimuli. For both production and perception, the overall direction of the observed effects and the results of regression modeling suggest that the task-dependent cue of closure voicing plays a major disambiguating role for the underlying voicing contrast in word-final obstruents in Russian.

The present findings corroborate the view that orthographic influences and exposure to full minimal pairs can strongly affect the outcome of the neutralization

process (e.g., Jassem & Richter, 1989; Piroth & Janker, 2004; Warner et al., 2004, 2006), yet the current results do not provide full support to the claim that the observed lack of discrete neutralization in experimental studies can be dismissed as an artefact of methodological bias (e.g., Fourakis & Iverson, 1984; Manaster Ramer, 1996a, 1996b). At the same time, presence of incompletely neutralized cues to underlying voicing is shown to not pose the type of serious challenge to formal phonology which is advocated for in a number of previous works (e.g., Port, 1996; Ernestus, 2000; Port & Leary, 2005; Ernestus & Baayen, 2007a, 2007b) and to favour a formal model of devoicing that uses a less restrictive definition of neutralization and allows the devoicing process to target the primary phonetic cue to the underlying contrast.

As will be pointed out in the remainder of the chapter, several important questions remain to be addressed in future work. For example, since test tokens were always randomized in the current study and unique randomization was performed for each participant both within and across experimental blocks, the group-level effect of item order could not be analyzed using the statistical model adopted in the present investigation. However, the overall strength of voicing cues may be predicted to increase throughout the experiment as speakers are presented with additional minimal pairs. Hence, a follow-up examination is needed to determine how the production of voicing cues varies throughout the recording session. In addition, the biasing effects of lexical competition are seen in the present investigation even though each token is encountered only once and minimal pairs constitute only 10% of all experimental items. In comparison, in many previous neutralization studies, minimal pairs represented 50% of

all test tokens and multiple repetitions of each item were recorded. The low proportion of minimal pairs that is found in the current study may explain why voicing-dependent differences were more prominent in some of the previous investigations, and future research will need to examine the extent to which the robustness of cues to underlying voicing depends on the exact percentage of minimal pairs among the stimuli.

The present study also demonstrates that, when methodological factors encourage maintenance of the voicing contrast, Russian speakers rely on the acoustic cue of phonetic voicing. However, cues other than (or in addition to) glottal pulsing may be task-dependent in other languages. For example, sensitivity of the release duration parameter to methodological biasing has already been proposed for Afrikaans (van Rooy et al., 2003), and the same release duration cue may be task-dependent in other Germanic languages, such as German and Dutch. Thus, methodological influences on the production of the voicing contrast need to be researched on a language-specific basis before any general conclusions can be offered. Furthermore, for Russian, a follow-up investigation is required to confirm the origin of the task-independent differences in closure duration and release duration. Such cues do not show sensitivity to methodological manipulation and, therefore, do not appear to be part of the conscious grammatical competence of Russian speakers for the voicing contrast in word-final obstruents. However, considering that differences in consonantal duration are affected by underlying voicing at a statistically significant level, such cues must have a relation with the phonological voicing contrast, and future research will need to investigate the exact role these non-primary differences play in non-neutralizing environments (e.g., in pre-

vocalic positions) as well as the extent to which the appearance of such cues in word-final obstruents may be motivated by paradigmatically-related wordforms.

Another important area of future research concerns the articulatory aspects of the voicing contrast. Due to the physiological requirements of maintaining voicing, the timing, place and surface area of the constriction created in the oral cavity as well as the overall volume of the supraglottal space are all known to differ for voiced versus voiceless targets (Ohala, 1983; Westbury, 1983; Shimizu, 1996). Such differences may be present even in the absence of acoustic cues to underlying voicing (e.g., after cessation of vocal fold vibration), yet articulatory evidence for the preservation of voicing-dependent cues in word-final consonants has so far been restricted to small-scale production experiments that did not take into account the effects of task-dependent biases. Fuchs (2005), for example, observed articulatory differences between phonologically voiced versus voiceless coronal consonants in EMA and EPG data from German. However, Fuchs' findings were based on a small sample of speakers (n=3, including the author herself) who performed a reading task.

With respect to speech perception, more research is needed to determine the exact nature of the place-dependent asymmetries observed in the identification experiment (i.e., higher identification accuracy for velar stops). As noted earlier, the same type of effect is not significant in production and cannot be attributed to the usage frequency of specific lexical items. Thus, a follow-up experiment is needed to investigate whether voicing-dependent cues are more salient in the case of velar plosives or whether the observed patterns may be attributable to such factors as listeners' awareness of the general

differences in statistical distribution of voiced versus voiceless consonants in the lexicon (Kleber et al., 2010). In addition, psycholinguistic research (priming studies, ERP experiments) is required to explore the degree to which devoiced forms activate their voiceless counterparts and to establish whether listeners make their identification decisions primarily at an early and purely acoustic level of processing or, alternatively, also rely on higher-level phonological knowledge that becomes available at a later stage of lexical access. These research questions are especially important considering that psycholinguistic studies of the voicing contrast in obstruents have traditionally concentrated on the word-initial position only (e.g., Andruski et al., 1994; Utman et al., 2000; van Alphen & McQueen, 2006; van Alphen, 2007).

On a more general note, the majority of experimental investigations of the voicing contrast in word-final consonants have so far relied on production data elicited in laboratory settings and have usually assumed that tokens recorded during linguistic experiments are representative of more spontaneously produced instances of the same wordforms. However, whereas casual speech often involves longer stretches of connected utterances and allows for considerable reduction and variability, experimentally-elicited tokens are usually short, less variable and more hyperarticulated (Krause & Braida, 2002). As such, although similar patterns have sometimes been found in both laboratory and conversational data (e.g., Purnell et al., 2005a, 2005b), the contrasts seen in experimentally-elicited speech from Russian may or may not be characteristic of the effects that exist in the case of non-laboratory communication, and any interpretation of the current experimental findings and evaluation of their consequences for phonological

theory requires a profound understanding of the relationship between speakers' linguistic performance when the goal of speech production and perception is successful completion of an experimental task versus their performance when the objective of communication is transfer of information. Therefore, follow-up research will need to compare the realization and perceptual salience of consonantal voicing in Russian words produced in various experimental tasks to how the same underlying contrast is manifested in more spontaneously produced utterances. Results of this future investigation will reveal whether any of the task-independent effects that are observed in experimentally-elicited speech remain significant in connected utterances or whether non-laboratory communication involves cues other than acoustic characteristics of the final consonant (e.g., manipulation of prosodic patterns or sentence structures). Addressing these questions will help determine the exact role of the linguistic activity in which the speaker is engaged on the production of the voicing contrast and will make it possible to narrow down the list of grammatical effects that must be accounted for in formal analyses.

Finally, in addition to the voicing contrast in word-final obstruents, preservation of acoustic cues to an underlying contrast has been investigated for devoicing of word-initial and word-medial consonants (Burton & Robblee, 1997; van de Velde & van Hout, 2001; Piroth, 2003; Beckman et al., 2009; Kleber et al., 2010), neutralization of manner of articulation (Kim & Jongman, 1996; Simonet et al., 2008), assimilatory processes (Nolan, 1992; Snoeren & Segui, 2003; Snoeren et al., 2006), consonantal reduction and deletion (Rudin, 1980; Manuel et al., 1992; Kilpatrick et al., 2007), aspiration (Bishop, 2007), adjustment of vowel length and vowel reduction (Lahiri et al., 1987; Padgett &

Tabain, 2005), and alveolar stop flapping and epenthesis (Fox & Terbeek, 1977; Charles-Luce, 1997; Charles-Luce et al., 1999; Herd, 2007; Braver, 2011). As in the case of obstruent voicing, some of these studies reported maintenance of the underlying contrast in production and perception (e.g., Fox & Terbeek, 1977; Charles-Luce, 1997; Bishop, 2007; Herd, 2007), whereas other investigations found no statistically significant differences that could potentially reveal the underlying form (e.g., Kim & Jongman, 1996; Kilpatrick et al., 2007). However, many of the methodological problems discussed in the present dissertation also apply to research on other types of neutralization, including reliance on small sample sizes, failure to control the stimuli for exogenous variability, and absence of by-item analyses. Hence, the role of task-dependent effects must be explored in more detail not only for the process of word-final devoicing but also for other phonological phenomena before the full extent of the implications of laboratory data can be determined and before a general model of neutralization can be proposed.

References

Abu Bakar, H., K. Durvasula, N. Pincus & T. McKinnon. (2007). Incomplete laryngeal neutralization in Kuala Lumpur Malay. Paper presented at the 11th International Symposium on Malay/Indonesian Linguistics, Manokwari, Indonesia.

Alphen, P. van. (2007). Prevoicing in Dutch initial plosives: production, perception, and word recognition. In J. van de Weijer & E. van der Torre (eds.), *Voicing in Dutch.* Amsterdam: Benjamins, 99-124.

Alphen, P. van & J. McQueen. (2006). The effect of voice onset time differences on lexical access in Dutch. *Journal of Experimental Psychology: Human Perception & Performance* 32: 178-196.

Andruski, J., S. Blumstein & M. Burton. (1994). The effect of subphonetic differences on lexical access. *Cognition* 52: 173-187.

Avanesov, R. (1956). *Fonetika Sovremennogo Russkogo Literaturnogo Jazyka* [Phonetics of the Modern Literary Russian Language]. Moscow: Moscow University Press.

Baayen, H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R.* Cambridge: Cambridge University Press.

Baumann, M. (1995). *The Production of Syllables in Connected Speech*. Ph.D. dissertation, University of Nijmegen.

Beckman, J., M. Jessen & C. Ringen. (2009). German fricatives: coda devoicing or positional faithfulness? *Phonology* 26: 231-268.

Benni, T. (1964). *Fonetyka Opisowa Języka Polskiego* [Descriptive Phonetics of the Polish Language]. Wrocław, Poland: Ossolineum.

Bishop, J. (2007). Incomplete neutralization in Eastern Andalusian Spanish: perceptual consequences of durational differences involved in s-aspiration. In J. Trouvain & W. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences.* Saarbrücken, Germany: Universität des Saarlandes, 1765-1768.

Blevins, J. (2004). *Evolutionary Phonology*. Cambridge University Press.

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32: 117-166.

Blevins, J. & A. Garrett. (2004). The evolution of metathesis. In B. Hayes, R. Kirchner & D. Steriade (eds.), *Phonetically Based Phonology*. Cambridge, UK: Cambridge University Press, 117-156.

Blevins, J. & A. Wedel. (2009). Inhibited sound change: an evolutionary approach to lexical competition. *Diachronica* 26: 143-183.

Blumstein, S. (1991). The relation between phonetics and phonology. *Phonetica* 48: 108-119.

Boersma, P. (1998). *Functional Phonology*. *Formalizing the Interaction Between Articulatory and Perceptual Drives*. The Hague: Holland Academic Graphics.

Boersma, P. & D. Weenink. (2009). Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved May 1, 2009, http://www.praat.org/.

Bogdanova, (2001). *Proiznoshenije i Transkriptsija* [Pronunciation and Transcription]. St. Petersburg: St. Petersburg State University Press.

Bondarko, L. (1977). *Zvukovoj Stroj Sovremennogo Russkogo Jazyka* [The Sound System of the Modern Russian Language]. Moscow: Prosveshchenie.

Booij, G. (1995). *The Phonology of Dutch*. Oxford, UK: Oxford University Press.

Braver, A. (2011). Incomplete neutralization in American English flapping: a production study. *University of Pennsylvania Working Papers in Linguistics* 17: 31-40.

Brockhaus, W. (1995). *Final Obstruent Devoicing in the Phonology of German*. Tübingen, Germany: Niemeyer.

Browman, C. & L. Goldstein. (1992). Articulatory phonology: an overview. *Phonetica* 49: 155-180.

Buben, V. (1935). *Influence de l'Orthographe sur la Prononciation du Français Modern* [Influence of Orthography on the Pronunciation of Modern French]. Bratislava: University Komenského.

Burton, M. & K. Robblee. (1997). A phonetic analysis of voicing assimilation in Russian. *Journal of Phonetics* 25: 97-114.

Bybee, J. (2001). *Phonology and Language Use.* Cambridge, MA: Cambridge University Press.

Byrd, D., S. Lee, D. Riggs & J. Adams. (2005). Interacting effects of syllable and phrase position on consonant articulation. *Journal of the Acoustical Society of America* 118: 3860-3873.

Charles-Luce, J. (1985). Word-final devoicing in German: effects of phonetic and sentential contexts. *Journal of Phonetics* 13: 309-324.

Charles-Luce, J. (1993). The effects of semantic context on voicing neutralization. *Phonetica* 50: 28-43.

Charles-Luce, J. (1997). Cognitive factors involved in preserving a phonemic contrast. *Language & Speech* 40: 229-247.

Charles-Luce, J. & D. Dinnsen. (1987). A reanalysis of Catalan devoicing. *Journal of Phonetics* 15: 187-190.

Charles-Luce, J., K. Dressler & E. Ragonese. (1999). The effects of semantics on children's preservation of a phonemic voice contrast. *Journal of Child Language* 26: 505-530.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22: 129-159.

Cho, T. & P. Ladefoged. (1999). Variations and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27: 207-229.

Chomsky, N. & M. Halle. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Clark, H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior* 12: 335-359.

Comrie, B. (1997). Turkish phonology. In A. Kaye (ed.), *Phonologies of Asia and Africa (Including the Caucasus)*. Winona Lake, IN: Eisenbrauns, 883-898.

Costa, P. & I. Mattingly. (1981). Production and perception of phonetic contrast during phonetic change. *Journal of the Acoustical Society of America* 69: S67.

Côté, M.-H. (2000). *Consonant Cluster Phonotactics: A Perception-based Approach*. Ph.D. dissertation, MIT.

Côté, M.-H. (2004a). On perception and sonority in cluster simplification: consonant deletion in Icelandic preterits. *Journal of Germanic Linguistics* 16: 203-243.

Côté, M.-H. (2004b). Syntagmatic distinctness in consonant deletion. *Phonology* 21: 1-41.

Côté, M.-H. & V. Kharlamov. (2011) .The impact of experimental task on syllabification judgments: a case study of Russian. In C. Cairns & E. Raimy (eds.), *Handbook of the Syllable*. Leiden: Brill, 273-294.

Crosswhite, K. (1999). Intra-paradigmatic homophony avoidance in two dialects of Slavic. In M. Gordon (ed.), *UCLA Working Papers in Linguistics 1: Papers in Phonology 2*. Los Angeles: UCLA Department of Linguistics, 48-67.

Crosswhite, K. (2001). *Vowel Reduction in Optimality Theory.* New York: Routledge.

Crystal, T. & A. House. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America* 88: 101-112.

Cubberley, P. (2002). *Russian: A Linguistic Introduction.* Cambridge, UK: Cambridge University Press.

Cycowicz, Y., D. Friedman, M. Rothstein & J. Snodgrass. (1997). Picture naming by young children: norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology* 65: 171-237.

Denes, P. (1963). *The Speech Chain: The Physics and Biology of Spoken Language.* New York: Bell Laboratories.

Dijkstra, T., A. Roelofs & S. Fieuws. (1995). Orthographic effects on phoneme monitoring. *Canadian Journal of Experimental Psychology* 49: 264-271.

Dinnsen, D. (1983). *On the Characterization of Phonological Neutralization*. Bloomington, IN: Indiana University Linguistics Club.

Dinnsen, D. (1985). A re-examination of phonological neutralization. *Journal of Linguistics* 21: 265-279.

Dinnsen, D. & J. Charles-Luce. (1984). Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics* 12: 49-60.

Dmitrieva, O., A. Jongman & J. Sereno. (2010). Phonological neutralization by native and non-native speakers: the case of Russian final devoicing. *Journal of Phonetics* 38: 483-492.

Donaldson, B. (1993). *A Grammar of Afrikaans.* New York: Mouton de Gruyter.

Donnenwerth-Nolan, S., M. Tanenhaus & M. Seidenberg. (1981). Multiple code activation in word recognition: evidence from rhyme monitoring. *Journal of Experimental Psychology: Human Learning & Memory* 7: 170-180.

Drager, K., J. Hay & A. Walker. (2010). Pronounced rivalries: attitudes and speech production. *Te Reo* 53: 27-53.

Duncan, S., M. Rosenberg & J. Finkelstein. (1969). The paralanguage of experimenter bias. *Sociometry* 32: 207-219.

Ernestus, M. (2000). *Voice Assimilation and Segment Reduction in Casual Dutch: A Corpus-based Approach of the Phonetics–Phonology Interface.* Utrecht, Holland: LOT.

Ernestus, M. & H. Baayen. (2003). Predicting the unpredictable: interpreting neutralized segments in Dutch. *Language* 70: 5-38.

Ernestus, M. & H. Baayen. (2004). Analogical effects in regular past tense production in Dutch. *Linguistics* 42: 873-903.

Ernestus, M. & H. Baayen. (2006). The functionality of incomplete neutralization in Dutch: the case of paste-tense formation. In L. Goldstein, D. Whalen & C. Best (eds.), *Laboratory Phonology 8*. Berlin: Mouton de Gruyter, 27-49.

Ernestus, M. & H. Baayen. (2007a). Intraparadigmatic effects on the perception of voice. In E. van der Torre. & J. van de Weijer (eds.), *Voicing in Dutch*. Amsterdam: Benjamins, 153-173.

Ernestus, M. & H. Baayen. (2007b). Paradigmatic effects in auditory word recognition: the case of alternating voice in Dutch. *Language & Cognitive Processes* 22: 1-24.

Ferme, A. (2006). Obstruent devoicing and voice assimilation in Slovene. *Folia Linguistica* 40: 191-205.

Flemming, E. (1995). *Perceptual Features in Phonology*, Ph.D. dissertation, UCLA.

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner & D. Steriade (eds.), *Phonetically Based Phonology.* Cambridge, UK: Cambridge University Press, 232-276.

Forster, K. & J. Forster. (2003). DMDX: a windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers* 35: 116-124.

Fourakis, M. & G. Iverson. (1984). On the 'incomplete neutralization' of German final obstruents. *Phonetica* 41: 140-149.

Fox, R. & D. Terbeek. (1977). Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics* 5: 27-34.

Fuchs, S. (2005). *Articulatory Correlates of the Voicing Contrast in Alveolar Obstruent Production in German.* Berlin: Zentrum für Allgemeine Sprachwissenschaft.

Gamkrelidze, T. (1975). On the correlation of stops and fricatives in a phonological system. *Lingua* 35: 231-361.

Goldsmith, J. (1990). *Autosegmental and Metrical Phonology*. Oxford, UK: Blackwell.

Hall, T. (1992). *Syllable Structure and Syllable-related Processes in German*. Tübingen, Germany: Niemeyer.

Halle, M. (1959). *The Sound Pattern of Russian: A Linguistic and Acoustical Investigation.* The Hague: Mouton.

Halle, M. & K. Stevens. (1967). On the mechanism of glottal vibration for vowels and consonants. *MIT Quarterly Progress Report* 85: 267-271.

Hallé, P., C. Chéreau & J. Segui. (2000). Where is the /b/ in ‘‘absurde’’ [apsyrd]? It is in French listeners’ minds. *Journal of Memory & Language* 43: 618-639.

Harris, M. & N. Umeda. (1974). Effect of speaking mode on temporal factors in speech. *Journal of the Acoustical Society of America* 56: 1016-1018.

Hay, J. & K. Drager. (2010). Stuffed toys and speech perception. *Linguistics* 48: 865-892.

Hay, J., K. Drager & P. Warren. (2009). Careful who you talk to: an effect of experimenter identity on the production of the NEAR/SQUARE merger in New Zealand English. *Australian Journal of Linguistics* 29: 269-285.

Hay, J., P. Warren & K. Drager. (2010). Short-term exposure to one dialect affects processing of another. *Language & Speech* 53: 447-471.

Heim, M. (1976). *Contemporary Czech*. Columbus, Ohio: Slavica Publishers.

Herd, W. (2007). Incomplete neutralization of /t/ and /d/ in flapping environments: an analysis of the North Midland dialect of American English. *Journal of the Acoustical Society of America* 121: 3167-3168.

Hock, H. (1999). Finality, prosody, and change. In O. Fujimura, B. Joseph & B. Palek (eds.), *Proceedings of LP ’98: Item Order in Language and Speech, Vol. 1*. Prague: Karolinum Press, 15-30.

Hockett, C. (1967). The quantification of functional load. *Word* 23: 320-339.

Hualde, J. (1992). *Catalan: Descriptive Grammar*. Oxon, UK: Routledge.

Hume, E. & K. Johnson. (2001). A model of the interplay of speech perception and phonology. In E. Hume & K. Johnson (eds.), *The Role of Speech Perception in Phonology*. New York: Academic Press, 3-26.

Ingram, D. (1989). *Phonological Disability in Children*. London: Cole & Whurr, Ltd.

Inozuka, E. (1991). The realization of the German neutralized word-final plosives /g,k/: an acoustic analysis. *Sophia Linguistica* 30: 119-134.

Iosad, P. (2010). Incomplete neutralization and unorthodox markedness in Breton laryngeal phonology. Paper presented at the 6th Celtic Linguistics Conference, University College Dublin, Dublin, Ireland.

Iverson, G. & J. Salmons. (1995). Aspiration and laryngeal representation in Germanic. *Phonology* 12: 369-396.

Iverson, G. & J. Salmons. (1999). Glottal spreading bias in Germanic. *Linguistische Berichte* 178: 135-151.

Iverson, G. & J. Salmons. (2007). Domains and directionality in the evolution of German final fortition. *Phonology* 24: 1-25.

Janson, T. & R. Schulman. (1983). Non-distinctive features and their use. *Journal of Linguistics* 19: 321-336.

Jassem, W. & L. Richter. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics* 17: 317-325.

Jessen, M. (1998). *Phonetics and Phonology of Tense and Lax Obstruents in German.* Amsterdam: Benjamins.

Jessen, M. (2001). Phonetic implementation of the distinctive auditory features [voice] and [tense] in stop consonants. In T. Alan Hall (ed.), *Distinctive Feature Theory*. Berlin: Mouton de Gruyter, 237-294.

Jessen, M. & C. Ringen. (2002). Laryngeal features in German. *Phonology* 19: 189-218.

Johnson, K. (1997). Speech perception without speaker normalization: an exemplar model. In K. Johnson & J. Mullennix (eds.), *Talker Variability in Speech Processing*. San Diego: Academic Press, 145-165.

Jongman, A. (2004). Phonological and phonetic representations: the case of neutralization. In A. Agwuele, W. Warren & S.-H. Park (eds.), *Proceedings of the 2003 Texas Linguistics Society Conference*. Somerville: Cascadilla Press, 9-16.

Jongman, A., J. Sereno, M. Raaijmakers & A. Lahiri. (1992). The phonological representation of [voice] in speech perception. *Language & Speech* 35: 137-152.

Kager, R. (1999). *Optimality Theory*. Cambridge, UK: Cambridge University Press.

Kahlen-Halstenbach, B. (1990) Zur psychologischen Realität der Auslautverhärtung im Deutschen [On the psychological reality of final devoicing in German]. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 43: 645-655.

Kallestinova, E. (2004). Voice and aspiration of stops in Turkish. *Folia Linguistica* 38: 117-143.

Kaye, J., J. Lowenstamm & J.-R. Vergnaud. (1985). The internal structure of phonological representations: a theory of charm and government. *Phonology Yearbook* 2: 305-328.

Keating, P., W. Linker & M. Huffman. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics* 11: 277-290.

Kemps, R., L. Wurm, M. Ernestus, R. Schreuder & H. Baayen. (2005). Prosodic cues for morphological complexity: comparatives and agent nouns in Dutch and English. *Language & Cognitive Processes* 20: 43-73.

Kenstowicz, M. & C. Kisseberth. (1979). *Generative Phonology: Description and Theory*. New York: Academic Press.

Kilpatrick, C., R. Shosted & A. Arvaniti. (2007). On the perception of incomplete neutralization. In J. Trouvain & W. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany: Universität des Saarlandes, 653-656.

Kim, H. & A. Jongman. (1996). Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics* 24: 295-312.

Kiparsky, P. (1976). Abstractness, opacity, and global rules. In A. Koutsoudas (ed.), *The Application and Ordering of Grammatical Rules*. The Hague: Mouton.

Kiparsky, P. (2008). Universals constraint change; change results in typological generalizations. In J. Good (ed.), *Linguistic Universals and Language Change.* Oxford, UK: Oxford University Press, 23-53.

Kirchner, R. (1998). *Lenition in Phonetically-based Optimality Theory*, Ph.D. dissertation, UCLA.

Kirchner, R. (2004). Consonant lenition. In B. Hayes, R. Kirchner & D. Steriade (eds.), *Phonetically Based Phonology.* Cambridge, UK: Cambridge University Press, 313-345.

Klatt, D. (1976). Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59: 1208-1221.

Kleber, F., T. John & J. Harrington. (2010). The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics* 38: 185-196.

Kohler, K. (1977). The production of plosives. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel* 8: 30-110.

Kohler, K. (1979). Dimensions in the perception of fortis and lenis plosives. *Phonetica* 36: 332-343.

Kohler, K. (1984). Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica* 41: 150-174.

Kopkalli, H. (1993). *A Phonetic and Phonological Analysis of Final Devoicing in Turkish.* PhD Dissertation, University of Michigan.

Krause, J. & L. Braida. (2002). Investigating alternative forms of clear speech: the effects of speaking rate and speaking mode on intelligibility. *Journal of the Acoustical Society of America* 112: 2165-2172.

Kuzla, C. & T. Dubeda. (2008). Incomplete neutralization and prosodic conditioning of voice assimilation in Czech. Paper presented at the Consonant Clusters and Structural Complexity Workshop, Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University, Munich, Germany.

Ladefoged, P. & I. Maddieson. (1996). *Sounds of the World's Languages*. Oxford, UK: Blackwell.

Lahiri, A., H. Schriefers & C. Kuijpers. (1987). Contextual neutralization of vowel length: evidence from Dutch. *Phonetica* 44: 91-102.

Laks, B. (2005). La liaison et l'illusion. *Langages* 158: 101-126.

Lass, R. (1971). Boundaries as obstruents: Old English voicing assimilation and universal strength hierarchies. *Journal of Linguistics* 7: 15-30.

Lass, R. (1984). *Phonology: An Introduction to Basic Concepts.* Cambridge, MA: Cambridge University Press.

Lees, R. (1961). *The Phonology of Modern Standard Turkish.* Bloomington, IN: Indiana University Press.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge, MA: MIT Press.

Lightner, T. (1972). *Problems in the Theory of Phonology.* Edmonton: Linguistic Research Inc.

Lindblom, B. (1967). Vowel duration and a model of lip mandible coordination. *Speech Transmission Laboratory*, *Quarterly Progress & Status Report* 4: 1-29.

Lindblom, B. (1983). Economy of speech gestures. In P. MacNeilage (ed.), *The Production of Speech*. New York: Springer, 217-245.

Lisker, L. (1974). On 'explaining' vowel duration variation. *Glossa* 8: 233-246.

Lisker, L. (1986). 'Voicing' in English: a catalogue of acoustic features signaling /b/ vs. /p/ in trochees. *Language & Speech* 29: 3-11.

Lisker, L. & A. Abramson. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20: 384-422.

Locker, L., L. Hoffman & J. Bovaird. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods* 39: 723-730.

Lombardi, L. (1994). *Laryngeal Features and Laryngeal Neutralization*. New York: Garland.

Lombardi, L. (1999). Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language & Linguistic Theory* 17: 267-302.

Lupker, S. (1982). The role of phonetic and orthographic similarity in picture–word interference. *Canadian Journal of Psychology* 36: 349-376.

Lyashevskaya, O. & S. Sharoff. (2009). *Chastotnyj Slovar Sovremennogo Russkogo Jazyka* [The Frequency Dictionary of the Modern Russian Language]. Moscow: Azbukovnik.

Maddieson, I. (1997). Phonetic universals. In J. Laver & W. Hardcastle (eds.), *The Handbook of Phonetic Sciences*. Oxford, UK: Blackwell, 619-639.

Manaster Ramer, A. (1996a). A letter from an incompletely neutral phonologist. *Journal of Phonetics* 24: 477-489.

Manaster Ramer, A. (1996b). Report on Alexis' dreams - bad as well as good. *Journal of Phonetics* 24: 513-519.

Manuel, S., S. Shattuck-Hufnagel, M. Huffman, K. Stevens, R. Carlson & S. Hunnicutt. (1992). Studies of vowel and consonant reduction. In J. Ohala, T. Nearey, B. Derwing, M. Hodge & G. Wiebe (eds.), *Proceedings of the 1992 International Conference on Spoken Language Processing*. Banff: University of Alberta, 943-946.

Mascaró, J. (1987). Underlying voicing recoverability of finally devoiced obstruents in Catalan. *Journal of Phonetics* 15: 183-186.

Matsui, M. (2011). The identifiability and disriminability between incompletely neutralized sounds: evidence from Russian. In W.-S. Lee & E. Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong: City University of Hong Kong, 1342-1345.

Mitleb, F. (1981). *Segmental and Non-segmental Structure in Phonetics. Evidence from Foreign Accent.* Ph.D. dissertation, University of Indiana.

Moulton, W. (1962). *The Sounds of English and German*. Chicago: University of Chicago Press.

Myers, S. (2012). Final devoicing: production and perception studies. In T. Borowsky, S. Kawahara, T. Shinya & M. Sugahara (eds.), *Prosody Matters: Essays in Honor of Elisabeth Selkirk*. London: Equinox Press.

Nearey, T. & B. Rochet. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association* 24: 1-18.

Nicolae, A. & A. Nevins. (2009). The phonetics and phonology of fricative non-neutralization in Turkish. Paper presented at NELS 40, MIT, Boston.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language & Social Psychology* 18: 62-85.

Nolan, F. (1992). The descriptive role of segments: evidence from assimilation. In G. Docherty & D. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press, 261-280.

O'Dell, M. & R. Port. (1983). Discrimination of word-final voicing in German. *Journal of the Acoustical Society of America* 73: S31.

Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In P. MacNeilage (ed.), *The Production of Speech*. Berlin: Springer, 189-216.

Ohala, J. (1990). There is no interface between phonology and phonetics: a personal view. *Journal of Phonetics* 18: 153-171.

Ohala, J. (1995a). Experimental phonology. In J. Goldsmith (ed.), *A Handbook of Phonological Theory*. Oxford, UK: Blackwell, 713-722.

Ohala, J. (1995b). The phonetics of phonology. In G. Bloothooft, V. Hazan, D. Huber & J. Llisterri (eds.), *European Studies in Phonetics and Speech Communication*. Utrecht, Holland: OTS Publications, 85-89.

Ohala, J. (1997). Phonetics in phonology. In *Proceedings of the 4th Seoul International Conference on Linguistics*. Seoul: Linguistic Society of Korea, 45-50.

Oostendorp, M. van. (2008). Incomplete devoicing in formal phonology. *Lingua* 118: 1362-1374.

Ozhegov, S. & N. Shwedova. (2000). *Tolkovij Slovar Russkogo Jazyka* [Dictionary of the Russian Language]. Moscow: Azbukovnik.

Padgett, J. & M. Tabain. (2005). Adaptive dispersion theory and phonological vowel reduction in Russian. *Phonetica* 62:14-54.

Peterson, G. & I. Lehiste. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32: 693-703.

Petrova, O., R. Platt, C. Ringen & S. Szentgyorgyi. (2006). Voice and aspiration: evidence from Russian, Hungarian, German, Swedish, and Turkish. *The Linguistic Review* 23: 1-35.

Pierrehumbert, J. (2001). Exemplar dynamics: word frequency, lenition and contrast. In J. Bybee & P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure.* Amsterdam: John Benjamins, 137-157.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (eds.), *Laboratory Phonology 7*. Berlin: Mouton de Gruyter, 101-139.

Piroth, H. (2003). Final devoicing and syllabification in German consonant clusters: a phonetic investigation. In M. Solé, D. Recasens & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences.* Barcelona: Casual Productions, 2749-2752.

Piroth, H. & P. Janker. (2004). Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics* 32: 81-109.

Piroth, H., L. Schiefer, P. Janker & B. Johne. (1991). Evidence for final devoicing in German? An experimental investigation. In *Proceedings of the 12th International Congress of Phonetic Sciences, Vol. 2*. Aix-en-Provence, France: Publications de l'Université de Provence, 138-141.

Pitt, M. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory & Language* 61: 19-36.

Polinsky, M. & O. Kagan. (2007). Heritage languages: in the 'wild' and in the classroom. *Language and Linguistics Compass* 1: 368-395.

Port, R. (1996). Phonetic discreteness and formal linguistics: reply to A. Manaster-Ramer. *Journal of Phonetics* 24: 491-511.

Port, R. & P. Crawford. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17: 257-282.

Port, R. & A. Leary. (2005). Against formal phonology. *Language* 81: 927-964.

Port, R. & M. O'Dell. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13: 455-471.

Port, R., F. Mitleb & M. O'Dell. (1981). Neutralization of voicing in German is incomplete. *Journal of the Acoustical Society of America* 70: S13.

Prince, A. & P. Smolensky. (2004). *Optimality Theory: constraint interaction in generative grammar.* Cambridge, MA: Blackwell.

Purnell, T., D. Tepeli & J. Salmons. (2005a). German substrate effects in Wisconsin English: evidence for final fortition. *American Speech* 80: 135-164.

Purnell, T., J. Salmons, D. Tepeli & J. Mercer. (2005b). Structured heterogeneity and change in laryngeal phonetics: Upper Midwestern final obstruents. *Journal of English Linguistics* 33: 307-338.

Pye, S. (1986). Word-final devoicing of obstruents in Russian. *Cambridge Papers in Phonetics and Experimental Linguistics* 5: 1-10.

Raaijmakers, J., J. Schrijnemakers & F. Gremmen. (1999). How to deal with the 'language-as-fixed-effect fallacy': common misconceptions and alternative solutions. *Journal of Memory & Language* 41: 416-426.

Raphael, L . (1981). Durations and contexts as cues to word-final cognate opposition in English. *Phonetica* 38: 126-147.

Revoile, S., J. Pickett & L. Holden. (1982). Acoustic cues to final stop voicing for impaired- and normal-hearing listeners. *Journal of the Acoustical Society of America* 72: 1146-1154.

Rooy, B. van, D. Wissing & D. Paschall. (2003). Demystifying incomplete neutralisation during final devoicing. *Southern African Linguistics and Applied Language Studies* 21: 49-66.

Rosnow, R. & R. Rosenthal. (1997). *People Studying People: Artifacts and Ethics in Behavioral Research.* Oxford, UK: W. H. Freeman.

Rossion, B. & G. Pourtois. (2004). Revisiting Snodgrass and Vanderwart's object databank: the role of surface detail in basic level object recognition. *Perception* 33: 217-236.

Röttger, T., B. Winter & S. Grawunder. (2011). The robustness of incomplete neutralization in German. In W.-S. Lee & E. Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong: City University of Hong Kong, 1342-1345.

Rubach, J. (1990). Final devoicing and cyclic syllabification in German. *Linguistic Inquiry* 21: 79-94.

Rubach, J. (1993). *The Lexical Phonology of Slovak*. Oxford, UK: Oxford University Press.

Rudin, C. (1980). Phonetic evidence for a phonological rule: g-deletion in Turkish. *Research in Phonetics* 1: 217-232.

Seidenberg, M. & M. Tanenhaus. (1979). Orthographic effects in rhyme monitoring. *Journal of Experimental Psychology: Human Learning & Memory* 5: 546-554.

Severens, E., S. van Lommel, E. Ratinckx & R. Hartsuiker. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica* 119: 159-187.

Shcherba, L. (1912). *Russkie Glasnye v Kachestvennom i Kolichestvennom Otnoshenii* [Russian Vowels with Respect to Quality and Quantity]. St. Petersburg: U.N. Erlikh Press.

Shimizu, K. (1996). *A Cross-language Study of Voicing Contrasts of Stop Consonants in Asian Languages.* Tokyo: Seibido Publishing.

Shrager, M. (2006). Neutralization of word-final voicing in Russian. Paper presented at the First Conference of The Slavic Linguistics Society, Indiana University, Bloomington, IN.

Silverman, D. (2010). Neutralization and anti-homophony in Korean. *Journal of Linguistics* 46: 453-482.

Simon, E. (2010). Phonological transfer of voicing and devoicing rules: evidence from L1 Dutch and L2 English conversational speech. *Language Sciences* 32: 63-86.

Simonet, M., M. Rohena-Madrazo & M. Paz. (2008). Preliminary evidence for incomplete neutralization of coda liquids in Puerto Rican Spanish. In L. Colantoni & J. Steele (eds.), *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*. Somerville, MA: Cascadilla Proceedings Project, 72-86.

Slis, I. (1986). Assimilation of voice in Dutch as a function of stress, word boundaries and sex of speaker and listener. *Journal of Phonetics* 14: 311-426.

Slis, I. & A. Cohen. (1969). On the complex regulating the voiced-voiceless distinction. *Language and Speech* 12: 80-102, 137-155.

Slowiaczek, L. & D. Dinnsen. (1985). On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13: 325-341.

Slowiaczek, L. & H. Szymanska. (1989). Perception of word-final devoicing in Polish. *Journal of Phonetics* 17: 205-212.

Smith, B., R. Hayes-Harb, M. Bruss & A. Harker. (2009). Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics* 37: 257-275.

Snoeren, N. & J. Segui. (2003). A voice for the voiceless: voice assimilation in French. In M. Sole, D. Recasens & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences.* Barcelona: Casual Productions, 2325-2328.

Snoeren, N., P. Halle & J. Segui. (2006). A voice for the voiceless: production and perception of assimilated speech in French. *Journal of Phonetics* 34: 241-268.

Stampe, D. (1969). The acquisition of phonetic representation. *Chicago Linguistic Society* 5: 443-454.

Steriade, D. (1999). Phonetics in phonology: the case of laryngeal neutralization. In M. Gordon (ed.) *UCLA Working Papers in Linguistics 2: Papers in Phonology 3*. Los Angeles: UCLA Department of Linguistics, 25-146.

Steriade, D. (2000). Paradigm uniformity and the phonetics/phonology boundary. In J. Pierrehumbert & M. Broe (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon.* Cambridge, UK: Cambridge University Press, 313-334.

Taft, M. & G. Hambly. (1985). The influence of orthography on phonological representations in the lexicon. *Journal of Memory & Language* 24: 320-335.

Tieszen, B. (1997). *Final Stop Devoicing in Polish: An Acoustic and Historical Account for Incomplete Neutralization*. Ph.D. dissertation, University of Wisconsin-Madison.

Tieszen, B. & C. Read. (1993). Final stop devoicing in Polish: incomplete neutralization. *Journal of the Acoustical Society of America* 94: 1866.

Timberlake, A. (2004). *A Reference Grammar of Russian*. Cambridge, UK: Cambridge University Press.

Treiman, R. & C. Danis. (1988). Syllabification of intervocalic consonants. *Journal of Memory & Language* 27: 87-104.

Trubetzkoy, N. (1939). Grundzüge der Phonologie [Principles of Phonology]. *Travaux du cercle linguistique de Prague* 7. Reprint: Berkeley: University of California Press [1969].

Umeda, N. (1975). Vowel duration in American English. *Journal of the Acoustical Society of America* 58: 434-455.

Utman, J., S. Blumstein & M. Burton. (2000). Effects of subphonetic and syllable structure on word recognition. *Perception & Psychophysics* 62: 1297-1311.

Vaux, B. & B. Samuels. (2005). Laryngeal markedness and aspiration. *Phonology* 22: 395-436.

Velde, H. van de & R. van Hout. (2001). The devoicing of fricatives in a reading task. In T. van der Wouden & H. Broekhuis (eds.), *Linguistics in the Netherlands 2001*. Amsterdam: John Benjamins, 225-236.

Vennemann, T. (1968). *German Phonology*. PhD dissertation, UCLA.

Volaitis, L. & J. Miller. (1992). Phonetic prototypes: influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America* 92: 723-735.

Vonesh, E. & V. Chinchilli. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements.* London: Chapman & Hall.

Wade, T. (2010). *A Comprehensive Russian Grammar.* Oxford, UK: Wiley-Blackwell.

Wardrip-Fruin, C. (1982). On the status of temporal cues to phonetic categories: preceding vowel duration as a cue to voicing in final stop consonants. *Journal of the Acoustical Society of America* 71: 187-195.

Warner, N., A. Jongman, J. Sereno & R. Kemps. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *Journal of Phonetics* 32: 251-276.

Warner, N., E. Good, A. Jongman & J. Sereno. (2006). Orthographic vs. morphological incomplete neutralization effects. *Journal of Phonetics* 34: 285-293.

Westbury, J. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America* 73: 1322-1336.

Westbury, J. & P. Keating. (1980). Central representation of vowel duration. *Journal of the Acoustical Society of America* 67: S37.

Whalen, D. (1992). Further results on the duration of infrequent and frequent words. *Journal of the Acoustical Society of America* 91: 2339-2340.

Wierzchowska, B. (1980). *Fonetyka i Fonologia Języka Polskiego* [Phonetics and Phonology of the Polish Language]. Wrocław, Poland: Zakład Narodowy im. Ossolińskich.

Wissing, D. & A. van Rooy. (1992). Onvolledige neutralisasie: die ontstemmingsregel in Afrikaans [Incomplete neutralization: the devoicing rule in Afrikaans.] *South African Journal of Linguistics* S13: 135-144.

Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics* 38: 329-336.

Ziegler, J. & L. Ferrand. (1998). Orthography shapes the perception of speech: the consistency effect in auditory word recognition. *Psychonomic Bulletin & Review* 5: 683-689.

Appendices

Appendix 1: List of test items

Num	Contrast	Orthogr. form (Cyrillic)	Phonolog. form*	Gloss	Orthogr. form (Cyrillic)	Phonolog. form*	Gloss
1	/p ~ b/	сноп	/snɔp/	sheaf	сноб	/snɔb/	snob
2	/p ~ b/	серп	/sɛrp/	sickle	серб	/sɛrb/	Serb
3	/p ~ b/	крап	/krap/	specks	краб	/krab/	crab
4	/p ~ b/	столп	/stɔlp/	pillar	столб	/stɔlb/	pole
5	/p ~ b/	грипп	/grip/	flu	гриб	/grib/	mushroom
6	/p ~ b/	суп	/sup/	soup	зуб	/zub/	tooth
7	/p ~ b/	кляп	/klʲap/	gag	штаб	/ʃtab/	headquarters
8	/p ~ b/	склеп	/sklɛp/	crypt	хлеб	/xlɛb/	bread
9	/p ~ b/	храп	/xrap/	snoring	раб	/rab/	slave
10	/p ~ b/	поп	/pɔp/	priest	лоб	/lɔb/	forehead
11	/p ~ b/	подкоп	/padkɔp/	sap	соскоб	/saskɔb/	scraping
12	/p ~ b/	сироп	/sirɔp/	syrup	сугроб	/sugrɔb/	snow-bank
13	/p ~ b/	циклоп	/tsiklɔp/	Cyclops	озноб	/aznɔb/	chill
14	/p ~ b/	этап	/ɛtap/	stage	ухаб	/uxab/	bump
15	/p ~ b/	укроп	/ukrɔp/	dill	микроб	/mikrɔb/	microbe
16	/t ~ d/	кот	/kɔt/	cat	код	/kɔd/	code
17	/t ~ d/	гот	/gɔt/	Goth	год	/gɔd/	year
18	/t ~ d/	рот	/rɔt/	mouth	род	/rɔd/	gender
19	/t ~ d/	плот	/plɔt/	raft	плод	/plɔd/	fruit
20	/t ~ d/	прут	/prut/	rod	пруд	/prud/	pond
21	/t ~ d/	шут	/ʃut/	jester	суд	/sud/	court
22	/t ~ d/	скат	/skat/	stingray	спад	/spad/	decline
23	/t ~ d/	свет	/svɛt/	light	след	/slɛd/	trace
24	/t ~ d/	цвет	/tsvɛt/	color	плед	/plɛd/	rug
25	/t ~ d/	брат	/brat/	brother	клад	/klad/	treasure
26	/t ~ d/	пират	/pirat/	pirate	парад	/parad/	parade
27	/t ~ d/	салат	/salat/	salad	доклад	/daklad/	report
28	/t ~ d/	живот	/ʒivɔt/	stomach	завод	/zavɔd/	factory
29	/t ~ d/	магнит	/magnit/	magnet	гибрид	/gibrid/	hybrid
30	/t ~ d/	пилот	/pilɔt/	pilot	комод	/kamɔd/	commode
31	/k ~ g/	рок	/rɔk/	fate	рог	/rɔg/	horn
32	/k ~ g/	блок	/blɔk/	block	блог	/blɔg/	blog
33	/k ~ g/	сток	/stɔk/	drain	стог	/stɔg/	haystack
34	/k ~ g/	мак	/mak/	poppy	маг	/mag/	magician
35	/k ~ g/	лук	/luk/	onion	луг	/lug/	meadow
36	/k ~ g/	злак	/zlak/	grass	флаг	/flag/	flag
37	/k ~ g/	грек	/grɛk/	Greek	снег	/snɛg/	snow

38	/k ~ g/	внук	/vnuk/	grandson	друг	/drug/	friend
39	/k ~ g/	толк	/tɔlk/	sense	долг	/dɔlg/	debt
40	/k ~ g/	франк	/frank/	frank	шланг	/ʃlang/	hose
41	/k ~ g/	урок	/urɔk/	lesson	пирог	/pirɔg/	pie
42	/k ~ g/	отсек	/atsɛk/	compartment	ковчег	/kavtʃɛg/	ark
43	/k ~ g/	клубок	/klubɔk/	roll	сапог	/sapɔg/	boot
44	/k ~ g/	барак	/barak/	barrack	овраг	/avrag/	ravine
45	/k ~ g/	каблук	/kabluk/	heel	супруг	/suprug/	spouse
46	/f ~ v/	граф	/graf/	count	нрав	/nrav/	temper
47	/f ~ v/	штраф	/ʃtraf/	fine	сплав	/splav/	alloy
48	/f ~ v/	шеф	/ʃɛf/	chief	лев	/lɛv/	lion
49	/f ~ v/	блеф	/blɛf/	bluff	хлев	/xlɛv/	stable
50	/f ~ v/	пуф	/puf/	pouf	клюв	/klʲuv/	beak
51	/f ~ v/	жираф	/ʒiraf/	giraffe	удав	/udav/	boa
52	/f ~ v/	рельеф	/rɛlʲjɛf/	topography	припев	/pripɛv/	chorus
53	/f ~ v/	шериф	/ʃɛrif/	sheriff	полив	/paliv/	watering
54	/f ~ v/	тариф	/tarif/	tariff	залив	/zaliv/	bay
55	/f ~ v/	мастиф	/mastif/	mastiff	мотив	/mativ/	motive
56	/s ~ z/	нос	/nɔs/	nose	воз	/vɔz/	carriage
57	/s ~ z/	вес	/vɛs/	weight	срез	/srɛz/	cut
58	/s ~ z/	рис	/ris/	rice	приз	/priz/	prize
59	/s ~ z/	трус	/trus/	coward	груз	/gruz/	load
60	/s ~ z/	час	/tʃas/	hour	джаз	/dʒaz/	jazz
61	/s ~ z/	индус	/indus/	Hindu	арбуз	/arbuz/	watermelon
62	/s ~ z/	поднос	/padnɔs/	tray	гипноз	/gipnɔz/	hypnosis
63	/s ~ z/	кокос	/kakɔs/	coconut	наркоз	/narkɔz/	anaesthesia
64	/s ~ z/	редис	/redis/	radish	круиз	/kruiz/	cruise
65	/s ~ z/	запас	/zapas/	stock	отказ	/atkaz/	rejection
66	/ʃ ~ ʒ /	душ	/duʃ/	shower	муж	/muʒ/	husband
67	/ʃ ~ ʒ /	куш	/kuʃ/	jackpot	уж	/uʒ/	adder
68	/ʃ ~ ʒ /	фарш	/farʃ/	mince	шарж	/ʃarʒ/	caricature
69	/ʃ ~ ʒ /	ёрш	/jɔrʃ/	ruffe	корж	/kɔrʒ/	layer
70	/ʃ ~ ʒ /	клёш	/klʲɔʃ/	flare	ёж	/jɔʒ/	hedgehog
71	/ʃ ~ ʒ /	гуляш	/gulʲaʃ/	goulash	муляж	/mulʲaʒ/	waxwork
72	/ʃ ~ ʒ /	шалаш	/ʃalaʃ/	hovel	стеллаж	/stɛlaʒ/	shelving
73	/ʃ ~ ʒ /	алкаш	/alkaʃ/	drunkard	монтаж	/mantaʒ/	assembly
74	/ʃ ~ ʒ /	лаваш	/lavaʃ/	pita	багаж	/bagaʒ/	luggage
75	/ʃ ~ ʒ /	чуваш	/tʃuvaʃ/	Chuvash	вираж	/viraʒ/	turn

* - Phonological forms reflect vowel reduction in unstressed syllables. Predictable palatalization is not shown.

Appendix 2: List of fill-in-the-blank descriptions

Num.	Final C	Orth. form (Cyrillic)	Fill-in-the-blank description
1	/p/	грипп	эпидемия, инфекция, свиной .....
2	/p/	кляп	пасть, затычка, тряпичный .....
3	/p/	крап	карты, метка, незаметный .....
4	/p/	подкоп	тюрьма, побег, ... под стену
5	/p/	сироп	сахар, концентрированный раствор, густой .....
6	/p/	склеп	подземная усыпальница, фамильный .....
7	/p/	столп	Дворцовая площадь, колонна, Александрийский .....
8	/p/	храп	сон, дыхание, громкий .....
9	/p/	этап	бег, эстафета, последний .....
10	/b/	озноб	высокая температура, дрожь, сильный .....
11	/b/	серб	Загреб - хорват, Белград - .....
12	/b/	сноб	высокомерный человек, название журнала, ...
13	/b/	соскоб	медицинский анализ - ..... на энтеробиоз
14	/b/	сугроб	зима, обочина дороги, высокий.....
15	/b/	ухаб	выбоина на дороге, глубокий .....
16	/b/	штаб	армия, главная ставка, генеральный .....
17	/t/	брат	дочь - сын, сестра - .....
18	/t/	гот	розовый - эмо, чёрный - .....
19	/t/	пилот	корабль - капитан, самолёт - .....
20	/t/	прут	металлический стержень, железный .....
21	/t/	рот	собака - пасть, человек - .....
22	/t/	салат	кола - напиток, оливье - .....
23	/t/	свет	солнце, лампа, яркий .....
24	/t/	цвет	духи - запах, краски - .....
25	/t/	шут	средневековье, клоун, придворный .....
26	/d/	гибрид	результат скрещивания - .....
27	/d/	год	високосный, прошлый, новый .....
28	/d/	доклад	конференция, выступление, научный .....
29	/d/	код	банк, карта, пин-.....
30	/d/	парад	День Победы, площадь, военный .....
31	/d/	плед	накидка, покрывало, шерстяной .....

32	/d/	плод	Адам, Ева, запретный .....
33	/d/	пруд	маленький водоем, искусственный .....
34	/d/	род	грамматика, склонение, средний .....
35	/d/	след	отпечаток обуви, ... ноги
36	/d/	спад	кризис, убытки, экономический .....
37	/d/	суд	мировой, гражданский, верховный .....
38	/k/	барак	жилье для рабочих, одноэтажный .....
39	/k/	блок	компьютер, процессор, системный .....
40	/k/	внук	отец - сын, дед - .....
41	/k/	грек	Мадрид - испанец, Афины - .....
42	/k/	злак	яблоко - фрукт, рожь - .....
43	/k/	отсек	корабль, трюм, грузовой .....
44	/k/	рок	судьба, несчастье, злой .....
45	/k/	сток	крыша, сливная труба, водо .....
46	/k/	толк	польза, прок, ... в деле
47	/k/	урок	университет - лекция, школа - .....
48	/k/	франк	Америка - доллар, Швейцария - .....
49	/g/	блог	интернет, Живой Журнал, популярный .....
50	/g/	долг	кредит, выплата, вернуть .....
51	/g/	друг	собака - ..... человека
52	/g/	ковчег	всемирный потоп, Ноев .....
53	/g/	луг	покос, пастбище, зеленый .....
54	/g/	маг	чародей, волшебник, белый .....
55	/g/	овраг	земля, глубокая впадина, длинный .....
56	/g/	снег	зима, осадки, падающий .....
57	/g/	супруг	жених, после свадьбы, официальный статус - .....
58	/f/	блеф	обман в покере - .....
59	/f/	граф	Балконский - князь, Монте-Кристо - .....
60	/f/	рельеф	земля, поверхность, ... местности
61	/f/	тариф	сотовая связь, акция, выгодный .....
62	/f/	шеф	ресторан, кухня, ... - повар
63	/f/	штраф	безбилетник, наказание, денежный .....
64	/v/	залив	Мексиканский, Финский, Персидский .....
65	/v/	мотив	причина, повод преступления - .....
66	/v/	нрав	человек, характер, крутой .....

67	/v/	полив	вода, газон, автоматический .....
68	/v/	припев	песня, повторяющаяся часть, ...
69	/v/	сплав	соединение металлов - прочный .....
70	/v/	хлев	деревня, сарай, коровий .....
71	/s/	вес	диета, килограмм, лишний .....
72	/s/	запас	знание слов - словарный .....
73	/s/	индус	Пекин - китаец, Нью-Дели - .....
74	/s/	рис	Россия - картофель, Китай - .....
75	/s/	трус	..... не играет в хоккей
76	/s/	час	секунда, минута, ...
77	/z/	воз	Крылов, басня, а ..... и ныне там
78	/z/	гипноз	внушение, погружение в транс - .....
79	/z/	груз	тяжелый чемодан, непосильный .....
80	/z/	джаз	музыка, саксофон, американский .....
81	/z/	круиз	теплоход, морское путешествие, недельный .....
82	/z/	наркоз	операция, искусственный сон, ...
83	/z/	отказ	просьба, отрицательный ответ - .....
84	/z/	приз	конкурс, победа, главный .....
85	/z/	срез	острый скальпель, поперечный .....
86	/ʃ/	алкаш	алкоголик, пьяница, беспробудный .....
87	/ʃ/	гуляш	мясо кубиками, подливка, говяжий .....
88	/ʃ/	ёрш	пресноводная рыба, колючий .....
89	/ʃ/	клёш	моряк, широкая штанина, брюки .....
90	/ʃ/	куш	деньги, выигрыш, сорвать .....
91	/ʃ/	лаваш	пресное тесто, лепёшка, армянский .....
92	/ʃ/	чуваш	Казань - татарин, Чебоксары - .....
93	/ʒ/	багаж	пассажир, вещи, ручной .....
94	/ʒ/	вираж	скорость, крутой поворот, вписаться в .....
95	/ʒ/	корж	пицца, тесто, тонкий .....
96	/ʒ/	монтаж	Фотошоп, изменённая фотография, фото-.....
97	/ʒ/	муж	невеста - жених, жена - .....
98	/ʒ/	муляж	модель предмета, восковой .....
99	/ʒ/	уж	маленькая змея, безобидный .....
100	/ʒ/	шарж	зарисовка, карикатура, дружеский .....

Appendix 3: List of vector images

Num.	Final C	Orth. form (Cyrillic)	Num.	Final C	Orth. form (Cyrillic)
1	/p/	поп	26	/k/	лук
2	/p/	серп	27	/k/	мак
3	/p/	сноп	28	/g/	пирог
4	/p/	суп	29	/g/	рог
5	/p/	укроп	30	/g/	сапог
6	/p/	циклоп	31	/g/	стог
7	/b/	гриб	32	/g/	флаг
8	/b/	зуб	33	/g/	шланг
9	/b/	краб	34	/f/	жираф
10	/b/	лоб	35	/f/	мастиф
11	/b/	микроб	36	/f/	пуф
12	/b/	раб	37	/f/	шериф
13	/b/	столб	38	/v/	клюв
14	/b/	хлеб	39	/v/	лев
15	/t/	живот	40	/v/	удав
16	/t/	кот	41	/s/	кокос
17	/t/	магнит	42	/s/	нос
18	/t/	пират	43	/s/	поднос
19	/t/	плот	44	/s/	редис
20	/t/	скат	45	/z/	арбуз
21	/d/	завод	46	/ʃ/	душ
22	/d/	клад	47	/ʃ/	фарш
23	/d/	комод	48	/ʃ/	шалаш
24	/k/	каблук	49	/ʒ/	ёж
25	/k/	клубок	50	/ʒ/	стеллаж

Appendix 4: Plosive-final monosyllabic MPs - Results of RM ANOVAs (production data)

Factor	Measure	df_1	By-Subject		df_2	By-Item		Meets F_1 x F_2 Criterion
			F_1	p		F_2	p	
UR-voice	V. dur.	1,74	18.53	.000	1,24	0.42	.522	
	Cl. dur.	1,74	40.07	.000	1,24	11.59	.002	v
	Rel. dur.	1,74	33.21	.000	1,24	9.75	.005	v
	Cl. voi.	1,74	73.94	.000	1,24	207.03	.000	v
C-place	V. dur.	1,74	64.57	.000	2,24	2.96	.071	
	Cl. dur.	1,74	22.18	.000	2,24	7.98	.002	v
	Rel. dur.	1,74	26.95	.000	2,24	17.56	.000	v
	Cl. voi.	1,74	11.26	.000	2,24	8.68	.001	v
Group	V. dur.	3,74	0.64	.588	1,24	8.78	.000	
	Cl. dur.	3,74	0.14	.938	1,24	1.06	.363	
	Rel. dur.	3.74	0.30	.826	1,24	2.52	.078	
	Cl. voi.	3,74	5.00	.003	1,24	52.86	.000	v
UR-voice x C-place	V. dur.	1,74	0.14	.851	2,24	0.01	.996	
	Cl. dur.	1,74	1.39	.252	2,24	0.28	.755	
	Rel. dur.	1,74	5.34	.006	2,24	1.92	.169	
	Cl. voi.	1,74	3.51	.033	2,24	2.58	.096	
UR-voice x Group	V. dur.	3,74	0.18	.912	1,24	0.15	.907	
	Cl. dur.	3,74	2.11	.106	1,24	2.30	.090	
	Rel. dur.	3,74	1.23	.305	1,24	1.31	.279	
	Cl. voi.	3,74	9.43	.000	1,24	44.13	.000	v
C-place x Group	V. dur.	3,74	0.70	.648	2,24	0.82	.542	
	Cl. dur.	3,74	1.33	.252	2,24	2.95	.015	
	Rel. dur.	3,74	0.79	.576	2,24	1.27	.291	
	Cl. voi.	3,74	0.58	.740	2,24	0.74	.595	
UR-voice x C-place x Group	V. dur.	3,74	0.21	.968	2,24	0.19	.978	
	Cl. dur.	3,74	0.94	.466	2,24	0.74	.609	
	Rel. dur.	3,74	0.36	.905	2,24	0.28	.921	
	Cl. voi.	3,74	0.78	.583	2,24	1.01	.422	

Appendix 5: Plosive-final monosyllabic non-MPs - Results of RM ANOVAs (production data)

Factor	Measure	df_1	By-Subject		df_2	By-Item		Meets F_1 x F_2 Criterion
			F_1	p		F_2	p	
UR-voice	V. dur.	1,74	3.47	.067	1,24	0.67	.421	
	Cl. dur.	1,74	3.27	.075	1,24	2.15	.156	
	Rel. dur.	1,74	13.26	.000	1,24	16.48	.000	v
	Cl. voi.	1,74	50.22	.000	1,24	327.87	.000	v
C-place	V. dur.	1,74	53.02	.000	2,24	17.68	.000	v
	Cl. dur.	1,74	14.95	.000	2,24	28.72	.000	v
	Rel. dur.	1,74	26.15	.000	2,24	69.09	.000	v
	Cl. voi.	1,74	17.63	.000	2,24	48.19	.000	v
Group	V. dur.	3,74	0.30	.821	1,24	3.60	.029	
	Cl. dur.	3,74	0.06	.982	1,24	1.10	.350	
	Rel. dur.	3.74	0.33	.802	1,24	5.21	.003	
	Cl. voi.	3,74	3.43	.021	1,24	82.77	.000	v
UR-voice x C-place	V. dur.	1,74	1.00	.368	2,24	0.16	.853	
	Cl. dur.	1,74	1.67	.510	2,24	0.77	.474	
	Rel. dur.	1,74	0.18	.827	2,24	0.11	.899	
	Cl. voi.	1,74	1.12	.327	2,24	3.20	.058	
UR-voice x Group	V. dur.	3,74	0.31	819	1,24	0.22	.826	
	Cl. dur.	3,74	1.44	.237	1,24	1.92	.144	
	Rel. dur.	3,74	0.91	.442	1,24	1.79	.159	
	Cl. voi.	3,74	9.54	.000	1,24	56.67	.000	v
C-place x Group	V. dur.	3,74	0.25	.958	2,24	0.48	.769	
	Cl. dur.	3,74	1.19	.317	2,24	3.01	.017	
	Rel. dur.	3,74	0.72	.629	2,24	2.98	.013	
	Cl. voi.	3,74	0.70	.645	2,24	1.57	.172	
UR-voice x C-place x Group	V. dur.	3,74	0.29	.933	2,24	0.31	.887	
	Cl. dur.	3,74	1.43	.208	2,24	1.80	.125	
	Rel. dur.	3,74	1.25	.284	2,24	1.76	.122	
	Cl. voi.	3,74	0.87	.509	2,24	1.79	.116	

Appendix 6: Plosive-final disyllabic non-MPs - Results of RM ANOVAs (production data)

Factor	Measure	df_1	By-Subject		df_2	By-Item		Meets F_1 x F_2 Criterion
			F_1	p		F_2	p	
UR-voice	V. dur.	1,74	5.28	.024	1,24	0.81	.377	
	Cl. dur.	1,74	0.60	.442	1,24	0.42	.521	
	Rel. dur.	1,74	0.83	.365	1,24	0.57	.459	
	Cl. voi.	1,74	72.54	.000	1,24	124.10	.000	v
C-place	V. dur.	1,74	65.24	.000	2,24	17.25	.000	v
	Cl. dur.	1,74	13.17	.000	2,24	18.39	.000	v
	Rel. dur.	1,74	23.51	.000	2,24	24.51	.000	v
	Cl. voi.	1,74	8.73	.000	2,24	8.97	.001	v
Group	V. dur.	3,74	0.68	.565	1,24	5.40	.002	
	Cl. dur.	3,74	0.02	.997	1,24	.49	.661	
	Rel. dur.	3.74	0.36	.780	1,24	2.58	.079	
	Cl. voi.	3,74	6.28	.001	1,24	106.93	.000	v
UR-voice x C-place	V. dur.	1,74	0.23	.797	2,24	0.04	.965	
	Cl. dur.	1,74	0.52	.580	2,24	0.14	.871	
	Rel. dur.	1,74	0.18	.808	2,24	0.07	.931	
	Cl. voi.	1,74	0.11	.891	2,24	0.10	.904	
UR-voice x Group	V. dur.	3,74	0.22	.883	1,24	0.11	.932	
	Cl. dur.	3,74	1.05	.376	1,24	2.33	.091	
	Rel. dur.	3,74	0.81	.493	1,24	0.78	.477	
	Cl. voi.	3,74	14.60	.000	1,24	55.30	.000	v
C-place x Group	V. dur.	3,74	1.54	.169	2,24	1.78	.132	
	Cl. dur.	3,74	0.99	.433	2,24	3.55	.006	
	Rel. dur.	3,74	0.28	.944	2,24	0.53	.735	
	Cl. voi.	3,74	1.14	.340	2,24	2.63	.023	
UR-voice x C-place x Group	V. dur.	3,74	0.12	.994	2,24	0.06	.997	
	Cl. dur.	3,74	0.74	.610	2,24	0.90	.492	
	Rel. dur.	3,74	0.16	.981	2,24	0.18	.960	
	Cl. voi.	3,74	0.42	.860	2,24	0.70	.633	

Appendix 7: Fricative-final monosyllabic non-MPs - Results of RM ANOVAs (production data)

Factor	Measure	df_1	By-Subject		df_2	By-Item		Meets F_1 x F_2 Criterion
			F_1	p		F_2	p	
UR-voice	V. dur.	1,74	12.62	.001	1,24	0.13	.717	
	F. dur.	1,74	79.32	.000	1,24	17.82	.000	v
	Cl. voi.	1,74	31.32	.000	1,24	75.52	.000	v
C-place	V. dur.	1,74	7.07	.001	2,24	0.13	.876	
	F. dur.	1,74	87.59	.000	2,24	29.65	.000	v
	Cl. voi.	1,74	1.14	.312	2,24	0.60	.554	
Group	V. dur.	3,74	0.53	.665	1,24	9.33	.000	
	F. dur.	3,74	0.28	.838	1,24	6.60	.001	
	Cl. voi.	3,74	5.10	.003	1,24	51.15	.000	v
UR-voice x C-place	V. dur.	1,74	13.75	.000	2,24	0.12	.890	
	F. dur.	1,74	4.59	.013	2,24	0.55	.582	
	Cl. voi.	1,74	1.57	.214	2,24	0.45	.643	
UR-voice x Group	V. dur.	3,74	3.27	.026	1,24	2.01	.130	
	F. dur.	3,74	2.40	.075	1,24	4.62	.006	
	Cl. voi.	3,74	5.68	.001	1,24	47.02	.000	v
C-place x Group	V. dur.	3,74	0.67	.670	2,24	0.93	.469	
	F. dur.	3,74	1.16	.331	2,24	2.80	.019	
	Cl. voi.	3,74	0.60	.691	2,24	1.06	.393	
UR-voice x C-place x Group	V. dur.	3,74	1.92	.082	2,24	1.11	.363	
	F. dur.	3,74	0.23	.961	2,24	0.35	.899	
	Cl. voi.	3,74	0.61	.690	2,24	0.89	.497	

Appendix 8: Fricative-final disyllabic non-MPs - Results of RM ANOVAs (production data)

Factor	Measure	df_1	By-Subject		df_2	By-Item		Meets F_1 x F_2 Criterion
			F_1	p		F_2	p	
UR-voice	V. dur.	1,74	15.19	.000	1,24	0.58	.455	
	F. dur.	1,74	23.78	.000	1,24	14.94	.001	v
	Cl. voi.	1,74	34.14	.000	1,24	345.98	.000	v
C-place	V. dur.	1,74	3.96	.022	2,24	0.21	.812	
	F. dur.	1,74	79.07	.000	2,24	53.82	.000	v
	Cl. voi.	1,74	1.45	.239	2,24	3.20	.059	
Group	V. dur.	3,74	0.63	.602	1,24	7.88	.000	
	F. dur.	3,74	0.23	.874	1,24	6.01	.002	
	Cl. voi.	3,74	5.27	.002	1,24	156.24	.000	v
UR-voice x C-place	V. dur.	1,74	4.98	.008	2,24	0.14	.869	
	F. dur.	1,74	3.75	.028	2,24	0.94	.405	
	Cl. voi.	1,74	0.74	.437	2,24	1.26	.301	
UR-voice x Group	V. dur.	3,74	2.25	.089	1,24	1.24	.302	
	F. dur.	3,74	1.61	.195	1,24	3.58	.025	
	Cl. voi.	3,74	5.62	.002	1,24	130.63	.000	v
C-place x Group	V. dur.	3,74	1.38	.229	2,24	0.83	.530	
	F. dur.	3,74	0.98	.437	2,24	2.48	.042	
	Cl. voi.	3,74	1.37	.246	2,24	6.27	.000	
UR-voice x C-place x Group	V. dur.	3,74	1.48	.189	2,24	0.75	.582	
	F. dur.	3,74	0.21	.969	2,24	0.34	.884	
	Cl. voi.	3,74	1.00	.412	2,24	3.95	.004	

Appendix 9: Plosive-final monosyllabic MPs - Results of RM ANOVAs (perception data)

Token type	Factor	Measure	df_1	By-Subject		df_2	By-Item		Meets F_1 x F_2 Criterion
				F_1	p		F_2	p	
Full wordform	UR-voice	%-voiced	1,51	660.09	.000	1,24	173.68	.000	
		RT	1,51	0.29	.593	1,24	0.01	.904	v
	C-place	%-voiced	2,51	0.76	.474	2,24	4.26	.026	
		RT	2,51	2.90	.064	2,24	23.57	.000	
	P-Group	%-voiced	2,51	15.25	.000	1,24	3.85	.016	v
		RT	2,51	7.30	.000	1,24	2.86	.052	
	UR-voice x C-place	%-voiced	2,51	45.16	.000	2,24	12.12	.000	v
		RT	2,51	3.89	.027	2,24	0.51	.605	
	UR-voice x P-group	%-voiced	1,51	66.13	.000	1,24	13.23	.000	v
		RT	1,51	0.34	.782	1,24	0.23	.848	
	C-place x P-group	%-voiced	2,51	3.91	.002	2,24	0.95	.453	
		RT	2,51	4.44	.001	2,24	2.26	.057	
	UR-voice x C-place x P- group	%-voiced	2,51	7.81	.000	2,24	1.68	.147	
		RT	2,51	1.99	.076	2,24	1.29	.278	
Final rhyme only	UR-voice	%-voiced	1,51	465.23	.000	1,24	197.63	.000	v
		RT	1,51	8.54	.005	1,24	1.20	.284	
	C-place	%-voiced	2,51	0.79	.459	2,24	7.36	.003	
		RT	2,51	0.79	.457	2,24	13.58	.000	
	P-group	%-voiced	2,51	17.12	.000	1,24	5.65	.002	v
		RT	2,51	12.40	.000	1,24	5.31	.005	v
	UR-voice x C-place	%-voiced	2,51	37.82	.000	2,24	16.87	.000	v
		RT	2,51	2.13	.129	2,24	0.30	.746	
	UR-voice x P-group	%-voiced	1,51	43.40	.000	1,24	13.98	.000	v
		RT	1,51	2.57	.067	1,24	1.78	.173	
	C-place x P-group	%-voiced	2,51	3.67	.002	2,24	1.00	.427	
		RT	2,51	1.38	.230	2,24	0.58	.703	
	UR-voice x C-place x P- group	%-voiced	2,51	1.30	.262	2,24	0.48	.808	
		RT	2,51	1.61	.161	2,24	1.15	.346	

Appendix 10: Plosive-final monosyllabic non-MPs - Results of RM ANOVAs (perception data)

Factor	Measure	df_1	By-Subject		df_2	By-Item		Meets F_1 x F_2 Criterion
			F_1	p		F_2	p	
UR-voice	%-voiced	1,51	393.60	.000	1,24	174.08	.000	∨
	RT	1,51	1.856	.179	1,24	0.28	.602	
C-place	%-voiced	2,51	0.68	.512	2,24	5.14	.014	
	RT	2,51	0.85	.432	2,24	11.46	.000	∨
P-group	%-voiced	2,51	7.37	.000	1,24	3.61	.024	∨
	RT	2,51	7.20	.000	1,24	5.35	.005	∨
UR-voice x C-place	%-voiced	2,51	27.61	.000	2,24	11.95	.000	∨
	RT	2,51	0.29	.748	2,24	0.07	.935	
UR-voice x P-group	%-voiced	1,51	18.28	.000	1,24	9.57	.000	∨
	RT	1,51	1.88	.145	1,24	1.15	.330	
C-place x P-group	%-voiced	2,51	2.15	.059	2,24	1.23	.307	
	RT	2,51	1.29	.267	2,24	0.80	.547	
UR-voice x C-place x P- group	%-voiced	2,51	0.56	.752	2,24	0.29	.918	
	RT	2,51	2.05	.075	2,24	1.48	.213	

Appendix 11: Plosive-final disyllabic MPs - Results of RM ANOVAs (perception data)

Factor	Measure	df_1	By-Subject		df_2	By-Item		Meets F_1 x F_2 Criterion
			F_1	p		F_2	p	
UR-voice	%-voiced	1,51	253.18	.000	1,24	94.95	.000	∨
	RT	1,51	0.96	.331	1,24	0.53	.472	
C-place	%-voiced	2,51	0.91	.411	2,24	6.03	.008	
	RT	2,51	0.78	.463	2,24	20.94	.000	
P-group	%-voiced	2,51	12.14	.000	1,24	5.76	.002	∨
	RT	2,51	2.42	.077	1,24	2.93	.053	
UR-voice x C-place	%-voiced	2,51	23.15	.000	2,24	8.28	.002	∨
	RT	2,51	2.56	.088	2,24	1.33	.284	
UR-voice x P-group	%-voiced	1,51	25.40	.000	1,24	12.12	.000	∨
	RT	1,51	2.28	.088	1,24	2.84	.057	
C-place x P-group	%-voiced	2,51	1.96	.082	2,24	1.07	.386	
	RT	2,51	0.39	.865	2,24	0.49	.776	
UR-voice x C-place x P-group	%-voiced	2,51	1.57	.167	2,24	0.82	.542	
	RT	2,51	1.18	.322	2,24	2.84	.057	

Appendix 12: Fricative-final monosyllabic non-MPs - Results of RM ANOVAs (perception data)

Factor	Measure	df_1	By-Subject		df_2	By-Item		Meets F_1 x F_2 Criterion
			F_1	p		F_2	p	
UR-voice	%-voiced	1,51	241.23	.000	1,24	38.60	.000	∨
	RT	1,51	0.79	.377	1,24	0.11	.748	
C-place	%-voiced	2,51	0.65	.525	2,24	1.01	.377	
	RT	2,51	0.43	.655	2,24	3.91	.034	
P-group	%-voiced	1,51	2.82	.044	1,24	1.23	.303	
	RT	1,51	2.11	.107	1,24	1.03	.375	
UR-voice x C-place	%-voiced	2,51	18.23	.000	2,24	2.81	.080	
	RT	2,51	0.37	.696	2,24	0.04	.959	
UR-voice x P-group	%-voiced	1,51	29.53	.000	1,24	10.45	.000	∨
	RT	1,51	1.23	.299	1,24	1.33	.273	
C-place x P-group	%-voiced	2,51	4.04	.001	2,24	2.06	.083	
	RT	2,51	1.94	.084	2,24	0.92	.471	
UR-voice x C-place x P-group	%-voiced	2,51	1.47	.199	2,24	0.49	.780	
	RT	2,51	0.53	.778	2,24	0.63	.676	

Appendix 13: Fricative-final disyllabic non-MPs - Results of RM ANOVAs (perception data)

Factor	Measure	df_1	By-Subject		df_2	By-Item		Meets F_1 x F_2 Criterion
			F_1	p		F_2	p	
UR-voice	%-voiced	1,51	232.28	.000	1,24	32.51	.000	V
	RT	1,51	2.50	.120	1,24	0.26	.611	
C-place	%-voiced	2,51	0.33	.717	2,24	0.27	.766	
	RT	2,51	0.40	.670	2,24	2.86	.077	
P-group	%-voiced	2,51	1.30	.276	1,24	0.75	.506	
	RT	2,51	2.89	.042	1,24	1.98	.131	
UR-voice x C-place	%-voiced	2,51	20.145	.000	2,24	2.73	.085	
	RT	2,51	2.05	.139	2,24	0.26	.773	
UR-voice x P-group	%-voiced	1,51	24.64	.000	1,24	10.30	.000	V
	RT	1,51	3.62	.016	1,24	2.62	.064	
C-place x P-group	%-voiced	2,51	7.20	.000	2,24	3.52	.007	V
	RT	2,51	1.01	.419	2,24	1.01	.421	
UR-voice x C-place x P-group	%-voiced	2,51	2.20	.050	2,24	1.08	.382	
	RT	2,51	1.74	.118	2,24	1.30	.273	